

INTRODUCCION
A LA
TEORIA DE LA
ESTADISTICA

ALEXANDER M. MOOD
Y
FRANKLIN A. GRAYBILL

INTRODUCCION
A LA
TEORIA DE LA
ESTADISTICA

Adaptación a la 2.^a edición norteamericana por
RAFAEL PRO BERMEJO
Profesor de la Escuela de Estadística de la Universidad de Madrid



AGUILAR

colección ciencia y técnica
sección matemáticas y estadística
obra incorporada con el asesoramiento
de luis bravo gala

la cuarta edición española se ha preparado
sobre la traducción de la primera edición original
realizada por francisco azorín poch

edición española
© aguilar s a de ediciones 1955 1969 juan bravo 38 madrid
deposito legal m 2053/1978
cuarta edición—cuarta reimpresión—1978
ISBN 84-03-20102-8
printed in spain impreso en españa por gráficas ema
miguel yuste 31 madrid

edición original
© mcgraw-hill book company 1952 1963
introduction to the theory of statistics (second edition)
mcgraw-hill book company inc new york

PROLOGOS

PROLOGO DEL TRADUCTOR A LA PRIMERA EDICION ESPAÑOLA

Aunque todavía no es muy grande el número de obras de Estadística, originales o traducidas al español, parece oportuno subrayar los relevantes méritos de este libro, que hicieron preferir su traducción a la de tantos textos publicados hasta la fecha en Inglaterra y en los Estados Unidos. Creemos que esta obra es, entre las de nivel medio, la que mejor tiene en cuenta las necesidades prácticas del estadístico de hoy, lo que ya se pone de manifiesto en la definición que de la Estadística da el autor al comienzo del capítulo primero, como «tecnología de la experimentación científica».

En cuanto al ingrato problema de la traducción de términos y conceptos, que la ausencia de un medio universal de expresión impone al científico moderno, hemos procurado seguir en lo posible la nomenclatura utilizada en obras anteriormente publicadas por esta Editorial, en particular las más recientes de Introducción a la Estadística matemática, traducida por J. Ros Jimeno, y Métodos matemáticos de Estadística, cuyo traductor es E. Cansado. Quizá la más discutida haya sido la de traducir las palabras inglesas test y testing por las españolas de dócima y docimasia. En la Nota del traductor de la última obra citada se justifica extensamente su empleo, así como el inconveniente de otras traducciones propuestas. En muchos casos, se ha traducido test y to test por contraste y contrastar, lo que en general no ofrece inconvenientes, pero en la teoría del diseño de experimentos se usa en inglés el término contrast para indicar una combinación de efectos, que permite el contraste (en el sentido de «oposición, contraposición o diferencia entre seres o cosas») entre los resultados de diferentes tratamientos o variedades, por lo que aquí parece más conveniente reservar la palabra española contraste para la traducción del término inglés contrast.

Confiamos en que esta obra podrá desempeñar un papel útil entre los textos estadísticos directamente accesibles a los lectores de habla española, y recordamos que gran parte de estos textos

han sido publicados por Aguilar, S. A. de Ediciones, que por ello merece nuestro reconocimiento. Deseamos que esta traducción ocupe un puesto honroso entre Introducción a la Estadística matemática, de Yule-Kendall, y Métodos matemáticos de Estadística, de H. Cramér, contribuyendo así a la ya iniciada formación de técnicos estadísticos en España y en Hispanoamérica.

FRANCISCO AZORÍN.

PREFACIO DEL AUTOR A LA PRIMERA EDICION

Este libro se ha desarrollado a partir de un conjunto de notas preparadas por mí en 1945. En aquella fecha no existía texto moderno especialmente dirigido a quienes empezaran a estudiar estadística matemática. Desde entonces la situación ha mejorado considerablemente, y si yo hubiera sabido por anticipado los libros que había en preparación, es probable que no me hubiera decidido a escribir esta obra. No obstante, como el libro parece ser suficientemente distinto a los demás en el modo de presentar las cosas, espero pueda proporcionar a profesores y estudiantes una útil posibilidad de elección.

Las notas antes mencionadas se emplearon durante tres años como texto en un curso para estudiantes ya graduados, de dos niveles diferentes, en el Iowa State College. Solo se exigía para este curso un año de cálculo, requerimiento que prejuzga el nivel del libro (la clase de cálculo en Iowa constaba de cuatro horas semanales e incluía un estudio detallado de desarrollos de Taylor, derivadas parciales e integración múltiple). No se suponían conocimientos previos de estadística.

Es este un libro de estadística, no de matemáticas, como cualquier matemático podrá ver fácilmente; no hay gran rigor matemático en sus desarrollos, por la sencilla razón de que resultaría pesado y supondría una pérdida de tiempo injustificada en este nivel de enseñanza; claro es que el rigor en los razonamientos resulta totalmente esencial en buena estadística y he procurado ponerlo así de manifiesto, haciendo que el lector se dé cuenta de su necesidad y subrayando varios fallos en argumentos rigurosos.

Aunque este texto se refiere primordialmente a la teoría de la estadística, se ha tenido plenamente en cuenta a aquellos estudiantes que temen perder un solo momento en *frivolidades* matemáticas. Toda cuestión nueva viene acompañada de un pequeño cortejo de cuestiones prácticas y, lo que es más importante, se ha llevado a efecto un serio esfuerzo para ilustrar por medio de problemas las diversas formas en que puede aplicarse la teoría.

Los problemas constituyen parte esencial del libro: se extienden desde simples ejemplos numéricos a teoremas que se necesitan en

capítulos posteriores, e incluyen materias tal vez más importantes que algunas de las estudiadas en el texto; el que una materia se haya tratado o no en los problemas se ha basado más en la conveniencia de hacerlo así que en su importancia; por ejemplo, casi todas las cuestiones de correlación se tratan en los problemas. Me pareció poco eficaz ocuparme dos veces de las cuestiones multivariantes: una desde el punto de vista de la regresión, y otra desde el de la correlación. En el texto se ha expuesto con mayor insistencia la regresión por su carácter más general.

El autor de un libro de texto ha de sentirse en deuda, prácticamente, con todos los que han tratado la materia correspondiente, y desde aquí me reconozco obligado a todos los estadísticos. No obstante, al reconocer explícitamente su contribución ha de fijar un límite, y yo he simplificado la situación trazando este muy alto; solo mencionaré, pues, a los más destacados.

Mi mayor deuda personal es con S. S. Wilks, quien despertó mi interés por la estadística y fue mi mentor durante mi época de estudiante. Cualquier mérito que pueda tener este libro deberá atribuirse en gran parte a sus meditadas explicaciones y a la comprensiva dirección de mis estudios.

Todos mis colegas en el Iowa State College han contribuido a mi comprensión y visión general de la estadística. Reconozco en particular lo mucho que debo a G. W. Brown, W. G. Cochran y G. W. Snedecor.

Entre los numerosos estudiantes que revisaron por completo las notas originales debo mencionar, por sus excelentes comentarios y sugerencias, a H. D. Block, quien, al final del manuscrito, hizo de este una cuidadosa y competente revisión. Margaret Kirwin y Ruth Burns tradujeron con esmero mis garabatos en una perfecta copia mecanográfica. Bernice Brown y miss Burns leyeron cuidadosamente todas las pruebas de imprenta.

Estoy en deuda también con Catherine Thompson y Maxime Merrington, y con E. S. Pearson, director de *Biometrika*, por haberme permitido incluir las tablas III y V, que son versiones abreviadas de tablas publicadas en *Biometrika*. Lo mismo digo respecto a los profesores R. A. Fisher y Frank Yates, y a la firma Oliver and Boyd, Ltd., de Edimburgo, por su autorización para reproducir la tabla IV de su libro *Tablas estadísticas para investigadores científicos*¹.

¹ Hay edición española de Aguilar, S. A. de Ediciones, 1954.

En el último capítulo hay algunas dójimas a libre distribución, que fueron desarrolladas conjuntamente por G. W. Brown y por mí en el Iowa State College, en un proyecto del Office of Naval Research. El profesor Brown me ha permitido generosa y amablemente incluir este material, que debiera haber aparecido impreso por primera vez con su nombre y el mío. Estas dójimas aparecen en las secciones 16-5 a 16-9.

ALEXANDER MCFARLANE MOOD.

PREFACIO DE LOS AUTORES A LA SEGUNDA EDICION

Dado que la primera edición de esta obra se publicó en 1950, muchas nuevas técnicas estadísticas se han creado desde entonces, y muchas otras, que eran solo del dominio de los estadísticos matemáticos, se conocen y utilizan ahora por los estadísticos aplicados. Para incluir parte de este material hemos tenido que eliminar otro, con el fin de no aumentar excesivamente el volumen del libro. El propósito general de exponer la teoría en conexión con problemas prácticos concretos contribuyó, aparentemente en gran medida, al éxito de la primera edición, y hemos procurado mantenerlo en la presente.

Para estudiar este libro no es preciso haber seguido un curso previo de estadística. La preparación matemática necesaria es la usual en un primer curso de cálculo. Aunque no esencial, es deseable poseer algún conocimiento de la aritmética de matrices; por otra parte, en el capítulo 9 se hace una breve introducción de las operaciones necesarias. Se han señalado con asterisco algunas de las secciones que utilizan recursos de álgebra matricial y que pueden omitirse sin interrumpir la continuidad del libro.

Los autores se sienten en deuda con el profesor Herman Chernoff, que dedicó mucho tiempo a revisar a fondo gran parte del manuscrito, incluso redactando de nuevo varias secciones.

También expresamos nuestra gratitud al Dr. David Weeks, que leyó la totalidad del manuscrito; a Terrence Connell, William Owen y Scott Urquhart, que nos ayudaron en la corrección de pruebas; así como a los doctores James Pachares y Leon Harter y a los directores de *Biometrika* por su amable autorización para reproducir determinado material en las tablas VI, VII y VIII.

ALEXANDER M. MOOD.
FRANKLIN A. GRAYBILL.

INDICE GENERAL

INDICE GENERAL

PRÓLOGO DEL TRADUCTOR A LA PRIMERA EDICIÓN ESPAÑOLA ...	Pág.	IX
PREFACIO DEL AUTOR A LA PRIMERA EDICIÓN ...		XI
PREFACIO DE LOS AUTORES A LA SEGUNDA EDICIÓN ...		XIV
CAP. 1.—INTRODUCCIÓN ...		3
1-1. Estadística, <i>pág.</i> 3.—1-2. Objeto y amplitud de este libro, 6.—1-3. Sistema de referencia, 7.—Bibliografía, 7.		
CAP. 2.—PROBABILIDAD ...		9
2-1. Introducción, <i>pág.</i> 9.—2-2. Probabilidad clásica o <i>a priori</i> , 9.—2-3. Probabilidad <i>a posteriori</i> o frecuencial, 12.—2-4. Modelos de probabilidad, 15.—2-5. Conjuntos de puntos, 17.—2-6. Desarrollo axiomático de la probabilidad, 21.—2-7. Espacio muestral discreto con un número finito de puntos, 22.—2-8. Permutaciones y combinaciones, 23.—2-9. Fórmula de Stirling, 29.—2-10. Notaciones de sumas y productos, 30.—2-11. Los teoremas binomial y polinomial, 30.—2-12. Funciones generatrices combinatorias, 33.—2-13. Probabilidad marginal, 37.—2-14. Probabilidad condicional, 40.—2-15. Dos leyes básicas de la probabilidad, 42.—2-16. Sucesos compuestos, 45.—2-17. Independencia, 51.—2-18. Variables aleatorias, 52.—Problemas, 54.—Bibliografía, 60.		
CAP. 3.—VARIABLES ALEATORIAS DISCRETAS ...		61
3-1. Introducción, <i>pág.</i> 61.—3-2. Funciones de cuantía, 65.—3-3. Distribuciones multivariantes, 66.—3-4. Distribución binomial, 75.—3-5. Distribución polinomial, 80.—3-6. Distribución de Poisson, 81.—3-7. Otras distribuciones discretas, 83.—Problemas, 84.—Bibliografía, 87.		
CAP. 4.—VARIABLES ALEATORIAS CONTINUAS ...		88
4-1. Introducción, <i>pág.</i> 88.—4-2. Variables aleatorias continuas, 88.—4-3. Distribuciones multivariantes, 96.—4-4. Distribuciones acumulativas, 99.—4-5. Distribuciones marginales, 104.—4-6. Distribuciones condicionales, 106.—4-7. Independencia, 107.—4-8. Muestra aleatoria, 108.—4-9. Distribuciones deducidas de otras, 110.—Problemas, 113.—Bibliografía, 117.		
CAP. 5.—VALORES ESPERADOS Y MOMENTOS ...		118
5-1. Valores esperados, <i>pág.</i> 118.—5-2. Momentos, 122.—5-3. Funciones generatrices de momentos, 130.—5-4. Momentos para distribuciones multivariantes, 133.—5-5. El problema de los momentos, 134.—5-6. Esperanzas condicionales, 134.—Problemas, 136.—Bibliografía, 139.		
CAP. 6.—DISTRIBUCIONES CONTINUAS ESPECIALES ...		140
6-1. Distribución uniforme, <i>pág.</i> 140.—6-2. La distribución normal, 141.—6-3. La distribución gamma, 145.—6-4. La distribución beta, 148.—6-5. Otras distribuciones, 151.—6-6. Funciones de densidad completas, 151.—Problemas, 155.—Bibliografía, 159.		
CAP. 7.—MUESTREO ...		160
7-1. Inferencia inductiva, <i>pág.</i> 160.—7-2. Poblaciones y muestras, 162.—7-3. Distribuciones muestrales, 164.—7-4. Momentos muestrales, 166.—7-5. Ley de los grandes números, 169.—7-6. El teorema central del límite, 172.—7-7. Aproximación normal a la distribución binomial, 176.—7-8. Papel de la distribución normal en estadística, 179.—Problemas, 180. Bibliografía, 183.		

CAP. 8.—ESTIMACIÓN PUNTUAL	185
8-1. Teoría de la decisión, <i>pág.</i> 185.—8-2. Estimación puntual, 190.—8-3. Estadísticos suficientes; caso de un solo parámetro, 193.—8-4. Estadísticos suficientes; más de un parámetro, 196.—8-5. Estimador insesgado, 198.—8-6. Estimador consistente, 199.—8-7. Estimadores asintóticamente eficientes, 200.—8-8. Estimadores insesgados de varianza mínima, 202.—8-9. Principio de máxima verosimilitud, 206.—8-10. Algunos estimadores máximo-verosímiles, 210.—8-11. Propiedades de los estimadores máximo-verosímiles, 213.—8-12. Estimación por el método de los momentos, 214.—8-13. Estimadores de Bayes, 215.—Problemas, 221.—Bibliografía, 226.	
CAP. 9.—DISTRIBUCIÓN NORMAL MULTIVARIANTE	228
9-1. La distribución normal bivalente, <i>pág.</i> 228.—9-2. Matrices y determinantes, 234.—9-3. Distribución normal multivariante, 238.—Problemas, 248. Bibliografía, 252.	
CAP. 10.—DISTRIBUCIONES EN EL MUESTREO	253
10-1. Distribuciones de funciones de variables aleatorias, <i>pág.</i> 253.—10-2. Distribución de la media muestral para densidades normales, 259.—10-3. Distribución χ^2 cuadrado, 259.—10-4. Independencia de la media y varianzas muestrales en densidades normales, 261.—10-5. La distribución «F», 265.—10-6. Distribución «t» de Student, 267.—10-7. Distribución de las medias muestrales en densidades binomiales y de Poisson, 268.—10-8. Distribución, en muestras grandes, de estimadores máximo-verosímiles, 270.—10-9. Distribución de estadísticos ordinales, 276.—10-10. Recorrido «studentizado», 279.—Problemas, 280.—Bibliografía, 284.	
CAP. 11.—ESTIMACIÓN POR INTERVALOS	285
11-1. Intervalos confidenciales, <i>pág.</i> 285.—11-2. Intervalos confidenciales para la media de una distribución normal, 289.—11-3. Intervalos confidenciales para la varianza de una distribución normal, 291.—11-4. Región confidencial para la media y la varianza de una distribución normal, 293.—11-5. Método general para la obtención de intervalos confidenciales, 295.—11-6. Intervalos confidenciales para el parámetro de una distribución binomial, 299.—11-7. Intervalos confidenciales para muestras grandes, 301.—11-8. Regiones confidenciales para muestras grandes, 303.—11-9. Intervalos confidenciales múltiples, 307.—Problemas, 312.—Bibliografía, 315.	
CAP. 12.—DOCIMASIA DE HIPÓTESIS	317
12-1. Introducción, <i>pág.</i> 317.—12-2. Dócima de una hipótesis simple contra una alternativa simple, 324.—12-3. Hipótesis compuestas, 335.—12-4. Docimasia de $\theta \leq \theta_0$ contra $\theta > \theta_0$ para densidades con un parámetro único θ , 339.—12-5. Docimasia de la hipótesis $H_1: \theta_1 \leq \theta \leq \theta_2$ contra la hipótesis alternativa $H_2: \theta > \theta_2$, $\theta < \theta_1$, 341.—12-6. Dócima de la razón de verosimilitud generalizada, 343.—12-7. Dócimas relativas a la media de una población normal, 347.—12-8. Diferencias entre las medias de dos poblaciones normales, 350.—12-9. Dócimas de la varianza de una distribución normal, 354.—12-10. Dócima de la bondad del ajuste, 356.—12-11. Dócimas de independencia en tablas de contingencia, 359.—Problemas, 368. Bibliografía, 376.	
CAP. 13.—REGRESIÓN E HIPÓTESIS LINEALES	378
13-1. Introducción, <i>pág.</i> 378.—13-2. Modelos lineales simples, 379.—13-3. Predicción, 386.—13-4. Discriminación, 389.—13-5. Estimación puntual. Caso B, 391.—13-6. El modelo lineal general, 394.—Problemas, 410. Bibliografía, 413.	
CAP. 14.—MODELOS DE DISEÑO EXPERIMENTAL	415
14-1. Introducción, <i>pág.</i> 415.—14-2. Modelo de diseño experimental, 417.—14-3. Modelo de clasificación simple, 431.—14-4. Modelo de clasificación doble, 433.—14-5. Otros modelos, 438.—Problemas, 438.—Bibliografía, 442.	
CAP. 15.—DÓCIMAS SUCESIONALES DE HIPÓTESIS	444
15-1. Análisis sucesionales, <i>pág.</i> 444.—15-2. Construcción de dócimas sucesionales, 445.—15-3. Funciones de potencia, 449.—15-4. Tamaño muestral medio, 453.—15-5. Inspección por muestreo, 456.—15-6. Inspección por muestreo sucesional, 459.—15-7. Dócima sucesional para la media de una población normal, 461.—Problemas, 463.—Bibliografía, 466.	

CAP. 16.—MÉTODOS NO PARAMÉTRICOS	467
16-1. Introducción, <i>pág.</i> 467.—16-2. Una distribución básica, 468.—16-3. Posición y dispersión, 470.—16-4. Comparación de dos poblaciones, 474.—16-5. Límites de tolerancia, 482.—16-6. Décima de rangos para dos muestras, 483.—16-7. Eficiencias asintóticas y décima de aleatorización, 486. Problemas, 489.—Bibliografía, 492.	
TABLAS	493
Descripción de las tablas, <i>pág.</i> 495.—Tabla I: Ordenadas de la función de densidad normal, 498.—Tabla II: Distribución normal acumulativa, 499.—Tabla III: Distribución ji cuadrado acumulativa, 500.—Tabla IV: Distribución acumulativa de «Student», 501.—Tabla V: Distribución <i>F</i> acumulativa, 502.—Tabla VI: Puntos porcentuales superiores del 1% del recorrido studentizado, 504.—Tabla VII: Puntos porcentuales superiores del 5% del recorrido studentizado, 505.—Tabla VIII: Puntos porcentuales superiores del 10% del recorrido studentizado, 506.	
SOLUCIONES DE LOS PROBLEMAS PROPUESTOS AL FINAL DE LOS CAPÍTULO.	509
INDICE ALFABÉTICO	531

INTRODUCCION
A LA
TEORIA DE LA ESTADISTICA

CAPITULO 1

INTRODUCCION

1-1. Estadística.—Para situar este libro en la perspectiva adecuada es necesario que empecemos por considerar qué es la estadística. La concepción profana de estadística suele incluir la recogida de grandes masas de datos y la presentación de estos en tablas o gráficos; puede incluir también el cálculo de totales, promedios, porcentajes, etc. En todo caso, estas operaciones, más o menos rutinarias, son una parte, pero solo una parte incidental de la estadística. Estadística es también el diseño de experimentos, el diseño de sobrevisiones muestrales, la reducción y el proceso de datos, y otras muchas cuestiones.

Describiremos la estadística como la tecnología del método científico. La estadística proporciona instrumentos para la toma de decisiones cuando prevalecen condiciones de incertidumbre. Estos instrumentos pueden ser de aplicación y utilidad completamente general en cualquier campo de la ciencia: físico, biológico, social, etcétera. Son aplicables no solo en el mundo científico, sino también en el de la empresa y en el de los asuntos cotidianos. Por otra parte, ciertos instrumentos pueden estar especialmente diseñados para campos especiales de la investigación.

La estadística puede dividirse en dos amplias ramas: 1) estadística descriptiva, que está relacionada con el resumen de datos y la descripción de estos; 2) estadística inferencial, relacionada con el proceso de utilizar datos para tomar decisiones en el caso más general del que forman parte estos datos. El proceso de tomar decisiones en situaciones generales, sobre la base de una información incompleta contenida en datos muestrales, es arriesgado y no puede realizarse con certeza; la probabilidad es una medida de esta incertidumbre. Hay dos tipos de incertidumbre con los que tenemos que enfrentarnos: 1) la incertidumbre debida a la aleatoriedad, y 2) la incertidumbre debida a nuestra ignorancia del verdadero estado del sistema. Lo aclararemos con un ejemplo.

La compañía A cultiva cierta clase de plantas, recolecciona las semillas y las envasa en paquetes de 25 semillas cada uno. Un almacén de venta al por menor adquiere algunos de los paquetes y garantiza a sus compradores que 22 al menos de las 25 semillas de cada paquete crecerán; en caso contrario, les dará otro paquete libre de todo gasto. El almacenista tiene dos tipos de incertidumbre

con que luchar: 1) no está seguro de qué proporción p_A de los paquetes que la compañía tiene en venta será aceptable (contienen al menos 22 semillas que crecerán), y 2) puesto que la compañía tiene del orden de un millón de paquetes de semillas en venta y el almacenista adquiere solo unos 200 paquetes, se enfrenta con otra incertidumbre; es decir, aun conociendo que la proporción p_A del millón de paquetes es aceptable, ¿cómo puede estar «seguro» o «razonablemente seguro» de que la proporción p_A de los 200 paquetes que *él adquiere* sea aceptable? Aunque p_A sea 0,99, es decir, aunque 990 000 paquetes del millón que la compañía tiene en venta sean aceptables, los 200 paquetes del almacenista podrían haberse seleccionado «accidentalmente» entre los 10 000 inaceptables y perdería mucho dinero.

El primer tipo de incertidumbre, no conocer p_A , proporción de paquetes aceptables que la compañía produce, tiene su origen en la ignorancia del estado del sistema (llamado, a veces, verdadero estado de la naturaleza). El segundo se debe a lo que se designa frecuentemente como «aleatoriedad».

Al almacenista se le ofrece la posibilidad de mejorar su situación por experimentación (o exigir a la compañía que realice pruebas de germinación), por la que puede tomar «decisiones» basadas en lo que él cree que es el estado de la naturaleza (dado por p_A). Aun así, nunca será capaz de determinar p_A exactamente y con certeza. Si conoce la pérdida en que incurrirá si determina que la proporción de paquetes aceptables es p'_A cuando realmente es p_A , necesitará experimentar y tomar decisiones de tal forma que de alguna manera haga mínima su pérdida.

Para complicar más las cosas, otra compañía (la compañía B) vende también la misma clase de semillas a idéntico precio por paquete, por lo que el almacenista debe decidir qué compañía será su proveedora. Si p_A es mayor que p_B , comprará a la compañía A ; en caso contrario, adquirirá las semillas de la compañía B .

El almacenista puede realizar un experimento (o exigir a cada compañía que efectúe pruebas de germinación) y elegir una de dos acciones: (a_1) comprar a la compañía A ; o (a_2) comprar a la compañía B , según los resultados del experimento y su evaluación de la pérdida que puede sufrir si toma una decisión errónea.

El diseño del experimento, determinar el número y clase de observaciones a realizar, y decidir cómo deben utilizarse los resultados para tomar «buenas» decisiones, son problemas estadísticos.

Otra división del campo de la estadística que merece una breve consideración es la que existe entre teoría y metodología. La teoría estadística es una rama de la matemática aplicada; tiene sus raíces en la rama de la matemática pura conocida con el nombre de teoría de la probabilidad, y en realidad la estructura completa

de la teoría estadística en sentido amplio puede considerarse que incluye la teoría de la probabilidad. Incluye también otras cuestiones que no forman parte de la teoría de la probabilidad propiamente dicha, como las consecuencias del principio de aleatorización, diversos principios de estimación y otros relativos a la docimasia de hipótesis, y, en general, un principio de toma de decisiones. Cabe considerar estos principios como axiomas que se integran en la axiomática de la teoría de la probabilidad.

El estadístico se ocupa, por supuesto, de la producción de instrumentos para uso de los investigadores. Al encontrarse con un problema experimental determinado, construye un modelo matemático que se ajuste, lo mejor posible, a la situación experimental; analiza el modelo por métodos matemáticos, y, finalmente, establece procedimientos para el estudio del problema. En sus trabajos se guía por los principios de la teoría estadística.

El estadístico se ocupa así mismo del desarrollo y extensión de la teoría estadística. Existen muchos problemas importantes del diseño experimental y de la inferencia estadística que permanecen todavía sin tocar, porque la teoría estadística no ha sido aún capaz de resolverlos. El gran avance de la aplicación de los métodos estadísticos durante las tres décadas pasadas fue posible gracias a los desarrollos teóricos de largo alcance que habían tenido lugar en la época que precedió inmediatamente a la citada.

Pueden ser de interés unas observaciones sobre los orígenes de la teoría estadística. Ciertas zonas de la experimentación biológica alcanzaron un punto en que, para su progreso, se hacía imperativo el uso de los ahora denominados métodos estadísticos. Fueron entonces los mismos biólogos los que desarrollaron lo esencial de esta teoría. Aunque este desarrollo ha sido paralelo al de casi todas las ramas del conocimiento abstracto, resulta curioso, sin embargo, en el caso de la estadística; la teoría estadística parece una evolución muy natural de la teoría de la probabilidad, que cuenta varios centenares de años de antigüedad; pero la verdad es que los investigadores probabilistas prescinden totalmente de la estadística. Conviene advertir, de paso, que la situación que dio lugar al nacimiento de la teoría estadística continúa existiendo; hay muchas zonas de la experimentación científica en espera de métodos aún no creados.

Además de la teoría, hay que considerar la práctica de la estadística. Hay un gran bagaje de instrumentos y técnicas para investigadores que crece apreciablemente en el transcurso de cada año. Hasta hace poco tiempo el estadístico tenía poco que ver con estos instrumentos, y se contentaba con ponerlos a disposición de quien quisiera hacer uso de los mismos. Pero al aumentar la complejidad de los experimentos por el progreso de la investigación científica, el

instrumental estadístico alcanzó análoga complejidad y especialización. Actualmente, al investigador en determinadas zonas le es imposible familiarizarse con todas las técnicas que pueden serle útiles. Además, a mayor especialización de un instrumento, menor flexibilidad de este; muchas veces hay que modificarlo para adaptarlo a un experimento determinado, y esto requiere conocimientos muy profundos de la teoría estadística.

El empleo del instrumental estadístico no es una simple cuestión de escoger la llave que mejor se adapte a un perno; más bien se trata de elegir entre varias, todas las cuales parecen adaptarse igual de bien, sin que ninguna de ellas se ajuste exactamente al mismo. Hay mucha diferencia entre una fórmula algebraica y, digamos, un experimento de nutrición con cerdos. No hay en la fórmula nada de tipo mágico; se trata simplemente de un instrumento, obtenido además a partir de un simple modelo matemático que, probablemente, no representará con gran precisión la situación verdadera. Para emplear dicho instrumento hay que hacer toda una serie de juicios relativos a la naturaleza y magnitud de los diversos errores engendrados por discrepancia entre el modelo y el experimento efectivo. Estos juicios no pueden hacerse por el estadístico o el experimentador, pues dependen a la vez de la naturaleza de la teoría estadística y de la del material experimentado.

Para resolver esta dificultad sale a escena el estadístico aplicado. Tiene su campo de acción en diversos centros de investigación académica e industrial, y su función es, desde luego, colaborar con los investigadores en sus experimentaciones y estudios. Debe estar muy familiarizado, tanto con la teoría como con la metodología de la estadística, aunque su trabajo no pertenezca al campo de esta ciencia, sino al de la aplicación de que se trate. Lo que nos interesa subrayar es que la estadística aplicada se ha desarrollado hasta tal punto que puede considerarse que constituye ya un campo de interés especial.

1-2. Objeto y amplitud de este libro.—Se ocupa este libro de la teoría, más que de las aplicaciones de la estadística. En su desarrollo se deducirán y analizarán diversos instrumentos. Un segundo propósito de esta obra es poner en claro las condiciones en que deben emplearse determinadas técnicas estadísticas importantes; pero nuestro propósito principal es la exposición de la teoría estadística.

El libro es una introducción en el sentido de que no supone conocimientos previos de estadística en el lector. Y es elemental por no presuponer más conocimientos matemáticos que los del cálculo elemental. Sin embargo, es deseable, aunque no esencial, alguna familiaridad con la aritmética de matrices.

Tal restricción del nivel matemático es necesariamente costosa. Habremos de omitir, p. ej., muchas cuestiones interesantes, pero de carácter más técnico; habrá que reducir la generalidad de los teoremas; será necesario de cuando en cuando prescindir de demostraciones; a veces se sacrificará el rigor matemático, y tendremos que usar en ocasiones razonamientos tediosos, prescindiendo de otros más directos, pero que requieren un nivel matemático más elevado. Sin embargo, estos sacrificios afectarán a nuestra obra menos de lo que pudiera suponerse. Los aspectos esenciales de la teoría son del todo comprensibles sin necesidad de matemáticas superiores.

Puesto que la teoría estadística se funda en la teoría de la probabilidad, empezaremos este estudio dando algunos conceptos y teoremas del cálculo de probabilidades que necesitaremos más adelante. A continuación, consideraremos algunos modelos matemáticos cuya aproximación a muchas situaciones experimentales corrientes ha sido puesta de manifiesto por la experiencia. Después será posible el estudio matemático de problemas de inferencia estadística, y de diseño y análisis de experimentos e investigaciones.

1-3. Sistema de referencia.—Los capítulos van divididos en secciones numeradas; la numeración empieza de nuevo en cada capítulo. Los teoremas, definiciones, ejemplos, etc., se numeran también por capítulos. Así, Sec. 5-3 indica la sección 3 del capítulo 5; teorema 5-1 indica el teorema 1 del capítulo 5, etc.

Las ecuaciones se numeran de nuevo en cada sección y los números de las ecuaciones se encierran siempre entre paréntesis. Al referirse a una ecuación de la misma sección se da solo el número de la ecuación; en caso contrario, se expresan primero los números del capítulo y sección. Así, ecuación (6) indica la sexta ecuación de la misma sección, y ecuación (9-1-12) la duodécima ecuación de la primera sección del capítulo 9.

Los números entre corchetes indican las referencias numeradas en la bibliografía dada al final de cada capítulo.

BIBLIOGRAFIA

1. ARROW, Kenneth J.: «Alternative approaches to the theory of choice in risktaking situations», *Econometrica*, vol. 19 (1951), págs. 404-437.
2. CHURCHMAN, C. West.: *Theory of Experimental Inference*, The Macmillan Company, Nueva York, 1948.
3. FISHER, R. A.: *Statistical Methods and Scientific Inference*, Hafner Publishing Company, Nueva York, 1956.
4. GOOD, I. J.: *Probability and the Weighing of Evidence*, Charles Griffin & Co., Ltd., Londres, y Hafner Publishing Company, Nueva York, 1950.

5. JEFFREYS, Harold: *Scientific Inference*, Cambridge University Press, Londres, 1957.
6. KOLMOGOROFF, A. N.: *Foundations of the Theory of Probability*, Chelsea Publishing Company, Nueva York, 1950.
7. LINDLEY, D. V.: «Statistical inference», *Journal of the Royal Statistical Society*, Serie B, vol. 15 (1953), págs. 30-76.
8. NEYMAN, Jerzy: «Outline of a theory of statistical estimation based on the classical theory of probability», *Philosophical Transactions of the Royal Society of London*, Serie A, vol. 236 (1937), págs. 333-380.
9. SAVAGE, Leonard J.: *The Foundations of Statistics*, John Wiley & Sons, Inc., Nueva York, 1954.

CAPITULO 2

PROBABILIDAD

2-1. Introducción.—Uno de los instrumentos fundamentales de la estadística es la probabilidad, que tuvo sus orígenes en los juegos de azar, en el siglo XVII.

Los juegos de azar, como implica su nombre, incluyen acciones tales como girar la rueda de una ruleta, lanzar dados, tirar al aire una moneda, extraer una carta, etc., en las cuales el resultado de una prueba es incierto. Sin embargo, es sabido que, aun cuando el resultado de una prueba en particular sea incierto, existe un resultado *que se puede predecir* a largo plazo. Se sabe, p. ej., que en muchas tiradas de una moneda ideal (equilibrada, simétrica), aproximadamente en la mitad de las pruebas se obtiene cara. Es una regularidad que puede predecirse a largo plazo y que permite hacer negocio a las casas de juego.

En la ciencia experimental se presenta también un tipo similar de incertidumbre y regularidad a largo plazo. Así, p. ej., en genética es incierto saber si un descendiente será macho o hembra, pero en un plazo largo se conoce aproximadamente el porcentaje de descendientes que serán machos y el de aquellos que serán hembras. Una compañía de seguros de vida no puede predecir qué personas de un país morirán a la edad de cincuenta años, pero sí puede predecir bastante satisfactoriamente *cuántas* personas de ese país morirán a esta edad.

Examinaremos en primer lugar la teoría clásica de la probabilidad, o sea de la probabilidad *a priori*; luego se expondrá la teoría frecuencial y, finalmente, desarrollaremos un modelo axiomático; este es el orden del desarrollo histórico de la teoría.

2-2. Probabilidad clásica o *a priori*.—Como se ha dicho en la sección anterior, en sus principios la teoría de la probabilidad estuvo íntimamente relacionada con los juegos de azar. Esta relación sugirió la definición clásica. Así, p. ej., supongamos que queremos hallar la probabilidad del suceso obtener cara al lanzar una moneda ideal. Razonamos de esta forma: Puesto que solo existen dos resultados, cara o cruz, y dado que la moneda está bien equilibrada, cabe esperar obtener cara y cruz con la misma frecuencia, aproximadamente; por tanto, en un gran número de pruebas, es de esperar que se obtendrá cara alrededor de la mitad de

las veces, y así, la probabilidad del suceso obtener cara estará dada por el valor $\frac{1}{2}$. Esta clase de razonamiento originó la siguiente definición clásica de probabilidad:

Definición 2-1.—*Definición clásica de probabilidad.* Si un suceso puede ocurrir de n maneras mutuamente excluyentes e igualmente verosímiles y si n_A de estas poseen un atributo A , la probabilidad de A es la fracción n_A/n .

Aplicaremos esta definición a algunos ejemplos sencillos, como ilustración de su significado.

Si se lanza un dado ordinario, hay seis resultados posibles: puede caer hacia arriba cualquiera de las seis caras numeradas. Estos seis resultados son *mutuamente excluyentes*, ya que no pueden aparecer dos o más caras simultáneamente. Si además suponemos que el dado está bien construido, los seis resultados son *igualmente verosímiles*; no hay por qué esperar una cara con preferencia a cualquier otra. Supongamos ahora que deseamos conocer la probabilidad de que el resultado de una tirada sea un número par. Tres de los seis resultados posibles tienen tal atributo. La probabilidad de que aparezca un número par al lanzar el dado es, por tanto, $\frac{3}{6}$ ó $\frac{1}{2}$. Análogamente, la probabilidad de que el resultado sea mayor que 2 es $\frac{2}{3}$.

Para dar otro ejemplo, supongamos que se saca una carta al azar de una baraja ordinaria¹. Se ve en seguida que la probabilidad de sacar espadas es $\frac{13}{52}$ ó $\frac{1}{4}$. La probabilidad de sacar un número entre 5 y 10, ambos inclusive, es $\frac{24}{52}$ ó $\frac{6}{13}$.

La aplicación de la anterior definición no siempre resulta tan inmediata como en estos casos sencillos. Examinemos cuidadosamente el sentido de «mutuamente excluyentes» y de «igualmente verosímiles». Supongamos que alguien deseara calcular la probabilidad de obtener dos caras lanzando una moneda dos veces. Podría razonarse que los resultados posibles en las dos tiradas son: dos caras, dos cruces, y una cara y una cruz; y como uno de estos resultados tiene el atributo que se desea, la probabilidad será $\frac{1}{3}$. Este razonamiento es falso, porque los tres resultados no son igualmente verosímiles o probables. El tercero puede ocurrir de dos maneras, ya que la cara puede aparecer en la primera tirada y la cruz en la segunda, o la cara en la segunda y la cruz en la primera. Por tanto, hay cuatro resultados igualmente probables: CC, CX, XC, XX. El primero de estos tiene el atributo deseado, mientras que los otros carecen de él. La probabilidad correcta es, por consiguiente, $\frac{1}{4}$. El mismo resultado se obtendría si se lanzaran dos monedas simultá-

¹ De ahora en adelante consideraremos como baraja ordinaria, siguiendo al autor, la de 52 cartas o baraja francesa, con corazones, tréboles, espadas o picas y diamantes o rombos. (N. del T.)

neamente. Supongamos ahora que alguien quisiera calcular la probabilidad de que una carta extraída de una baraja ordinaria sea un as o una espada. Al enumerar los resultados favorables podría contar 4 ases y 13 espadas, y razonar que hay 17 resultados posibles con el atributo deseado. Claro que esto es incorrecto, porque los sucesos no son mutuamente excluyentes; el hecho de que una carta sea un as no impide que sea también una espada.

Observemos que la probabilidad es siempre un número comprendido entre 0 y 1. La razón n_A/n debe ser una fracción propia, ya que el total de resultados posibles no puede ser menor que el número de resultados con un determinado atributo. Si un suceso ha de ocurrir con seguridad, su probabilidad es 1. Si es seguro que no ha de ocurrir, su probabilidad es 0. Así, la probabilidad de obtener 8 al lanzar un dado es 0. La probabilidad de que el resultado sea menor que 10 es 1.

Las probabilidades determinadas mediante la definición clásica se denominan probabilidades *a priori*. Cuando se dice que la probabilidad de obtener una cara lanzando una moneda es $1/2$, se llega a este resultado por puro razonamiento deductivo. El resultado no requiere el lanzamiento de moneda alguna, ni siquiera disponer de ella. Decimos que si la moneda está bien construida, la probabilidad de obtener cara es $1/2$; pero esto es poco más que decir una misma cosa de dos maneras distintas. Nada se dice de cómo puede determinarse si una moneda en particular está bien construida.

No debe preocuparnos el hecho de que al desarrollar la teoría de la probabilidad hayamos de tratar de objetos ideales, porque esta condición es común a todos los sistemas matemáticos. La geometría, p. ej., trata conceptualmente de círculos perfectos, líneas de anchura cero, etc., y es, sin embargo, una rama útil del conocimiento que puede aplicarse a diversos problemas prácticos.

Existen varios inconvenientes en esta manera clásica, *a priori*, de abordar el problema. Es obvio que la definición de probabilidad deberá modificarse de algún modo cuando el total de resultados posibles sea infinito.

Podría buscarse, p. ej., la probabilidad de que un número natural extraído al azar sea par. La respuesta intuitiva a esta cuestión es $1/2$. Si hubiera de justificarse este resultado basándose en la definición, podría razonarse del siguiente modo: supongamos que solo se consideran los 20 primeros números naturales; como 10 de estos son pares, la razón de sucesos favorables al total de posibles es $10/20$ ó $1/2$. Si consideramos los 200 primeros, 100 de estos son pares y la razón es también $1/2$. En general, los $2N$ primeros números naturales contienen N números pares; si formamos la razón $N/2N$ y hacemos tender N a infinito, de modo que comprenda todo el conjunto de los números naturales, la razón sigue siendo $1/2$.

El argumento anterior es plausible y también la respuesta, pero su justificación rigurosa no es cosa sencilla. Observemos que el razonamiento depende de la ordenación natural de los números enteros y positivos, y una ordenación distinta podría dar lugar a un resultado diferente; p. ej., podrían ordenarse los números naturales de este modo: 1, 3, 2; 5, 7, 4; 9, 11, 6; ..., tomando la primera pareja de números impares, seguida del primer número par; la segunda pareja de números impares, seguida del segundo número par, y así sucesivamente. Con esta ordenación podría decirse que la probabilidad de obtener un número par es $\frac{1}{3}$. También pueden ordenarse los números naturales de modo que la razón n/N aumente y disminuya, oscilando sin tender a valor alguno, cuando N crezca.

La definición clásica de probabilidad suscita otra dificultad más grave aún que la que se presenta en el caso de un número infinito de resultados posibles. Supongamos una moneda de la que sabemos que tiene un sesgo a favor de las caras (esto es, una distribución tal de masa que hace más probable que aparezca cara que cruz). Los dos resultados posibles al lanzar la moneda no son igualmente probables: ¿Cuál es la probabilidad de obtener cara? La definición clásica nos deja sin posible respuesta.

Nos encontramos aún con otra dificultad de la definición clásica cuando queremos responder a preguntas tales como la siguiente: ¿cuál es la probabilidad de que un niño nacido en Chicago sea varón?, o ¿cuál es la probabilidad de que un varón muera antes de los cincuenta años?, o ¿cuál es la probabilidad de que una torta para té comprada en cierta pastelería contenga al menos tres cacahuetes?, o ¿cuál es la probabilidad de que una lámpara luzca al menos durante cien horas? Deseamos que todos estos problemas tengan respuesta dentro del marco de la teoría de la probabilidad. Sin embargo, las cuestiones de «simetría», «igualmente verosímiles», etc., no pueden considerarse como lo serían en un juego de azar. Por tanto, tendremos que alterar o extender nuestra definición para incluir problemas análogos a los anteriores en la estructura de la teoría. Esta probabilidad, aplicable más extensamente, se llama probabilidad *a posteriori* o probabilidad frecuencial y será analizada en la sección siguiente.

2-3. Probabilidad *a posteriori* o frecuencial.—Una moneda, que suponemos bien equilibrada y simétrica, fue lanzada 100 veces; los resultados se recogen en la tabla 2-1. Un hecho importante que debe observarse es que la frecuencia relativa de caras tiende a estabilizarse en torno al valor $\frac{1}{2}$. Esto no es sorprendente, ya que la moneda es simétrica y de antemano sabíamos que, en una larga serie de tiradas, se obtendrían aproximadamente tantas

caras como cruces. En otro ejemplo, un único dado fue lanzado 300 veces, recogién dose los resultados en la tabla 2-2. Obsérvese cómo la frecuencia relativa de obtener un uno se aproxima a $\frac{1}{6}$; análogamente para un dos, un tres, un cuatro, un cinco y un seis. Estos resultados no son inesperados puesto que el dado que se empleó era suficientemente simétrico y equilibrado; se esperaba que, en una larga serie de pruebas, cada cara del dado apareciera con, aproximadamente, la misma frecuencia. Esto sugiere que la frecuencia relativa de la tabla 2-1 podría utilizarse como una aproximación de la probabilidad de obtener cara, con la moneda empleada, o cabría utilizar las frecuencias relativas de la tabla 2-2 como aproximaciones de las probabilidades de que aparezcan los diferentes números con ese dado.

En el experimento de la moneda es razonable suponer que existe un número, que designaremos con p , que es la probabilidad de obtener cara. Si la moneda parece verdaderamente bien equilibrada y simétrica, podemos emplear la definición 2-1 y establecer que p es aproximadamente igual a $\frac{1}{2}$. Decir que p es igual a $\frac{1}{2}$ es solo una aproximación, puesto que para esta moneda particular no podemos estar seguros de que los dos resultados, cara y cruz, sean con exactitud igualmente verosímiles. Pero comprobados el equilibrio y la simetría de la moneda, parece bastante razonable suponer que lo son. Alternativamente, podría lanzarse la moneda un gran número de veces, anotando los resultados como en la tabla 2-1 y utilizar la frecuencia relativa de una cara como una apro-

TABLA 2-1.—RESULTADOS DEL LANZAMIENTO DE UNA MONEDA
100 VECES

Resultado	Frecuencia	Frecuencia relativa observada	Frecuencia relativa esperada en una larga serie de pruebas con una moneda equilibrada
C	56	0,56	0,50
X	44	0,44	0,50
Total	100	1,00	1,00

ximación de p . En el experimento del dado, podría aproximarse la probabilidad p_2 de obtener un dos utilizando la definición 2-1 o la frecuencia relativa de la tabla 2-2. Lo importante es que postulamos la existencia de un número p que se define como la probabilidad de obtener cara con la moneda, o un número p_2 que es la probabilidad de obtener un dos al lanzar el dado. En los ejemplos

TABLA 2-2.—RESULTADOS DEL LANZAMIENTO DE UN DADO
300 VECES

Resultado	Frecuencia	Frecuencia relativa	Frecuencia relativa es- perada en una larga serie de pruebas con un dado equilibrado
1	51	0,170	0,1667
2	54	0,180	0,1667
3	48	0,160	0,1667
4	51	0,170	0,1667
5	49	0,163	0,1667
6	47	0,157	0,1667
Total	300	1,000	1,0000

citados parece poco importante el que utilicemos la definición 2-1 o la frecuencia relativa para hallar la probabilidad p .

Supongamos, como se dijo anteriormente, que la moneda está desequilibrada, de tal forma que después de un examen estamos completamente seguros de que los dos sucesos, cara y cruz, *no* son igualmente verosímiles. Aun en estos casos puede postularse la existencia de un número p como probabilidad de obtener cara, pero para hallar el valor de p no podremos aplicar la definición clásica. Tendremos que utilizar la teoría frecuencial.

En muchas investigaciones científicas se realizan observaciones que tienen un elemento de incertidumbre o que no pueden predecirse. Como ejemplo muy simple, supongamos que se desea predecir si el próximo niño que nazca en cierta localidad será varón o hembra. Este suceso individual es incierto, pero los resultados de grupos de nacimientos pueden ser tratados satisfactoriamente. Observamos que existe cierta regularidad en una gran serie de observaciones, semejante a la regularidad de la razón frecuencial de una cara cuando lanzábamos una moneda. Si, p. ej., al examinar los registros observamos que un 51 por 100 de los nacidos son varones, es razonable postular que la probabilidad de que nazca un varón en esa localidad es igual a un número p , y tomar, como aproximación de él, 0,51. Este método de definición se denomina a veces probabilidad estadística.

Para hacer más concreta esta idea, supongamos que pueden hacerse observaciones (o experimentos) bajo condiciones completamente uniformes. Es decir, hecha la observación, se repite el suceso en condiciones análogas y se hace otra observación; esto se repite muchas veces y, aunque las condiciones sean siempre si-

milares, existe una variación incontrolable que es «casual» o «aleatoria», de tal forma que no es posible predecir el resultado de las observaciones individualmente. En muchos de estos casos, las observaciones caen dentro de ciertas clases, en las que las frecuencias relativas son bastante estables. Esto sugiere que postulemos un número p , llamado probabilidad del suceso, y aproximar p por la frecuencia relativa con que aparece dicho suceso en las repetidas observaciones. Así, p. ej., supongamos que el experimento consiste en muestrear la población de una gran ciudad para ver cuántos votantes se pronunciarán a favor de cierto candidato. Los resultados son «a favor» o «no a favor» y no es predecible la respuesta de cada votante, pero es razonable postular un número p como probabilidad de que una respuesta dada sea «a favor». La frecuencia relativa de respuestas «a favor» puede utilizarse como valor aproximado de p .

Como otro ejemplo, imaginemos que el experimento u observación consiste en el muestreo de transistores en una gran colección de estos. Postularemos que la probabilidad de que un transistor dado sea defectuoso es p . Podemos aproximar p seleccionando «al azar» varios transistores del conjunto dado y calculando la frecuencia relativa del número de defectuosos.

Lo importante es la posibilidad de *imaginar* una serie de observaciones o experimentos realizados en condiciones bastante uniformes. Puede postularse entonces un número p como probabilidad de que ocurra el suceso A , y p puede ser aproximado por la frecuencia relativa del suceso A en una serie de experimentos.

2-4. Modelos de probabilidad.—Uno de los objetivos de la ciencia consiste en predecir y describir sucesos del mundo en que vivimos. Una manera de hacerlo es construir modelos matemáticos que describan adecuadamente el mundo real. Así, p. ej., la ecuación $s = \frac{1}{2}gt^2$ expresa cierta relación entre los símbolos s , g y t . Con el fin de utilizar la ecuación $s = \frac{1}{2}gt^2$ en una experiencia del mundo real para predecir s , distancia recorrida por un cuerpo que cae, como función del tiempo t , tiene que conocerse la constante gravitatoria g . Esta es una constante física que debe ser medida por experimentación si se desea que la ecuación $s = \frac{1}{2}gt^2$ sea útil. La razón de haber mencionado esta ecuación es que en la teoría de la probabilidad hacemos algo muy parecido: construimos un modelo probabilístico que pueda utilizarse para describir sucesos del mundo real. Así, p. ej., puede desearse hallar una ecuación adecuada para predecir el sexo de cada nacido en cierta localidad. La ecuación sería muy compleja y no se ha descubierto ninguna. Sin embargo, puede construirse un modelo de probabilidad que, aunque no sea muy útil para tratar un suceso individual, sirva

perfectamente para ocuparse de grupos de sucesos. Cabe postular la existencia de un número p que represente la probabilidad de que un nacido sea varón. A partir de esta probabilidad fundamental, podemos responder a preguntas tales como: ¿Cuál es la probabilidad de que de diez nacidos, al menos tres sean varones?, o ¿cuál es la probabilidad de que haya tres varones consecutivos en los próximos cinco nacimientos? Para contestar a preguntas tales como estas y a muchas otras análogas, desarrollaremos un modelo idealizado de probabilidad.

Consideraremos una teoría de la probabilidad adecuada solo para aquellas situaciones que pueden ser descritas por los resultados de experimentos conceptuales. Es decir, consideraremos únicamente aquellos sucesos cuya repetición sea concebible bajo condiciones semejantes. Así pueden tratarse los nacimientos de varones, el lanzamiento de una moneda, el número de automóviles, etc.; pero no se incluyen problemas tales como ¿cuál es la probabilidad de que mi esposa me ame?, o ¿cuál es la probabilidad de que no hubiera ocurrido la segunda guerra mundial?

También necesitamos que pueda enumerarse cada posible resultado de un experimento. Así, p. ej., al lanzar una moneda existen dos resultados posibles: cara y cruz. Asociaremos probabilidades solamente con estos resultados. Añadiremos, sin embargo, que aun cuando un resultado sea imposible puede ser incluido (su probabilidad es 0). Lo fundamental es recordar que ha de incluirse cada resultado que *puede ocurrir*.

Cada resultado imaginable de un experimento conceptual, que puede repetirse bajo condiciones similares, será denominado un punto muestral, y la totalidad de los resultados imaginables (o puntos muestrales) se llamará el espacio muestral.

Antes de proceder al desarrollo de la teoría, daremos algunos ejemplos.

Ejemplo 2-1.—Si un experimento aleatorio consiste en lanzar una moneda dos veces, existen cuatro resultados imaginables: (C, C), (C, X), (X, C), (X, X). Por tanto, hay cuatro puntos muestrales que forman el espacio muestral.

Ejemplo 2-2.—Si un experimento aleatorio consiste en observar el sexo de los nacidos en cierta población, existen dos resultados imaginables: varón y hembra; por tanto, hay dos puntos muestrales en el espacio muestral.

Ejemplo 2-3.—Supongamos que se selecciona una muestra de 50 semillas de un saco, para ver cuántas germinan. El experimento aleatorio consiste en extraer 50 semillas del saco. Los resultados posibles son las cantidades que germinan de las 50. Puede haber 0, 1, 2, ..., ó 50 que germinen, por lo que existen 51 puntos muestrales que forman el espacio muestral.

Ejemplo 2-4.—Imaginemos que en una gran ciudad se seleccionan 500 personas al azar para ver cuántas consumen cierta marca de leche. El número imaginable de las que consumen tal marca de leche entre las 500 personas es 0, 1, 2, ..., 500. Cada uno de estos 501 números es un punto muestral del espacio muestral.

Ejemplo 2-5.—Supongamos que un experimento aleatorio consiste en preguntar a los espectadores de la televisión de cierta ciudad si presencian regularmente tres programas especificados. Hay ocho resultados imaginables: (S, S, S), (S, S, N), (S, N, S), (N, S, S), (S, N, N), (N, S, N), (N, N, S), (N, N, N), donde (S, N, S) significa «sí» presencia el primer programa, «no» el segundo programa y «sí» el tercer programa, etc. El espacio muestral está formado por ocho puntos.

Ejemplo 2-6.—En los ejemplos anteriores el espacio muestral está formado por un número finito de puntos. Daremos ahora un ejemplo de espacio muestral que contiene un número infinito de puntos. Supongamos que se desea determinar el número de tiradas de una moneda que deberá hacerse hasta que aparezca la primera cara. Esta puede aparecer en la tirada 1.^a, 2.^a, ..., n -ésima, ... Aquí el espacio muestral está formado por una infinidad numerable de puntos (tantos como números enteros positivos).

Ejemplo 2-7.—En este ejemplo, el espacio muestral estará formado por tantos puntos (llamado un continuo de puntos) como números reales positivos. Sea el experimento aleatorio consistente en seleccionar una muestra aleatoria de estudiantes de sexto curso en determinada ciudad y registrar su peso. El resultado puede ser cualquier número positivo. Cabe objetar que no habrá estudiantes que pesen menos de un kilogramo o más de 1000. Es cierto, pero no es objeción si se incluyen los resultados imposibles al enumerar cada resultado imaginable. Por tanto, este espacio muestral estará formado por todos los números positivos.

2-5. Conjuntos de puntos.—Definiremos ciertas operaciones sobre el conjunto de puntos que forman el espacio muestral y que son necesarias para posteriores desarrollos de la teoría. Un conjunto de puntos, llamado a veces simplemente un conjunto, es una colección de elementos que tienen ciertas propiedades específicas. Un conjunto podría ser el de los 10 primeros números naturales o una colección de automóviles o de cualesquiera otros objetos. Si s es un punto o un elemento que pertenece al conjunto S , escribiremos $s \in S$.

Definición 2-2.—Dos conjuntos S_1 y S_2 se dice que son iguales si cada elemento o punto de S_1 es también un punto de S_2 , y cada punto de S_2 es así mismo un punto de S_1 ; es decir, si S_1 y S_2 con-

tienen exactamente los mismos puntos. Indicaremos la igualdad escribiendo $S_1 = S_2$.

Definición 2-3.—Si cada elemento (o punto) de un conjunto S_1 es también un elemento de S , llamaremos a S_1 un subconjunto de S y escribiremos $S_1 \subset S$.

Ejemplo 2-8.—Sea S el conjunto de los enteros x tales que $x = 1, 2, 3, 4, 5, 6, 7$. Escribiremos $S = \{x: x = 1, 2, 3, 4, 5, 6, 7\}$. Sea $S_1 = \{y: y = 1, 2, 3\}$. Entonces S_1 es un subconjunto de S . Si $S_2 = \{z: z = 1, 3, 5, 7\}$, S_2 no es un subconjunto de S_1 , pero sí un subconjunto de S .

Definición 2-4.—En cada aplicación de la teoría, existirá un conjunto universal, el espacio muestral S , tal que todos los otros conjuntos que intervengan en el análisis son subconjuntos de S .

Algunas veces puede que no se indique explícitamente el espacio muestral, pero generalmente estará implícito en el contexto de la discusión.

Definición 2-5.—El complemento de un conjunto S_1 , respecto al espacio muestral S , será el conjunto de puntos que están en S pero no en S_1 . Se indicará por $S - S_1$ o a veces por $\bar{S}_1$.

En el ejemplo 2-8, el conjunto $\bar{S}_1$ está dado por

$$\bar{S}_1 = \{x: x = 4, 5, 6, 7\} \quad \text{y} \quad \bar{S}_2 = \{z: z = 2, 4, 6\}.$$

Definición 2-6.—Si un conjunto S_1 no contiene puntos, se denomina el conjunto nulo, y será indicado por $\emptyset$.

Adoptaremos los convenios de que el conjunto nulo es un subconjunto de todo conjunto, y de que cada conjunto es un subconjunto de sí mismo.

Después de haber enumerado los resultados de un experimento aleatorio y definido un espacio muestral S , necesitaremos considerar ciertos subconjuntos del espacio muestral. Así, p. ej., en el ejemplo 2-3, cabe preguntarse, ¿cuál es la probabilidad de que germinen más de 30 semillas? Estamos haciendo una pregunta relacionada con un subconjunto de S ; el subconjunto contiene los elementos 31, 32, ..., 50. Con nuestra terminología diremos, ¿cuál es la probabilidad de que ocurra el suceso A , donde A es el subconjunto 31, 32, ..., 50? Esto nos lleva a la siguiente definición:

Definición 2-7.—Un suceso A está definido en el espacio muestral S como un subconjunto A de puntos de S , y cuando decimos «probabilidad de que ocurra el suceso A » queremos decir probabilidad de que aparezca cualquier punto de A .

Así, p. ej., cuando preguntamos ¿cuál es la probabilidad de que germinen más de 30 semillas? estamos preguntando, en esencia, cuál es la probabilidad de que se presente cualquier punto del su-

ceso A , donde $A = \{x: x=31, 32, \dots, 50\}$; si «ocurre» cualquier punto de A , diremos que se ha producido el suceso A (germinar más de 30 semillas).

Si el espacio muestral contiene un continuo de puntos, como en el ejemplo 2-7, no definiremos todo subconjunto como un suceso, sino sólo subconjuntos *medibles*. El término medible está tomado de las matemáticas superiores y el lector no necesita preocuparse de él en este libro, ya que los subconjuntos que consideremos serán medibles y, por tanto, serán llamados sucesos.

En el ejemplo 2-7 podemos preguntar ¿cuál es la probabilidad de que el estudiante pese entre 75 Kg y 85 Kg? Con nuestra terminología preguntaríamos ¿cuál es la probabilidad del suceso A , donde $A = \{x: 75 < x < 85\}$?

En relación con el espacio muestral S , supongamos que S_1 y S_2 son dos sucesos, es decir, dos subconjuntos de S . Pueden definirse otros dos sucesos: 1) el conjunto de puntos que están en ambos, S_1 y S_2 , y 2) el conjunto de puntos que están en S_1 o en S_2 o en ambos, S_1 y S_2 . Estos conjuntos se precisan en las dos definiciones siguientes.

Definición 2-8.—Sean S_1 y S_2 dos sucesos cualesquiera del espacio muestral S ; el suceso formado por todos los puntos que están en S_1 o en S_2 , o en ambos, se llama *unión* de S_1 y S_2 y se denota por $S_1 \cup S_2$.

Ejemplo 2-9.—Consideremos de nuevo el ejemplo 2-5. Sea S_1 el suceso definido por la condición de que la respuesta para el primer programa sea «sí»; es decir, S_1 contiene los cuatro puntos (S, S, S), (S, S, N), (S, N, S), (S, N, N). Sea S_2 el suceso definido por la condición de que la respuesta para el tercer programa sea «no»; esto es, S_2 contiene los cuatro puntos (S, S, N), (S, N, N), (N, S, N), (N, N, N). Entonces, el conjunto $S_1 \cup S_2$ contiene los seis puntos (S, S, S), (S, S, N), (S, N, S), (S, N, N), (N, S, N), (N, N, N).

Definición 2-9.—Sean S_1 y S_2 dos sucesos cualesquiera del espacio muestral S ; el suceso formado por todos los puntos que están en ambos, S_1 y S_2 , se llama *intersección* de S_1 y S_2 y se denota por $S_1 \cap S_2$, o a veces por $S_1 S_2$.

Ejemplo 2-10.—Si se definen S_1 y S_2 como en el ejemplo 2-9, el suceso $S_1 \cap S_2$ contiene los dos puntos (S, S, N), (S, N, N).

De las definiciones anteriores se desprenden los resultados siguientes, donde S es el espacio muestral, y S_1 y S_2 , sucesos de S .

1. $\overline{S} = \emptyset$.
2. Si S_1 y S_2 no tienen puntos comunes (conjuntos mutuamente excluyentes), $S_1 \cap S_2 = \emptyset$.
3. $S_1 \cap S = S_1$.
4. $S_1 \cup S = S$.

$$5. S \cap \bar{S}_1 = S - S_1 = \bar{S}_1.$$

$$6. S_1 \cup S_1 = S_1.$$

$$7. S_1 \cap S_1 = S_1.$$

Existe la posibilidad de establecer otras muchas relaciones entre sucesos, pero estas bastarán para nuestros propósitos inmediatos.

Clasificaremos el espacio muestral en dos tipos, *discreto* y *continuo*. Lo hacemos con el propósito de aligerar las explicaciones de posteriores desarrollos de la teoría. En realidad, con instrumentos de las matemáticas superiores que están fuera de los objetivos de este libro, pueden tratarse ambos tipos en una teoría unificada.

Definición 2-10.—*Un espacio muestral S se dice que es discreto si contiene: 1) un número finito de puntos, o 2) un número infinito de puntos (infinito numerable) que pueden ponerse en correspondencia uno a uno con los números naturales.*

En los ejemplos 2-1 a 2-5, el espacio muestral contiene un número finito de puntos y, por tanto, es discreto. En el ejemplo 2-6 hay un número infinito de puntos, pero pueden ordenarse en una sucesión (en correspondencia uno a uno con los números naturales); por tanto, ese espacio muestral es también discreto. Sin embargo, en el ejemplo 2-7 el espacio muestral está formado por todos los números reales x , donde $x > 0$, y no es posible poner estos números en correspondencia uno a uno con los naturales. Por tanto, existen espacios muestrales que no son discretos, sino que contienen lo que se llama un continuo de puntos.

Definición 2-11.—*Un espacio muestral S se llama espacio muestral continuo, si contiene un continuo de puntos.*

Concluiremos esta sección con algunos ejemplos adicionales, y daremos luego un desarrollo axiomático de la probabilidad.

Ejemplo 2-11.—Consideremos un experimento aleatorio consistente en observar la duración de la vida de tubos electrónicos. El resultado puede ser cualquier número positivo y, por tanto, el espacio muestral es continuo.

Ejemplo 2-12.—Sea un experimento aleatorio que consiste en seleccionar al azar tres personas de entre los empleados de cierta compañía y registrar la renta anual de cada persona. El espacio muestral está formado por las ternas (x_1, x_2, x_3) , donde x_1 , x_2 y x_3 son las rentas respectivas de las tres personas, y cada una de ellas puede tomar cualquier valor mayor que cero. Definamos el suceso A por la condición de que la renta anual total de las tres personas que se muestrean exceda de 15 000 dólares. Esto puede escribirse de la forma siguiente:

$$A = \{(x_1, x_2, x_3) : x_1 > 0, x_2 > 0, x_3 > 0; x_1 + x_2 + x_3 > 15\,000\}$$

Ejemplo 2-13.—Un experimento aleatorio consiste en lanzar dos dados y observar los números que salen. El espacio muestral consta de 36 puntos, que son $(1, 1), (1, 2), \dots, (6, 6)$, donde los pares ordenados de números representan los resultados respectivos del primero y del segundo dado. Definimos el suceso A por la condición de que la suma de las dos puntuaciones sea igual a siete; tal suceso está formado por seis puntos, que son $(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)$.

Ejemplo 2-14.—En un experimento agrícola, se examina el rendimiento de cinco variedades de trigo. Las cinco variedades se desarrollan bajo condiciones uniformes. El resultado es una colección de cinco números $(y_1, y_2, y_3, y_4, y_5)$, donde y_i representa el rendimiento de la i -ésima variedad, en quintales por hectárea. El espacio muestral es continuo, ya que cada y_i puede ser cualquier número real mayor o igual que 0. Definamos el suceso A en este ejemplo por las condiciones de que y_2, y_3, y_4 e y_5 sean cada una 10 o más quintales por hectárea mayores que y_1 , variedad estándar.

Con nuestra notación, escribiremos

$$A = \{(y_1, y_2, y_3, y_4, y_5) : y_j \geq y_1 + 10; j = 2, 3, 4, 5; 0 \leq y_1\}$$

2-6. Desarrollo axiomático de la probabilidad.—En las secciones anteriores hemos dado los conceptos de probabilidad clásica y frecuencial que pueden ayudarnos a resolver importantes problemas de la ciencia experimental. Para coadyuvar a la solución de estos problemas, desarrollaremos una teoría matemática de la probabilidad y mostraremos luego cómo puede utilizarse este modelo idealizado en los problemas del mundo real.

En primer lugar enunciaremos los axiomas que se emplearán para desarrollar la teoría.

Sea S un espacio muestral, y A , cualquier suceso de S ; es decir, A es cualquier subconjunto de S . Diremos que P es una *función de probabilidad* en el espacio muestral S si se satisfacen los tres axiomas siguientes.

Axioma 1. $P(A)$ es un número real tal que $P(A) \geq 0$ para todo suceso A de S .

Axioma 2. $P(S) = 1$.

Axioma 3. Si $S_1, S_2, \dots$ es una sucesión de sucesos mutuamente excluyentes de S , es decir, si $S_i \cap S_j = \emptyset$ para $i \neq j = 1, 2, \dots$,

$$P(S_1 \cup S_2 \cup \dots) = P(S_1) + P(S_2) + \dots$$

Estos axiomas, que se utilizarán para desarrollar un modelo idealizado, están motivados por las definiciones de probabilidad clásica y frecuencial. Demostraremos ahora algunos teoremas que son resultados directos de los axiomas.

Teorema 2-1.—Sea S un espacio muestral y P una función de probabilidad en S ; la probabilidad de que no ocurra el suceso A es $1 - P(A)$. Con la notación de los conjuntos de puntos, esto se escribe en la forma $P(\bar{A}) = 1 - P(A)$.

Demostración.—En virtud de la definición 2-6, $A \cap \bar{A} = \emptyset$; también $A \cup \bar{A} = S$, y, por el axioma 2, se tiene $1 = P(S) = P(A \cup \bar{A})$. Por el axioma 3, $1 = P(S) = P(A \cup \bar{A}) = P(A) + P(\bar{A})$, con lo que queda demostrado.

Teorema 2-2.—Sea S un espacio muestral con función de probabilidad P ; en tal caso, $0 \leq P(A) \leq 1$ para cualquier suceso A de S .

Demostración.—Por el axioma 1, $P(A) \geq 0$, con lo que solo será necesario demostrar que $P(A) \leq 1$. Por el teorema 2-1, $P(A) + P(\bar{A}) = 1$; pero $P(\bar{A}) \geq 0$ por el axioma 1, luego $P(A) = 1 - P(\bar{A}) \leq 1$.

Teorema 2-3.—Sea S un espacio muestral con una función de probabilidad P . Si S_0 es el conjunto nulo, $P(S_0) = 0$.

Demostración.—De los resultados de la sección 2-5, se deduce que $\bar{S} = S_0 = \emptyset$. Por el axioma 3, $P(S \cup \bar{S}) = P(S) + P(\bar{S}) = P(S) + P(S_0)$. Pero $S \cup \bar{S} = S$ y $P(S) = 1$, luego $P(S_0) = 0$.

Si estos axiomas y los teoremas de ellos resultantes han de ayudarnos a desarrollar un modelo útil, debemos tener una regla o función que nos permita calcular la probabilidad de cada suceso A (subconjunto) del espacio muestral S . Explicaremos cómo se construye tal función en las próximas secciones. Lo haremos para tres espacios muestrales diferentes: 1) un espacio muestral discreto con un número finito de puntos, donde cada uno de ellos tiene la misma probabilidad; 2) un espacio muestral discreto general, que será estudiado en el capítulo 3; 3) un espacio muestral continuo, que estudiaremos en el capítulo 4.

2-7. Espacio muestral discreto con un número finito de puntos.—En ciertos tipos de problemas, entre los cuales los juegos de azar constituyen ejemplos notables, el espacio muestral contiene un número finito n de puntos, y la probabilidad asignada a cada punto es $1/n$.

En otras palabras, en ciertos problemas existe un número finito de ordenaciones (n) y es totalmente realista suponer que la probabilidad de cada ordenación es $1/n$. En general, es suficiente para estos problemas la definición clásica, y pueden emplearse los métodos combinatorios para la enumeración de las ordenaciones. Veremos cómo este espacio muestral especial (número finito de puntos con igual probabilidad para cada uno de ellos) encaja en la teoría general y a continuación indicaremos varios métodos que pueden utilizarse para resolver estos problemas.

Definición 2-12.—Sean $s_1, s_2, \dots, s_n$ los n puntos muestrales de un espacio muestral discreto S ; se dice que la función P es una función de probabilidad de sucesos igualmente verosímiles si satisface las condiciones siguientes:

a) $P(s_1) = P(s_2) = \dots = P(s_n) = 1/n$.

b) Si es A un suceso que contiene n_A cualesquiera de los puntos muestrales s_i , entonces $P(A) = n_A/n$.

La condición a) afirma que cada uno de los n puntos es igualmente verosímil y, por tanto, que su probabilidad es $1/n$. La condición b) establece que la probabilidad de un suceso que contenga n_A de los n puntos muestrales es n_A/n . Es fácil comprobar que esta función satisface los axiomas 1, 2 y 3 y es, por tanto, una función de probabilidad. La función de probabilidad de sucesos igualmente verosímiles es exactamente lo mismo que el concepto dado en la definición 2-1.

Ejemplo 2-15.—Supongamos que un experimento aleatorio consiste en lanzar una moneda simétrica equilibrada dos veces. El espacio muestral S (resultados) está formado por cuatro puntos: $(C, C) = s_1, (C, X) = s_2, (X, C) = s_3, (X, X) = s_4$, donde (C, X) significa cara en la primera tirada y cruz en la segunda, etc. Parece totalmente razonable asignar a cada punto muestral la probabilidad $1/4$. Imaginemos que A es el suceso «el resultado de la primera tirada es cara»; entonces $A = s_1 \cup s_2$. Este suceso (subconjunto) contiene dos puntos, de donde $P(A) = 2/4 = 1/2$. Definamos el suceso B por la condición de que aparezca al menos una cara; entonces $B = s_1 \cup s_2 \cup s_3$. B contiene tres puntos, luego $P(B) = 3/4$. Supongamos ahora que se desea hallar la probabilidad de que no ocurra B , es decir, $P(\bar{B})$. Por la definición 2-5, $s_4 = \bar{B}$, y por el teorema 2-1, $1 - P(B) = P(\bar{B}) = P(s_4) = 1/4$; luego $P(\bar{B}) = 1/4$.

2-8. Permutaciones y combinaciones.—Para calcular la probabilidad de un suceso A cuando es aplicable la definición 2-1 o su equivalente 2-12 (supondremos aplicables estas definiciones en las Secs. 2-8 a 2-13), necesitamos calcular el número total n de ordenaciones mutuamente excluyentes e igualmente verosímiles, y el número n_A con el atributo A . Esta enumeración se facilita a menudo mediante ciertas fórmulas combinatorias que desarrollaremos a continuación y que se basan en los dos principios fundamentales siguientes:

a) Si un suceso A puede ocurrir de m maneras, y un suceso diferente B puede ocurrir de n maneras, el suceso A o B puede ocurrir de $m + n$ maneras, siempre que A y B no puedan ocurrir simultáneamente.

b) Si un suceso A puede ocurrir de m maneras, y un suceso diferente B puede ocurrir de n maneras, el suceso A y B puede ocurrir de mn maneras.

Estas dos ideas pueden ilustrarse haciendo corresponder A a la extracción de una espada de una baraja y B a la extracción de un corazón. Cada uno de estos sucesos puede realizarse de 13 maneras. El número de maneras en que puede sacarse un corazón o una espada es, evidentemente,

$$13 + 13 = 26$$

Para aclarar el segundo principio supongamos dos cartas extraídas de la baraja, de modo que una sea una espada y la otra un corazón. Hay $13 \times 13 = 169$ maneras de hacer esto, ya que con el as de espadas podemos poner cualquiera de los 13 corazones; con el rey de espadas, también cualquiera de los 13 corazones, y así sucesivamente para las demás espadas, hasta 13.

Estos dos principios pueden, evidentemente, generalizarse teniendo en cuenta más de dos sucesos. Así, si tres sucesos A , B y C , mutuamente excluyentes, pueden ocurrir de m , n y p maneras, respectivamente, el suceso A o B o C puede ocurrir de $m + n + p$ maneras, y el suceso A y B y C puede ocurrir de mnp maneras.

Usaremos ahora el segundo de estos principios para contar el número de disposiciones de un conjunto de objetos. Consideremos el número de disposiciones posibles con las letras a , b y c . Podemos tomar cualquiera de las tres y colocarla la primera; cualquiera de las otras dos podrá colocarse la segunda, y la tercera posición la ocupará la letra restante. La ocupación de la primera posición es un suceso que puede ocurrir de tres maneras; la de la segunda, de dos, y la de la tercera, de una. Luego los tres sucesos juntos pueden presentarse de $3 \times 2 \times 1 = 6$ maneras. Las seis disposiciones, denominadas *permutaciones*, son

$$abc, acb, bac, bca, cab, cba$$

En este ejemplo apenas merecía emplear el método del recuento, ya que es suficientemente sencillo escribir las seis permutaciones. Pero si quisiéramos buscar el número de permutaciones de seis letras, habríamos tenido que escribir

$$6 \times 5 \times 4 \times 3 \times 2 \times 1 = 720$$

permutaciones.

Es evidente que, en general, el número de permutaciones de n objetos diferentes es

$$n(n-1)(n-2)(n-3) \dots (2)(1)$$

Este producto de un número natural por todos los números naturales menores que él, se denota corrientemente con el símbolo $n!$ (léase factorial de n). Así, $2! = 2$, $3! = 6$, $4! = 24$, $5! = 120$, etc. Puesto que

$$(n-1)! = \frac{n!}{n}$$

es corriente definir $0!$ como 1, de modo que la relación sigue aplicándose cuando $n=1$.

Vamos a contar ahora el número de permutaciones que pueden hacerse con n objetos cuando sólo se emplean r objetos en una permutación dada. Por el razonamiento anterior, la primera posición puede ocuparse de n maneras; la segunda, de $n-1$ maneras, y así sucesivamente; al llegar a la r -ésima posición, habremos usado $r-1$ de los objetos, de modo que nos quedarán $n-(r-1)$, entre los cuales podremos elegir. El número de permutaciones de n objetos tomando r cada vez es, por tanto, $n(n-1)(n-2) \dots (n-r+1)$. Este número se designa con el símbolo $P_{n,r}$:

$$P_{n,r} = n(n-1)(n-2) \dots (n-r+1) = \frac{n!}{(n-r)!} \quad (1)$$

Así, el número de permutaciones de las letras a, b, c, d , tomando dos cada vez, es $P_{4,2} = 4 \times 3 = 12$. Haciendo $r=n$ en la ecuación (1), obtenemos el resultado enunciado anteriormente: el número de permutaciones de n objetos tomados de n en n es $n!$.

Con ayuda de la ecuación (1) podemos resolver ahora el siguiente problema: ¿De cuántas maneras distintas pueden elegirse r objetos de entre n objetos? $P_{n,r}$ representa el número de selecciones posibles, así como todas las disposiciones de cada selección o *combinación*. Dos combinaciones son distintas si no están formadas por el mismo conjunto de objetos. Así, abc y abd son combinaciones diferentes de tres letras, mientras que abc y bca son permutaciones diferentes de la misma combinación. Representemos por el símbolo $\binom{n}{r}$ el número de combinaciones diferentes. Es evidente que $P_{n,r}$ es igual a $\binom{n}{r}$ veces $r!$, ya que cada combinación de r objetos tiene $r!$ disposiciones. Por tanto,

$$\binom{n}{r} = \frac{P_{n,r}}{r!} = \frac{n(n-1)(n-2) \dots (n-r+1)}{r!} = \frac{n!}{r!(n-r)!} \quad (2)$$

Otro símbolo corriente para este número es $C_{n,r}$ pero no lo usaremos en este texto. El número de combinaciones de cinco objetos

tomados tres a tres es

$$\binom{5}{3} = \frac{5 \times 4 \times 3}{3!} = \frac{60}{6} = 10$$

Puede darse a $\binom{n}{r}$ una interpretación diferente. Es el número de maneras en que pueden dividirse n objetos en dos grupos, uno de ellos de r objetos, y el otro grupo, de $n-r$ objetos. Supongamos ahora que queremos dividir n objetos en tres grupos que contengan n_1, n_2, n_3 objetos, respectivamente, siendo

$$n_1 + n_2 + n_3 = n$$

Empezaremos por dividirlos en dos grupos que contengan n_1 y $n_2 + n_3$ objetos. Esto puede hacerse de $\binom{n}{n_1}$ maneras. Entonces podemos dividir el segundo grupo en otros dos grupos que contengan n_2 y n_3 objetos, lo que puede hacerse de $\binom{n_2 + n_3}{n_2}$ maneras. Mediante el segundo principio de enumeración, el número total de maneras de hacer a la vez ambas divisiones es

$$\binom{n}{n_1} \binom{n_2 + n_3}{n_2} = \frac{n!}{n_1!(n_2 + n_3)!} \frac{(n_2 + n_3)!}{n_2!n_3!} = \frac{n!}{n_1!n_2!n_3!}$$

Este tipo de razonamiento puede extenderse para hallar el número de maneras de dividir n objetos en k grupos que contengan $n_1, n_2, \dots, n_k$ objetos, siendo $n_1 + n_2 + \dots + n_k = n$. Se halla en seguida que este número es

$$\frac{n!}{n_1!n_2! \dots n_k!} \quad (3)$$

Por tanto, el número de maneras de dividir cuatro objetos en tres grupos que contengan 1, 1 y 2 objetos es

$$\frac{4!}{1!1!2!} = 12$$

La expresión (3) tiene una segunda interpretación. Es el número de permutaciones diferentes de n objetos cuando n_1 de los objetos son iguales entre sí y de una clase, n_2 iguales y de otra clase, y así sucesivamente. Si nos referimos al ejemplo numérico anterior, hay 12 permutaciones de las letras a, b, c, c . Para ver que la expre-

sión (3) da el número correcto, consideremos n objetos diferentes (p. ej., las letras $a, b, c, \dots, p$) dispuestos en un orden dado. Consideremos ahora una división de este conjunto de objetos en k grupos, de los cuales el primero contiene n_1 objetos; el segundo, n_2 , y así sucesivamente. Sustituyamos ahora en la disposición original de los objetos todos los elegidos para el primer grupo por unos, todos los elegidos para el segundo grupo por doses, y así sucesivamente. El resultado será una permutación de n_1 unos, n_2 doses, ..., n_k símbolos k . Reflexionando un poco, nos convenceremos de que cada división de las letras en los k grupos corresponde a una permutación diferente de los números considerados y que este es el conjunto total de permutaciones, pues si hubiera otro habría otra división de las letras en k grupos.

Hemos obtenido tres fórmulas en esta sección, no solo por su utilidad, sino porque su obtención sirve para ilustrar la aplicación de los dos principios de recuento que dimos al principio de la sección. Lo más importante es considerar los métodos; las fórmulas servirán para resolver muchos problemas, pero resultan inútiles en muchos otros, y entonces hay que recurrir a los principios elementales.

Ejemplo 2-16.—Si se extraen dos cartas de una baraja ordinaria, ¿cuál es la probabilidad de que una sea espada y otra corazón?

Puesto que nada se dice sobre el orden de presentación de la espada y el corazón, se trata de un problema de combinaciones. Para calcular la probabilidad, hemos de hallar el total de resultados posibles en una extracción de dos cartas, y después, el número de esos resultados que tienen el atributo en cuestión. El total de combinaciones de dos cartas que pueden hacerse con 52 cartas es $\binom{52}{2} = 1326 = n$. Ya hemos visto antes que hay $13 \times 13 = 169 = n_A$ combinaciones distintas con el atributo requerido. Por tanto,

$$P(A) = \frac{n_A}{n} = \frac{169}{1326} = \frac{13}{102}$$

Este problema puede resolverse también considerando el número de permutaciones posibles; el denominador de la fracción sería entonces $P_{52,2} = 2652$. Para obtener el numerador basta considerar que cada una de las 169 combinaciones de dos cartas tiene dos permutaciones, lo que da $2 \times 169 = 338$ para el número de permutaciones que tienen el atributo requerido. O también podemos empezar del siguiente modo: El número de permutaciones donde aparece primero una espada y después un corazón es $13 \times 13 = 169$, según el principio b), y el número de aquellas en las que se presenta el corazón primero y la espada después es el mismo. Cualquiera

de estos grupos de permutaciones satisface nuestro requerimiento, y según el principio *a*), el número requerido será $169 + 169 = 338$; por tanto, volvemos a obtener para la probabilidad el valor $13/102$.

Ejemplo 2-17.—¿Cuál es la probabilidad de obtener 3 o más espadas en una extracción de 4 cartas de una baraja ordinaria?

Se trata también, en este caso, de combinaciones. El total de combinaciones posibles con cuatro cartas es $\binom{52}{4} = 270\,725$. En consecuencia, el espacio muestral S está formado por $270\,725 = n$ puntos, y la probabilidad asignada a cada punto es $1/n = 1/270\,725$. Para hallar el numerador n_A , consideremos lo siguiente: La condición, al menos tres espadas, significa tres o cuatro espadas. El número de grupos de 4 cartas que contienen exactamente 3 espadas es $\binom{13}{3} \times 39 = 11\,154$; el primer factor es el número de combinaciones de las 13 espadas de 3 en 3, y el segundo, el número de maneras en que puede elegirse una carta de los otros tres palos; se toma el producto, de acuerdo con el principio *b*). El número de combinaciones que pueden obtenerse con las 13 espadas de 4 en 4 es $\binom{13}{4} = 715$. Por el principio *a*), el número de grupos que poseen el atributo en cuestión es $11\,154 + 715 = 11\,869 = n_A$. La probabilidad pedida es

$$P(A) = n_A/n = 11\,869/270\,725.$$

También cabe hallar el numerador por el siguiente método: El número de combinaciones con las 13 espadas de 3 en 3 es $\binom{13}{3} = 286$. La cuarta carta puede ocurrir que sea o no espada, y si las 3 primeras son espadas, habrá que elegir esa cuarta carta a partir del grupo de las 49 restantes. Por consiguiente, el número de grupos requerido será $49 \times 286 = 14\,014$. Este razonamiento no es válido, porque cuenta más de una vez los grupos de 4 espadas. Una combinación específica de espadas es as-rey-reina, y extrayendo la sota de espadas de entre las restantes 49, tenemos la combinación as-rey-reina-sota. Pero contamos a la vez la combinación que considera as-reina-sota, extrayendo el rey de las 49 cartas restantes. Es evidente que se han contado 4 veces los grupos de 4 espadas. Puede obtenerse el resultado correcto restando 3 veces el número de grupos con 4 espadas. El resultado es

$$14\,014 - 3 \binom{13}{4} = 11\,869$$

igual que anteriormente.

Ejemplo 2-18.—Se echan 7 bolas en 4 cajas numeradas, de modo que cada bola tenga que caer en una caja y todas tengan igual probabilidad de caer en cualquiera de las 4 cajas. ¿Cuál es la probabilidad de que en la primera caja caigan precisamente 2 bolas?

Puesto que la primera bola puede caer en cualquiera de las 4 cajas, la segunda lo mismo, etc., el total de resultados posibles es, por el principio *b*), 4^7 . Para enumerar cuántos resultados habrá con el atributo deseado, empezamos por dividir las 7 bolas en dos grupos, uno de los cuales contenga 2 bolas, y el otro, 5 bolas. Esto puede hacerse de $\binom{7}{2}$ maneras. Ahora pondremos el grupo de dos en la primera caja y distribuiremos las otras cinco entre las 3 cajas restantes. Esto puede hacerse, por el mismo razonamiento anterior, de 3^5 maneras. El número de resultados favorables es, por tanto, $\binom{7}{2} 3^5$, y la probabilidad deseada es

$$P(A) = \frac{n_A}{n} = \frac{\binom{7}{2} 3^5}{4^7} = \frac{7 \times 3^5}{4^7} \cong 0,3115$$

(El símbolo $\cong$ se emplea para denotar igualdad aproximada.)

2.9. Fórmula de Stirling.—Al hallar valores numéricos de las probabilidades nos encontramos con la necesidad de evaluar largas expresiones factoriales que son de cálculo laborioso por multiplicación directa. Si se dispone de una máquina de sumar y no hay un gran número de factores en la expresión, suele resultar conveniente emplear logaritmos. No obstante, si los factores son numerosos, hasta este método resulta pesado, y puede ahorrarse mucho trabajo utilizando la fórmula de Stirling, que da un valor aproximado de $n!$. Este es

$$n! \cong \sqrt{2\pi} e^{-n} n^{n+1/2} \quad (1)$$

en donde e es la base de los logaritmos neperianos 2,71828... Una aproximación mucho mejor puede obtenerse sustituyendo el factor e^{-n} por $e^{-(n-(1/12n))}$, pero este refinamiento se emplea solo en raras ocasiones. Para indicar la aproximación obtenida por medio de esta fórmula, podemos calcular $10!$, cuyo valor exacto es 3 628 800. La fórmula (1), utilizando logaritmos con cinco cifras decimales, da

$$10! \cong 3\,599\,000$$

mientras que la fórmula más refinada da

$$10! \cong 3\,629\,000$$

El error en (1) para $n=10$ es algo menor del 1 %; el porcentaje de error disminuye al aumentar n .

2-10. Notaciones de sumas y productos.—Una suma de términos tales como $n_3 + n_4 + n_5 + n_6 + n_7$ suele designarse con el símbolo

$\sum_{i=3}^7 n_i$. Σ es la letra griega mayúscula sigma, que en estas ocasiones suele denominarse *signo sumatorio*, y la letra i recibe el nombre de *índice sumatorio*. El término que sigue al símbolo Σ se denomina *sumando*. La $i=3$ debajo de la Σ indica que el primer término de la suma se obtiene haciendo $i=3$ en el sumando. El 7 encima de Σ indica que el término final de la suma se obtiene haciendo $i=7$. Los otros términos de la suma se obtienen dando a i los valores enteros comprendidos entre los límites 3 y 7. Así tenemos

$$\sum_{j=2}^5 (-1)^{j-2} j x^{2j} = 2x^4 - 3x^6 + 4x^8 - 5x^{10}$$

Análoga notación se obtiene para un producto utilizando la letra griega mayúscula Π . En este caso los términos que resultan de sustituir valores enteros en lugar del índice se multiplican en vez de sumarse. Así tenemos

$$\prod_{a=1}^5 \left[c + (-1)^a \frac{a}{b} \right] = \left(c - \frac{1}{b} \right) \left(c + \frac{2}{b} \right) \left(c - \frac{3}{b} \right) \left(c + \frac{4}{b} \right) \left(c - \frac{5}{b} \right)$$

Utilizando esta notación, la expresión (2-8-3) antes deducida puede escribirse

$$n! / \prod_{i=1}^k n_i!$$

2-11. Los teoremas binomial y polinomial¹.—El desarrollo de la expresión binomial $(x+y)^n$ puede verse en los libros de álgebra elemental, y suele obtenerse por inducción la demostración de su desarrollo. Utilizaremos aquí para obtener el desarrollo binomial un método combinatorio que se generaliza fácilmente al caso polinomial. Si escribimos la expresión binomial en la forma $(x+y)(x+y)\dots(x+y)$, con n factores, el problema de hallar el coeficiente de uno de los términos, p. ej., $x^n - a y^a$, se reduce al de hallar el número de maneras de dividir n factores en dos grupos. El primer término del des-

¹ Está muy generalizado el uso impreciso de la palabra *multinomial* por *polinomial*. (N. del T.)

arrollo es x^n , que se obtiene eligiendo la x en cada uno de los factores. El término siguiente es un cierto coeficiente multiplicado por $x^{n-1}y$. Este término se obtiene eligiendo la x en $n-1$ de los factores y la y en el restante. El factor del cual se toma y puede elegirse entre n , y, por tanto, el coeficiente de $x^{n-1}y$ es n . En general, para obtener el coeficiente de $x^{n-a}y^a$ tenemos que contar el número de maneras de dividir los n factores en dos grupos, de modo que uno de ellos contenga a factores y el otro $n-a$; y se elige de cada uno de los factores del primer grupo y x de cada uno de los factores del segundo grupo. El número de maneras de dividir los n factores en dichos dos grupos es $\binom{n}{a}$, que es el coeficiente deseado. El desarrollo binomial es, por tanto,

$$\begin{aligned}(x+y)^n &= x^n + nx^{n-1}y + \binom{n}{2}x^{n-2}y^2 + \dots + y^n \\ &= \sum_{i=0}^n \binom{n}{i} x^{n-i}y^i\end{aligned}\quad (1)$$

El teorema polinomial se deduce inmediatamente; desarrollando la expresión

$$(x_1 + x_2 + \dots + x_k)^n$$

se obtienen términos de la forma

$$Cx_1^{n_1} x_2^{n_2} \dots x_k^{n_k}$$

en donde C representa un cierto coeficiente y los exponentes satisfacen la relación

$$\sum_{i=1}^k n_i = n$$

Se trata ahora de determinar C : los términos de la forma dada se presentan cuando se elige x_1 en n_1 de los n factores, x_2 en n_2 de los restantes factores, y así sucesivamente. El número de maneras de obtener dicho término es igual al de formas de dividir los n factores en k grupos que contengan $n_1, n_2, \dots, n_k$ factores. Esta es la expresión (2-8-3). Por tanto, el término general del desarrollo polinomial es

$$\frac{n!}{n_1! n_2! \dots n_k!} x_1^{n_1} x_2^{n_2} \dots x_k^{n_k} \quad \text{o} \quad n! \prod_{i=1}^k \frac{x_i^{n_i}}{n_i!}$$

y podemos escribir

$$(x_1 + x_2 + \dots + x_k)^n = \sum_{n_1, n_2, \dots, n_k} n! \prod_{i=1}^k \frac{x_i^{n_i}}{n_i!} \quad (2)$$

Hemos indicado solamente que la suma se extiende a los índices $n_1, n_2, \dots, n_k$. Cada índice toma valores desde 0 hasta n , pero no pueden ser sumados independientemente con estos valores, por-

que debe verificarse $\sum_{i=1}^k n_i = n$. La suma se extiende a todos los conjuntos de valores de $n_1, n_2, \dots, n_k$ tales que su suma sea n y tales que n_i sea un entero que puede tomar cualquier valor desde 0 hasta n , ambos inclusive. La suma es de desarrollo muy laborioso cuando n es grande. Como aclaración, consideraremos un caso sencillo:

$$(x_1 + x_2 + x_3)^4 = \sum_{n_1, n_2, n_3} \frac{4!}{n_1! n_2! n_3!} x_1^{n_1} x_2^{n_2} x_3^{n_3}$$

El conjunto de valores (n_1, n_2, n_3) que satisfacen a $n_1 + n_2 + n_3 = 4$ es

$(4, 0, 0), (3, 1, 0), (3, 0, 1), (2, 2, 0), (2, 1, 1), (2, 0, 2), (1, 3, 0), (1, 2, 1),$
 $(1, 1, 2), (1, 0, 3), (0, 4, 0), (0, 3, 1), (0, 2, 2), (0, 1, 3), (0, 0, 4);$

por tanto, la suma tiene 15 términos, los primeros de los cuales son

$$\begin{aligned} (x_1 + x_2 + x_3)^4 &= \frac{4!}{4!} x_1^4 + \frac{4!}{3!} x_1^3 x_2 + \frac{4!}{3!} x_1^3 x_3 + \frac{4!}{2! 2!} x_1^2 x_2^2 + \dots + \frac{4!}{4!} x_3^4 \\ &= x_1^4 + 4x_1^3 x_2 + 4x_1^3 x_3 + 6x_1^2 x_2^2 + \dots + x_3^4 \end{aligned}$$

Un conjunto de números como $(3, 1, 0)$ se denomina *partición* de 4 en 3 partes; $(2, 6)$ es una partición de 8 en 2 partes. Los 15 tríos de números antes enunciados forman el conjunto completo de *particiones ordenadas* de cuatro en tres partes. Estas particiones se llaman *ordenadas* porque se consideran distintas dos particiones que consten de las mismas partes si difieren en el orden.

Si no se especifica que las particiones deben ordenarse, se supone que son sin ordenar; así, las particiones de cuatro en tres partes son simplemente $(4, 0, 0), (3, 1, 0), (2, 2, 0), (2, 1, 1)$. La suma polinomial (2) puede describirse del siguiente modo en función de particiones: se extiende la suma a todas las particiones ordenadas de n en k partes, siendo las partes $(n_1, n_2, \dots, n_k)$.

2-12. Funciones generatrices combinatorias.—Las enumeraciones de resultados posibles y de resultados que presentan un cierto atributo pueden llegar a constituir un problema de gran complicación. En realidad, es sencillo enunciar problemas en los cuales la enumeración resulta prácticamente imposible. Un recurso muy eficaz en la resolución de problemas de enumeración consiste en el uso de las llamadas *funciones generatrices*. Las funciones generatrices combinatorias constituyen en sí mismas una sección especial de las matemáticas, y nos limitaremos a considerar un pequeño número de casos sencillos. Se trata simplemente de indicar la naturaleza de este método de análisis.

Estudiemos de nuevo el ejemplo 2-18, en el que se echaban 7 bolas en 4 cajas, y consideremos la función

$$(x_1 + x_2 + x_3 + x_4)^7$$

El coeficiente de un término como el $x_1^2 x_2^4 x_3$ en el desarrollo de este polinomio viene dado por $7!/2!4!1!0!$ [fórmula (2-8-3)], que es precisamente el número de maneras de dividir 7 objetos en 4 grupos, de modo que el primero contenga 2 objetos; el segundo, 4, y así sucesivamente. De este modo, cualquier término del desarrollo polinomial describe un posible resultado; un factor tal como el x_i^5 indica que 5 bolas han caído en la caja i -ésima, y su coeficiente numérico da el número de maneras como puede ocurrir tal resultado. Si ahora sustituimos las x por unos, el término queda reducido a

$$7! / \prod_{i=1}^4 n_i!,$$

y para obtener todo el conjunto de resultados posibles tenemos que sumar esta expresión para todos los conjuntos de las n_i cuya suma es 7. Esta suma es precisamente por el teorema polinomial

$$(1 + 1 + 1 + 1)^7 = 4^7$$

Si queremos hallar la probabilidad de que la primera caja contenga 2 bolas, tendremos que sumar $7!/\prod n_i!$ para todos los conjuntos de n_i que tienen $n_1=2$. Escribamos de nuevo el término en la forma

$$\frac{7!}{2!5!} \frac{5!}{n_2!n_3!n_4!}$$

Se trata ahora de sumar estos términos para todos los valores tales que $n_2 + n_3 + n_4 = 5$. Si multiplicamos $5!/n_2!n_3!n_4!$ por $1^{n_2}1^{n_3}1^{n_4}$,

tenemos el término general de $(1+1+1)^5$; por tanto, la suma deseada es $7!/2!5!$ multiplicado por 3^5 .

El polinomio $(x_1 + x_2 + x_3 + x_4)^7$ es un tipo sencillo de función generatriz; se trata de una expresión algebraica a la que puede darse una interpretación respecto al problema físico de que se trate. Puede utilizarse como respuesta a cualquiera de las cuestiones relativas al problema físico con el que se relaciona. Así, si se pide el número de maneras en que las dos primeras cajas pueden contener 2 o más bolas cada una, añadiríamos los coeficientes de todos los términos de la función generatriz que tienen potencias de x_1 y x_2 iguales o superiores a 2.

Consideremos ahora otro problema: una urna contiene 5 bolas negras y 4 blancas; todas las bolas se extraen, una a una, de la urna, y las tres primeras extraídas se colocan en una caja negra, mientras que las seis últimas se colocan en una caja blanca. ¿Cuál es la probabilidad de que el número de bolas negras en la caja negra más el número de bolas blancas en la caja blanca sea igual a 5?

Podemos resolver este problema considerando las bolas de cada color que hay que numerar; el total de maneras de dividir los nueve objetos en dos grupos, de tal modo que el primero contenga tres, y el segundo seis, es $\binom{9}{3}$. Para obtener 5 bolas de igual color que la caja que las contiene, es evidente que deberemos tener 2 bolas negras en la caja negra y 3 bolas blancas en la caja blanca. La caja negra puede llenarse de $\binom{5}{2} \binom{4}{1}$ maneras, puesto que hay $\binom{5}{2}$ maneras de separar 2 bolas negras de entre las 5, completando las 3 extraídas, y $\binom{4}{1}$ maneras de separar una bola blanca entre dichas 3. La probabilidad es $\binom{5}{2} \binom{4}{1} / \binom{9}{3}$.

Puede relacionarse con este problema la siguiente función generatriz:

$$(x_1 t + x_2)^5 (x_1 + x_2 t)^4$$

donde x_1 corresponde a la caja negra y x_2 a la caja blanca. El primer factor corresponde a las 5 bolas negras y el segundo a las 4 blancas. Consideraremos el coeficiente del término en $x_1^3 x_2^6$. Este será un polinomio en t , y si hacemos t igual a uno, el polinomio tomará el valor $\binom{9}{3}$, ya que entonces tendremos el coeficiente de $x_1^3 x_2^6$ en $(x_1 + x_2)^9$. El coeficiente de t^r en el polinomio es el número de maneras en que r bolas pueden caer en cajas del mismo color que las bolas. Al formar un término en $x_1^3 x_2^6$ podemos elegir algunas de las

x_1 del factor $(x_1t + x_2)^5$ y las restantes del otro factor. Las elegidas del primer factor representan bolas negras, y las elegidas del segundo, blancas. De modo que cuando asociamos una bola negra con la caja negra, obtenemos un factor t y cuando asociamos una bola blanca con la caja blanca obtenemos también un factor t . La potencia de t proporciona entonces el total de veces que una bola está asociada con una caja de su mismo color. Al desarrollar la función generatriz encontraríamos que el coeficiente de $x_1^3x_2^6t^5$ es $\binom{5}{2} \binom{4}{1}$, como antes.

La función generatriz carece de valor en problema tan sencillo, pero resulta útil cuando se consideran más de dos colores. Supongamos, p. ej., una urna que contiene n_1 bolas de un cierto color, n_2 de otro y n_3 de un tercer color; y supongamos que extraemos m_1 bolas y las colocamos en una caja del primer color, después extraemos m_2 y las colocamos en una caja del segundo color, y las restantes m_3 bolas en una caja del tercer color. Sea n el número total de bolas; entonces

$$n = n_1 + n_2 + n_3 = m_1 + m_2 + m_3$$

El coeficiente de $x_1^{m_1}x_2^{m_2}x_3^{m_3}t^r$ en la función

$$(x_1t + x_2 + x_3)^{n_1}(x_1 + x_2t + x_3)^{n_2}(x_1 + x_2 + x_3t)^{n_3}$$

nos da el número de maneras en que r bolas son del mismo color que la caja que las contiene. En este caso, es difícil calcular el coeficiente; pero el procedimiento para hallarlo es inmediato, y su cálculo es considerablemente más laborioso si se prescinde de la función generatriz.

Consideraremos otro tipo de función generatriz. Si lanzamos 5 dados, ¿cuál es la probabilidad de que la suma de puntos obtenida sea 15?

Puesto que el primer dado puede caer de seis maneras, y el segundo también, y así sucesivamente, el total de resultados posibles es 6^5 . Necesitamos ahora el número de estos resultados cuya suma es igual a 15. En el caso de dos dados, es sencillo escribir todas las combinaciones posibles que dan una suma especificada. Así, para obtener una suma de cinco, los dos dados pueden presentarse de la forma (1, 4), (2, 3), (3, 2), (4, 1). Estas son las particiones ordenadas de cinco en dos partes, cuando no se considera el cero como parte. En nuestro problema tenemos que enumerar todas las particiones ordenadas de 15 en 5 partes que comprenden del 1 al 6, ambos inclusive.

En los problemas en que intervienen particiones de números,

hay una función generatriz que generalmente simplifica la enumeración. Para el problema particular de contar las maneras de obtener 15 con 5 dados, consideremos la función

$$(x + x^2 + x^3 + x^4 + x^5 + x^6)^5 \quad (1)$$

Se trata de un polinomio en x en el que el término de grado inferior es x^5 y el de grado superior, x^{30} . Supongamos que escribimos la función como producto de cinco factores en lugar de como quinta potencia. Asociaremos el primer factor con el primer dado, el segundo factor con el segundo dado, y así sucesivamente. En el desarrollo de la función habrá un cierto número de términos x^{15} ; uno, p. ej., se presentará cuando se elija x de cada uno de los tres primeros factores y x^6 de los otros dos. Esta situación corresponde a la aparición de un 1 en los tres primeros dados y de un 6 en los otros dos. Se ve en seguida que existe correspondencia biunívoca entre las maneras como puede presentarse x^{15} en su desarrollo y las maneras en que los puntos de los cinco dados pueden sumar 15. Por tanto, el número que buscamos es el coeficiente de x^{15} del desarrollo de la función. Este coeficiente puede encontrarse del modo más sencillo utilizando la siguiente identidad:

$$\frac{1 - x^n}{1 - x} = 1 + x + x^2 + \dots + x^{n-1} \quad (2)$$

que puede comprobarse multiplicando ambos miembros por $1 - x$. Mediante esta identidad, la función generatriz puede escribirse en la forma

$$\frac{x^5(1 - x^6)^5}{(1 - x)^5}$$

Podemos omitir el factor x^5 y hallar el coeficiente de x^{10} en lo que resta. Necesitamos ahora de otra identidad:

$$\begin{aligned} \frac{1}{(1 - x)^n} &= 1 + \binom{n}{1}x + \binom{n+1}{2}x^2 + \binom{n+2}{3}x^3 + \dots \quad |x| < 1 \\ &= \sum_{i=0}^{\infty} \binom{n+i-1}{i} x^i \end{aligned} \quad (3)$$

que reduce nuestro problema a hallar el coeficiente de x^{10} en

$$(1 - x^6)^5 \sum_{i=0}^{\infty} \binom{4+i}{i} x^i$$

Si se desarrolla el primer factor, todos los términos, menos los dos primeros, tienen potencias de x superiores a 10 y puede prescindirse de ellos. El problema consiste ahora en hallar el coeficiente de x^{10} en

$$(1 - 5x^6) \sum_{i=0}^{\infty} \binom{4+i}{i} x^i$$

que tiene dos términos en x^{10} : uno, resulta al multiplicar 1 por el término que corresponde a $i=10$ en la suma, y el otro, al multiplicar $-5x^6$ por el término correspondiente a $i=4$. El coeficiente es, por tanto, $\binom{14}{10} - 5\binom{8}{4}$ y la probabilidad que se trataba de hallar es

$$\frac{\binom{14}{10} - 5\binom{8}{4}}{6^5} = \frac{651}{7776} \\ \cong 0,0837$$

Estos ejemplos sirven para indicar cómo pueden abordarse los problemas de enumeración por medio de funciones generatrices. Se trata de un método muy potente, pero no podemos desarrollarlo aquí; solo nos proponíamos indicar su existencia.

2-13. Probabilidad marginal.—En esta sección suponemos que el espacio muestral S , formado por n puntos con probabilidad $1/n$, es particionado en r subconjuntos mutuamente excluyentes (disjuntos) $A_1, A_2, \dots, A_r$. Sea $B_1, B_2, \dots, B_s$ otra partición de S en s subconjuntos mutuamente excluyentes (disjuntos). Los n puntos de S pueden clasificarse en una tabla de doble entrada, tal como la tabla 2-3.

TABLA 2-3

	B_1	B_2	$\dots$	B_s
A_1	n_{11}	n_{12}	$\dots$	n_{1s}
A_2	n_{21}	n_{22}	$\dots$	n_{2s}
$\vdots$	$\vdots$	$\vdots$		$\vdots$
$\vdots$	$\vdots$	$\vdots$		$\vdots$
A_r	n_{r1}	n_{r2}		n_{rs}

El cuadro indica que de los n resultados, n_{11} tienen a la vez el atributo A_1 y el atributo B_1 ; n_{12} , el A_1 y el B_2 ; y, en general, n_{ij} , los atributos A_i y B_j . La suma de todas las n_{ij} es n . Como ejemplo, podemos considerar la extracción de una carta de una baraja ordinaria. Los 52 resultados pueden clasificarse bien por el palo (A_1, A_2, A_3, A_4), bien por la denominación ($B_1, B_2, B_3, B_4, \dots, B_{13}$). En este ejemplo todas las n_{ij} son iguales a la unidad

La probabilidad del suceso A_i y B_j , por ejemplo, se representará a veces por $P(A_i, B_j)$ en lugar de por $P(A_i \cap B_j)$, y el valor de esta probabilidad es, evidentemente, n_{ij}/n . En general,

$$P(A_i, B_j) = \frac{n_{ij}}{n}$$

Puede interesarnos solamente uno de los criterios de clasificación, A , y sernos indiferente la clasificación B . En este caso prescindimos de B en el símbolo, y la probabilidad de un valor cualquiera A_2 se designa por $P(A_2)$, y se tiene

$$\begin{aligned} P(A_2) &= \frac{n_{21} + n_{22} + n_{23} + \dots + n_{2s}}{n} \\ &= \sum_{j=1}^s \frac{n_{2j}}{n} \end{aligned}$$

que recibe el nombre de *probabilidad marginal*; la calificación de marginal se emplea siempre que se prescinde de uno o más criterios de clasificación. Es evidente que

$$P(A_i) = \sum_{j=1}^s \frac{n_{ij}}{n}$$

o bien

$$P(A_i) = \sum_{j=1}^s P(A_i, B_j) \quad (1)$$

ya que $n_{ij}/n = P(A_i, B_j)$. La probabilidad marginal de B_j es, análogamente,

$$P(B_j) = \sum_{i=1}^r P(A_i, B_j) \quad (2)$$

Con la terminología de los conjuntos de puntos, hemos particionado el espacio muestral S en rs subconjuntos disjuntos, designando el subconjunto general por $A_i \cap B_j$. Ahora bien:

$$A_i = (A_i \cap B_1) \cup (A_i \cap B_2) \cup (A_i \cap B_3) \cup \dots \cup (A_i \cap B_s)$$

Puesto que

$$(A_i \cap B_j) \cap (A_i \cap B_{j'}) = \emptyset$$

cuando $j \neq j'$, podemos aplicar el axioma 3 y obtener

$$P(A_i) = P(A_i \cap B_1) + P(A_i \cap B_2) + \dots + P(A_i \cap B_s)$$

o, con notación más sencilla,

$$P(A_i) = \sum_{j=1}^s P(A_i, B_j) = \sum_{j=1}^s \frac{n_{ij}}{n}$$

En el ejemplo de las cartas, la probabilidad de que una de ellas sea un as es la suma de las probabilidades de que sea el as de espadas, el as de corazones o cualquiera de los otros dos ases.

Considerando un caso más general, supongamos tres criterios de clasificación A , B y C . Sea n_{ijk} el número de resultados de entre los n que presentan los caracteres A_i , B_j , C_k , y sean $C_1, C_2, \dots, C_t$ las clases que constituyen C , siendo A y B las mismas clases anteriores. La clasificación completa sería una tabla de triple entrada compuesta de t capas de tablas de doble entrada, correspondiendo cada capa a un C_k . La probabilidad marginal de, p. ej., A_i y C_k es

$$P(A_i, C_k) = \sum_{j=1}^s P(A_i, B_j, C_k) \quad (3)$$

y la probabilidad marginal de C_k es

$$P(C_k) = \sum_{i=1}^r \sum_{j=1}^s P(A_i, B_j, C_k) \quad (4)$$

$$= \sum_{i=1}^r P(A_i, C_k) \quad (5)$$

$$= \sum_{j=1}^s P(B_j, C_k) \quad (6)$$

La generalización de estas ideas a más de tres criterios de clasificación es inmediata.

2-14. Probabilidad condicional.—Volviendo a la clasificación de doble entrada de la tabla 2-3, imaginemos que se examina el resultado de un experimento aleatorio para un atributo, pero no para el otro. Deseamos hallar la probabilidad de que el otro atributo tenga un valor determinado. Supongamos que se ha observado que el suceso tiene el atributo B_3 . ¿Cuál es la probabilidad de que tenga así mismo el atributo A_2 ? El total de resultados de A cuando ha ocurrido B_3 , es $\sum_{i=1}^r n_{i3}$, y el número de resultados favorables a A_2 es n_{23} . Así, la probabilidad de A_2 , cuando se sabe que ha ocurrido B_3 , es $n_{23} / \sum_{i=1}^r n_{i3}$, que se designa con el nombre de *probabilidad condicional* y cuyo símbolo es $P(A_2|B_3)$. En general (suponiendo que los denominadores no son nulos),

$$P(A_i|B_j) = \frac{n_{ij}}{\sum_{i=1}^r n_{ij}}$$

$$P(B_j|A_i) = \frac{n_{ij}}{\sum_{j=1}^s n_{ij}}$$

Dividiendo el numerador y el denominador de la fracción del segundo miembro por n , tenemos

$$P(A_i|B_j) = \frac{P(A_i, B_j)}{P(B_j)} \quad (1)$$

$$P(B_j|A_i) = \frac{P(A_i, B_j)}{P(A_i)} \quad (2)$$

o bien

$$P(A_i, B_j) = P(A_i|B_j)P(B_j) \quad (3)$$

$$= P(B_j|A_i)P(A_i) \quad (4)$$

Esta última ecuación puede enunciarse del siguiente modo: La probabilidad de que un resultado tenga los atributos A_i y B_j es

igual a la probabilidad marginal de A_i multiplicada por la probabilidad condicional de B_j cuando ha ocurrido A_i .

La idea de probabilidad condicional admite una generalización inmediata a situaciones donde interviene más de un criterio de clasificación; p. ej., en el caso de tres criterios, se ve inmediatamente que

$$P(A_i, B_j | C_k) = \frac{P(A_i, B_j, C_k)}{P(C_k)} \quad (5)$$

$$P(A_i | B_j, C_k) = \frac{P(A_i, B_j, C_k)}{P(B_j, C_k)} \quad (6)$$

y también que

$$P(A_i, B_j, C_k) = P(A_i, B_j | C_k) P(C_k) \quad (7)$$

$$= P(A_i | B_j, C_k) P(B_j, C_k) \quad (8)$$

$$= P(A_i | B_j, C_k) P(B_j | C_k) P(C_k) \quad (9)$$

Podrían obtenerse otras relaciones análogas permutando las letras A, B, C . Así

$$P(B_j | A_i, C_k) = \frac{P(A_i, B_j, C_k)}{P(A_i, C_k)} \quad (10)$$

y

$$P(A_i, B_j, C_k) = P(B_j | A_i, C_k) P(A_i | C_k) P(C_k) \quad (11)$$

o bien

$$P(A_i, B_j, C_k) = P(B_j | A_i, C_k) P(C_k | A_i) P(A_i) \quad (12)$$

No disponemos de espacio para escribir todas las relaciones análogas posibles, pero aconsejamos al lector que lo haga. Estas relaciones son fundamentales en la teoría de la estadística y deben ser bien entendidas. Las probabilidades condicionales anteriores no están definidas si el denominador es 0. Esto es válido también en lo que sigue.

Al describir la probabilidad condicional hemos utilizado un espacio muestral bastante especial, espacio que contiene un número finito n de puntos, cada uno con probabilidad $1/n$. Sin embargo, la idea es completamente general y puede definirse para espacios muestrales discretos y continuos de la siguiente forma:

Definición 2-13.—Sean A y B dos sucesos de un espacio muestral S tal que $P(B) > 0$. La probabilidad condicional del suceso A cuando ha ocurrido el suceso B , y que se designa por $P(A|B)$, es

$$P(A|B) = \frac{P(A, B)}{P(B)}$$

Ejemplo 2-19.—Una urna contiene seis bolas rojas y cuatro negras. Se extraen dos bolas sin reemplazamiento. ¿Cuál es la probabilidad de que la segunda bola sea roja si se sabe que la primera es roja? Emplearemos la fórmula de la definición 2-13. Sea B el suceso «la primera bola es roja» y A el suceso «la segunda bola es roja». Por tanto, $P(A, B)$ es la probabilidad de que tanto la primera como la segunda bola sean rojas. Hay $\binom{10}{2}$ maneras de extraer dos bolas de la urna; luego el espacio muestral contiene $\binom{10}{2}$ puntos, cada uno de ellos con probabilidad $1/\binom{10}{2}$. El número de maneras de sacar dos bolas rojas es $\binom{6}{2}$ y, por tanto,

$$P(A, B) = \frac{\binom{6}{2}}{\binom{10}{2}} = \frac{1}{3}$$

$P(B)$ es la probabilidad de que la primera bola extraída sea roja, la cual vale, evidentemente, $6/10$; luego

$$P(A|B) = \frac{1}{3} \bigg/ \frac{6}{10} = \frac{5}{9}$$

Esta probabilidad podría haberse calculado directamente, puesto que si la primera bola es roja, quedan en la urna cinco bolas rojas y cuatro negras, por lo que $P(A|B) = \frac{5}{9}$.

2-15. Dos leyes básicas de la probabilidad.—Si A y B son dos subconjuntos mutuamente excluyentes (lo que significa que $A \cap B = \emptyset$ o, en otras palabras, que A y B no pueden presentarse simultáneamente), el axioma 3 establece que

$$P(A \cup B) = P(A) + P(B)$$

o, con notación más sugestiva,

$$P(A \text{ o } B) = P(A) + P(B).$$

Podemos obtener una fórmula análoga para sucesos cualesquiera A y B que sean o no mutuamente excluyentes.

Teorema 2-4.—Sea S un espacio muestral con función de probabilidad P . Si A y B son dos sucesos cualesquiera de S ,

$$P(A \cup B) = P(A) + P(B) - P(A, B)$$

o, en otras palabras, la probabilidad de que ocurra el suceso A o el suceso B , o ambos, es igual a la probabilidad de que se produzca el suceso A , más la probabilidad de que ocurra el suceso B , menos la probabilidad de que ocurran A y B simultáneamente.

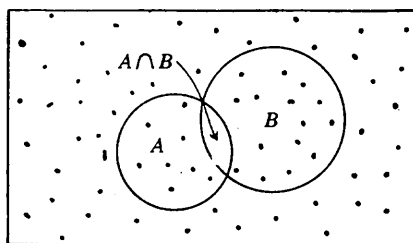


FIG. 2-1.

Demostración. El álgebra de conjuntos permite establecer que

$$A \cup B = A \cup (\bar{A} \cap B)$$

(véase la Fig. 2-1). Pero A y $\bar{A} \cap B$ son disjuntos; luego, aplicando el axioma 3, se deduce que

$$P(A \cup B) = P[A \cup (\bar{A} \cap B)] = P(A) + P(\bar{A} \cap B).$$

Ahora bien:

$$B = (A \cap B) \cup (\bar{A} \cap B)$$

y los conjuntos $(A \cap B)$ y $(\bar{A} \cap B)$ son disjuntos. Aplicando de nuevo el axioma 3, tenemos:

$$P(B) = P[(A \cap B) \cup (\bar{A} \cap B)] = P(A \cap B) + P(\bar{A} \cap B)$$

o bien,

$$P(\bar{A} \cap B) = P(B) - P(A \cap B)$$

Sustituyendo $P(\bar{A} \cap B)$ en la ecuación anterior, se obtiene

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

como se quería demostrar.

Para demostrar el teorema en un espacio muestral finito S , donde cada punto tiene la probabilidad $1/n$, nos referiremos a la ta-

bla 2-3 y calcularemos $P(A_1 \cup B_2)$. La probabilidad de que ocurran los sucesos A_1 o B_2 se calcula sumando todas las n_{ij} de la primera fila y la segunda columna, y dividiendo por n . Así,

$$\begin{aligned} P(A_1 \text{ o } B_2) &= \frac{\sum_{j=1}^s n_{1j} + \sum_{i=2}^r n_{i2}}{n} \\ &= \frac{\sum_{j=1}^s n_{1j} + \sum_{i=1}^r n_{i2} - n_{12}}{n} \\ &= P(A_1) + P(B_2) - P(A_1, B_2) \end{aligned}$$

La situación está también representada en la figura 2-1, donde el espacio muestral se ha representado por puntos de un plano, y los dos sucesos A y B por círculos. El suceso $A \cap B$ es la región lenticular común a ambos círculos; al sumar los puntos de estos, contamos dos veces los de esta región, que deben, por tanto, restarse una vez.

Cabe generalizar esta ley para más de dos subconjuntos; en este caso

$$\begin{aligned} P(A \text{ o } B \text{ o } C) &= P(A) + P(B) + P(C) - P(A, B) - \\ &\quad - P(A, C) - P(B, C) + P(A, B, C) \end{aligned}$$

como se comprueba fácilmente dibujando un esquema análogo al de la figura 2-1, con tres círculos que se corten entre sí limitando una región común a los tres. La ley general para h subconjuntos, que puede demostrarse por inducción, es

$$\begin{aligned} P(A_1 \text{ o } A_2 \dots \text{ o } A_h) &= \sum_{i=1}^h P(A_i) - \sum_{\substack{ij \\ i < j}} P(A_i, A_j) + \\ &\quad + \sum_{\substack{ij,k \\ i < j < k}} P(A_i, A_j, A_k) - \dots \pm P(A_1, A_2, \dots, A_h) \end{aligned}$$

donde la segunda suma se extiende a todas las combinaciones de los números 1, 2, ..., h tomados dos a dos; la tercera, a todas las combinaciones de estos números tomados tres a tres, y así sucesivamente. Si todos los subconjuntos son mutuamente excluyentes,

las probabilidades correspondientes a sumas posteriores a la primera son nulas.

Al definir (Def. 2-13) la probabilidad condicional, hemos deducido en esencia la ley multiplicativa de las probabilidades. Esta ley dice que *la probabilidad de los sucesos A y B es igual a la probabilidad condicional de B, en el supuesto de que ha ocurrido A, multiplicada por la probabilidad marginal de A*. En símbolos,

$$P(A, B) = P(A)P(B|A) \quad (1)$$

$$= P(B)P(A|B) \quad (2)$$

Nos apoyaremos en la figura 2-1 para demostrarla en un espacio muestral finito donde cada punto tiene probabilidad $1/n$. Sea n el número de puntos de la figura 2-1; m_1 el número de puntos de A (incluyendo los que son comunes a B); m_2 el número de puntos de B, y m_3 el número de los comunes a A y B. Entonces suponiendo $m_1 > 0$, $m_2 > 0$)

$$P(A, B) = \frac{m_3}{n}$$

$$P(A) = \frac{m_1}{n}$$

$$P(B) = \frac{m_2}{n}$$

$$P(A|B) = \frac{m_3}{m_2}$$

$$P(B|A) = \frac{m_3}{m_1}$$

de donde se deducen inmediatamente (1) y (2).

En general, puede demostrarse por inducción que

$$P(A_1, A_2, \dots, A_h) = P(A_1)P(A_2|A_1)P(A_3|A_1, A_2)P(A_4|A_1, A_2, A_3) \\ \dots P(A_h|A_1, A_2, \dots, A_{h-1}) \quad (3)$$

y hay $h!$ relaciones como esta, que se obtienen permutando las letras en el segundo miembro.

2-16. Sucesos compuestos.—La ley multiplicativa de las probabilidades resulta especialmente útil por simplificar el cálculo de las probabilidades de sucesos compuestos. Un suceso compuesto consiste en dos o más sucesos simples, como ocurre cuando se lanza un dado dos veces o se extraen sucesivamente tres cartas. El siguiente ejemplo servirá de ilustración. Se extraen 2 bolas, una tras

otra, de una urna que contiene 2 bolas negras, 3 blancas y 4 rojas. ¿Cuál es la probabilidad de que la primera bola extraída sea roja y la segunda blanca? (La primera no se reemplaza en la urna antes de la extracción de la segunda). Los resultados de este suceso compuesto pueden clasificarse según dos criterios: el color de la primera bola y el color de la segunda; podemos, pues, construir un cuadro análogo a la tabla 2-3. La clasificación A se basa en el color de la primera bola, y hacemos corresponder A_1 , A_2 y A_3 a los colores negro, blanco y rojo, respectivamente. Análogamente, las clases B_1 , B_2 y B_3 corresponderán a los mismos colores para la segunda bola. El número total de resultados es $n=9 \times 8=72$. No es $\binom{9}{2}=36$ porque estamos considerando permutaciones y no combinaciones; esto es, no decimos que una bola sea roja y otra blanca: exigimos que los colores aparezcan en un orden determinado. La tabla completa de resultados es

	B_1	B_2	B_3
A_1	2	6	8
A_2	6	6	12
A_3	8	12	12

y la probabilidad pedida

$$P(A_3, B_2) = 12/72 = 1/6$$

Utilizando la ley multiplicativa de las probabilidades, solo necesitamos considerar los dos sucesos simples por separado. En este caso se debe utilizar dicha ley en la forma

$$P(A_3, B_2) = P(A_3)P(B_2|A_3)$$

Ahora bien: $P(A_3)$ es simplemente la probabilidad de obtener bola roja en una sola extracción, que es igual a $4/9$, y $P(B_2|A_3)$, la probabilidad de extraer una bola blanca cuando se ha extraído ya una roja, probabilidad que vale $3/8$. El producto de estos dos números da la probabilidad pedida,

$$P(A_3, B_2) = 4/9 \times 3/8 = 1/6$$

La validez de la técnica anterior no es evidente; no resulta inmediato que la probabilidad marginal $P(A_3)$ pueda calcularse pres-

cindiendo por completo del segundo suceso, ni que la probabilidad condicional corresponda al suceso físico simple que antes hemos descrito.

Para un suceso compuesto, que conste de dos sucesos simples, basta considerar una tabla de 2 por 2 (dos filas y dos columnas). Hacemos corresponder A_1 a un acierto en el primer suceso, y A_2 a un fallo en el mismo, y sea m_1 el número de maneras en que puede presentarse con acierto el primer suceso y m_2 el número de veces en que dicho suceso puede fallar. Supongamos que B_1 y B_2 se definen análogamente para el segundo suceso. Sean m_{11} y m_{12} los números de maneras en que el segundo suceso puede ser un acierto y un fallo, suponiendo que se haya obtenido acierto en el primero, y m_{21} y m_{22} , los números de maneras en que el segundo puede resultar acierto o fallo cuando se sabe que ha fallado el primero. La tabla 2×2 es

	B_1	B_2
A_1	$m_1 m_{11}$	$m_1 m_{12}$
A_2	$m_2 m_{21}$	$m_2 m_{22}$

El número total de resultados posibles es

$$n = m_1 m_{11} + m_1 m_{12} + m_2 m_{21} + m_2 m_{22}$$

La probabilidad buscada es

$$P(A_1, B_1) = \frac{m_1 m_{11}}{n} \quad (1)$$

La probabilidad marginal $P(A_1)$ es

$$\frac{m_1 m_{11}}{n} + \frac{m_1 m_{12}}{n} = \frac{m_1(m_{11} + m_{12})}{m_1(m_{11} + m_{12}) + m_2(m_{21} + m_{22})} \quad (2)$$

Ahora bien: la probabilidad de obtener un acierto en el primer suceso sin tener en cuenta el segundo es $m_1/(m_1 + m_2)$, que no coincide con la expresión anterior, salvo que

$$m_{11} + m_{12} = m_{21} + m_{22};$$

esto es, a menos que el número total de resultados del segundo suceso sea el mismo sin tener en cuenta si el primero ha sido acierto o fallo. La probabilidad condicional es $m_{11}/(m_{11} + m_{12})$ y da la pro-

babilidad de un acierto para el segundo suceso, en el supuesto de que el primero ha sido un acierto. Podríamos inclinarnos a deducir que este uso de la probabilidad condicional solo es correcto cuando el número de resultados para el segundo suceso es independiente del resultado del primero; pero precisamente ocurre lo contrario. La probabilidad correcta es

$$P(A_1, B_1) = \frac{m_{11}}{m_{11} + m_{12}} \quad (3)$$

y no el valor $m_1 m_{11}/n$ dado por la ecuación (1).

El valor calculado por este método condicional es siempre correcto, mientras que el obtenido por enumeración de resultados lo es solo cuando el número de ellos para el segundo suceso es independiente del resultado del primero.

Un ejemplo sencillo aclarará la situación. Supongamos que se lanza una moneda, y que si sale cara, se coloca una bola negra en una urna, mientras que si aparece cruz, se introducen en la urna una bola negra y otra blanca. A continuación, se extrae una bola de la urna; si salió cara, la bola tendrá que ser negra. Si representamos cara, cruz, negra y blanca por C, X, N, B , los tres resultados posibles serán CN, XN, XB . Estos tres resultados no son, por supuesto, igualmente probables.

Si el experimento se repitiese cierto número de veces, cabría esperar que el resultado CN ocurriera doble número de veces que cualquiera de los otros dos; $P(CN) = 1/2$, y no $1/3$.

En general, los resultados posibles de un suceso compuesto no son igualmente probables si el número de resultados del segundo suceso depende del resultado del primero; por tanto, no es aplicable la definición de probabilidad. No obstante, si la definición puede aplicarse por separado a los sucesos componentes, será posible calcular la probabilidad del suceso compuesto utilizando el método de las probabilidades condicionales. Desgraciadamente, no es posible dar una demostración formal de estas afirmaciones. Tenemos que limitarnos a confiar en nuestra intuición, o más bien en el valor de cualquier testimonio experimental que poseamos.

Tal testimonio puede obtenerse, p. ej., ejecutando el experimento antes descrito un cierto número de veces.

Ejemplo 2-20.—Como nueva aclaración del método de las probabilidades condicionales, calculemos la probabilidad de que, de 5 cartas extraídas de una baraja ordinaria (de 52 cartas), dos exactamente sean ases.

Supondremos que la baraja consta de 4 cartas A , con lo que representamos los ases, y 48 cartas N , símbolo que representa los no ases. Para usar probabilidades condicionales, tenemos que suponer

que solo sacamos una de las 5 cartas cada vez, y en un orden determinado; p. ej., A, A, N, N, N . Utilizaremos la ecuación (2-15-3) con $h=5$:

$$P(A, A, N, N, N) = P(A)P(A|A)P(N|A, A)P(N|A, A, N)P(N|A, A, N, N)$$

En este caso, $P(A) = 4/52$. Al quitar un as de la baraja, $P(A|A) = 3/51$; al quitar dos ases, $P(N|A, A) = 48/50$. Siguiendo así,

$$P(A, A, N, N, N) = 4/52 \times 3/51 \times 48/50 \times 47/49 \times 46/48$$

Esta probabilidad corresponde a un orden determinado, pero el problema no especificaba orden alguno, de modo que tenemos que considerar todos los órdenes posibles. Hay $5!/(2!3!) = 10$ permutaciones de dos cartas A y tres cartas N , de manera que tenemos que calcular 10 probabilidades, y, por la ley aditiva, la probabilidad es la suma de estas 10 probabilidades. Se ve en seguida, sin embargo, que todas las probabilidades son iguales. Así, p. ej.,

$$P(N, A, N, N, A) = 48/52 \times 4/51 \times 47/50 \times 46/49 \times 3/48$$

que es igual que el número anterior, con la diferencia de que los numeradores aparecen permutados. Naturalmente que lo mismo ocurrirá con todas las permutaciones. Por tanto, la probabilidad buscada es

$$10P(A, A, N, N, N) = \frac{10 \times 4 \times 3 \times 47 \times 46}{52 \times 51 \times 50 \times 49} \\ \cong 0,0399$$

Ejemplo 2-21.—Se extraen 6 cartas con reemplazamiento de una baraja ordinaria; ¿cuál es la probabilidad de que cada uno de los cuatro palos venga representado al menos una vez entre las 6 cartas? Resolveremos el problema empezando por hallar la probabilidad de que no aparezcan todos los palos. Representemos por A la aparición de todos los palos y por B la no aparición de uno de los palos, por lo menos; por tanto, $B = \bar{A}$. Puesto que es seguro que ha de ocurrir A o B ,

$$P(A \text{ o } B) = 1$$

y dado que A y B son mutuamente excluyentes

$$P(A \text{ o } B) = P(A) + P(B) = 1$$

y

$$P(A) = 1 - P(B)$$

Por tanto, si sabemos hallar $P(B)$, podremos determinar inmediatamente $P(A)$.

Para obtener $P(B)$, clasifiquemos los resultados favorables a B en cuatro conjuntos: B_1 es el conjunto de todos los resultados en que no aparecen espadas; B_2 aquel donde faltan corazones; B_3 y B_4 son los conjuntos en que faltan, respectivamente, cada uno de los otros dos palos. Estos conjuntos se solapan; un resultado que solo tenga espadas y corazones pertenecerá a B_3 y B_4 . Evidentemente

$$P(B) = P(B_1 \cup B_2 \cup B_3 \cup B_4)$$

y

$$P(B) = \sum P(B_i) - \sum P(B_i, B_j) + \sum P(B_i, B_j, B_k) - P(B_1, B_2, B_3, B_4)$$

donde la suma se extiende a todas las combinaciones de los sub-índices. La probabilidad $P(B_1)$ de que no aparezcan espadas en las seis extracciones es $(3/4)^6$ y este valor es el mismo para todos los B_i ; por tanto,

$$\sum P(B_i) = 4(3/4)^6$$

La probabilidad $P(B_1, B_2)$ de que no aparezcan espadas ni corazones en las seis extracciones es $(1/2)^6$, y es la misma para cada uno de los seis pares de los cuatro palos tomados dos a dos; por tanto,

$$\sum P(B_i, B_j) = 6(1/2)^6$$

Análogamente

$$\sum P(B_i, B_j, B_k) = 4(1/4)^6$$

y

$$P(B_1, B_2, B_3, B_4) = 0$$

ya que es imposible la falta simultánea de todos los palos. La probabilidad buscada es, por consiguiente,

$$\begin{aligned} P(A) &= 1 - 4(3/4)^6 + 6(1/2)^6 - 4(1/4)^6 \\ &\approx 0,381 \end{aligned}$$

Una pequeña alteración de este ejemplo servirá para aclarar otra técnica utilizable.

Ejemplo 2-22.—Se sacan cartas de una en una, con reemplazamiento, de una baraja ordinaria hasta que todos los palos hayan aparecido al menos una vez. ¿Cuál es la probabilidad de que se necesiten seis extracciones?

Refiriéndonos al ejemplo anterior, representemos por P_n la probabilidad de que todos los palos aparezcan al menos una vez en la extracción de n cartas. Evidentemente

$$P_n = 1 - 4(3/4)^n + 6(1/2)^n - 4(1/4)^n$$

Supongamos ahora que conociésemos la respuesta a este problema para un valor cualquiera de n . Designemos por p_n esta probabilidad; esto es, la de que hagan falta exactamente n extracciones para obtener todos los palos.

Si se extraen n cartas, la presentación de los cuatro palos podrá ocurrir por primera vez en la cuarta extracción, en la quinta o en la sexta, y así sucesivamente. Puesto que estos resultados son mutuamente excluyentes, tenemos

$$P_n = p_4 + p_5 + p_6 + \dots + p_n$$

De esta relación se deduce que

$$p_n = P_n - P_{n-1}$$

y, en particular, que

$$\begin{aligned} p_6 &= 1 - 4(3/4)^6 + 6(1/2)^6 - 4(1/4)^6 - [1 - 4(3/4)^5 + 6(1/2)^5 - 4(1/4)^5] \\ &= (3/4)^5 - 3(1/2)^5 + 3(1/4)^5 \\ &= 0,147 \end{aligned}$$

2-17. Independencia.—Si $P(A|B)$ no depende del suceso B , diremos que los sucesos A y B son independientes. Esto se expresa por la siguiente definición:

Definición 2-14.—Sean A y B dos sucesos de un espacio muestral S . Se dice que estos dos sucesos son independientes si se satisface cualquiera de las siguientes igualdades:

$$\begin{aligned} a) \quad & P(A|B) = P(A) \\ b) \quad & P(B|A) = P(B) \\ c) \quad & P(A, B) = P(A)P(B) \end{aligned} \tag{1}$$

Así, p. ej., supongamos que se lanza un dado dos veces y que deseamos hallar la probabilidad de que los resultados sean dos y tres, en este orden:

$$P(2, 3) = P(2)P(3|2) = 1/6 \times 1/6 = P(2)P(3)$$

de modo que los dos sucesos son independientes.

En el ejemplo 2-20, en el que considerábamos dos ases en cinco cartas, los cinco sucesos componentes del suceso estudiado son independientes si se exige que cada carta extraída vuelva a introducirse entre las demás, barajando a continuación antes de la extracción de la próxima carta. La probabilidad de que la segunda carta sea un as es entonces $4/52$ en vez de $3/51$. La probabilidad de obtener dos ases al sacar cinco cartas con reemplazamiento es

$$10(4/52)^2(48/52)^3 = 0,0465$$

2-18. Variables aleatorias.—En los ejemplos 2-1, 2-2 y 2-5, el espacio muestral es un conjunto de puntos arbitrarios, mientras que en los 2-3, 2-4, 2-6 y 2-7 el espacio muestral es un conjunto de números. A menudo es ventajoso asociar un conjunto de números reales con el resultado de un experimento aleatorio (con el espacio muestral).

Podemos hacer esto definiendo una variable aleatoria.

Definición 2-15.—Representemos por S un espacio muestral en el que se define una función de probabilidad. Sea x una función de valores reales definida en S (la función x transforma puntos de S en puntos del eje x). Se dice que x es una variable aleatoria (variable aleatoria unidimensional).

Si es s un punto del espacio muestral S y x una variable aleatoria, $x(s)$ es el valor de la variable aleatoria en s . Como aclaración, consideremos el ejemplo 2-1, donde el espacio muestral está formado por cuatro puntos:

$$s_1 = (C, C); \quad s_2 = (C, X); \quad s_3 = (X, C); \quad s_4 = (X, X)$$

Sea x una variable aleatoria que puede tomar los valores x_1, x_2, x_3, x_4 , definidos por x_i = número de caras en s_i . Esto puede representarse por la siguiente tabla:

Puntos de S	s_1	s_2	s_3	s_4
$x(s)$	2	1	1	0

Si se supone que S es un espacio muestral en el que está definida una función de probabilidad de sucesos igualmente verosímiles, escribimos

$$P(x=0) \text{ para designar } P[\{s: x(s)=0\}] = P(s_4) = 1/4$$

$$P(x=1) \text{ para designar } P[\{s: x(s)=1\}] = P(s_2 \cup s_3) = 1/2$$

$$P(x=2) \text{ para designar } P[\{s: x(s)=2\}] = P(s_1) = 1/4$$

Corrientemente no nos preocupamos de la función x o del espacio muestral S , sino de la probabilidad de que el valor de la variable aleatoria pertenezca a cierto conjunto A , lo que escribimos $P(x \text{ en } A)$.

Para ver cómo pueden calcularse las probabilidades de que el valor de una variable aleatoria esté en cierto conjunto, se utilizará una función de densidad de la variable aleatoria. De ello se tratará en los capítulos 3 y 4. Para los propósitos de este libro, las variables aleatorias se dividirán en dos clases: 1) una variable aleatoria se dice *discreta* si toma solamente un número finito o una infinidad numerable de valores (se estudian en el Cap. 3); 2) una variable aleatoria se denomina continua si toma un continuo de valores (se estudiarán en el Cap. 4).

A veces, una variable aleatoria unidimensional no resultará adecuada para nuestros propósitos, por lo que generalizaremos la definición.

Definición 2-16.—*Sea S un espacio muestral en el que está definida una función de distribución de probabilidad. Sean x e y dos funciones de valores reales definidas en S . El par (x, y) recibe el nombre de variable aleatoria bidimensional y se dice que las variables x e y están distribuidas conjuntamente. Así, una variable aleatoria bidimensional transforma puntos de S en puntos del plano xy .*

Para cualquier conjunto A del plano xy estaremos interesados en $P[(x, y) \text{ esté en } A]$, con lo que representamos

$$P[\{s: (x(s), y(s)) \text{ esté en } A\}]$$

Una definición análoga es válida para la distribución conjunta de n variables aleatorias, siendo n cualquier número natural. Como en el caso de una variable aleatoria unidimensional, generalmente no nos interesará el espacio muestral S o las funciones x e y , sino que definiremos una función de densidad que pueda utilizarse para calcular $P[(x, y) \text{ esté en } A]$. Diremos que « x e y son variables aleatorias distribuidas conjuntamente» y, a veces, emplearemos indistintamente las denominaciones «variable aleatoria» y «valor de una variable aleatoria». Así, p. ej., cuando digamos «una variable aleatoria x es igual a 1» o escribamos $x=1$, queremos decir «el valor de una variable aleatoria x es igual a 1».

Supongamos que, en el ejemplo 2-1, la función x está definida por

x : número de caras en la primera tirada

Puntos de S	(C, C)	(C, X)	(X, C)	(X, X)
$x(s)$	1	1	0	0

y definamos y por

y : número de caras en la segunda tirada

Puntos de S	(C, C)	(C, X)	(X, C)	(X, X)
$y(s)$	1	0	1	0

La variable aleatoria bidimensional (x, y) está definida por

Puntos de S	(C, C)	(C, X)	(X, C)	(X, X)
Valores correspondientes de (x, y)	(1, 1)	(1, 0)	(0, 1)	(0, 0)

Utilizaremos letras en negrita U, V, W, u, v, w , etc., o estas letras con subíndices, para indicar variables aleatorias. A menudo emplearemos la palabra variante en lugar de variable aleatoria.

PROBLEMAS

Cuando se pida una probabilidad en los problemas que siguen, supondremos que todas las ordenaciones son igualmente verosímiles. Defínase muy cuidadosamente en cada caso el espacio muestral. En los problemas en que se pide una probabilidad es instructivo imaginar el experimento aleatorio, e interpretar la probabilidad resultante como una razón frecuencial. En algunos de los problemas cabe realizar un experimento aleatorio y comparar la probabilidad calculada con la razón frecuencial observada.

1. Sean A, B, C sucesos (subconjuntos) de un espacio muestral S , definidos por

$$S = \{x: x=1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$$

$$A = \{x: x=1, 3, 5, 7, 9\}$$

$$B = \{x: x=1, 2, 3, 4, 5, 6\}$$

$$C = \{x: x=5, 6, 7, 8, 9, 10\}$$

Hallar los sucesos

a) $A \cap (B \cap C)$

c) $A \cup \bar{A}$

e) $\bar{C} \cap \bar{B}$

b) $\overline{A \cup B}$

d) $A \cup B \cup C$

f) $\bar{S}$

Utilizando las definiciones de A, B, C y S , demostrar que

g) $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$

k) $A \cup S = S$

h) $\overline{A \cup B} = \bar{A} \cap \bar{B}$

l) $A \cap \bar{A} = \emptyset$

i) $A \cup A = A$

m) $\overline{A \cap B} = \bar{A} \cup \bar{B}$

j) $A \cap A = A$

n) $A \cap (B \cap \bar{A}) = \emptyset$

2. Sean A, B, C sucesos de un espacio muestral S definidos por

$$S = \{x: 0 \leq x \leq 20\}$$

$$A = \{x: 0 \leq x \leq 5\}$$

$$B = \{x: 3 \leq x \leq 10\}$$

$$C = \{x: 7 \leq x \leq 15\}$$

Hallar los sucesos:

- $A \cup B$; es decir, ocurre A o B , o ambos.
- $\bar{A}$; es decir, no ocurre A .
- $A \cap \bar{B}$; es decir, ocurre A y no ocurre B .
- $A \cup (B \cap \bar{C})$; es decir, ocurre A u ocurre B pero no C .
- $\overline{A \cup C}$: es decir, no ocurre ni A ni C .
- $\bar{B}$: es decir, no ocurre B .

3. Sean A y B sucesos de un espacio muestral S definidos por

$$S = \{(x, y): x \geq 0, y \geq 0\}$$

$$A = \{(x, y): x \geq 0, 0 \leq y \leq 5\}$$

$$B = \{(x, y): x \geq 0, 5 \leq y \leq 10\}$$

Hallar los sucesos:

a) A	d) $\overline{A \cap B}$	f) $\bar{A} \cap B$
b) $\bar{B}$	e) $\bar{A} \cap \bar{B}$	g) $A \cup \bar{B}$
c) $A \cap \bar{B}$		

4. Sean A y B sucesos de un espacio muestral S . Exprésese con palabras el significado de los siguientes símbolos:

a) $\bar{A} \cup B$	c) $A \cap B$	e) $\bar{A} \cap \bar{B}$
b) $\bar{A} \cup \bar{B}$	d) $\bar{A} \cap B$	f) $\bar{A} \cap A$

5. Una urna contiene 10 bolas, 6 rojas y 4 verdes. Se extraen dos bolas al azar con reemplazamiento. Describir el espacio muestral S y la función de probabilidad P . ¿Cuántos puntos contiene el espacio muestral?

6. Una urna contiene 4 bolas blancas y 6 negras. ¿Cuál es la probabilidad de que una bola extraída al azar sea blanca?

7. Si se lanzan dos monedas, ¿cuál es la probabilidad de obtener una cara y una cruz?

8. Si se colocan en un estante en orden aleatorio 4 volúmenes de una cierta obra, ¿cuál será la probabilidad de que el orden sea perfecto?

9. ¿Cuál es la probabilidad de obtener tres caras lanzando tres monedas? ¿Cuál es la probabilidad de obtener dos o más caras? Lanzar

100 veces tres monedas y observar la frecuencia con que aparecen al menos dos caras.

10. Una urna contiene tres bolas blancas y dos negras. ¿Cuál es la probabilidad de que 2 bolas extraídas sean ambas negras?

11. ¿Cuántos números de dos cifras pueden formarse con los guarismos 1, 2, 3, 4, 5, suponiendo que no pueden repetirse estos? ¿Y si se permite la repetición (duplicación) de los guarismos?

12. ¿Cuántos números de tres cifras pueden formarse con 0, 1, 2, 3, 4, si no se permite la repetición? ¿Cuántos de estos serán pares?

13. ¿De cuántas maneras puede formarse con 9 hombres una comisión de tres?

14. Hay 6 caminos que van de A a B y 3 de B a C. ¿De cuántas maneras se puede ir de A a C pasando por B?

15. ¿Cuántas cantidades diferentes de dinero pueden formarse con seis monedas de valores distintos?

16. ¿De cuántas maneras pueden dividirse 6 niñas y 4 niños en dos grupos de 2 niños y 3 niñas?

17. En un campeonato de liga de pelota base con 8 equipos, ¿cuántos encuentros serán necesarios si cada equipo ha de jugar dos veces en su campo con cada uno de los demás?

18. ¿Cuántos equipos de fútbol pueden formarse con 12 hombres que puedan ocupar cualquier posición delantera y 10 hombres que puedan ocupar cualquiera de las demás posiciones?

19. ¿Cuántas señales puede transmitir un barco con 5 banderas diferentes si cada bandera puede ocupar 5 posiciones?

20. ¿Cuántas placas de matrícula de cinco símbolos pueden hacerse siendo los dos primeros letras y los tres últimos números?

21. ¿Cuántas diagonales tiene un polígono de 12 lados?

22. ¿Cuántas fichas componen un dominó que incluya desde la blanca doble al 12 doble?

23. ¿Cuál es la probabilidad de obtener un 7 con 2 dados?

24. ¿Cuál es la probabilidad de que 2 cartas extraídas de una baraja ordinaria sean espadas?

25. ¿Cuál es la probabilidad de que un grupo de 5 cartas contenga exactamente 2 ases? ¿Y de que contenga 2 o más ases?

26. ¿Cuál es la probabilidad de que una mano de *bridge* sea una serie completa?

27. Una urna contiene 5 bolas blancas, 4 rojas y 3 negras. Otra contiene 5 blancas, 6 rojas y 7 negras. Se elige una bola de cada urna. ¿Cuál es la probabilidad de que todas sean del mismo color?

28. Demostrar que $\binom{n}{r} = \binom{n}{n-r}$.

29. ¿De cuántas maneras pueden dividirse n objetos diferentes en k grupos que contengan $n_1, n_2, \dots, n_k$ objetos, si

$$n_1 + n_2 + \dots + n_k = n - m \quad \text{donde } n > m$$

30. Una urna contiene m bolas blancas y n negras. Se extraen k bolas y se colocan a un lado, sin fijarse en el color. Se saca otra bola. ¿Cuál es la probabilidad de que sea blanca?

31. Se lanzan 6 dados. ¿Cuál es la probabilidad de que aparezcan cada uno de los números posibles?

32. Se lanzan 7 dados. ¿Cuál es la probabilidad de que aparezcan cada uno de los números posibles?

33. ¿Cuál es la probabilidad de obtener con 3 dados un total de 4 puntos?

34. Una urna contiene 10 bolas numeradas del 1 al 10. Se extraen 4 bolas y se supone que x es el segundo en orden ascendente de magnitud de los cuatro números extraídos. ¿Cuál es la probabilidad de que $x=3$?

35. Si se echan n bolas en k cajas de modo que cada bola tenga igual probabilidad de caer en cualquiera de las cajas, ¿cuál es la probabilidad de que una caja determinada contenga m bolas?

36. Demostrar que

$$\sum_{i=1}^n CX_i = C \sum_{i=1}^n X_i.$$

37. Demostrar que

$$\prod_{i=1}^n CX_i^a = C^n \left(\prod_{i=1}^n X_i \right)^a.$$

38. Demostrar que

$$\left(\sum_{i=1}^n X_i \right)^2 = \sum_{i=1}^n \sum_{j=1}^n X_i X_j = \sum_{i=1}^n X_i^2 + \sum_{\substack{i=1 \\ i \neq j}}^n \sum_{j=1}^n X_i X_j$$

39. Demostrar que

$$\prod_{i=1}^{2n+1} (X + n + 1 - i) = X \prod_{i=1}^n (X^2 - i^2).$$

40. Hallar el coeficiente de $x^6 y^3$ en el desarrollo del binomio $(x^2 - ay)^5$.

41. Hallar el coeficiente de $x^2 y^2 z^3$ en el desarrollo del trinomio $(2x - y - z)^7$.

42. Si se lanzan 6 bolas en 3 cajas de modo que todas tengan igual probabilidad de caer en cualquiera de las cajas, ¿cuál es la probabilidad de que las 3 cajas queden ocupadas?

43. Se numeran los vértices de un tetraedro regular con los números 1, 2, 3, 4. Se lanzan 5 tetraedros. ¿Cuál es la probabilidad de que la suma de los vértices superiores sea 12?

44. Se retiran las espadas y los corazones de una baraja, colocándolos en fila y descubiertos. Se barajan las cartas restantes y se colocan descubiertas en una fila debajo de la anterior. ¿Cuál es la probabilidad de que todos los tréboles queden situados debajo de las espadas? ¿Cuál es la probabilidad de que entre los 26 pares de cartas, 16 consten de cartas de igual color?

45. Seis cartas se extraen de una baraja ordinaria. ¿Cuál es la probabilidad de que consten de una pareja (dos ases, o dos cincos, p. ej.) y 4 cartas que no formen pareja? ¿Y de que haya dos parejas y dos que no formen pareja?

46. Se retiran de una baraja ordinaria todas las figuras y se divide el resto en los cuatro palos. De cada palo se extrae una carta al azar. ¿Cuál es la probabilidad de que el total obtenido con las cuatro cartas sea 20?

47. Una urna contiene 3 bolas negras, 3 blancas y 2 rojas. Se extraen 3 bolas que se colocan en una caja negra, luego otras 3 que se colocan en una caja blanca, y las 2 restantes en una caja roja. ¿Cuál es la probabilidad de que todas las bolas menos 2 caigan en cajas de su mismo color?

48. Una urna contiene 5 bolas negras y 4 blancas; otra urna contiene 4 negras y 5 blancas. Se traslada una bola de la primera a la segunda urna; a continuación se extrae una bola de la segunda urna, ¿cuál es la probabilidad de que sea blanca?

49. En el problema anterior supongamos que se trasladan dos bolas, en lugar de una, de la primera a la segunda urna. Hallar la probabilidad de que una bola extraída a continuación de la segunda urna sea blanca.

50. Si se sabe que al lanzar 5 monedas aparecieron al menos 2 caras, ¿cuál es la probabilidad de que el número exacto de caras fuese 3?

51. Si un jugador de *bridge* tiene 7 espadas, ¿cuál es la probabilidad de que su compañero tenga al menos una espada? ¿Y de que tenga al menos dos espadas?

52. Si un jugador de *bridge* y su compañero tienen entre ambos 8 espadas, ¿cuál es la probabilidad de que las otras 5 espadas estén en dos grupos de 3 y 2 entre los dos jugadores contrarios?

53. Un jugador de *bridge* y su compañero tienen todas las espadas, excepto el rey, el tres y el dos. ¿Cuál es la probabilidad de que el rey lo tenga uno y el tres y el dos otro de los jugadores contrarios? ¿Cuál es la probabilidad de que uno determinado de los jugadores contrarios tenga el rey, o el rey y el dos, o el rey y el tres, o el rey, el tres y el dos?

54. Una persona lanza repetidas veces dos dados. Gana si saca un 8 antes de obtener un 7. ¿Cuál es la probabilidad de ganar?

NOTA: $1 + x + x^2 + x^3 + \dots = 1/(1-x)$, si $|x| < 1$.

55. En un juego de dados un jugador lanza dos veces un par de dados. Gana si los dos números obtenidos no difieren en más de 2, con las siguientes excepciones: si obtiene un 3 en la primera tirada deberá sacar un 4 en la segunda; si obtiene un 11 en la primera tirada, deberá obtener un 10 en la segunda. ¿Cuál es la probabilidad de ganar?

56. Consideremos el siguiente juego con dos dados: una persona lanza los dados, y gana si en la primera tirada obtiene 7 u 11; pierde si en la primera tirada obtiene 2, 3 ó 12. Si obtiene 4, 5, 6, 8, 9 ó 10 en la primera tirada, continúa arrojando los dados hasta que obtenga un 7 o el primero de los números que obtuvo; en el segundo caso gana y en el primero pierde. ¿Cuál es la probabilidad de ganar?

57. En la herencia mendeliana simple, cada característica física de una planta o animal viene determinada por un solo par de genes. Un ejemplo es el color de los guisantes. Representemos por a y v los colores amarillo y verde; los guisantes serán verdes si la planta tiene el par de genes (v, v) determinantes del color; serán amarillos si el par de genes es (a, a) o (a, v) . En vista de esto último, se dice que el color amarillo domina sobre el verde. La progenie recibe un gene de cada uno de los dos ascendientes y tiene igual probabilidad de obtener cualquiera de los dos genes de cada ascendiente. Si se cruza un guisante (a, a) con un guisante (v, v) , todos los guisantes resultantes del cruce son (a, v) y amarillos por la dominancia. Si se cruza un guisante (a, v) con uno (v, v) , la probabilidad de que cada guisante que resulta sea amarillo es 0,5 y la de que sea verde, 0,5 también. Si se efectúa un número grande de tales cruces es de esperar que aproximadamente la mitad resulten amarillos y la mitad verdes. Si se cruzan guisantes (a, v) y (a, v) , ¿cuál será la proporción esperada de amarillos? ¿Y cuál será la proporción esperada de (a, a) entre los amarillos?

58. Los guisantes pueden ser lisos o rugosos y esto constituye un carácter mendeliano simple. Liso domina sobre rugoso, de modo que los guisantes (l, l) y (l, r) son lisos y los (r, r) , rugosos. Si se cruzan guisantes (a, v) (l, r) con guisantes (v, v) (r, r) , ¿cuáles son los resultados posibles y cuáles sus probabilidades correspondientes? ¿Y para el cruce (a, v) (l, r) con (v, v) (l, r) ? ¿Y para el cruce (a, v) (l, r) con (a, v) (l, r) ?

59. El albinismo en los seres humanos es un carácter mendeliano simple. Representemos por a y n albino y no albino; el segundo es dominante, de modo que padres normales no pueden tener un hijo albino a menos que ambos sean (n, a) . Supongamos que en una población grande la proporción de genes n sea p y la proporción de genes a sea $q=1-p$, de modo que q^2 de los individuos sean albinos. Suponiendo que el albinismo no sea un factor en la selección matrimonial ni en el número de hijos de un matrimonio determinado, ¿qué proporción de individuos de la generación siguiente es de esperar que sean albinos? Si los albinos se casaran solo con albinos y tuvieran tantos hijos por promedio como los no albinos, ¿qué proporción de individuos de la generación siguiente es de esperar que sean albinos? ¿Qué ocurriría, finalmente, a la población si los albinos continuaran generación tras generación casándose con albinos? (Se supone que el número de individuos es el mismo en cada generación.)

60. Se sabe que una urna se ha llenado lanzando un dado y colocando bolas blancas en ella en número igual al obtenido con el lanzamiento del dado. A continuación se han añadido bolas negras en número determinado por una segunda tirada del dado. Se sabe también que el número total de bolas en la urna es 8. ¿Cuál es la probabilidad de que la urna contenga exactamente 5 bolas blancas?

61. Una urna A contiene 2 bolas blancas y 2 negras; una urna B contiene 3 bolas blancas y 2 negras. Se traslada una bola de la urna A a la B ; después se extrae una bola de la B , que resulta blanca. ¿Cuál es la probabilidad de que la bola trasladada fuese blanca?

62. Hay 6 urnas que contienen 12 bolas blancas y negras; una tiene 8 bolas blancas; dos, 6 bolas blancas, y tres, 4 bolas blancas. Se elige una urna al azar, y se extraen 3 bolas (sin reemplazamiento) de dicha urna; de estas, 2 son blancas y 1 negra. ¿Cuál es la probabilidad de que la urna elegida contuviera 6 bolas blancas y 6 negras?

63. En una ciudad se publican 3 periódicos, A , B y C . Realizada una encuesta, se estima que de la población adulta:

- 20% lee A
- 16% lee B
- 14% lee C
- 8% lee A y B
- 5% lee A y C
- 4% lee B y C
- 2% lee los tres.

¿Qué porcentaje lee al menos uno de estos periódicos? De los que leen al menos un periódico, ¿qué porcentaje lee A y B ?

64. Se lanzan 12 dados. ¿Cuál es la probabilidad de que cada una de las 6 caras aparezca al menos una vez?

65. Se lanza repetidamente un dado hasta que cada una de las 6 caras aparezca al menos una vez. ¿Cuál es la probabilidad de que haya que lanzarlo 10 veces?

BIBLIOGRAFIA

1. CRAMÉR, H.: *Elementos de la teoría de probabilidades y algunas de sus aplicaciones*, Aguilar, 6.ª ed., Madrid, 1968.
2. CRAMÉR, H.: *Métodos matemáticos de estadística*, Aguilar, 4.ª ed., Madrid, 1968.
3. FELLER, W.: *An Introduction to Probability Theory and Its Applications*, John Wiley & Sons, Inc., Nueva York, 1950.
4. KOLMOGOROFF, A. N.: *Foundations of the Theory of Probability*, 2.ª ed., Chelsea Publishing Company, Nueva York, 1950.
5. MUNROE, M.: *Theory of Probability*, McGraw-Hill Book Company, Inc., Nueva York, 1951.
6. PARZEN, E.: *Modern Probability Theory and Its Applications*, John Wiley & Sons, Inc., Nueva York, 1960.
7. USPENSKY, J.: *Introduction to Mathematical Probability*, McGraw-Hill Book Company, Inc., Nueva York, 1937.

CAPITULO 3

VARIABLES ALEATORIAS DISCRETAS

3-1. Introducción.—En el capítulo 2 hemos dado los axiomas para una función de probabilidad y explicado con detalle cómo se construye esta para un espacio muestral especial con un número finito de puntos, cada uno de los cuales tiene la misma probabilidad. En este capítulo nos ocuparemos de una variable aleatoria discreta.

Definición 3-1.—*Se dice que x es una variable aleatoria discreta unidimensional si es una variable aleatoria que toma solo un número finito o infinito numerable de valores del eje x . Supongamos que x toma únicamente los valores $x_1, x_2, \dots, x_n, \dots$ con probabilidades $f(x_1), f(x_2), \dots, f(x_n), \dots$ e imaginemos que A es cualquier subconjunto de los puntos $x_1, x_2, \dots, x_n, \dots$. La probabilidad $P(A)$ del suceso A (probabilidad de que x esté en A), se define como*

$$P(A) = \sum_A f(x)$$

donde $\sum_A f(x)$ representa la suma de $f(x)$ para aquellos valores x_i que pertenecen a A .

Así, p. ej., $P(x=2)$ quiere decir probabilidad de que el valor de la variable aleatoria sea igual a 2. $P(3 < x < 5)$ significa probabilidad de que el valor de la variable aleatoria esté comprendido entre 3 y 5, etc.

Naturalmente, en una situación experimental, $f(x)$ se toma de tal forma que se ajuste al problema particular considerado. Llamaremos a $f(x)$ *función de densidad de una variable discreta* o, simplemente, *función de cuantía* y a veces diremos « x se distribuye según $f(x)$ », o « $f(x)$ es la distribución de x ». Cualquier función puede ser una función de densidad si satisface a

$$\begin{aligned} f(x_i) &\geq 0 & i=1, 2, \dots \\ \sum f(x_i) &= 1 \end{aligned}$$

Como ejemplo, supongamos que un experimento aleatorio consiste en lanzar cuatro monedas simétricas y registrar el número de caras. Sea el resultado el valor de una variable aleatoria x ; por tanto, x toma los valores 0, 1, 2, 3, 4. Para calcular la función de densidad $f(x)$, probabilidad de que aparezcan x caras, ob-

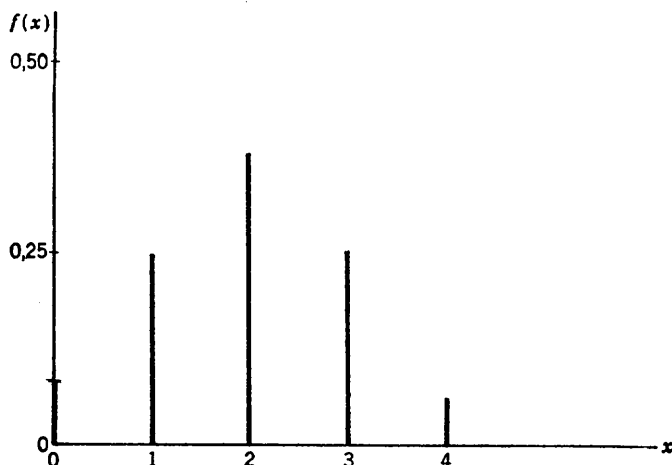


FIG. 3-1.

servamos que el número de formas en que pueden caer las cuatro monedas es 2^4 , ya que cada una de ellas puede caer de dos maneras distintas. El número de formas en que pueden aparecer x caras es $\binom{4}{x}$; por tanto,

$$f(x) = \frac{\binom{4}{x}}{2^4} \quad x = 0, 1, 2, 3, 4$$

Puesto que

$$\sum_{x=0}^4 f(x) = \frac{1}{2^4} \sum_{x=0}^4 \binom{4}{x} = \frac{2^4}{2^4} = 1$$

y $f(x) \geq 0$ para $0 \leq x \leq 4$, $f(x)$ es una función de densidad. A veces escribiremos $0 \leq x \leq 4$ para expresar que $x = 0, 1, 2, 3, 4$.

Dando a x cada uno de sus valores posibles, podemos calcular $f(x)$ y representar la función como en la figura 3-1, utilizando segmentos verticales de longitud igual a $f(x)$ según una cierta escala.

Conviene considerar que $f(x)$ da las frecuencias relativas o frecuencias con que se presenta cada uno de los valores de x . Así, si suponemos que las cuatro monedas se lanzan un gran número de veces, esperaremos que no aparezcan caras ($x=0$) en $1/16$ aproximadamente de las tiradas; esperaremos que aparezca una cara ($x=1$) en la cuarta parte aproximadamente de las tiradas, y así sucesivamente. La representación gráfica de esta función hace evidente en seguida varias cosas: que el número más probable de caras es dos; que es de esperar que se presente una cara con frecuencia aproximadamente cuatro veces mayor que la correspondiente a ninguna cara; que es de esperar que tres caras ocurran con la misma frecuencia aproximada que una cara, y así sucesivamente. Decimos *aproximadamente* porque ya estamos familiarizados con las fluctuaciones que acompañan a los sucesos aleatorios. Así, si tiramos diez veces una sola moneda, esperamos por término medio cinco caras y cinco cruces; pero, en realidad, es muy probable que en una prueba dada sean distintos el número de caras y el de cruces.

Los resultados de un experimento real de lanzamiento de 4 monedas pueden verse en la siguiente tabla. Se lanzaron 4 monedas 160 veces, contando el número de caras aparecidas en cada prueba.

RESULTADOS DEL LANZAMIENTO DE 4 MONEDAS 160 VECES

Número de caras	Ocurrencias efectivas	Ocurrencias esperadas
0	6	10
1	41	40
2	56	60
3	45	40
4	12	10
	160	160

La concordancia entre las ocurrencias efectivas y las esperadas no es demasiado buena (conviene recordar que la probabilidad de obtener cara tal vez no fuera exactamente $1/2$ para cada una de las monedas que se utilizaron); sin embargo, el carácter general

de la distribución de los resultados efectivos queda representado bastante bien por la función de cuantía $f(x)$.

Conocida la función de densidad de una variable aleatoria x , podemos dar respuesta a cualquier cuestión probabilística relativa a x . Refiriéndonos de nuevo a nuestro ejemplo particular, la probabilidad de obtener 2 caras es

$$P(x=2)=f(2)=\frac{\binom{4}{2}}{2^4}=\frac{3}{8}$$

La probabilidad de que el número de caras sea inferior a 3 es

$$P(x < 3) = \sum_{x=0}^2 f(x) = 11/16$$

La probabilidad de que el número de caras esté entre 1 y 3, ambos inclusive, es.

$$P(1 \leq x \leq 3) = \sum_{x=1}^3 f(x) = 7/8$$

Imaginemos que deseamos calcular la probabilidad condicional de que el número de caras sea menor que tres cuando se sabe que dicho número es menor que cuatro. Sea A el suceso «aparecen menos de tres caras»; es decir,

$$A = \{x: x=0, 1, 2\}$$

Sea B el suceso «aparecen menos de cuatro caras»; esto es,

$$B = \{x: x=0, 1, 2, 3\}$$

Deseamos calcular $P(A|B)$. Por definición de probabilidad condicional,

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Ahora bien:

$$A \cap B = \{x: x=0, 1, 2\}$$

luego

$$P(A \cap B) = \sum_{x=0}^2 f(x) = \frac{\sum_{x=0}^2 \binom{4}{x}}{2^4} = \frac{11}{16}$$

También

$$P(B) = \sum_{x=0}^3 f(x) = \frac{\sum_{x=0}^3 \binom{4}{x}}{2^4} = \frac{15}{16}$$

de donde

$$P(A|B) = \cancel{P}(x < 3 | x \leq 4) = \frac{^{11/16}}{^{15/16}} = \frac{11}{15}$$

La interpretación frecuencial es la siguiente: Supongamos que cuatro monedas ideales se lanzan un gran número de veces y se registra el número de caras de cada tirada *solamente* en los casos en que aparecen menos de cuatro caras. La fracción de estos casos (donde aparecen menos de cuatro caras) en que aparecen menos de tres caras será aproximadamente $^{11/15}$. Resultará instructivo para el lector realizar tal experimento y comprobar este resultado.

3-2. Funciones de cuantía.—Las propiedades esenciales de estas funciones ya han sido indicadas en la sección anterior, y nos basta considerarlas ahora de manera algo más general.

El conjunto de resultados posibles de un suceso aleatorio se divide en cierto número de clases mutuamente excluyentes en relación con determinado atributo. A cada clase se le asocia un valor de una *variable aleatoria*, o *variante* x . La función de cuantía es una función que da la probabilidad de que ocurra un valor determinado de x .

La variante x puede, naturalmente, describir un atributo, como era el caso en el lanzamiento de monedas, o puede ser simplemente el resultado de una codificación. Así, al extraer bolas de una urna, pueden clasificarse según su color, y podríamos definir una variable aleatoria x estableciendo arbitrariamente una correspondencia entre los valores de x y los colores: haciendo corresponder $x=1$, a negro; $x=2$, a rojo, y así sucesivamente. Al extraer una bola roja, la variante tomaría el valor 2.

La función de cuantía puede ser una expresión matemática, como en el caso de la sección anterior, o bien reducirse a una tabla de valores. Así, si una urna contiene 3 bolas negras, 2 rojas y 5 blan-

cas, podemos codificar los colores con 1, 2, 3, respectivamente. No nos molestamos en construir una expresión matemática, sino que nos limitamos a tabular la función:

$$\begin{array}{rcccl} x: & 1 & 2 & 3 & \\ f(x): & 0,3 & 0,2 & 0,5 & \end{array}$$

Utilizamos el calificativo *discreta* para distinguir una variante de este tipo de las variantes *continuas*, que se discutirán en el capítulo siguiente.

Distribuciones acumulativas.—A menudo es necesario calcular probabilidades del tipo $P(x < 3)$, $P(1 \leq x \leq 4)$, etc. En estos casos y también en otras situaciones es conveniente definir una nueva función, llamada *función de distribución acumulativa*. Para una función de cuantía $f(x_i)$, $i=1, 2, \dots$, la distribución acumulativa $F(x)$ se define por

$$F(x) = \sum f(x_i)$$

donde se suma para todos los valores de i tales que $x_i \leq x$. Es fácil ver que

$$F(x) = P(x \leq x)$$

y que

$$P(a < x \leq b) = F(b) - F(a)$$

Por consiguiente, puede demostrarse que para una variable aleatoria discreta es posible obtener la distribución acumulativa a partir de la función de densidad, y viceversa.

3.3. Distribuciones multivariantes.—Cuando el resultado de un suceso aleatorio pueda clasificarse de más de una forma, la función de cuantía es una función de más de una variable. Así, al extraer una carta de una baraja ordinaria, cabe caracterizarla según su palo y según su denominación. Sea x una variable aleatoria que toma los valores 1, 2, 3, 4 que hacemos corresponder a los palos en determinado orden (p. ej., picas, corazones, diamantes y tréboles), y sea y una variable aleatoria que toma los valores 1, 2, ..., 13, correspondientes a las denominaciones as, dos, ..., diez, J, Q y K. Entonces (x, y) es una variable aleatoria bidimensional. La probabilidad de extraer una carta determinada se representa por $f(x, y)$ y si cada carta tiene la misma probabilidad de ser extraída, la función de cuantía de (x, y) es, evidentemente,

$$f(x, y) = 1/52 \quad 1 \leq x \leq 4, 1 \leq y \leq 13 \quad (1)$$

Esta función puede representarse sobre un plano como en la figura 3-2; las probabilidades vienen representadas por segmentos verticales en los puntos (x, y) del plano horizontal en que están definidas. En este caso, los segmentos son de igual altura. Consideremos otro ejemplo: supongamos que se extraen 4 bolas de una urna

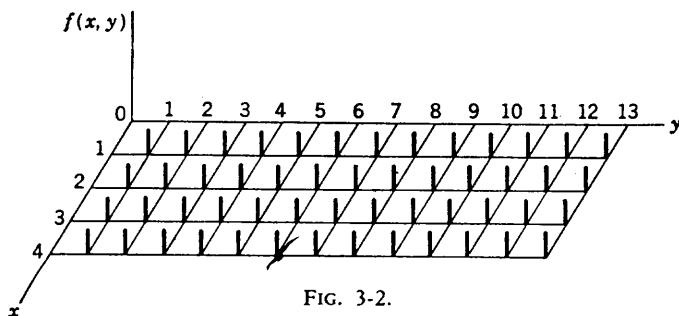


FIG. 3-2.

que contiene 5 bolas negras, 6 blancas y 7 rojas. Sea x el número de bolas blancas extraídas e y el de bolas rojas. La función de cuantía de la variable bidimensional (x, y) es

$$f(x, y) = \frac{\binom{6}{x} \binom{7}{y} \binom{5}{4-x-y}}{\binom{18}{4}} \quad 0 \leq x+y \leq 4 \quad (2)$$

y su representación gráfica puede verse en la figura 3-3. En este ejemplo cabría considerar una tercera variable z , número de bolas negras extraídas, obteniendo así una distribución tri variante. Pero z viene determinado exactamente por x e y , ya que $z = 4 - x - y$. No proporciona nueva información el añadir z al conjunto de variables aleatorias que caracterizan los resultados y, en realidad, si se incluye z en la función de cuantía, el conjunto de probabilidades representado por tal función, $f(x, y, z)$, será exactamente el mismo que el obtenido utilizando x e y .

Un ejemplo más sencillo de dependencia funcional es el del lanzamiento de una moneda un cierto número de veces. Supongamos que se lanza 4 veces y sea x el número de caras e y el número de cruces. Puesto que $x+y$ debe ser igual a 4, las variables son funcionalmente dependientes; si se conoce una, la otra queda determinada con exactitud. La función es

$$f(x, y) = \binom{4}{x} \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^y \quad x+y=4; \quad x, y=0, 1, 2, 3, 4$$

y está representada en la figura 3-4. No proporciona más información que la utilizada en el ejemplo de la sección 3-1; el conjunto de probabilidades es exactamente el mismo que antes.

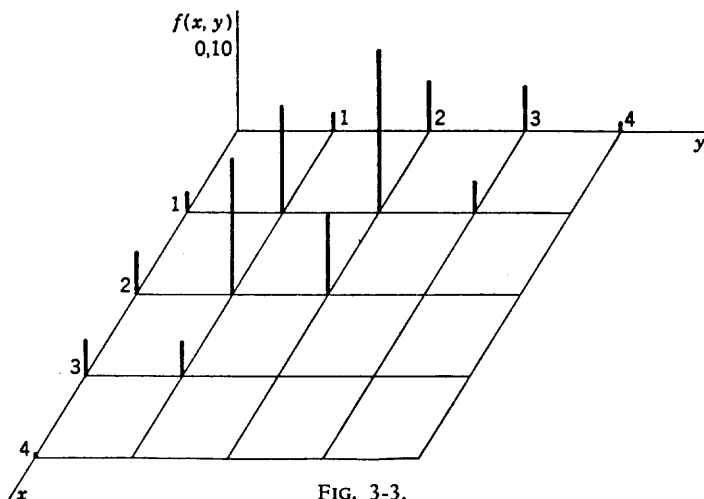


FIG. 3-3.

Hemos usado los términos dependiente e independiente desde dos puntos de vista totalmente distintos. En el capítulo 2 definimos

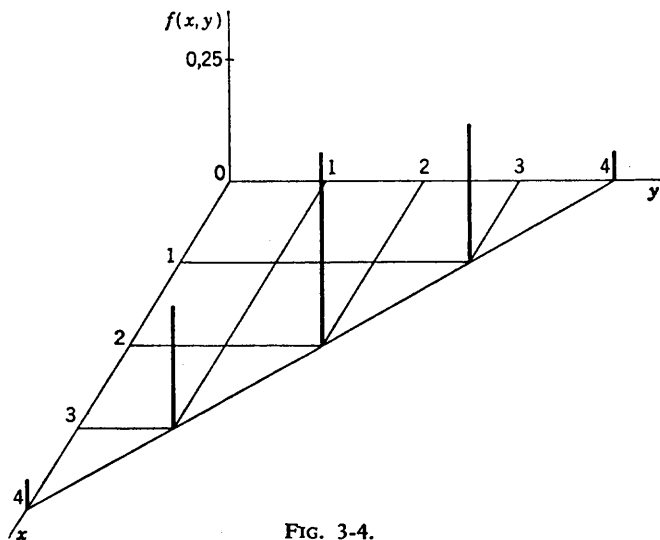


FIG. 3-4.

dos sucesos como independientes si la probabilidad condicional de uno, dado el otro, es igual a la probabilidad marginal del primero. En lo que sigue nos referiremos a este tipo de independencia como *independencia en sentido probabilístico*. Volviendo al ejemplo de la urna: x e y son *funcionalmente independientes* (ya que y no está determinada unívocamente cuando se conoce x), pero son *dependientes en sentido probabilístico* (como veremos más adelante).

En general, la variable aleatoria k -dimensional $(x_1, x_2, \dots, x_k)$ es discreta si solo puede tomar valores en un número finito o infinito numerable de puntos $(x_1, x_2, \dots, x_k)$ del espacio real k -dimensional. Designemos por $f(x_1, x_2, \dots, x_k)$ la probabilidad de que el valor de la variable aleatoria sea $(x_1, x_2, \dots, x_k)$, es decir,

$$P(x_1=x_1, x_2=x_2, \dots, x_k=x_k)=f(x_1, x_2, \dots, x_k)$$

para cada valor que puede tomar la variable aleatoria; $f(x_1, x_2, \dots, x_k)$ se denomina función de cuantía (conjunta) de la variable aleatoria k -dimensional. Sea A un subconjunto del conjunto de valores que puede tomar la variable aleatoria; entonces

$$P[(x_1, x_2, \dots, x_k) \text{ esté en } A] = \sum_A f(x_1, x_2, \dots, x_k)$$

donde $\sum_A f(x_1, x_2, \dots, x_k)$ significa: súmese la función de cuantía para todos los puntos de A . Sea $x_{i_1}, x_{i_2}, \dots, x_{i_t}$ cualquier subconjunto de las variables aleatorias discretas $x_1, x_2, \dots, x_k$. La distribución marginal de la variable aleatoria t -dimensional $(x_{i_1}, x_{i_2}, \dots, x_{i_t})$ es

$$f_{i_1, i_2, \dots, i_t}(x_{i_1}, x_{i_2}, \dots, x_{i_t}) = \sum f(x_1, x_2, \dots, x_k)$$

donde la suma se extiende a todos los valores, a excepción de $x_{i_1}, x_{i_2}, \dots, x_{i_t}$. Sean $x_{i_1}, x_{i_2}, \dots, x_{i_t}$ y $x_{j_1}, x_{j_2}, \dots, x_{j_s}$ dos subconjuntos disjuntos de las variables aleatorias discretas $x_1, x_2, \dots, x_k$. La distribución condicional de la variable aleatoria t -dimensional $(x_{i_1}, x_{i_2}, \dots, x_{i_t})$ dado el valor $(x_{j_1}, x_{j_2}, \dots, x_{j_s})$ de $(x_{j_1}, x_{j_2}, \dots, x_{j_s})$ es

$$\begin{aligned} g(x_{i_1}, x_{i_2}, \dots, x_{i_t} | x_{j_1}, \dots, x_{j_s}) &= \\ &= \frac{f_{i_1, i_2, \dots, i_t, j_1, j_2, \dots, j_s}(x_{i_1}, x_{i_2}, \dots, x_{i_t}, x_{j_1}, x_{j_2}, \dots, x_{j_s})}{f_{j_1, j_2, \dots, j_s}(x_{j_1}, x_{j_2}, \dots, x_{j_s})} \end{aligned}$$

para todos los valores de x para los cuales no se anula el denominador. Las variables aleatorias discretas $x_1, x_2, \dots, x_k$ son independientes (mutuamente) si y solamente si

$$f(x_1, x_2, \dots, x_k) = f_1(x_1)f_2(x_2) \dots f_k(x_k)$$

para todos los valores $(x_1, x_2, \dots, x_k)$ en que está definida la variable aleatoria $(x_1, x_2, \dots, x_k)$.

En el ejemplo de la urna, la distribución marginal de x es

$$f_1(x) = \sum_{y=0}^{4-x} f(x, y) = \frac{\binom{6}{x} \binom{12}{4-x}}{\binom{18}{4}} \quad 0 \leq x \leq 4 \quad (3)$$

La suma puede obtenerse mediante una identidad algebraica, pero aquí es más sencillo considerar el problema de nuevo como si se tratase de 6 bolas blancas y 12 no blancas. Análogamente, la distribución marginal de y es

$$f_2(y) = \sum_{x=0}^{4-y} f(x, y) = \frac{\binom{7}{y} \binom{11}{4-y}}{\binom{18}{4}} \quad 0 \leq y \leq 4 \quad (4)$$

Se ha representado esta función en la figura 3-5. La altura correspondiente a $y=0$, representación de $f_2(0)$, es igual a la suma de las alturas correspondientes a las verticales que se encuentran al recorrer el eje x en la figura 3-3; $f_2(1)$ es la suma de las alturas de las verticales al recorrer la línea $y=1$ en la figura 3-3, y así sucesivamente.

La distribución condicional de x , dado y , en el problema anterior de la urna es

$$g(x|y) = \frac{f(x, y)}{f_2(y)} = \frac{\binom{6}{x} \binom{5}{4-x-y}}{\binom{11}{4-y}} \quad \begin{array}{l} 0 \leq x \leq 4-y \\ y=0, 1, 2, 3, 4 \end{array}$$

Análogamente,

$$h(y|x) = \frac{\binom{7}{y} \binom{5}{4-x-y}}{\binom{12}{4-x}} \quad \begin{array}{l} 0 \leq y \leq 4-x \\ x=0, 1, 2, 3, 4 \end{array}$$

Si se da a x un valor determinado, p. ej., $x=1$, puede representarse la función $h(y|1)$ dando a y los valores sucesivos 0, 1, 2, 3. Las verticales tendrían las mismas alturas relativas que las situadas a lo largo de la recta $x=1$ en la figura 3-3; sus lon-

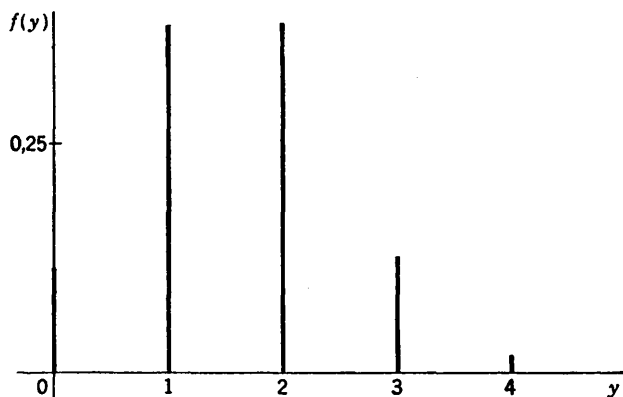


FIG. 3-5.

gitudes vienen multiplicadas por el factor $1/f_1(x)$ correspondiente a $x=1$, de modo que la suma de todas estas longitudes sea igual a la unidad. Observamos que $h(y|x)$ no es igual a la distribución marginal de y , de modo que y y x no son independientes en el sentido probabilístico. Desde luego, el hecho de que $h(y|x)$ contenga a x es suficiente para poner de manifiesto que ambas variables dependen una de otra en sentido probabilístico. No obstante, aunque tuviéramos un caso en que no interviniera x en $h(y|x)$ sería posible, sin embargo, que las variables dependieran una de otra por depender de x los límites de y . Ahora bien: si x no interviene ni en $h(y|x)$ ni en los límites de y , las dos variables serán evidentemente independientes en sentido probabilístico.

Como ejemplo de una distribución de más de dos variantes, supongamos que se extraen 12 cartas, sin reemplazamiento, de una baraja ordinaria, y sea x_1 el número de ases, x_2 el número de do- ses, x_3 el número de treses, y x_4 el número de cuatros. La distri-

bución de estas variantes viene dada por una función de cuatro variables, a saber:

$$f(x_1, x_2, x_3, x_4) = \frac{\binom{4}{x_1} \binom{4}{x_2} \binom{4}{x_3} \binom{4}{x_4} \binom{36}{12-x_1-x_2-x_3-x_4}}{\binom{52}{12}}$$

siendo el recorrido de cada variante $0 \leq x_i \leq 4$, con la restricción $\sum x_i \leq 12$. Hay un gran número de distribuciones marginales y condicionales asociadas con esta distribución. Algunos ejemplos son:

$$f_{23}(x_2, x_3) = \frac{\binom{4}{x_2} \binom{4}{x_3} \binom{44}{12-x_2-x_3}}{\binom{52}{12}} \quad 0 \leq x_i \leq 4$$

$$f_4(x_4) = \frac{\binom{4}{x_4} \binom{48}{12-x_4}}{\binom{52}{12}} \quad 0 \leq x_4 \leq 4$$

$$g(x_2, x_4 | x_1, x_3) = \frac{\binom{4}{x_2} \binom{4}{x_4} \binom{36}{12-x_1-x_2-x_3-x_4}}{\binom{44}{12-x_1-x_3}} \quad \begin{array}{l} 0 \leq x_i \leq 4 \\ x_2 + x_4 \leq 12 - x_1 - x_3 \end{array}$$

donde las dos primeras son distribuciones marginales y la tercera una distribución condicional. La misma distribución $f(x_1, x_2, x_3, x_4)$ puede, a su vez, considerarse como distribución marginal de alguna distribución más detallada, como, p. ej., la distribución en las seis variantes $x_1, x_2, x_3, x_4, x_5, x_6$, donde x_5 y x_6 son los números de cincos y seises que aparecen en la extracción de las 12 cartas.

No podemos representar la distribución en cuatro variantes; en realidad, hemos utilizado ya las tres dimensiones del espacio ordinario para representar las distribuciones bivariantes. Podríamos haber evitado esto utilizando otro procedimiento; sería posible utilizar circuitos de tamaños distintos en lugar de rectas verticales, representando de este modo en dos dimensiones las distribuciones bivariantes. Este método no hubiera proporcionado una representación tan clara de la magnitud relativa de las probabilidades. Con los circuitos podríamos conseguir la representación

gráfica de una distribución trivariante; pero para más de tres variantes ya no es posible encontrar una representación gráfica sencilla.

La probabilidad de un suceso determinado se obtiene sumando la función de cuantía para todos los puntos de la región definida por aquel. Supongamos la función bivalente $f(x, y)$ definida para $x=0, 1, 2, \dots, r$ e $y=0, 1, 2, \dots, s$. La probabilidad de que $x \leq 5$ e

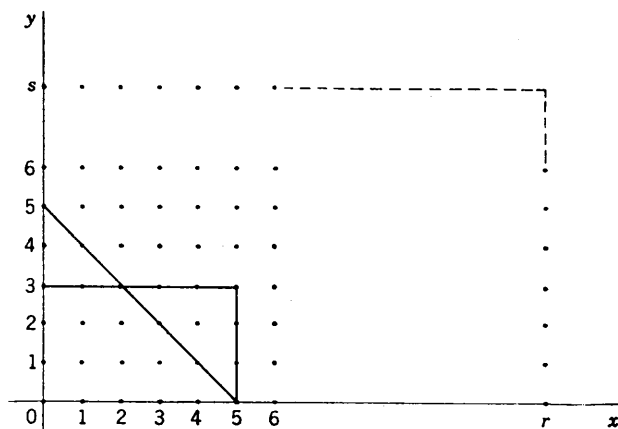


FIG. 3-6.

$y \leq 3$ se obtiene sumando $f(x, y)$ en la región definida por las desigualdades (rectángulo de la figura 3-6)

$$P(x \leq 5, y \leq 3) = \sum_{x=0}^5 \sum_{y=0}^3 f(x, y)$$

La probabilidad de que la suma de x e y sea inferior a 5 es igual a la suma de $f(x, y)$ extendida a todos los puntos interiores al triángulo limitado por la recta $x+y=5$:

$$\begin{aligned} P(x+y < 5) = & f(0, 0) + f(1, 0) + f(2, 0) + f(3, 0) + f(4, 0) \\ & + f(0, 1) + f(1, 1) + f(2, 1) + f(3, 1) \\ & + f(0, 2) + f(1, 2) + f(2, 2) \\ & + f(0, 3) + f(1, 3) \\ & + f(0, 4) \end{aligned}$$

$$= \sum_{x=0}^4 \sum_{y=0}^{4-x} f(x, y) = \sum_{y=0}^4 \sum_{x=0}^{4-y} f(x, y)$$

Otros ejemplos son

$$P(x+y=5) = \sum_{x=0}^5 f(x, 5-x)$$

$$P(x \leq 2 | y = 3) = \sum_{x=0}^2 g(x|3)$$

$$\sum_{x=0}^2 f(x, 3)$$

$$= \frac{\sum_{x=0}^r f(x, 3)}{\sum_{x=0}^r f(x, 3)}$$

$$P(x \leq 2 | y > 3) = \frac{\sum_{x=0}^2 \sum_{y=4}^s f(x, y)}{\sum_{x=0}^r \sum_{y=4}^s f(x, y)}$$

$$P(x+y=2 | x^2+y^2 \leq 5) =$$

$$= \frac{f(0, 2) + f(1, 1) + f(2, 0)}{f(0, 0) + f(0, 1) + f(0, 2) + f(1, 0) + f(1, 1) + f(1, 2) + f(2, 0) + f(2, 1)}$$

Para tres variables puede ocurrir que sea difícil obtener la imagen de dichas regiones, y para más de tres variables tenemos que conformarnos con la descripción analítica de la región para determinar las sumas buscadas. Algunos ejemplos relativamente sencillos son

$$P(x \leq 3, y \leq 4, 2 \leq z \leq 6) = \sum_{x=0}^3 \sum_{y=0}^4 \sum_{z=2}^6 f(x, y, z)$$

$$P(x+y=4 | z=2) = \sum_{x=0}^4 f(x, 4-x | 2)$$

$$P(x+y+z \leq 6) = \sum_{x=0}^6 \sum_{y=0}^{6-x} \sum_{z=0}^{6-x-y} f(x, y, z)$$

$$P(x+y+z=6) = \sum_{x=0}^6 \sum_{y=0}^{6-x} f(x, y, 6-x-y)$$

3-4. Distribución binomial.—La distribución binomial es, probablemente, la de uso más frecuente, entre las distribuciones discretas, en las aplicaciones de la teoría de la estadística. Es la distribución asociada con las pruebas repetidas de un mismo suceso. Supongamos que se representa por p la probabilidad de éxito o de acierto en determinado suceso. Este puede ser la ocurrencia de una cara en el lanzamiento de una moneda, en cuyo caso $p=1/2$; o bien la aparición de un 7 al arrojar dos dados, en cuyo caso $p=1/6$; también puede ser la salida de 2 o más ases al extraer 5 cartas de una baraja ordinaria, en cuyo caso

$$p = \frac{\binom{4}{2} \binom{48}{3} + \binom{4}{3} \binom{48}{2} + \binom{4}{4} \binom{48}{1}}{\binom{52}{5}}$$

o, más generalmente, p puede representar la probabilidad de que se produzca cualquier suceso efectivo al que no pueda asignársele una probabilidad numérica *a priori*.

Cualquiera que sea el suceso, si la probabilidad de que ocurra es p , la probabilidad de que no ocurra es $1-p$, ya que no es posible que el suceso ocurra y no ocurra a la vez en una prueba dada. Resulta cómodo designar $1-p$ por q , y al hablar de una prueba determinada diremos que la probabilidad de un acierto es p y la de obtener un fallo es q , donde $p+q=1$.

Definición 3-2.—Se dice que la variable aleatoria x se distribuye según una binomial puntual si la función de cuantía está dada por

$$f(x) = p^x q^{1-x} \quad x=0, 1; \quad 0 \leq p \leq 1$$

La variable x , que solo toma los valores 0 y 1, puede interpretarse de la siguiente manera: $f(1)$ representa la probabilidad de un éxito; $f(0)$ representa la probabilidad de un fallo. Evidentemente,

$$f(0) = q = 1 - p; \quad f(1) = p.$$

Así, p. ej., si se lanzan dos dados y la probabilidad de un siete se toma igual a $1/6$, la función de cuantía de x , número de sietes que aparecen en una tirada, es

$$f(x) = (1/6)^x (5/6)^{1-x} \quad x=0, 1$$

La probabilidad de un siete es $f(1)=1/6$; la probabilidad de ningún siete es $f(0)=5/6$. A continuación se define otra importante distribución.

Definición 3-3.—Sea una variable aleatoria x con función de cuantía dada por

$$f(x) = \binom{n}{x} p^x q^{n-x} \quad x=0, 1, 2, \dots, n$$

Esta es la distribución binomial.

Para ver cómo puede originarse esta distribución, procederemos como sigue: El resultado de la i -ésima prueba forma una variable aleatoria unidimensional que designaremos por x_i , siendo $x_i=0$ si el resultado de la i -ésima prueba constituye un fallo y $x_i=1$ si dicho resultado es un éxito. Sea $f(x_1, x_2, \dots, x_n)$ la probabilidad de que la variable aleatoria $x_i=x_i$, $i=1, 2, \dots, n$, siendo $x_i=0$ ó 1. Puesto que las variables aleatorias x_i son independientes, empleamos la definición de independencia para obtener

$$f(x_1, x_2, \dots, x_n) = f(x_1) f(x_2) \dots f(x_n)$$

Pero por la hipótesis de que cada variable aleatoria admite una distribución binomial puntual, se tiene

$$\begin{aligned} f(x_1, x_2, \dots, x_n) &= f(x_1) f(x_2) \dots f(x_n) \\ &= p^{x_1} q^{1-x_1} p^{x_2} q^{1-x_2} \dots p^{x_n} q^{1-x_n} \\ &= p^{\sum x_i} q^{n-\sum x_i} \quad x_i=0, 1 \end{aligned}$$

Así, p. ej., para el conjunto particular 0 0 1 0 1 1 1, tendríamos

$$f(0, 0, 1, 0, 1, 1, 1) = q q p q p p p$$

Evidentemente, para cualquier ordenación de k éxitos (1) y $n-k$ fracasos (0), la probabilidad es $p^k q^{n-k}$, puesto que $\sum x_i = k$ si hay k unos y $n-k$ ceros. Pero el número total de formas posibles de ordenar k unos y $n-k$ ceros es $\binom{n}{k}$; por tanto, la probabilidad de k éxitos exactamente es

$$\binom{n}{k} p^k q^{n-k} \quad k=0, 1, 2, \dots, n$$

En resumen, se tiene:

1. Si la probabilidad de que ocurra un suceso es p y la probabilidad de que no ocurra es $q=1-p$, la función de cuantía de la variable aleatoria x (número de ocurrencias) es

$$f(x) = p^x q^{1-x} \quad x=0, 1$$

Esta es la distribución binomial puntual.

2. En n pruebas, sea p la probabilidad (en cada prueba) de que ocurra un suceso, y supongamos independientes todas las pruebas. La función de cuantía de la variable aleatoria x (número de ocurrencias del suceso en n pruebas) es

$$f(x) = \binom{n}{x} p^x q^{n-x} \quad x = 0, 1, 2, \dots, n$$

Esta es la *distribución binomial*.

Esta función contiene otras dos variables p y n (no se cuenta q porque viene determinada por p) de carácter distinto; su variación corresponde a distribuciones binomiales distintas; para una distribución binomial determinada, p y n deben tener valores numéricos dados. Las variables de este tipo reciben el nombre de *parámetros*.

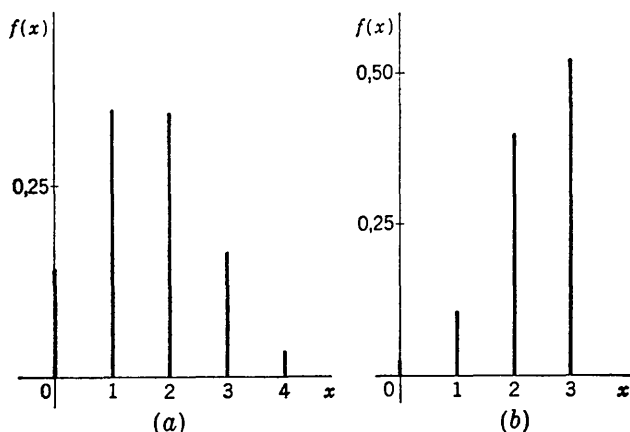


FIG. 3-7.

La función representa, pues, una familia de distribuciones con dos parámetros, y se obtiene un miembro determinado de esta familia al especificar los valores de p y n . El parámetro n recibe el nombre de *parámetro discreto*, ya que sólo puede tomar los valores aislados 1, 2, 3, ...; y carecería de sentido hablar de, p. ej., 2,53 pruebas. En cambio, p es un *parámetro continuo*, ya que puede concebirse para el mismo cualquier valor desde 0 hasta 1. Así, es posible que p valga 0,5 en el caso de una moneda bien construida, o, p. ej., 0,5000037 en el caso de una moneda ligeramente sesgada. Cualquier número arbitrariamente elegido entre 0 y 1 es un valor posible de p .

En la figura 3-7 se han representado dos casos particulares de la distribución binomial. En *a*) la que tiene por parámetros $p=0,4$ y $n=4$, y en *b*) la correspondiente a $p=0,8$ y $n=3$. En general, la distribución binomial tendrá un valor máximo que se determina como sigue: Sea m la parte entera del número $(n+1)p$ y sea e la parte fraccionaria. Por tanto, si $n=7$ y $p=0,3$, tendremos $m=2$ y $e=0,4$. El valor mayor de $f(x)$ se presenta al hacer $x=m$; m recibe el nombre de *valor modal* o simplemente *moda* de x . Para demostrar que este valor de x da el valor máximo de $f(x)$, supongamos por un mo-

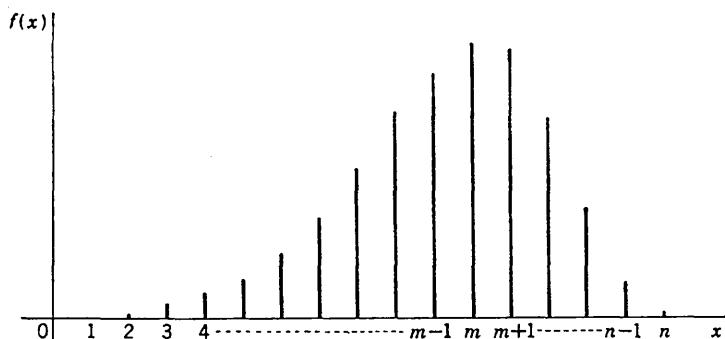


FIG. 3-8.

mento que e es distinto de cero, y formemos la razón $f(x+1)/f(x)$. Vamos a ver que esta razón es inferior a uno cuando x es mayor o igual que m , y superior a uno cuando x es inferior a m . Se trata de una situación como la que se ilustra en la figura 3-9. Ahora bien:

$$\frac{f(x+1)}{f(x)} = \frac{p}{q} \frac{n-x}{x+1}$$

y si x es igual o mayor que m tenemos

$$\frac{p}{q} \frac{n-x}{x+1} \leq \frac{p}{q} \frac{n-m}{m+1}$$

Sustituyendo m por $(n+1)p - e$, el segundo miembro puede escribirse

$$\frac{p}{q} \frac{n-m}{m+1} = \frac{(n+1) - [(1-e)/q]}{(n+1) + [(1-e)/p]}$$

que es ciertamente menor que la unidad. Si x es menor que $m-1$,

$$\begin{aligned}\frac{p}{q} \frac{n-x}{x+1} &> \frac{p}{q} \frac{n-(m-1)}{m} \\ &> \frac{p}{q} \frac{(n+1)q+e}{(n+1)p-e} \\ &> \frac{n+1+e/q}{n+1-e/p}\end{aligned}$$

que es, por tanto, mayor que 1. Hemos omitido el caso

$$x=m-1$$

para el cual

$$\begin{aligned}\frac{f(x+1)}{f(x)} &= \frac{p}{q} \frac{n-m+1}{m} \\ &= \frac{(n+1)+e/q}{(n+1)-e/p}\end{aligned}$$

que es también superior a 1 si e es distinto de 0. Si $e=0$, la razón es igual a la unidad, y $f(m)=f(m-1)$; hay dos valores máximos de $f(x)$ que son iguales y que corresponden a $x=m$ y a $x=m-1$. Esta situación es la representada en la figura 3-7 (a) en donde $(n+1)p=2$ es un número entero exacto, de modo que $f(1)$ y $f(2)$ son dos valores máximos iguales de $f(x)$.

Para valores grandes de n el aspecto de la distribución binomial es en general como el de la figura 3-8. En la figura 3-7 (b) la moda corresponde a $x=n$, para $q=0,8$ y $n=3$; pero al aumentar n , la moda se aleja del extremo derecho del recorrido; así, para $n=100$, tenemos

$$101 \times 0,8 = 80,8$$

de forma que la moda es 80 y está muy alejada del valor extremo $x=100$.

El cálculo de probabilidades binomiales resulta laborioso cuando n es grande. Pueden emplearse métodos aproximados para el cálculo de $\binom{n}{x} p^x q^{n-x}$; pero omitiremos estos porque raras veces se necesita el cálculo de términos aislados. En la mayor parte de las aplicaciones, lo que se precisan son sumas parciales; así, podemos ne-

cesitar la probabilidad de que x sea superior a un número entero a ,

$$P(x > a) = \sum_{x=a+1}^n f(x)$$

En los capítulos 7 y 11 daremos métodos de cálculo para dichas sumas.

3-5. Distribución polinomial¹.—La distribución polinomial está asociada con las pruebas repetidas de un suceso con más de dos resultados posibles. Así, el resultado del lanzamiento de un dado puede ser cualquiera de los seis números 1, 2, 3, ..., 6. Si el suceso se refiere a la aparición de ases en la extracción de, p. ej., 7 cartas, hay cinco resultados posibles: 0, 1, 2, 3 ó 4 ases.

En general, supongamos que hay k resultados posibles de un suceso aleatorio, y designemos las probabilidades respectivas de estos resultados por $p_1, p_2, \dots, p_k$. Evidentemente, debe verificarse

$$\sum_{i=1}^k p_i = 1 \quad (1)$$

lo mismo que $p + q = 1$ en el caso binomial. Supongamos que el suceso se repite n veces, y sea x_1 el número de veces que ocurre el resultado asociado a p_1 ; x_2 el número de veces que ocurre el resultado asociado a p_2 , y así sucesivamente. La función de cuantía para las variables aleatorias $x_1, x_2, \dots, x_{k-1}$ es

$$f(x_1, x_2, \dots, x_{k-1}) = \frac{n!}{\prod_{i=1}^k x_i!} \prod_{i=1}^k p_i^{x_i} \quad x_i = 0, 1, \dots, n; \quad \sum_{i=1}^k x_i = n \quad (2)$$

Hemos escrito la función de modo que solo intervengan en ella $k-1$ de las variantes x_i , ya que solo $k-1$ de estas son funcionalmente independientes; x_k queda determinada exactamente por la

relación $\sum_{i=1}^k x_i = n$, una vez especificados los valores de $x_1, \dots, x_{k-1}$.

Se trata, por tanto, de una distribución multivariante con $k-1$ variantes. El x_k que aparece en el segundo miembro de (2) debe interpretarse simplemente como un símbolo de la expresión

$$n - x_1 - x_2 - \dots - x_{k-1}$$

¹ Más conocida por *multinomial*.

La expresión (2) es una familia de distribuciones con k parámetros, a saber: $n, p_1, p_2, \dots, p_{k-1}$. La otra variable p_k , como la q en la distribución binomial, está determinada exactamente por

$$p_k = 1 - p_1 - p_2 - \dots - p_{k-1}$$

Un caso particular de la distribución polinomial se obtiene haciendo, p. ej., $n=3, k=3, p_1=0,2, p_2=0,3$, y resulta:

$$f(x_1, x_2) = \frac{3!}{x_1! x_2! (3 - x_1 - x_2)!} (0,2)^{x_1} (0,3)^{x_2} (0,5)^{3-x_1-x_2}$$

Se ha representado esta función en la figura 3-9.

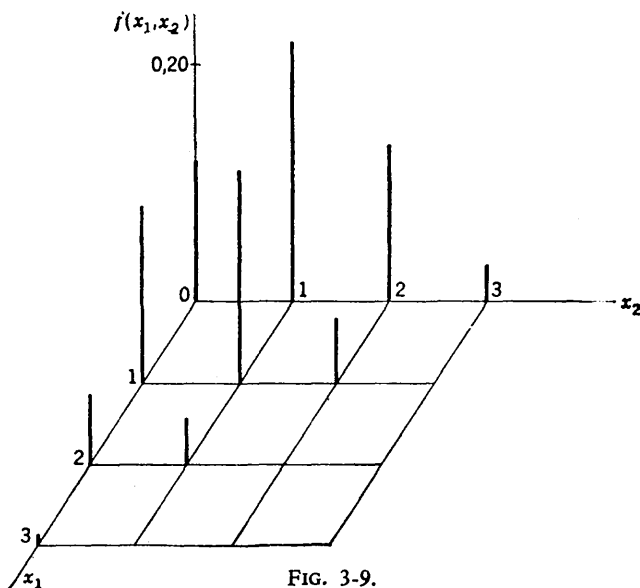


FIG. 3-9.

3-6. Distribución de Poisson.—La función de cuantía de Poisson se da en la siguiente definición.

Definición 3-4.—Diremos que la variable aleatoria x se distribuye según una distribución de Poisson si la función de cuantía es

$$f(x) = \frac{e^{-m} m^x}{x!} \quad x = 0, 1, 2, 3, \dots \quad (1)$$

donde m es cualquier número positivo.

Puesto que la exponencial e^m tiene el desarrollo en serie

$$e^m = 1 + m + \frac{m^2}{2!} + \dots + \frac{m^x}{x!} + \dots$$

se deduce que

$$\sum_{x=0}^{\infty} f(x) = 1$$

Esta distribución se aplica útilmente en aquellas situaciones en que un gran número de objetos se encuentran distribuidos sobre un gran recinto de considerable extensión. Para considerar un ejemplo concreto, supongamos un volumen V de un fluido que contiene un gran número N de pequeños organismos. Se supone que estos organismos no tienen instintos sociales y que la probabilidad de que aparezcan en cualquier parte del fluido es la misma para un determinado volumen. Supongamos ahora que se examina una gota de volumen D al microscopio: ¿cuál es la probabilidad de que se hallen x organismos en la gota? Se supone que V es mucho mayor que D . Puesto que se supone que los organismos están distribuidos con probabilidad uniforme por todo el fluido, se deduce que la probabilidad de encontrar uno cualquiera de ellos en D es D/V . Y como hemos supuesto que carecen de instintos sociales, la presencia de uno de ellos en D no influye en la de cualquiera de los otros. Por tanto, la probabilidad de que haya x organismos en D es

$$\binom{N}{x} \left(\frac{D}{V}\right)^x \left(\frac{V-D}{V}\right)^{N-x} \quad (2)$$

Suponemos también ahora que los organismos son tan pequeños que puede prescindirse del espacio que ocupan; los N reunidos no ocuparían parte apreciable del volumen D . La función de Poisson es una aproximación de la expresión anterior, que es simplemente una función binomial en la que $p = D/V$ es muy pequeña.

La distribución de Poisson se obtiene haciendo que V y N tiendan a infinito, de tal modo que la densidad de organismos $N/V = d$ permanezca constante. Si se escribe de nuevo el producto (2) en la forma

$$\begin{aligned} & \frac{N(N-1)(N-2) \dots (N-x+1)}{x! N^x} \left(\frac{ND}{V}\right)^x \left(1 - \frac{ND}{NV}\right)^{N-x} = \\ & = \frac{\left(1 - \frac{1}{N}\right) \left(1 - \frac{2}{N}\right) \dots \left(1 - \frac{x-1}{N}\right) (Dd)^x \left(1 - \frac{Dd}{N}\right)^{N-x}}{x!} \end{aligned}$$

se ve en seguida que el límite, cuando N tiende a infinito, es

$$\frac{e^{-Dd}(Dd)^x}{x!}$$

que coincide con (1) si sustituimos m por Dd . Esto nos demuestra que m es el valor medio de x , ya que D , volumen de la porción examinada, multiplicado por la densidad general d , da el promedio esperado en el volumen D .

Hemos estudiado esta distribución con algún detalle porque a menudo se la aplica equivocadamente a datos que no cumplen los supuestos requeridos por la distribución. Así, no puede utilizarse en el estudio de distribuciones de larvas de insectos en una gran extensión de cultivo, porque los insectos depositan sus huevos en grupos, de modo que de encontrar uno en una pequeña área dada, lo probable es que se encuentren también otros.

Tal vez lo mejor sea considerar esta función de Poisson como una aproximación de la binomial, $\binom{N}{x} p^x q^{N-x}$, cuando Np es grande respecto a p , y N lo es con relación a Np . Resulta particularmente útil cuando se desconoce N .

3-7. Otras distribuciones discretas.—La función de cuantía hipergeométrica es

$$f(x) = \frac{\binom{m}{x} \binom{n}{r-x}}{\binom{m+n}{r}} \quad x=0, 1, \dots, r \quad (1)$$

La ecuación (3-3-3) es un caso particular, y la (3-3-2) constituye un ejemplo de distribución hipergeométrica bivalente.

La función de cuantía uniforme es

$$f(x) = \frac{1}{n} \quad x=1, 2, \dots, n \quad (2)$$

El lanzamiento de un dado proporciona un ejemplo.

La función de cuantía binomial negativa es

$$f(x) = p^r \binom{x+r-1}{r-1} q^x \quad x=0, 1, 2, \dots; p+q=1 \quad (3)$$

y $\sum f(x) = 1$, puesto que

$$\sum_{x=0}^{\infty} \binom{x+r-1}{r-1} q^x = \frac{1}{(1-q)^r} = \frac{1}{p^r}$$

Se obtiene un ejemplo de esta distribución representando por p la probabilidad de acierto y por q la probabilidad de fallo de un determinado suceso. Sea $f(x)$ la probabilidad de que se necesiten exactamente $x+r$ pruebas para obtener r aciertos. La última prueba deberá ser un acierto, y su probabilidad es p . Entre las otras $x+r-1$ pruebas deberá haber $r-1$ aciertos, y la probabilidad de que así ocurra es

$$\binom{x+r-1}{r-1} p^{r-1} q^x$$

El producto de estas dos probabilidades nos da la probabilidad deseada $f(x)$, que es la expresada por (3).

PROBLEMAS

Para cada distribución deberá especificarse el recorrido. No es necesario obtener respuestas numéricas que exijan cálculos laboriosos.

1. Se extraen cinco cartas de una baraja ordinaria. ¿Cuál es la función de cuantía del número de espadas?

2. Se echan 10 bolas en 4 cajas, de modo que cada bola tenga la misma probabilidad de caer en cualquiera de las cajas, ¿cuál es la función de cuantía del número de bolas que cae en la primera caja?

3. Se lanza una moneda hasta que aparece una cara, ¿cuál es la función de cuantía para el número de tiradas?

4. ¿Cuál es la función de cuantía para el número que aparece al arrojar un dado?

5. Se lanzan dos dados, ¿cuál es la función de cuantía de la suma de los dos números que aparecen?

6. Se extraen cartas, sin reemplazamiento, de una baraja ordinaria hasta que aparece una espada; ¿cuál es la función de cuantía para el número de extracciones?

7. Se lanzan 6 dados; ¿cuál es la función de cuantía para el número de unos y de doses?

8. Una urna contiene m bolas negras y n bolas blancas. Se extraen k bolas sin reemplazamiento, ¿cuál es la función de cuantía del número de bolas blancas? Especificar el recorrido para los diferentes tamaños relativos de m , n y k .

9. Se lanzan tres monedas n veces. Hallar la función de cuantía conjunta de x , número de veces en que no aparecen caras; y , número de veces en que aparece una cara; z , número de veces en que aparecen dos caras.

10. Una máquina fabrica clavos con un promedio del 1% de defectuosos. ¿Cuál es la función de cuantía del número de defectuosos en una muestra de 60 clavos?

11. Una urna contiene 8 bolas blancas y 12 negras. Se extraen las bolas una a una sin reemplazamiento, hasta que hayan aparecido 5 blancas. Hállese la función de cuantía del total de bolas extraídas.

12. Se extraen 6 cartas, sin reemplazamiento, de una baraja ordinaria. Hállese la función de cuantía conjunta del número de ases y del número de reyes.

13. Demostrar que

$$\sum_{i=0}^c \binom{a}{i} \binom{b}{c-i} = \binom{a+b}{c}$$

igualando los coeficientes de x^c en

$$(1+x)^a (x+1)^b = (1+x)^{a+b}$$

A partir de este resultado, compruébese algebraicamente que la suma de la función de cuantía hipergeométrica es la unidad.

14. Utilícese el resultado del problema 13 para hallar la distribución marginal del número de ases a que se refiere el problema 12.

15. En una ciudad con 5000 adultos se pregunta a una muestra de 100 cuál es su opinión sobre una propuesta de proyecto municipal; se obtienen 60 respuestas a favor del proyecto y 40 en contra. Si en realidad los adultos de la ciudad estuvieran divididos en dos grupos iguales respecto a dicha propuesta, ¿cuál sería la probabilidad de obtener una mayoría de 60 o más a favor, en una muestra de tamaño 100?

16. Un distribuidor de semillas ha determinado a partir de numerosos ensayos que el 5% de un grupo grande de semillas no germina; vende las semillas en paquetes de 200, garantizando la germinación del 90%. ¿Cuál es la probabilidad de que un paquete dado no cumpla la garantía?

17. Se trata de utilizar un proceso de fabricación para la obtención de conmutadores con un tanto por ciento de defectuosos no superior a 1%. Se comprueba el proceso cada hora, ensayando 10 conmutadores elegidos aleatoriamente entre los obtenidos en una hora. Si fallan uno o más de los 10, se detiene el proceso y se procede a un examen cuidadoso. Si la probabilidad real de producir un conmutador defectuoso es 0,01, ¿cuál es la probabilidad de que el proceso sea examinado sin necesidad en un caso determinado?

18. Con respecto al problema anterior, ¿cuántos conmutadores (en vez de 10) deberán ensayarse si el fabricante desea que la probabilidad de que el proceso sea examinado cuando se produzca un 10% de defectuosos sea 0,95?

19. A tiene dos peniques, B tiene uno; juegan hasta que uno de ellos gana los tres. ¿Cuál es la función de cuantía del número de pruebas necesarias para terminar el juego?

20. Teniendo en cuenta el problema anterior, ¿cuál es la función de cuantía del número de pruebas, suponiendo que gana A?

21. Se lanza un dado 10 veces. ¿Cuál es la probabilidad de que el número de unos y doses no difiera en más de 2 del valor modal?

22. Una distribución de Poisson tiene una moda doble en $x=1$ y $x=2$. ¿Cuál es la probabilidad de que x tome uno u otro de estos dos valores?

23. La escasez de glóbulos rojos puede determinarse examinando al microscopio una muestra de sangre. Suponiendo que un volumen pequeño determinado contenga por término medio 20 glóbulos rojos en personas normales, ¿cuál es la probabilidad de que una muestra de persona normal contenga menos de 15 glóbulos rojos?

24. Una compañía de seguros halla que el 0,005 % de la población fallece cada año de un cierto tipo de accidente. ¿Cuál es la probabilidad de que la compañía tenga que pagar a más de 3 de los 10 000 asegurados contra tales accidentes en un año dado?

25. Un cuadro de teléfonos atiende un promedio de 600 llamadas durante una hora de aglomeración. El cuadro puede hacer un máximo de 20 conexiones por minuto. Utilícese la distribución de Poisson para estimar la probabilidad de que el cuadro quede rebasado durante un minuto dado.

26. Se lanza un dado hasta que aparece un 6. ¿Cuál es la probabilidad de que haya que lanzarlo más de cinco veces?

27. Se lanzan dos dados 10 veces. Sea x el número de veces en que no aparecen unos e y el número de veces en que aparecen 2 unos. ¿Cuál es la probabilidad de que x e y sean cada uno inferior a 3?

28. En el problema 27, ¿cuál es la probabilidad de que $x+y$ sea igual a 4? ¿Cuál es la probabilidad de que $x+y$ esté comprendido entre 2 y 4, ambos inclusive?

29. Se lanza un dado 20 veces. ¿Cuál es la probabilidad de que haya al menos doble número de unos y doses que de treses?

30. Se sacan 10 cartas, sin reemplazamiento, de una baraja ordinaria. ¿Cuál es la probabilidad de que el número de picas exceda al número de tréboles?

31. Supongamos que al pasar un neutrón a través de plutonio pueda con igual probabilidad dejar libres 1, 2 ó 3 neutrones, y supongamos que esta segunda generación de neutrones pueda, a su vez, con igual probabilidad, dejar libres 1, 2 ó 3 neutrones de la tercera generación? ¿Cuál es la función de cuantía del número de neutrones de la tercera generación?

32. Utilizando la función de cuantía del problema 12, hállese la distribución condicional del número x de ases, dado el número y de reyes.

33. Utilizando la función de cuantía del problema 9, hállese la distribución condicional de x y z , dado y .

Determinense las sumas precisas para calcular las probabilidades siguientes, utilizando funciones de cuantía con tantas variantes como sean necesarias. Supóngase que todas las variantes toman los valores 0, 1, 2, ..., m .

34. $P(2x + y \leq 3)$.

35. $P(x^2 + y^2 = 25)$.

36. $P(x^2 < 5 | 1 \leq y \leq 6)$.
37. $P(x > 2y - a), 0 < a < m$.
38. $P(x > y > z)$.
39. $P(x + y = 5 | y = 3)$.
40. $P(x + y = 5 | z = 3)$.
41. $P(x \leq 3, y \leq 4, z \geq 5, w \geq 6)$.
42. $P(a \leq x \leq b | y = z), 0 < a < b < m$.
43. $P(x > 2y | x > z)$.

BIBLIOGRAFIA

1. CLARK, C.: *An Introduction to Statistics*, John Wiley & Sons, Inc., Nueva York, 1953.
2. FELLER, W.: *An Introduction to Probability Theory and Its Applications*, John Wiley & Sons, Inc., Nueva York, 1950.
3. FRASER, D. A. S.: *Statistics—An Introduction*, John Wiley & Sons, Inc., Nueva York, 1958.
4. FRYER, H.: *Elements of Statistics*, John Wiley & Sons, Inc., Nueva York, 1954.
5. HOEL, P. G.: *Elementary Statistics*, John Wiley & Sons, Inc., Nueva York, 1960.
6. National Bureau of Standards: «Tables of the Binomial Probability Distribution», *Applied Mathematics Series* 6, 1950.
7. RAO, C. R.: *Advanced Statistical Methods in Biometric Research*, John Wiley & Sons, Inc., Nueva York, 1952.

CAPITULO 4

VARIABLES ALEATORIAS CONTINUAS

4-1. Introducción.—Una variable aleatoria continua puede tomar cualquier valor de cierto intervalo o colección de intervalos sobre el eje x , el plano xy , etc., sin la restricción de que aquellos sean números aislados.

Como ejemplo, supongamos un rifle perfectamente apuntado al centro de un blanco de forma cuadrada; el rifle se dispara varias veces después de fijado en esta posición. No todas las balas darán en el centro, ya que las pequeñas variaciones en el peso de las balas, la forma de estas, la humedad y temperatura de la pólvora y otros factores causarán variaciones en las trayectorias. Después de varios disparos, el blanco tendrá un aspecto parecido al que se representa en la figura 4-1.

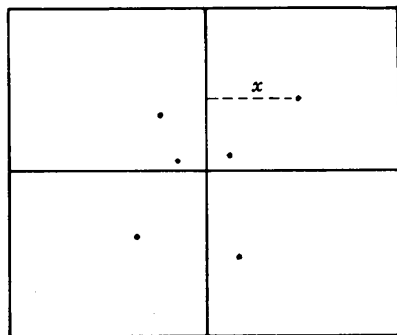


FIG. 4-1.

Sea x una variable aleatoria, definida como desviación horizontal del centro de un impacto, respecto a la vertical que pasa por el centro del blanco. Claro es que x puede tomar un número infinito no numerable de valores.

Esto constituye una diferencia entre las variables aleatorias discretas y continuas: Una variable aleatoria discreta puede tomar un número finito o infinito numerable de valores; en cambio, una variable continua puede asumir un número infinito no numerable de valores.

4-2. Variables aleatorias continuas.—En el caso de variantes discretas es posible asociar una probabilidad finita a cada punto del recorrido, aunque el número de puntos que constituyen este sea infinito, verificándose, sin embargo, que la suma de las probabilidades es igual a la unidad. Así, si x es el número de tiradas necesarias para obtener cara con una moneda, hemos visto que la función de cuantía de x es

$$f(x) = (1/2)^x \quad x = 1, 2, 3, 4, \dots$$

y

$$\sum_{x=1}^{\infty} f(x) = 1$$

En el caso de una variante continua, esto no es posible. Las probabilidades no sumarán uno, a menos que prácticamente a todos sus puntos (todos menos un conjunto numerable) se les haga corresponder probabilidad cero. Volviendo a referirnos a las desviaciones horizontales de los disparos de un rifle sobre un blanco, es claro que todos los valores de x dentro de un pequeño intervalo serán aproximadamente equiprobables, y no puede suponerse razonablemente que la mayor parte de ellos tienen probabilidad cero mientras que solo a algunos les corresponden probabilidades no nulas.

Conviene observar que la dificultad que hemos señalado es puramente lógica. Desde el punto de vista práctico, tal dificultad queda oculta por el hecho de que no es posible distinguir entre una desviación de 0,5 pulg y otra de 0,500003 pulg. Estamos sometidos a los límites de precisión correspondientes al procedimiento de medida utilizado, y una desviación solo puede identificarse dentro de un determinado intervalo. Así, en el caso de que solo podamos medir con error menor de una centésima de pulgada, si medimos una desviación de 4,26 pulg, habrá que interpretar este resultado en el sentido de que la desviación estará comprendida en el intervalo de 4,25 a 4,27 pulg y será mejor escribirlo así: $4,26 \pm 0,01$, para indicar lo que ocurre.

El problema lógico se aborda considerando intervalos en lugar de puntos. Empecemos por examinar varias probabilidades empíricas para intervalos. Supongamos que el rifle se dispara 100 veces hacia el blanco de la figura 4-1, y que la superficie del blanco está dividida en zonas por medio de rectas verticales distanciadas 1 pulg (véase Fig. 4-2). Considerando negativas las desviaciones de x a

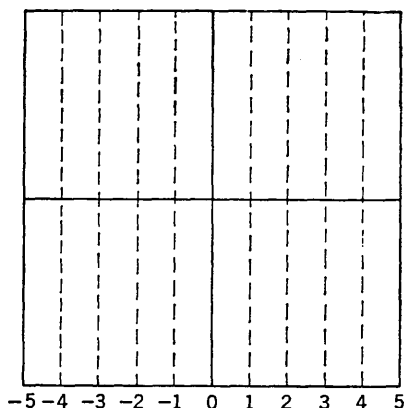


FIG. 4-2.

la izquierda de la línea central, supongamos las verticales trazadas por $x = \pm 1, \pm 2, \pm 3$, etc. Ahora bien: para una zona dada, como la $0 < x \leq 1$, el número de disparos contenidos en ella dividido por 100

será la frecuencia relativa de que una desviación esté comprendida entre 0 y 1. Podemos tabular una distribución hipotética de disparos, calculando las probabilidades empíricas como en la tabla 4-1.

TABLA 4-1

Zona	Número de disparos	Probabilidad empírica
$-5 < x \leq -4$	1	0,01
$-4 < x \leq -3$	1	0,01
$-3 < x \leq -2$	6	0,06
$-2 < x \leq -1$	13	0,13
$-1 < x \leq 0$	24	0,24
$0 < x \leq 1$	27	0,27
$1 < x \leq 2$	16	0,16
$2 < x \leq 3$	7	0,07
$3 < x \leq 4$	3	0,03
$4 < x \leq 5$	2	0,02

La distribución empírica contenida en esta tabla puede representarse utilizando rectas verticales como en el caso de las distribuciones discretas. No obstante, no trazaremos rectas en los puntos medios, p. ej., de cada intervalo, sino que preferiremos usar rectángulos de altura igual a la frecuencia relativa dividida por la am-

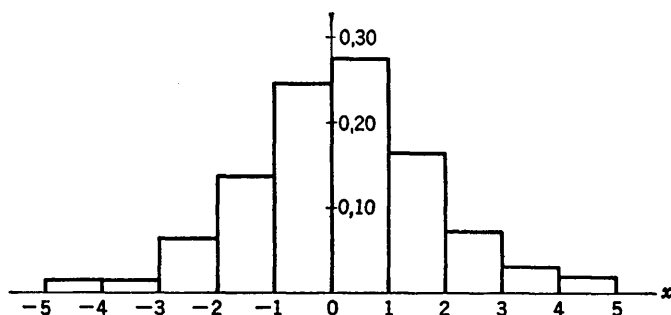


FIG. 4-3.

plitud del intervalo, y de base igual a dicha amplitud. Esto se hace para indicar que la frecuencia se refiere a todo el intervalo y no a un punto único del mismo. El resultado es el que se muestra en la figura 4-3.

Refiriéndonos a la figura 4-3, observemos que el área de cada rectángulo es igual a la frecuencia relativa del intervalo correspondiente, ya que la altura del rectángulo es igual a dicha frecuencia y su base es la unidad. Atenderemos a las áreas más que a las alturas. La suma de las áreas de todos los rectángulos es la unidad. Será útil considerar estas frecuencias relativas como estimaciones de las probabilidades.

Para intervalos distintos de los elegidos originariamente, podemos estimar también probabilidades. Así estimaríamos la probabilidad de que $0 < x \leq 2$ sumando las áreas de los dos rectángulos sobre tal intervalo, lo que nos da 0,43. Para estimar la probabilidad de, p. ej., $-0,25 < x \leq 1,5$, calcularíamos el área correspondiente a tal intervalo, obteniendo

$$0,06 + 0,27 + 0,08 = 0,41$$

Si se hicieran otros 100 disparos sobre el mismo blanco, obtendríamos otra distribución empírica, que, muy probablemente, sería distinta de la primera aunque de aspecto general parecido. Al cons-

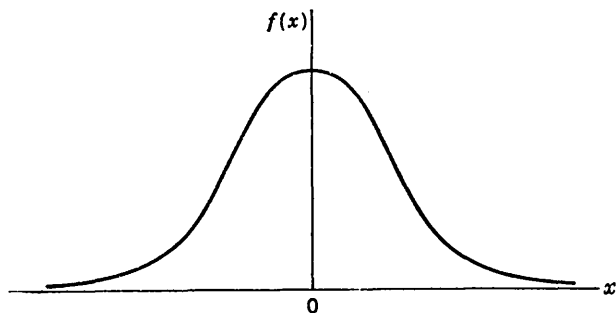


FIG. 4-4.

truir una teoría de la probabilidad, se acostumbra considerar estas probabilidades empíricas como estimaciones de una cierta probabilidad *verdadera*. A este fin, suponemos la existencia de una función f tal como la representada en la figura 4-4. Puede ocurrir que no nos sea posible especificar la función; pero suponemos que existe cierta función que da la probabilidad correcta para todo intervalo. Las probabilidades vienen dadas por áreas limitadas por la curva y no por valores de la función. Así,

$$P(0 < x \leq 1) = \int_0^1 f(x) dx$$

que es la probabilidad estimada por el área del rectángulo correspondiente al intervalo $0 < x \leq 1$ en la figura 4-3. (A lo largo de este libro, todas las integrales serán integrales de Riemann.)

Consideramos el gráfico de la función f como una curva continua mejor que como una función escalonada, por las siguientes razones: En primer lugar, se reconoce que la elección de intervalos en un experimento efectivo es totalmente arbitraria. En el experimento del rifle podríamos igualmente haber utilizado intervalos de $1/2$ de pulg de amplitud o intervalos que tuvieran por extremos 1,2; 2,2; 3,2; o intervalos de amplitudes distintas, p. ej., de 0, a 0,5, de 0,5 a 1,5, de 1,5 a 3. De modo que los escalones de la distribución empírica no tienen especial significado. En segundo lugar, imaginemos que se consideran dos intervalos pequeños con un extremo común, como $1,9 < x \leq 2$ y $2 < x \leq 2,1$. Puesto que el segundo intervalo está más alejado del centro que el primero, es razonable esperar que su probabilidad sea algo menor, pero no tanto como para suponer que es dos veces más probable que una desviación aparezca en el primer intervalo que en el segundo, como se indica en la figura 4-3. La curva continua proporciona una relación más razonable entre ambas probabilidades. En tercer lugar, cuando se realizan experimentos que comprenden gran número de pruebas, estos indican que no existen cambios bruscos en la curva que representa la distribución. Así, en el caso del rifle, si este se dispara 1000 veces y se utilizan intervalos de amplitud 0,1 de pulg, lo probable es que los escalones sean mucho menores que los de la figura 4-3 y se aproximen a una curva continua, tal como la de la figura 4-4.

En la discusión precedente, hemos pretendido dar una descripción de cómo se pueden originar variables aleatorias continuas y de cómo la frecuencia relativa da una noción de probabilidad. A continuación trataremos estas cuestiones de forma más concreta.

Definición 4-1.—*Se dice que x es una variable aleatoria continua (unidimensional) si existe una función f tal que $f(x) \geq 0$ para todo x del intervalo $-\infty < x < \infty$ y tal que para cualquier suceso A*

$$P(A) = P(x \text{ esté en } A) = \int_A f(x) dx$$

$f(x)$ se denomina función de densidad de x , y diremos a veces que « x se distribuye según $f(x)$ » o que « $f(x)$ es la distribución de x ». El único suceso que consideraremos en este libro para variables aleatorias continuas es un intervalo o una colección de un número finito de intervalos no superpuestos. Así, en el ejemplo del tiro con un

rifle, que se ha visto al principio de este capítulo, definamos dos intervalos A_1 y A_2 (véase Fig. 4-5) de la siguiente forma:

$$A_1 = \{x: -3 \leq x \leq 2\}, \quad A_2 = \{x: -2 \leq x \leq 4\}$$

El suceso « A_1 o A_2 » es

$$A_1 \cup A_2 = \{x: -3 \leq x \leq 4\}$$

El suceso « A_1 no ocurre» es

$$\bar{A}_1 = B_1 \cup B_2$$

donde

$$B_1 = \{x: -\infty < x < -3\}, \quad B_2 = \{x: 2 < x < \infty\}$$

El suceso « A_1 y A_2 » es

$$A_1 \cap A_2 = \{x: -2 \leq x \leq 2\}$$

En este ejemplo particular, el suceso (A_1 o A_2) es un intervalo, como también lo es el suceso (A_1 y A_2); el suceso $\bar{A}_1$ no es un intervalo, pero sí la unión de dos intervalos no superpuestos y, por tanto, es un suceso según nuestra definición.

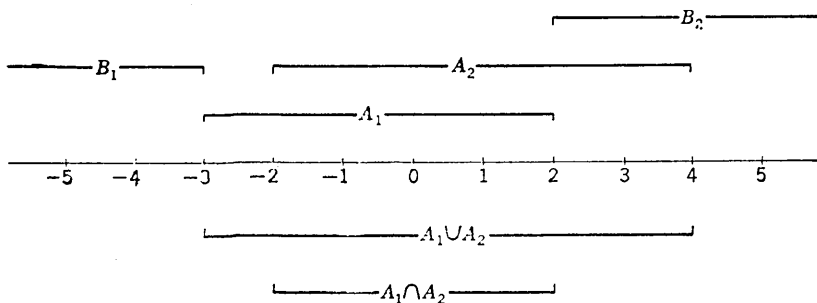


FIG. 4-5.

En cuanto sigue supondremos que las variables aleatorias continuas tienen función de densidad también continua, salvo, a lo sumo, en un número finito de puntos.

Definamos A por $A = \{x: a < x < b\}$. Para una variable aleatoria continua x con función de densidad $f(x)$

$$P(a < x < b) = P(A) = \int_a^b f(x) dx = \int_a^b f(x) dx$$

Puesto que $\int_a^b f(x) dx$ toma el mismo valor, ya sea el intervalo abierto, cerrado, o abierto a la derecha o a la izquierda, tenemos

$$P(a < x < b) = P(a \leq x < b) = P(a < x \leq b) = P(a \leq x \leq b)$$

Así, la integral en un punto es 0; por tanto, $P(x=a)=0$ para cualquier número a . Si A no es un intervalo, sino la unión de un número finito de intervalos no superpuestos ($A=A_1 \cup A_2 \cup \dots \cup A_k$, donde $A_i \cap A_j = \emptyset$ para todo $i \neq j$, y donde $A_i = \{x: a_i < x < b_i\}$), resulta que

$$\begin{aligned} P(A) &= P(A_1 \cup A_2 \cup \dots \cup A_k) = \int_A f(x) dx \\ &= \int_{A_1} f(x) dx + \int_{A_2} f(x) dx + \dots + \int_{A_k} f(x) dx \\ &= \int_{a_1}^{b_1} f(x) dx + \int_{a_2}^{b_2} f(x) dx + \dots + \int_{a_k}^{b_k} f(x) dx \end{aligned}$$

da la probabilidad de que el valor de la variable aleatoria esté comprendido entre a_1 y b_1 o entre a_2 y b_2 ... o entre a_k y b_k . En esencia esto establece que la probabilidad de que el valor de una variable aleatoria pertenezca a un conjunto A es el área comprendida entre $f(x)$ y el eje x sobre el conjunto. La función de densidad $f(x)$ puede ser una función que se aproxime a un histograma de frecuencias relativas, como en el ejemplo de las figuras 4-3 y 4-4, o bien obtenerse a partir de algún razonamiento teórico. Si el suceso A es todo el eje x , $P(A)$ debe ser igual 1, y se tiene

$$\int_{-\infty}^{\infty} f(x) dx = 1$$

Así, cualquier función f puede servir como función de densidad de una variable aleatoria continua x si satisface la condición

$$\begin{aligned} \int_{-\infty}^{\infty} f(x) dx &= 1 \\ f(x) &\geq 0 \quad -\infty < x < \infty \end{aligned}$$

Naturalmente, en un problema particular de aplicación, se elegirá f de tal forma que, para todo a y b ($a < b$),

$$P(a < x < b) = \int_a^b f(x) dx$$

represente la probabilidad de que el valor de la variable aleatoria x se halle comprendido entre a y b .

Cualquier función positiva en un dominio elegido arbitrariamente puede considerarse como una función de densidad de una variable aleatoria, siempre que la función esté multiplicada por una cons-

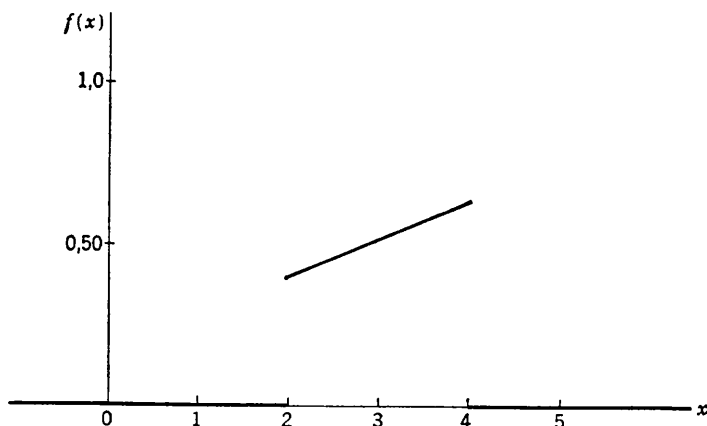


FIG. 4-6.

tante que haga que su integral sea igual a 1. Así, p. ej., la siguiente función es una función de densidad:

$$\begin{aligned} f(x) &= 0 & x &\leq 2 \\ &= 1/18(3 + 2x) & 2 < x < 4 \\ &= 0 & x &\geq 4 \end{aligned}$$

El valor de la función es positivo o cero, y

$$\begin{aligned} \int_{-\infty}^{\infty} f(x) dx &= \int_{-\infty}^2 0 dx + \int_2^4 1/18(3 + 2x) dx + \int_4^{\infty} 0 dx \\ &= 0 + 1 + 0 \\ &= 1 \end{aligned}$$

La probabilidad de que una variante que tenga esta densidad caiga, p. ej., en el intervalo $2 < x < 3$ es

$$\begin{aligned} P(2 < x < 3) &= \int_2^3 1/18(3 + 2x) dx \\ &= 4/9 \end{aligned}$$

Esta función se ha representado en la figura 4-6.

4-3. Distribuciones multivariantes.—Volviendo al experimento del rifle, podemos caracterizar cada disparo, no solo por su desviación horizontal x , sino por su desviación vertical y , medida perpendicularmente a partir de una horizontal que pasa por el centro del blanco. Supongamos efectuado un gran número de disparos, y dividido el blanco en cuadrados de 1 pulg² mediante rectas horizontales y verticales trazadas a distancias de 1 pulg. Podríamos contar el número de impactos por cada cuadrado y calcular la frecuencia relativa de cada uno de estos. Trazando paralelepípedos de alturas iguales a las frecuencias relativas de cada cuadrado, se obtendría un resultado análogo al que se representa en la figura 4-7. El volumen de un paralelepípedo estima la probabilidad de que un disparo caiga en el cuadrado sobre el que está construido dicho paralelepípedo.

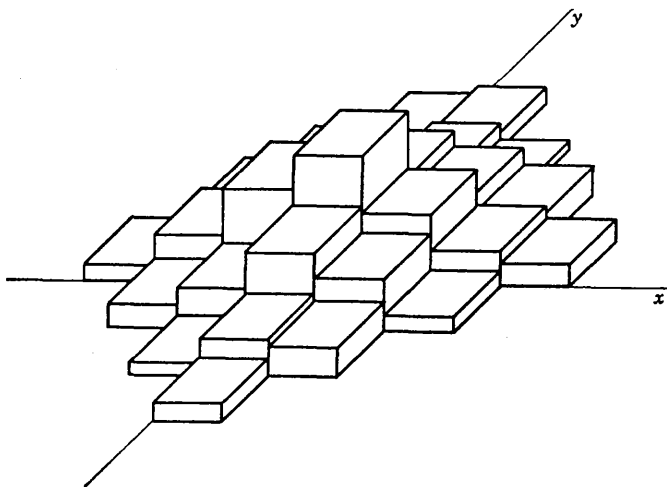


FIG. 4-7.

Resulta natural idealizar esta situación postulando la existencia de una función $f(x, y)$, cuya representación sería una superficie continua sobre el plano x, y . La probabilidad de que un disparo caiga en una región dada está representada por el volumen limitado por la superficie sobre tal región. En la figura 4-8 se representa un cuarto de dicha superficie. La probabilidad de que x e y caigan en la región rectangular $0 < x < a$, $0 < y < b$ ilustrada en la figura es

$$P(0 < x < a, 0 < y < b) = \int_0^a \int_0^b f(x, y) dy dx \quad (1)$$

Se dice que x e y son variables aleatorias continuas distribuidas conjuntamente si existe una función f tal que $f(x, y) \geq 0$ para todo $-\infty < x < \infty$, $-\infty < y < \infty$, y tal que para cualquier suceso A

$$P(A) = P[(x, y) \text{ esté en } A] = \int_A f(x, y) dy dx$$

Vemos que se precisa que

$$f(x, y) \geq 0 \quad \text{para} \quad -\infty < x < \infty, \quad -\infty < y < \infty \quad (2)$$

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dy dx = 1 \quad (3)$$

$f(x, y)$ recibe el nombre de *función de densidad conjunta* de las variables aleatorias x e y .

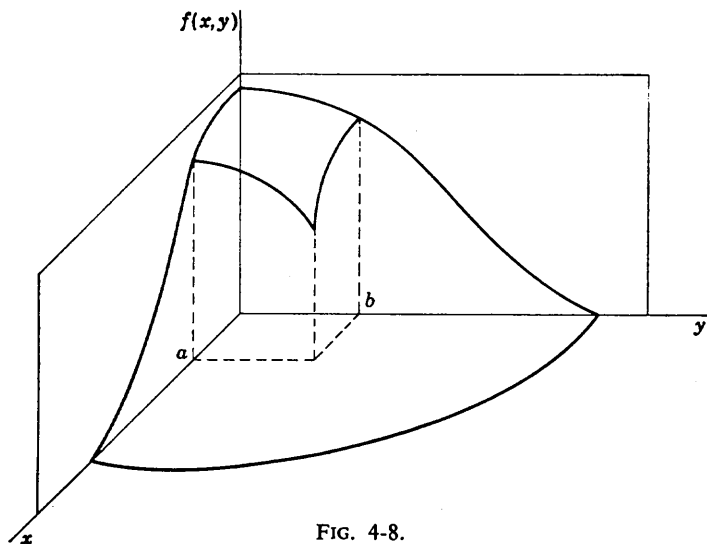


FIG. 4-8.

En el caso de k variables aleatorias continuas, la definición de distribución conjunta es análoga.

Como ilustración puede verse que la función $6 - x - y$ es positiva sobre el rectángulo $0 < x < 2$, $2 < y < 4$; puede utilizarse, por tanto, para definir una función de densidad conjunta sobre dicha región. Ya que

$$\int_0^2 \int_2^4 (6 - x - y) dy dx = 8$$

tenemos la función de densidad siguiente:

$$f(x, y) = \begin{cases} 1/8 (6 - x - y) & 0 < x < 2, 2 < y < 4 \\ 0 & \text{en otro caso} \end{cases} \quad (4)$$

Si x e y son variables aleatorias que tienen esta densidad, la probabilidad de que caigan, p. ej., en la región $x < 1$, $y < 3$ es

$$\begin{aligned} P(x < 1, y < 3) &= \int_{-\infty}^1 \int_{-\infty}^3 f(x, y) dy dx \\ &= \int_0^1 \int_2^3 1/8 (6 - x - y) dy dx \\ &= 3/8 \end{aligned}$$

La probabilidad de que $x + y$ sea inferior a 3 es

$$\begin{aligned} P(x + y < 3) &= \int_0^1 \int_2^{3-x} 1/8 (6 - x - y) dy dx \\ &= 5/24 \end{aligned}$$

La probabilidad de que $x < 1$ cuando se sabe que $y < 3$ es

$$P(x < 1 | y < 3) = \frac{P(x < 1, y < 3)}{P(y < 3)}$$

Hemos calculado anteriormente el numerador de esta expresión y el denominador es

$$\begin{aligned} P(y < 3) &= \int_0^2 \int_2^3 1/8 (6 - x - y) dy dx \\ &= 5/8 \end{aligned}$$

de donde

$$P(x < 1 | y < 3) = \frac{3/8}{5/8} = \frac{3}{5}$$

La generalización de estas ideas al caso de más de dos variables es inmediata. En general, cualquier función f puede considerarse como una función de densidad de k variables aleatorias, siempre que

$$\begin{aligned} f(x_1, x_2, \dots, x_k) &\geq 0 & -\infty < x_i < \infty \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(x_1, x_2, \dots, x_k) dx_1 dx_2 \dots dx_k &= 1 \end{aligned} \quad (5)$$

La probabilidad de que la variante $(x_1, x_2, \dots, x_k)$ esté en una región dada de un espacio k -dimensional se obtiene integrando la función de densidad sobre dicha región.

La función

$$f(x_1, x_2, x_3, x_4) = 16x_1x_2x_3x_4 \quad 0 < x_i < 1 \quad (6)$$

$$= 0 \quad \text{en otro caso}$$

es una función de densidad, por satisfacer las dos condiciones dichas. La probabilidad de que un punto caiga en la región $x_1 < 1/2$, $x_4 > 1/3$ es

$$P(x_1 < 1/2, x_4 > 1/3) = \int_{1/3}^1 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{1/2} f(x_1, x_2, x_3, x_4) dx_1 dx_2 dx_3 dx_4$$

$$= \int_{1/3}^1 \int_0^1 \int_0^1 \int_0^{1/2} 16x_1x_2x_3x_4 dx_1 dx_2 dx_3 dx_4$$

$$= 2/9$$

4-4. Distribuciones acumulativas.—Puesto que en el caso de variantes continuas las probabilidades vienen dadas por integrales, resulta a menudo conveniente considerar las integrales de las densidades con preferencia a las densidades mismas. Sea $f(x)$ una función de densidad de una variante (como, p. ej., la representada en la Fig. 4-4) y sea

$$F(x) = \int_{-\infty}^x f(t) dt \quad (1)$$

Esta función $F(x)$ es la probabilidad de que el valor de una observación sea inferior a x . Así

$$F(x) = P(x \leq x) \quad -\infty < x < \infty \quad (2)$$

$F(x)$ recibe el nombre de *función de distribución acumulativa* de x . En la figura 4-9 puede verse la representación gráfica de una distribución acumulativa. Si una función F es la función de distribución acumulativa de una variable aleatoria,

$$F \text{ es una función no decreciente} \quad (3)$$

$$F(-\infty) = 0 \quad (4)$$

$$F(\infty) = 1 \quad (5)$$

$$F \text{ es continua} \quad (6)$$

La función de densidad, si existe, puede hallarse a partir de la función de distribución acumulativa derivando F en los puntos donde esta tiene derivada; es decir,

$$f(x) = \frac{dF(x)}{dx}$$

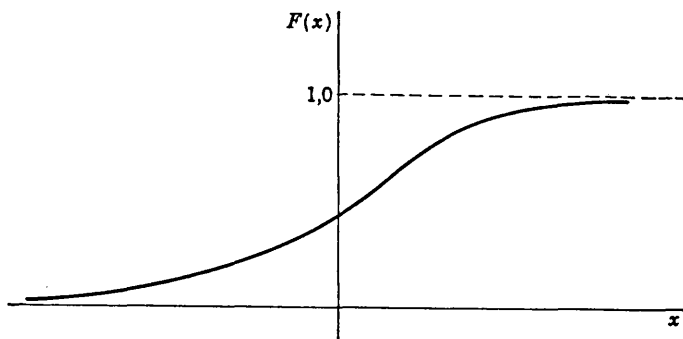


FIG. 4-9.

La probabilidad de que x caiga en un intervalo $a < x \leq b$ puede expresarse del siguiente modo con la función acumulativa:

$$\begin{aligned} P(a < x \leq b) &= P(x \leq b) - P(x \leq a) \\ &= F(b) - F(a) \end{aligned} \quad (7)$$

Refiriéndonos al ejemplo del final de la sección 4-2, en donde

$$\begin{aligned} f(x) &= 1/18 (3 + 2x) & 2 < x < 4 \\ &= 0 & \text{en otro caso} \end{aligned}$$

hallamos

$$\begin{aligned} F(x) &= 0 & x \leq 2 \\ &= \int_2^x 1/18 (3 + 2t) dt = 1/18 (x^2 + 3x - 10) & 2 < x < 4 \\ &= 1 & x \geq 4 \end{aligned}$$

y

$$\begin{aligned} P(2 < x < 3) &= F(3) - F(2) \\ &= 1/18 (9 + 9 - 10) - 0 \\ &= 4/9 \end{aligned}$$

La función está representada en la figura 4-10.

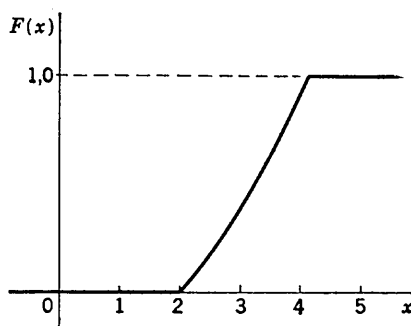


FIG. 4-10.

Para más de dos variantes, la función acumulativa se define análogamente:

$$F(x_1, x_2, \dots, x_k) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \dots \int_{-\infty}^{x_k} f(t_1, t_2, \dots, t_k) dt_k dt_{k-1} \dots dt_1 \quad (8)$$

siendo $f(x_1, x_2, \dots, x_k)$ la función de densidad. El valor de la función acumulativa en el punto $(a_1, a_2, \dots, a_k)$ es la probabilidad

$$\begin{aligned} P(x_1 \leq a_1, x_2 \leq a_2, \dots, x_k \leq a_k) &= F(a_1, a_2, \dots, a_k) = \\ &= P(x_1 < a_1, x_2 < a_2, \dots, x_k < a_k) \end{aligned} \quad (9)$$

ya que para variables aleatorias continuas

$$F(x) = P(x \leq x) = P(x < x)$$

Dada la función acumulativa F , puede hallarse la función de densidad derivando F respecto a cada una de sus variantes, en el supuesto de que las derivadas existan:

$$f(x_1, x_2, \dots, x_k) = \frac{\partial}{\partial x_1} \cdot \frac{\partial}{\partial x_2} \dots \frac{\partial}{\partial x_k} F(x_1, x_2, \dots, x_k) \quad (10)$$

Como ejemplo de una distribución acumulativa con dos variantes, podemos utilizar la función de densidad dada por la ecuación (4-3-4):

$$\begin{aligned} f(x, y) &= \frac{1}{8} (6 - x - y) & 0 < x < 2, 2 < y < 4 \\ &= 0 & \text{en otro caso} \end{aligned} \quad (11)$$

Hay nueve regiones en el plano x, y que deben tenerse en cuenta al definir $F(x, y)$; en la figura 4-11 se indican estas nueve regiones, con las coordenadas de los puntos de intersección de las rectas que las limitan. (La vertical izquierda coincide con el eje de las y). Esta

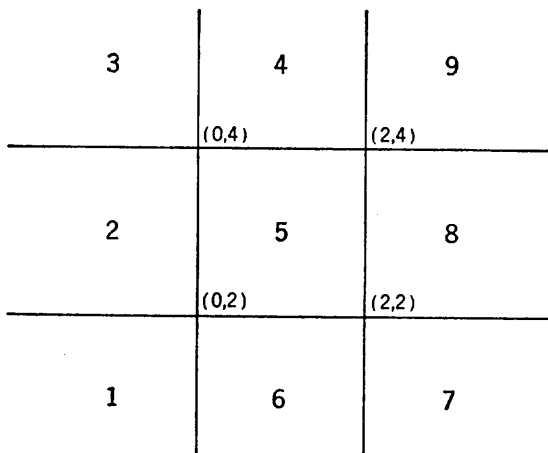


FIG. 4-11.

complicación se debe a la definición por partes de $f(x, y)$. Podríamos establecer simplemente que

$$F(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(s, t) dt ds \quad (12)$$

pero para que resulte útil es necesaria una caracterización más detallada de esta función. En la región 1 de la figura 4-11, $f(x, y)$ es cero; por tanto,

$$F(x, y) = 0 \quad x \leq 0, y \leq 2$$

En la región 2, aunque y es mayor que 2, tenemos $x \leq 0$, de modo que (12) sigue siendo nula, ya que $f(s, t)$ nunca es positiva en el campo de integración. Lo mismo ocurre en las regiones 3, 6 y 7. Para x, y en la región 5, el integrando no es cero si $0 < s < x$, $2 < t < y$, y tenemos

$$\begin{aligned} F(x, y) &= \int_0^x \int_2^y \frac{1}{8} (6 - s - t) dt ds \\ &= \int_0^x \frac{1}{8} \left[(6 - s)(y - 2) - \frac{y^2}{2} + 2 \right] ds \\ &= \frac{1}{16} x(y - 2)(10 - y - x) \quad 0 < x < 2, 2 < y < 4 \end{aligned} \quad (13)$$

Para cualquier punto en la región 4, el integrando de (12) es positivo cuando $0 < s < x$, $2 < t < 4$; por tanto,

$$F(x, y) = \int_0^x \int_2^4 f(s, t) dt ds$$

y esta integral puede calcularse haciendo $y=4$ en (13); lo que da

$$F(x, y) = \frac{1}{8}x(6-x) \quad 0 < x < 2, y \geq 4$$

Análogamente, en la región 8, $F(x, y) = F(2, y)$ si $x \geq 2$, de modo que

$$F(x, y) = \frac{1}{8}(y-2)(8-y) \quad x \geq 2, 2 < y < 4$$

y en la región 9, $F(x, y) = 1$. Combinando estos resultados

$$\begin{aligned} F(x, y) &= 0 & x \leq 0 \text{ ó } y \leq 2 \\ &= \frac{1}{16}x(y-2)(10-y-x) & 0 < x < 2, 2 < y < 4 \\ &= \frac{1}{8}x(6-x) & 0 < x < 2, y \geq 4 \\ &= \frac{1}{8}(y-2)(8-y) & x \geq 2, 2 < y < 4 \\ &= 1 & x \geq 2, y \geq 4 \end{aligned} \quad (14)$$

La función se ha representado en la figura 4-12.

La probabilidad de que (x, y) caiga en un rectángulo cualquiera $a_1 < x < b_1$, $a_2 < y < b_2$ puede escribirse mediante la función acumulativa:

$$\begin{aligned} P(a_1 < x < b_1, a_2 < y < b_2) &= P(x < b_1, y < b_2) - P(x < a_1, y < b_2) \\ &\quad - P(x < b_1, y < a_2) \\ &\quad \quad + P(x < a_1, y < a_2) \\ &= F(b_1, b_2) - F(a_1, b_2) - F(b_1, a_2) \\ &\quad \quad + F(a_1, a_2) \end{aligned} \quad (15)$$

Por consiguiente, en el ejemplo anterior, se tiene

$$\begin{aligned} P(0 < x < 1, 3 < y < 4) &= F(1, 4) - F(0, 4) - F(1, 3) + F(0, 3) \\ &= \frac{5}{8} - 0 - \frac{3}{8} + 0 \\ &= \frac{1}{4} \end{aligned}$$

Estas distribuciones pueden complicarse bastante cuando se trata de más de dos variantes y muchos problemas importantes de estadística aplicada están sin resolver por la excesiva complicación de las integraciones necesarias para resolverlos.

En este libro utilizaremos ordinariamente letras minúsculas para designar las funciones de densidad, y las mayúsculas correspondientes para representar su forma acumulativa. Así, tendremos.

$$G(x) = \int_{-\infty}^x g(t) dt$$

o, si la variante es discreta,

$$G(x) = \sum_{t \leq x} g(t)$$

Cuando hablemos de *densidad* nos referiremos específicamente a $g(x)$, y cuando hablemos de *distribución acumulativa*, a $G(x)$. La

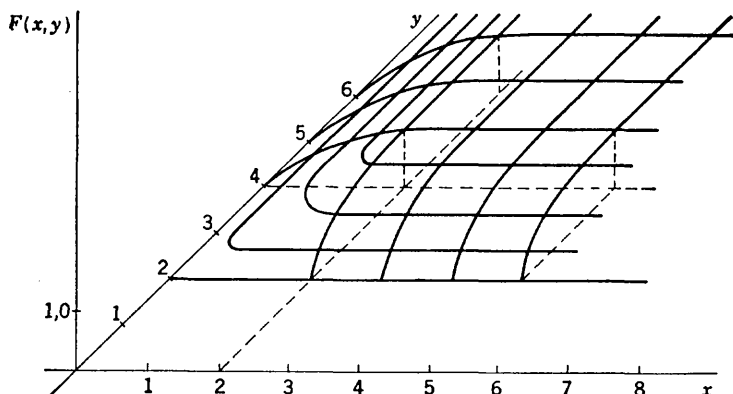


FIG. 4-12.

palabra *distribución* se utilizará como término más general que puede referirse tanto a la densidad como a la forma acumulativa.

4-5. Distribuciones marginales.—Cada distribución de más de una variable tiene asociadas varias distribuciones marginales. Sea $f(x, y)$ la densidad correspondiente a dos variantes continuas. Puede ocurrir que solo nos interese una de las variantes, p. ej., x . Buscaremos entonces una función de x tal que al integrarla sobre un intervalo, como $a < x < b$, nos dé la probabilidad de que x esté situada en este intervalo. Cada uno de tales intervalos corresponde en el plano x, y a una faja como la representada en la figura 4-13. Cualquier punto de esta faja satisface la condición $a < x < b$; por tanto,

$$P(a < x < b) = \int_a^b \int_{-\infty}^{\infty} f(x, y) dy dx \quad (1)$$

Cualquiera que sea la especificación de x , los límites de integración de y son de $-\infty$ a $+\infty$; de modo que podemos definir una función

$$f_1(x) = \int_{-\infty}^{\infty} f(x, y) dy \quad (2)$$

y esta será la función de densidad marginal, ya que

$$P(a < x < b) = \int_a^b f_1(x) dx \quad (3)$$

para cualquier par de valores a y b . Análogamente, la función de densidad marginal de y es

$$f_2(y) = \int_{-\infty}^{\infty} f(x, y) dx \quad (4)$$

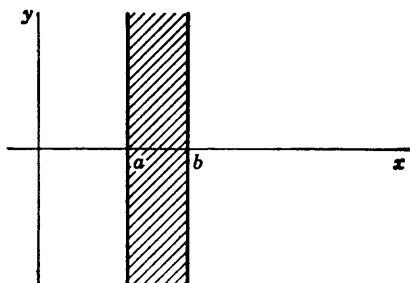


FIG. 4-13.

En general, dada una función de densidad $f(x_1, x_2, \dots, x_k)$, puede hallarse la densidad marginal de cualquier subconjunto de las variantes integrando la función respecto a todas las demás variantes entre los límites $-\infty$ y $+\infty$. Así, p. ej., la función de densidad de x_1, x_2 y x_4 es

$$\begin{aligned} f_{124}(x_1, x_2, x_4) &= \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(x_1, x_2, \dots, x_k) dx_3 dx_5 dx_6 \dots dx_k \end{aligned} \quad (5)$$

Refiriéndonos a la distribución definida por la ecuación (4-3-4), la función de densidad marginal de x es

$$\begin{aligned} f_1(x) &= \int_{-\infty}^{\infty} f(x, y) dy & -\infty < x < \infty \\ &= \int_2^4 \frac{1}{8} (6 - x - y) dy & 0 < x < 2 \\ &= \frac{1}{4} (3 - x) & 0 < x < 2 \\ &= 0 & x \leq 0 \text{ ó } x \geq 2 \end{aligned} \quad (6)$$

La función de distribución marginal acumulativa se encuentra fácilmente a partir de la función de distribución acumulativa. Para dos variables, esta función acumulativa marginal es

$$\begin{aligned} F_1(x) &= \int_{-\infty}^x \int_{-\infty}^{\infty} f(x, y) dy dx = \int_{-\infty}^x f_1(x) dx \\ &= F(x, \infty) \end{aligned} \quad (7)$$

Por tanto, basta con hacer infinita la variable que no nos interesa en la función acumulativa conjunta. Y en general, si $F(x_1, x_2, \dots, x_k)$ es una función acumulativa en k variantes, la función acumulativa marginal de x_1, x_2, x_4 , p. ej., es

$$F_{124}(x_1, x_2, x_4) = F(x_1, x_2, \infty, x_4, \infty, \dots, \infty) \quad (8)$$

En nuestro ejemplo particular podemos hallar la función acumulativa marginal de x , integrando $f_1(x)$; así

$$\begin{aligned} F_1(x) &= \int_{-\infty}^x f_1(t) dt \\ &= 0 & x \leq 0 \\ &= 1/8 x(6-x) & 0 < x < 2 \\ &= 1 & x \geq 2 \end{aligned} \quad (9)$$

El mismo resultado se obtiene haciendo que y tienda a infinito en $F(x, y)$, según queda definida por las ecuaciones (4-4-14).

4-6. Distribuciones condicionales.—Empezaremos por considerar una función de densidad bivalente, $f(x, y)$, que podría estar representada, p. ej., por la superficie de la figura 4-8. Imaginemos obtenido un punto (x, y) (p. ej., haciendo un disparo contra un blanco), y supongamos que se observa la segunda variante y , pero no la primera. Buscamos una función $f(x|y)$ que nos dé la densidad de x cuando se conozca y ; esto es, una función tal que

$$P(a < x < b|y) = \int_a^b f(x|y) dx \quad (1)$$

para valores cualesquiera de a y b .

Definimos $f(x|y)$ solo para aquellos valores de y tales que $f_2(y) > 0$. La definición es

$$f(x|y) = \frac{f(x, y)}{f_2(y)} \quad \text{cuando } f_2(y) > 0 \quad (2)$$

Por un razonamiento análogo, si $f_1(x)$ es la función de densidad marginal de x , la función de densidad condicional de y , dado x , será

$$f(y|x) = \frac{f(x, y)}{f_1(x)} \quad \text{cuando } f_1(x) > 0 \quad (3)$$

La función de densidad $f(x|y)$ es una función de densidad de la variante x ; y es un parámetro que tendrá un valor numérico

determinado para cada función de densidad condicional dada. Por tanto, $f_X(c)$ debe considerarse como constante. La función de densidad conjunta $f(x, y)$ está representada por una superficie sobre el plano x, y . Un plano perpendicular al x, y , que corte a este último según la recta $y=c$, cortará a la superficie según la curva $f(x, c)$. El área limitada por esta curva es

$$\int_{-\infty}^{\infty} f(x, c) dx = f_X(c)$$

En consecuencia, si dividimos $f(x, c)$ por $f_X(c)$, obtenemos una función de densidad, que es precisamente $f(x|c)$.

Para la función particular

$$f(x, y) = \frac{1}{8}(6 - x - y) \quad 0 < x < 2, 2 < y < 4$$

$$= 0 \quad \text{en otro caso}$$

hemos hallado en la sección anterior que la función de densidad marginal de x es

$$f_1(x) = \frac{1}{4}(3 - x) \quad 0 < x < 2$$

$$= 0 \quad \text{en otro caso}$$

En vista de (3), la función de densidad condicional de y para x fija es, por tanto,

$$f(y|x) = \frac{6 - x - y}{2(3 - x)} \quad 2 < y < 4, 0 < x < 2$$

Las distribuciones condicionales se definen de manera análoga para las distribuciones multivariantes; así, para cinco variantes que tengan como función de densidad $f(x_1, x_2, x_3, x_4, x_5)$, la función de densidad condicional de x_1, x_2, x_4 , para valores dados de x_3, x_5 , es

$$f(x_1, x_2, x_4|x_3, x_5) = \frac{f(x_1, x_2, x_3, x_4, x_5)}{f_{35}(x_3, x_5)} \quad \text{cuando } f_{35}(x_3, x_5) > 0$$

representando por $f_{35}(x_3, x_5)$ la función de densidad marginal de x_3 y x_5 .

4-7. Independencia.—Si en la función de densidad condicional $f(x|y)$ no interviene y , x es independiente de y en sentido probabilístico. Supongamos que sea este el caso y representemos $f(x|y)$ por $g(x)$. Puesto que según la sección 4-6

$$f(x|y) = g(x) = \frac{f(x, y)}{f_Y(y)} \quad \text{cuando } f_Y(y) > 0 \quad (1)$$

se deduce que

$$f(x, y) = g(x) f_2(y) \quad (2)$$

Por tanto, la función de densidad conjunta de x e y es el producto de dos funciones, en una de las cuales interviene solamente x , y en la otra solamente y . Si integramos (2) respecto a y , hallamos que $g(x)$ es sencillamente la función de densidad marginal de x .

Definición 4-2.—*Las k variantes $x_1, \dots, x_k$ son independientes en sentido probabilístico si (y solamente si) su distribución conjunta es igual al producto de sus distribuciones marginales.*

En general, si la distribución condicional de un subconjunto de cualquier conjunto de variantes es independiente de las demás variantes supuestas fijas, este subconjunto se dice que es independiente en sentido probabilístico de las restantes variantes. La función definida por la ecuación (4-3-6) nos procura un ejemplo de este caso:

$$\begin{aligned} f(x_1, x_2, x_3, x_4) &= 16x_1x_2x_3x_4 & 0 < x_i < 1 \text{ para todo } i \\ &= 0 & \text{en otro caso} \end{aligned}$$

La función de densidad marginal de, p. ej., x_2 y x_4 es,

$$\begin{aligned} f_{24}(x_2, x_4) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x_1, x_2, x_3, x_4) dx_1 dx_3 \\ &= 4x_2x_4 & 0 < x_2 < 1, 0 < x_4 < 1 \\ &= 0 & \text{en otro caso} \end{aligned}$$

Por tanto, la función de densidad condicional de x_1 y x_3 es

$$\begin{aligned} f(x_1, x_3 | x_2, x_4) &= 4x_1x_3 & 0 < x_1 < 1, 0 < x_3 < 1 \\ &= 0 & \text{en otro caso} \end{aligned}$$

Ni en esta función ni en sus límites intervienen x_2 y x_4 , de modo que el par de variantes (x_1, x_3) es independiente del par (x_2, x_4) en sentido probabilístico. En realidad, las cuatro variantes de esta distribución son mutuamente independientes, como puede deducirse del hecho de que la función puede descomponerse en cuatro funciones en cada una de las cuales interviene una sola de las variantes, y los límites son independientes.

4-8. Muestra aleatoria.—Consideremos el siguiente experimento: De una bolsa que contiene seis bolas rojas y cuatro negras, se extrae una bola y se anota su color. Sea x_1 el número de bolas rojas extraídas; la variable aleatoria x_1 puede tomar solo dos valores:

0 cuando la bola extraída no es roja (se extrae una bola negra) y 1 cuando se extrae una bola roja. Así, x_1 es una variable aleatoria cuya función de cuantía, si la extracción es tal que todas las bolas tienen la misma posibilidad de ser seleccionadas, es

$$f(x_1) = \left(\frac{6}{10}\right)^{x_1} \left(\frac{4}{10}\right)^{1-x_1} \quad x_1 = 0, 1$$

Supongamos ahora que la bola extraída es devuelta a la bolsa después de anotar su color, y que se repite el experimento. Designemos por x_2 el resultado de este segundo experimento; la función de cuantía de x_2 es

$$f(x_2) = \left(\frac{6}{10}\right)^{x_2} \left(\frac{4}{10}\right)^{1-x_2} \quad x_2 = 0, 1$$

Supongamos además que, en vez de considerar por separado estos dos resultados, nos interesa la distribución conjunta de la variable aleatoria bidimensional (x_1, x_2) . El experimento físico realizado es tal que la función de cuantía conjunta de la variable aleatoria bidimensional (x_1, x_2) será

$$\begin{aligned} g(x_1, x_2) &= f(x_1)f(x_2) = \left(\frac{6}{10}\right)^{x_1} \left(\frac{4}{10}\right)^{1-x_1} \left(\frac{6}{10}\right)^{x_2} \left(\frac{4}{10}\right)^{1-x_2} \\ &= \left(\frac{6}{10}\right)^{x_1+x_2} \left(\frac{4}{10}\right)^{2-(x_1+x_2)} \quad \begin{matrix} x_1 = 0, 1 \\ x_2 = 0, 1 \end{matrix} \end{aligned}$$

Cuando la función de cuantía conjunta de una variable aleatoria bidimensional es igual al producto de las funciones de cuantía de cada variable y estas funciones coinciden, decimos que se ha obtenido una muestra aleatoria de extensión 2 procedente de la función de cuantía $f(x)$. Estas ideas pueden generalizarse a más de dos variables aleatorias y también a funciones de densidad continuas. Así, formulamos la siguiente definición:

Definición 4-3.—Sean n variables aleatorias $x_1, x_2, \dots, x_n$ independientes conjuntamente, todas con la misma función de densidad $f(x)$. Decimos que $x_1, x_2, \dots, x_n$ es una muestra aleatoria de extensión n procedente de $f(x)$. La densidad conjunta de las n variables aleatorias $x_1, x_2, \dots, x_n$ es

$$g(x_1, x_2, \dots, x_n) = f(x_1)f(x_2) \dots f(x_n)$$

Así, p. ej., supongamos que se extrae una muestra aleatoria de extensión 3 procedente de la densidad $f(x) = 1, 0 < x < 1$. ¿Cuál es la probabilidad de que el valor de cada variable esté comprendido entre 0 y $1/2$? Deseamos

$$\begin{aligned} P(0 < x_1 < 1/2, 0 < x_2 < 1/2, 0 < x_3 < 1/2) &= \\ &= \int_0^{1/2} \int_0^{1/2} \int_0^{1/2} g(x_1, x_2, x_3) dx_1 dx_2 dx_3 \end{aligned}$$

pero, puesto que la muestra es aleatoria,

$$g(x_1, x_2, x_3) = f(x_1)f(x_2)f(x_3) = 1 \cdot 1 \cdot 1 = 1 \quad 0 < x_i < 1$$

$$= 0 \quad \text{en otro caso}$$

y, por tanto,

$$P(0 < x_1 < 1/2, 0 < x_2 < 1/2, 0 < x_3 < 1/2) =$$

$$= \int_0^{1/2} \int_0^{1/2} \int_0^{1/2} 1 \cdot dx_1 dx_2 dx_3 = 1/8$$

Supongamos que queremos hallar la probabilidad de que al menos una de las tres variables aleatorias tenga un valor comprendido entre $1/2$ y $3/4$. Esta probabilidad puede calcularse utilizando la distribución binomial. Cabe considerar las tres variables aleatorias como tres pruebas independientes, siendo p la probabilidad de obtener éxito en una prueba determinada:

$$p = P(1/2 < x_2 < 3/4) = P(1/2 < x_3 < 3/4) = P(1/2 < x_1 < 3/4)$$

$$= \int_{1/2}^{3/4} 1 \cdot dx = 1/4$$

La probabilidad de exactamente k éxitos en tres pruebas está dada por la distribución binomial,

$$P(k) = \binom{3}{k} p^k (1-p)^{3-k} = \binom{3}{k} (1/4)^k (3/4)^{3-k} \quad k=0, 1, 2, 3$$

La probabilidad de obtener al menos un éxito (al menos una variable aleatoria toma un valor del intervalo $1/2$ a $3/4$) es

$$P(1) + P(2) + P(3) = \sum_{k=1}^3 P(k) = 1 - P(0) = 1 - (3/4)^3 = 37/64$$

4-9. Distribuciones deducidas de otras.—Algunas veces es importante saber cómo deducir la función de densidad de una variable aleatoria y a partir de la función de densidad conocida de otra variable aleatoria x , cuando y es función de x ; es decir, $y=u(x)$. Supongamos, p. ej., que la función de densidad de la variable aleatoria x es

$$f(x) = 1 \quad 0 < x < 1$$

$$= 0 \quad \text{en otro caso}$$

e imaginemos que deseamos hallar la función de densidad de y , siendo $y=8x-2$. La distribución de y se llama distribución deducida de la distribución de x ; estas distribuciones deducidas desem-

peñan un papel sumamente importante en estadística. Hay muchas maneras de obtener las funciones de densidad de distribuciones deducidas, pero aquí nos limitaremos a ver un solo método, el que utiliza la distribución acumulativa. En el capítulo 10 se explicarán con detalle otros métodos.

Sea $g(y)$ la función de densidad de y y sea $G(y)$ la distribución acumulativa de y . Puesto que $f(x)=1$ para $0 < x < 1$ y 0 en otro caso, se tiene

$$\begin{aligned} F(x) &= P(x \leq x) = 0 && \text{para } x \leq 0 \\ F(x) &= P(x \leq x) = \int_0^x f(t) dt = \int_0^x 1 \cdot dt = x && \text{para } 0 < x < 1 \\ F(x) &= P(x \leq x) = 1 && \text{para } x \geq 1 \end{aligned} \quad (1)$$

Por definición de distribución acumulativa de y , resulta:

$$G(y) = P(y \leq y) \quad (2)$$

Sustituyendo en (2) $y=8x-2$, tenemos

$$G(y) = P(8x-2 \leq y) = P\left(x \leq \frac{y+2}{8}\right) \quad (3)$$

Pero, en virtud de (1), si reemplazamos x por $\frac{y+2}{8}$, se obtiene:

$$G(y) = \begin{cases} P\left(x \leq \frac{y+2}{8}\right) = 0 & \text{si } \frac{y+2}{8} \leq 0 \\ P\left(x \leq \frac{y+2}{8}\right) = \frac{y+2}{8} & \text{si } 0 < \frac{y+2}{8} < 1 \\ P\left(x \leq \frac{y+2}{8}\right) = 1 & \text{si } \frac{y+2}{8} \geq 1 \end{cases} \quad (4)$$

Por (4),

$$\begin{aligned} G(y) &= 0 && \text{si } y \leq -2 \\ &= \frac{y+2}{8} && \text{si } -2 < y < 6 \\ &= 1 && \text{si } 6 \leq y \end{aligned}$$

Para hallar la función de densidad, derivamos $G(y)$ en los puntos en que existe derivada; obtenemos así

$$\begin{aligned} g(y) &= 1/8 && -2 < y < 6 \\ &= 0 && \text{en otro caso} \end{aligned}$$

Puede comprobarse, para estar seguros, que la función satisface las condiciones que la califican como función de densidad.

Como otro ejemplo, sea

$$f(x) = e^{-x} \quad x > 0$$

$$= 0 \quad \text{en otro caso}$$

la función de densidad de la variable aleatoria x .

Supongamos que se extrae una muestra aleatoria de extensión 2, x_1, x_2 , procedente de $f(x)$ y que deseamos hallar la función de den-

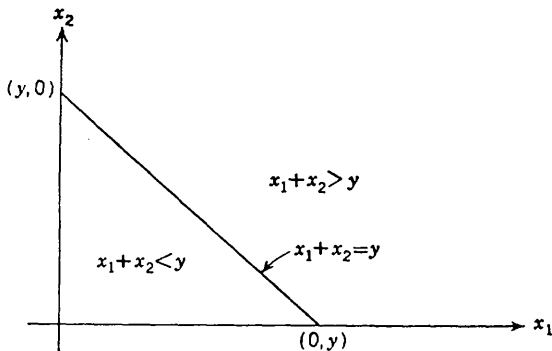


FIG. 4-14.

sidad de la variable aleatoria y , siendo $y = x_1 + x_2$. Por definición de muestra aleatoria, la densidad conjunta de las variables aleatorias x_1 y x_2 es

$$h(x_1, x_2) = e^{-x_1} e^{-x_2} \quad x_1 > 0, x_2 > 0$$

$$= 0 \quad \text{en otro caso}$$

Para obtener la función de densidad de y , que designaremos por $g(y)$, hallamos en primer lugar la función acumulativa $G(y)$ y la derivamos. Ahora bien:

$$G(y) = P(y \leq y) = P(x_1 + x_2 \leq y)$$

Pero $P(x_1 + x_2 \leq y)$ es el volumen de $h(x_1, x_2)$ en la región $x_1 + x_2 \leq y$. Puesto que $h(x_1, x_2) = 0$, excepto cuando x_1 y x_2 son ambos positivos, solo necesitamos considerar el primer cuadrante del plano $x_1 x_2$ (véase Fig. 4-14).

El volumen es

$$G(y) = P(y \leq y) = 0 \quad \text{para } y \leq 0$$

$$G(y) = P(y \leq y) = P(x_1 + x_2 \leq y) = \int_0^y \int_0^{y-x_1} e^{-(x_1+x_2)} dx_2 dx_1$$

$$= \int_0^y (e^{-x_1} - e^{-y}) dx_1 = 1 - e^{-y} - ye^{-y} \quad \text{para } y > 0$$

Así, tenemos

$$G(y) = 0 \quad y \leq 0$$

$$= 1 - (1+y)e^{-y} \quad y > 0$$

y

$$g(y) = \frac{dG(y)}{dy} = \begin{cases} 0 & y \leq 0 \\ ye^{-y} & y > 0 \end{cases}$$

o

$$g(y) = ye^{-y} \quad y > 0$$

$$= 0 \quad \text{en otro caso}$$

Al definir una función de densidad, algunas veces la especificaremos solo para el conjunto en el que es positiva. Así, p. ej., a veces escribiremos

$$f(x) = 2x \quad 0 < x < 1$$

$$= 0 \quad \text{en otro caso}$$

como

$$f(x) = 2x \quad 0 < x < 1$$

PROBLEMAS

En los problemas que siguen, $f(x)$ y $g(x)$ designarán funciones de densidad de la variable aleatoria x ; $f(x, y)$ representará la función de densidad conjunta de las dos variables aleatorias x, y ; etc.

1. Sea x una variable aleatoria con función de densidad

$$f(x) = \frac{1}{(x+1)^2} \quad x > 0$$

$$= 0 \quad \text{en otro caso}$$

Definamos los sucesos A_1, A_2, A_3 por

$$A_1 = \{x : -\infty < x < 0\}$$

$$A_2 = \{x : 0 \leq x < \infty\}$$

$$A_3 = \{x : 0 \leq x \leq 1\}$$

$$A_4 = \{x : -6 \leq x \leq 0\}$$

Hallar la probabilidad de los siguientes sucesos:

a) $\overline{A_1}$	i) $A_1 \cup A_2$
b) $\overline{A_1} \cap A_2$	j) $\overline{A_3}$
c) $\overline{A_1} \cup A_2$	k) $\overline{A_4} \cap A_1$
d) $\overline{A_2} \cap A_1$	l) $A_3 \cap A_4$
e) $A_3 \cap A_2$	m) $(A_3 \cap A_4) \cap (A_1 \cup A_2)$
f) $A_3 \cup A_2$	n) $\overline{A_2}$
g) $A_3 \cup A_1$	o) $A_1 \cup A_3$
h) $A_1 \cup A_4$	p) $\overline{A_1 \cup A_3}$

2. Sea

$$f(x) = e^{-x} \quad x > 0$$

$$= 0 \quad x \leq 0$$

la función de densidad de la variable aleatoria x . Hallar la probabilidad de cada uno de los sucesos a), b), ..., p) del problema 1.

3. Si la variable aleatoria x tiene como función de densidad $f(x) = 2x$ cuando $0 < x < 1$ e igual a 0 en otro caso, hállese la probabilidad de que: a) $x < 1/2$; b) $1/4 < x < 1/2$; c) $x > 3/4$ dado $x > 1/2$.

4. Defínase una función de densidad utilizando $x(2-x)$ sobre el conjunto $0 < x < 2$. Hállese la probabilidad de que $a < x < b$, si

$$0 < a < b < 2$$

y para $a < 0$, $2 < b$.

5. Si $f(x) = 4x^3$ cuando $0 < x < 1$ e igual a 0 en otro caso, hállese el número a tal que x tenga igual probabilidad de ser mayor que de ser menor que a . Hállese el número b tal que la probabilidad de que x exceda a b sea igual a 0,05.

6. Una variante x tiene por función de densidad $f(x) = x/2$ cuando $0 < x < 2$ e igual a 0 en otro caso. Si se obtienen dos valores de x , ¿cuál es la probabilidad de que ambos sean mayores que 1? Si se obtienen tres, ¿cuál es la probabilidad de que precisamente dos de ellos sean mayores que 1?

7. Una variante x tiene como función de densidad $f(x) = 1$ para $0 < x < 1$ y 0 en otro caso. Determínese el número a tal que sea 0,9 la probabilidad de que al menos uno de cuatro valores de x extraídos al azar exceda a a .

8. Supongamos que la vida en horas de un cierto tipo de lámpara de radio tenga por función de densidad $f(x) = 100/x^2$ cuando $x > 100$ y cero para $x \leq 100$. ¿Cuál es la probabilidad de que, de 3 de estas lámparas, ninguna tenga que ser sustituida en un determinado aparato de radio durante las primeras 150 horas de uso? ¿Cuál es la probabilidad de que las 3 lámparas tengan que reemplazarse durante las primeras 150 horas?

9. Una máquina fabrica pernos cuyos diámetros se distribuyen con función de densidad $f(x) = K(x-0,24)^3(x-0,26)^2$ si $0,24 < x < 0,26$, y 0 en otro caso; K es un número tal que hace $\int_{-\infty}^{\infty} f(x)dx = 1$. Los pernos se desechan si sus diámetros se desvían de 0,25 en más de 0,008. ¿Cuál es la proporción de pernos que es de esperar sean desechados?

10. Un avión de bombardeo, en vuelo directo sobre un ferrocarril, lleva 3 bombas. Si una bomba cae a menos de 40 pies de la vía, esta quedará suficientemente destruida para que haya de interrumpirse el tráfico. La densidad de impactos de una bomba viene dada por la función

$$\begin{aligned} f(x) &= (100+x)/10\,000 & -100 < x < 0 \\ &= (100-x)/10\,000 & 0 \leq x < 100 \\ &= 0 & \text{en otro caso} \end{aligned}$$

donde x representa la desviación en sentido perpendicular a la vía. Si se lanzan las 3 bombas, ¿cuál es la probabilidad de que quede destruida la vía?

11. Con los datos del problema anterior, se supone que el avión puede llevar 8 bombas más pequeñas, pero que cualquiera de estas debe caer a menos de 15 pies de la vía para destruirla. ¿Deberían emplearse las bombas pesadas o las ligeras en esta misión?

12. Una estación de suministro recibe gasolina una vez por semana. Si su volumen semanal de ventas, en miles de galones, se distribuye con función de densidad $f(x) = 5(1-x)^4$, $0 < x < 1$, ¿cuál deberá ser la capacidad de su depósito a fin de que la probabilidad de que se agote en una semana determinada sea 0,01?

13. Una remesa de munición de pequeño calibre se acepta como satisfactoria si en una muestra de 5 disparos ninguno se aleja más de 2 pies del centro de un blanco a distancia prefijada. Si la densidad de r , distancia del centro del blanco a un impacto dado, viene dada por la función

$$f(r) = \frac{2re^{-r^2}}{(1-e^{-9})}$$

$0 < r < 3$ para una remesa dada, ¿cuál es la probabilidad de que se acepte la remesa?

14. Si $f(x, y) = 1$, cuando $0 < x < 1$, $0 < y < 1$, y 0 en otro caso, hállese la probabilidad de que: a) $x < 1/2$, $y < 1/2$; b) $x+y < 1$; c) $x+y > 1$; d) $x > 2y$; e) $x > 1/3$; f) $x^2+y^2 < 1/4$; g) $x=y$; h) $x > 1/2$ dado $y < 1/2$; i) $x > y$ dado $y > 1/2$.

15. Si $f(x, y) = e^{-(x+y)}$ cuando $x > 0$, $y > 0$, y cero en otro caso, hállese $P(x > 1)$; $P(a < x+y < b)$ si $0 < a < b$; $P(x < y | x < 2y)$.

16. Para la misma distribución del problema 15, hállese el número a tal que $P(x+y < a) = 1/2$.

17. Si se eligen tres puntos (x, y) al azar, procedentes de la densi-

dad dada en el problema 15, ¿cuál es la probabilidad de que al menos uno de ellos caiga en el cuadrado $0 < x < 1$, $0 < y < 1$?

18. Una máquina fabrica ejes de diámetro x y otra máquina fabrica cojinetes de diámetro interior y . Supongamos que la densidad de x e y es $f(x, y) = 2500$, $0,49 < x < 0,51$, $0,51 < y < 0,53$, y cero en otro caso. Un cojinete se adapta satisfactoriamente a un eje si su diámetro excede al del eje al menos en 0,004, pero no en más de 0,036. ¿Cuál es la probabilidad de que un cojinete y un eje elegidos al azar ajusten bien?

19. Hallar y representar aproximadamente la distribución acumulativa correspondiente al problema 8. Empleése esta distribución para hallar $P(150 < x < 250)$.

20. Hallar y representar aproximadamente la distribución acumulativa de la función dada en el problema 15, y empleése el resultado para hallar $P(1 < x < 2, 3 < y < 4)$.

21. Hállese la función de densidad marginal de x para la distribución del problema 15: a) por integración respecto a y ; b) utilizando el resultado del problema 20 para obtener la distribución marginal acumulativa, derivando después esta.

22. Hállese la función de densidad condicional de x , dado y , para la distribución del problema 15. ¿Cuál es $P(0 < x < 1 | y = 2)$?

23. Si $f(x, y) = (n-1)(n-2)/(1+x+y)^n$ cuando $x > 0$ e $y > 0$, y cero en otro caso, hállese $F(x, y)$, $f_1(x)$, $F_1(x)$, $f(y|x)$; supóngase $n > 2$.

24. Si $f(x, y) = 24y(1-x-y)$ en el triángulo limitado por los ejes y la recta $x+y=1$, y cero en otro caso, hállese $f(x|y)$.

25. Si $f(x, y) = 3x$, $0 < y < x$, $0 < x < 1$, hállese la función de densidad condicional de x , dado y .

26. Si $f(x|y) = 3x^2/y^3$, $0 < x < y$, y $f_2(y) = 5y^4$, $0 < y < 1$, hállese $P(x > 1/2)$.

27. Si $f(x, y, z) = 8xyz$, $0 < x < 1$, $0 < y < 1$, $0 < z < 1$, hállese $P(x < y < z)$.

28. Si $f(x) = 1/(1+x)^2$, $x > 0$, hállese la función de densidad de x sabiendo que $x > 1$.

29. Si $f(x, y) = 1$, $0 < x < 1$, $0 < y < 1$, hállese la función de densidad condicional de x e y sabiendo que $y < x^n$, $n > 0$.

30. Si $f(x) = 1$, $0 < x < 1$, determínese la función de densidad de $y = 3x + 1$. (Hállese primero la función de distribución acumulativa de y , y después derívese esta.)

31. Si $f(x) = 2xe^{-x^2}$, $x > 0$, hállese la función de densidad de $y = x^2$.

32. Si $f(x, y) = 1$, $0 < x < 1$, $0 < y < 1$, hállese la función de densidad de $z = x + y$.

33. Si $f(x, y) = e^{-(x+y)}$, $x > 0$, $y > 0$, hállese la función de densidad de

$$z = \frac{(x+y)}{2}$$

34. Si $f(x, y) = 4xye^{-(x^2+y^2)}$, $x > 0$, $y > 0$, hállese la función de densidad de $z = \sqrt{x^2 + y^2}$.

35. Si $f(x, y) = 4xy$, $0 < x < 1$, $0 < y < 1$, hállese la función de densidad conjunta de $u = x^2$, $v = y^2$.

36. Si $f(x, y) = 3x$, $0 < y < x$, $0 < x < 1$, hállese la función de densidad de $z = x - y$.

37. Si $f(x) = (1+x)/2$, $-1 < x < 1$, hállese la función de densidad de $y = x^2$.

38. Si $f(x, y) = 1$, $0 < x < 1$, $0 < y < 1$, hállese la función de densidad de z definida por $z = x + y$, si $x + y < 1$, y $z = x + y - 1$ si $x + y > 1$.

39. Si $f(x, y) = e^{-(x+y)}$, $x > 0$, $y > 0$, hállese la función de densidad conjunta de $u = x + y$ y $v = x$. ¿Cuál es la función de densidad marginal de v ?

40. Si $f(x, y, z) = e^{-(x+y+z)}$, $x > 0$, $y > 0$, $z > 0$, hállese la función de densidad de su media $u = (x + y + z)/3$.

41. Si $f(x, y) = 4x(1-y)$, $0 < x < 1$, $0 < y < 1$, hállese la función de densidad de x sabiendo que $y < 1/2$.

42. Si la distribución de x viene dada por $f(x)$, $x > 0$, hállese la función de densidad de $y = ax^2 + b$, $a > 0$.

43. Si la distribución de x viene dada por $f(x)$, $-\infty < x < \infty$, y si u es una función de x definida por $y = u(x)$, tal que su derivada primera u' sea positiva y continua para todo x , hállese la función de densidad de $y = u(x)$.

44. Si $f(x, y) = g(x)g(y)$, $x > 0$, $y > 0$, hállese $P(x > y)$.

45. Si $f(x, y, z) = g(x)g(y)g(z)$, $x > 0$, $y > 0$, $z > 0$, ¿cuál es la probabilidad de que las coordenadas de un punto (x, y, z) obtenido al azar no satisfagan la condición $x > y > z$ o $x < y < z$.

46. ¿En qué distribuciones de las definidas en los problemas 23, 24, 32, 33 y 34 son las variantes independientes en sentido probabilístico?

BIBLIOGRAFIA

- BRUNK, H.: *An Introduction to Mathematical Statistics*, Ginn & Company, Boston, 1960.
- CLARK, C.: *An Introduction to Statistics*, John Wiley & Sons, Inc., Nueva York, 1953.
- FRASER, D. A. S.: *Statistics—An Introduction*, John Wiley & Sons, Inc., Nueva York, 1958.
- FRYER, H.: *Elements of Statistics*, John Wiley & Sons, Inc., Nueva York, 1954.
- HOEL, P. G.: *Elementary Statistics*, John Wiley & Sons, Inc., Nueva York, 1960.
- HOGG, R., y A. CRAIG: *Introduction to Mathematical Statistics*, The Macmillan Company, Nueva York, 1959.
- RAO, C. R.: *Advanced Statistical Methods in Biometric Research*, John Wiley & Sons, Inc., Nueva York, 1952.

CAPITULO 5

VALORES ESPERADOS Y MOMENTOS

5-1. Valores esperados.—El valor esperado de una variable aleatoria o de cualquier función de una variable aleatoria se obtiene hallando el valor medio de la función para todos los valores posibles de la variable. Estudiemos un ejemplo concreto. Si se lanzan tres monedas, la distribución del número de caras obtenidas es la binomial

$$f(x) = \binom{3}{x} \left(\frac{1}{2}\right)^3 \quad x=0, 1, 2, 3 \quad (1)$$

Para un valor determinado de x , como $x=2$, consideramos $f(2) = \frac{3}{8}$ como la frecuencia relativa con que aparecerían dos caras en un gran número de tiradas. Así, en 1000 tiradas es de esperar que no aparezcan caras en unas $1000 \times \frac{1}{8} = 125$ tiradas; que aparezca una cara en $1000 \times \frac{3}{8} = 375$ tiradas; dos caras en 375 tiradas, y tres caras en 125 tiradas. Hallemos ahora el número medio de caras en 1000 tiradas. Es de esperar que el número total de caras sea

$$125 \times 0 + 375 \times 1 + 375 \times 2 + 125 \times 3 = 1500$$

en las 1000 tiradas; por tanto, la media esperada es 1,5 caras por tirada. Este es el *valor esperado*, o *valor medio*, de x . Está claro que el mismo resultado se obtendría sin más que multiplicar todos los valores de x por sus probabilidades y sumar después los resultados; esto es,

$$0 \times \frac{1}{8} + 1 \times \frac{3}{8} + 2 \times \frac{3}{8} + 3 \times \frac{1}{8} = 1,5$$

El valor esperado es una media teórica o ideal. No es que esperemos efectivamente que x tome su valor esperado en una tirada determinada; en realidad, en el ejemplo anterior ello sería imposible. Sin embargo, sí que cabe esperar razonablemente que el valor medio de x en un gran número de tiradas se acerque al valor esperado de x .

Definición 5-1.—Sea x una variable aleatoria con densidad $f(x)$. El valor esperado de x , $E(x)$, es

$$E(x) = \sum_x x f(x)$$

si x es discreta, y

$$E(x) = \int_{-\infty}^{\infty} xf(x) dx$$

si x es continua.

Estas cantidades se definen como el valor esperado solo si son absolutamente convergentes. Si no lo son, decimos que el valor esperado de x no existe. Así, en el ejemplo anterior, tenemos

$$E(x) = \sum_{x=0}^3 xf(x) = 1,5$$

A veces nos interesará la media o valor esperado de cierta función de la variable aleatoria x , en lugar del correspondiente a la propia x . Así, p. ej., podemos estar interesados en el valor esperado de $2x$ o de $x^2 + 1$, etc.

Teorema 5-1.—Sea x una variable aleatoria con densidad $f(x)$. El valor esperado de una función u de la variable aleatoria x es

$$E[u(x)] = \sum_x u(x)f(x)$$

si x es discreta, y

$$E[u(x)] = \int_{-\infty}^{\infty} u(x)f(x) dx$$

si x es continua.

De nuevo observaremos que estas cantidades se definen como valores esperados solo si son absolutamente convergentes. La misma observación debe hacerse para todos los valores esperados de variables aleatorias y no la volveremos a repetir en lo sucesivo. La demostración del teorema 5-1 cae fuera del objetivo de este libro.

Ejemplo 5-1.—Supongamos $f(x)$ definida por (1); si $u(x) = x^2 + 1$, se tiene que

$$\begin{aligned} E[u(x)] &= E[x^2 + 1] = \sum_{x=0}^3 (x^2 + 1)f(x) \\ &= 1 \times \frac{1}{8} + 2 \times \frac{3}{8} + 5 \times \frac{3}{8} + 10 \times \frac{1}{8} = 4 \end{aligned}$$

A continuación definiremos el valor esperado para variables aleatorias multivariantes.

Definición 5-2.—Sea $(x_1, x_2, \dots, x_k)$ una variable aleatoria k -dimensional con densidad $f(x_1, x_2, \dots, x_k)$. El valor esperado de x_i para

cualquier $i=1, 2, \dots, k$ se designa por $E(x_i)$ y se define por

$$E(x_i) = \sum_{x_1} \sum_{x_2} \dots \sum_{x_k} x_i f(x_1, \dots, x_k)$$

si la variable aleatoria es discreta, y por

$$E(x_i) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} x_i f(x_1, \dots, x_k) dx_1 dx_2 \dots dx_k$$

si la variable aleatoria es continua.

Para hallar el valor esperado de una función de variables aleatorias, utilizaremos el siguiente teorema, cuya demostración omitimos.

Teorema 5-2.—Sea $(x_1, x_2, \dots, x_k)$ una variable aleatoria k -dimensional con densidad $f(x_1, x_2, \dots, x_k)$. El valor esperado de una función u de la variable aleatoria es

$$E[u(x_1, \dots, x_k)] = \sum_{x_1} \sum_{x_2} \dots \sum_{x_k} u(x_1, \dots, x_k) f(x_1, \dots, x_k)$$

si la variable aleatoria es discreta y

$$\begin{aligned} E[u(x_1, \dots, x_k)] &= \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} u(x_1, \dots, x_k) f(x_1, \dots, x_k) dx_1 \dots dx_k \end{aligned}$$

si la variable aleatoria es continua.

Consideremos algunos ejemplos.

Ejemplo 5-2.—Sea la variable aleatoria bidimensional (x_1, x_2) con densidad de probabilidad

$$\begin{aligned} f(x_1, x_2) &= p^{x_1+x_2}(1-p)^{2-(x_1+x_2)} & x_1 &= 0, 1; \quad x_2 = 0, 1 \\ & & 0 &< p < 1 \end{aligned}$$

Supongamos que queremos calcular el valor esperado de: a) $x_1 + x_2$; b) $x_1 x_2$; c) x_1 . Tenemos:

$$\begin{aligned} a) \quad E(x_1 + x_2) &= \sum_{x_1=0}^1 \sum_{x_2=0}^1 (x_1 + x_2) p^{x_1+x_2} (1-p)^{2-(x_1+x_2)} = \\ &= 2p(1-p) + 2p^2 = 2p \end{aligned}$$

$$b) \quad E(x_1 x_2) = \sum_{x_1=0}^1 \sum_{x_2=0}^1 (x_1 x_2) p^{x_1+x_2} (1-p)^{2-(x_1+x_2)} = p^2$$

$$c) \quad E(x_1) = \sum_{x_1=0}^1 \sum_{x_2=0}^1 (x_1) p^{x_1+x_2} (1-p)^{2-(x_1+x_2)} = p(1-p) + p^2 = p$$

Ejemplo 5-3.—Sea x una variable aleatoria con densidad

$$\begin{aligned} f(x) &= 2x & 0 < x < 1 \\ &= 0 & \text{en otro caso} \end{aligned}$$

Supongamos que deseamos hallar el valor esperado de $3x^2 - 1$. Así, $u(x) = 3x^2 - 1$, y

$$E[u(x)] = E[3x^2 - 1] = \int_0^1 (3x^2 - 1)2x \, dx = 1/2$$

Ejemplo 5-4.—Sea la variable aleatoria bidimensional (x_1, x_2) con densidad

$$\begin{aligned} f(x_1, x_2) &= 4x_1x_2 & 0 < x_1 < 1, 0 < x_2 < 1 \\ &= 0 & \text{en otro caso} \end{aligned}$$

Supongamos que queremos: a) $E(3x_1 + 2x_2)$; b) $E(x_1x_2)$; c) $E(x_1)$. Para a) obtenemos $u(x_1, x_2) = 3x_1 + 2x_2$ y

$$E[u(x_1, x_2)] = E(3x_1 + 2x_2) = \int_0^1 \int_0^1 (3x_1 + 2x_2)4x_1x_2 \, dx_1 \, dx_2 = 10/3$$

Para b) tenemos

$$E(x_1x_2) = \int_0^1 \int_0^1 (x_1x_2)(4x_1x_2) \, dx_1 \, dx_2 = 4/9$$

Para c) resulta

$$E(x_1) = \int_0^1 \int_0^1 x_1(4x_1x_2) \, dx_1 \, dx_2 = 2/3$$

Para evitar confusiones en la notación del valor esperado con la de las funciones, no utilizaremos nunca la letra E para representar una función. $E(g)$ representará siempre el valor esperado de g y nunca el valor de la función E en g . En el resto de este capítulo, no distinguiremos entre funciones de densidad discretas y continuas. Los valores esperados se darán siempre por medio de integrales, pero se entiende que las integrales deben reemplazarse por sumas en aquellos problemas concretos donde intervengan funciones de densidad discretas.

A continuación se enuncian algunos teoremas sobre valores esperados.

Teorema 5-3.—Sea x una variable aleatoria con densidad $f(x)$. Si c es una constante,

$$E(c) = c$$

Teorema 5-4.—Sea x una variable aleatoria con densidad $f(x)$. Si c es una constante,

$$E[cu(x)] = cE[u(x)]$$

Teorema 5-5.—Sea x una variable aleatoria con densidad $f(x)$. El valor esperado de la suma de dos funciones de x es la suma de los valores esperados; es decir,

$$E[u(x) + v(x)] = E[u(x)] + E[v(x)]$$

Estos resultados son casos especiales del siguiente teorema más general.

Teorema 5-6.—Sea $(x_1, x_2, \dots, x_k)$ una variable aleatoria k -dimensional con densidad $f(x_1, x_2, \dots, x_k)$, y sean $c_1, c_2, \dots, c_t$ constantes y $u_1, u_2, \dots, u_t$ funciones de la variable aleatoria. Entonces,

$$\begin{aligned} E[c_1 u_1(x_1, x_2, \dots, x_k) + c_2 u_2(x_1, x_2, \dots, x_k) + \dots \\ + c_t u_t(x_1, x_2, \dots, x_k)] = c_1 E[u_1(x_1, x_2, \dots, x_k)] \\ + c_2 E[u_2(x_1, x_2, \dots, x_k)] + \dots + c_t E[u_t(x_1, x_2, \dots, x_k)] \end{aligned}$$

Dejamos al cuidado del lector la demostración de estos teoremas.

Ejemplo 5-5.—Consideremos el apartado a) del ejemplo 5-4. Por el teorema 5-6, tenemos

$$E[3x_1 + 2x_2] = E[3x_1] + E[2x_2] = 3E(x_1) + 2E(x_2) = 3 \cdot 2/3 + 2 \cdot 2/3 = 10/3$$

5-2. Momentos.—Los momentos de una distribución son los valores esperados de las potencias de la variable aleatoria que tiene dicha distribución. El momento r -ésimo de x suele designarse por μ'_r y es

$$\mu'_r = E(x^r) = \int_{-\infty}^{\infty} x^r f(x) dx \quad (1)$$

El primer momento μ'_1 recibe el nombre de *media* de x . Los momentos respecto a un punto arbitrario cualquiera se definen por medio de la fórmula

$$E[(x - a)^r] = \int_{-\infty}^{\infty} (x - a)^r f(x) dx \quad (2)$$

y haciendo a igual a la media, obtenemos los momentos respecto a la media, que suelen designarse por μ_r :

$$\mu_r = E[(x - \mu'_1)^r] = \int_{-\infty}^{\infty} (x - \mu'_1)^r f(x) dx \quad (3)$$

Tenemos

$$\begin{aligned}\mu_1 &= \int_{-\infty}^{\infty} xf(x) dx - \mu'_1 \int_{-\infty}^{\infty} f(x) dx \\ &= \mu'_1 - \mu'_1 = 0\end{aligned}\quad (4)$$

y

$$\begin{aligned}\mu_2 &= \int_{-\infty}^{\infty} (x - \mu'_1)^2 f(x) dx \\ &= \int_{-\infty}^{\infty} [x^2 - 2x\mu'_1 + (\mu'_1)^2] f(x) dx \\ &= \mu'_2 - 2\mu'_1\mu'_1 + (\mu'_1)^2 \\ &= \mu'_2 - (\mu'_1)^2\end{aligned}\quad (5)$$

Este segundo momento respecto a la media recibe el nombre de *varianza* de x .

El valor medio de una variante localiza el centro de esta distribución en el siguiente sentido: si se considera el eje de las x como una barra de densidad variable, y se supone que la densidad en cualquier punto viene dada por $f(x)$, puede verse mediante cálculos elementales que el valor $x = \mu'_1$ es el centro de gravedad de la barra. Por tanto, la media puede considerarse como un valor central de la variante. Por esta razón se dice que la media es un parámetro de posición, ya que indica dónde está situado el centro de la distribución (en el sentido de centro de gravedad) sobre el eje de las x . A veces se utilizan otros valores centrales para indicar la posición o situación de una distribución. Uno es la *mediana*, que se define como el punto en que una recta vertical biseca el área limitada por la curva $f(x)$. La mediana es, por tanto, el punto μ'' tal que

$$\int_{-\infty}^{\mu''} f(x) dx = 1/2 = \int_{\mu''}^{\infty} f(x) dx \quad (6)$$

Otro valor central para distribuciones que tengan un máximo es la *moda*, que es el punto en que $f(x)$ alcanza su máximo. Sería fácil proponer otros valores centrales; los anteriores son los de uso corriente, y de los tres, la media es, con mucho, el más utilizado. A menudo utilizaremos el símbolo μ sin acento ni subíndice para designar la media.

La varianza μ_2 de una distribución es una medida de su dispersión o esparcimiento. Si la mayor parte del área limitada por la curva está próxima a la media, la varianza será pequeña. Mientras que si el área se extiende sobre un recorrido considerable, la va-

rianza será grande. En la figura 6-3 se representan distribuciones con diferentes varianzas. La varianza es necesariamente positiva o nula, ya que es la integral o suma de cantidades no negativas. Solo se anula cuando la distribución está concentrada en un punto, esto es cuando la distribución es discreta y hay solo un resultado posible. Suele emplearse el símbolo σ^2 para designar la varianza; la raíz cuadrada positiva de la varianza, σ , recibe el nombre de *desviación estándar* o *desviación típica*.

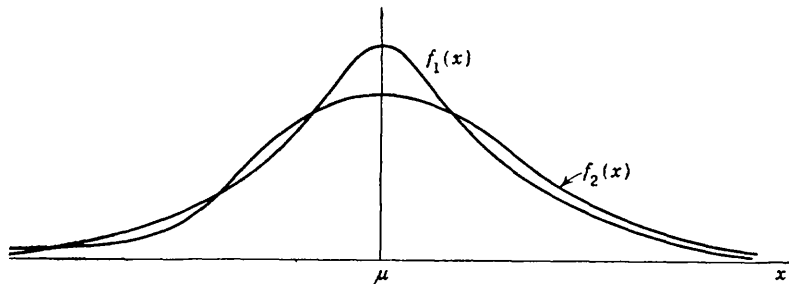


FIG. 5-1.

Veamos con mayor detalle de qué manera la varianza caracteriza a la distribución. Supongamos que $f_1(x)$ y $f_2(x)$ son dos densidades con la misma media, tales que

$$\int_{\mu-a}^{\mu+a} [f_1(x) - f_2(x)] dx \geq 0 \quad (7)$$

para cada valor de a . En la figura 5-1 se representan estas dos funciones de densidad. Puede demostrarse que en este caso la varianza σ_1^2 de la primera distribución es menor que la varianza σ_2^2 de la segunda. No demostraremos esto rigurosamente; pero el razonamiento es aproximadamente el siguiente: sea

$$g(x) = f_1(x) - f_2(x)$$

en donde $f_1(x)$ y $f_2(x)$ satisfacen a (7). Puesto que

$$\int_{-\infty}^{\infty} g(x) dx = 0,$$

el área positiva entre $g(x)$ y el eje x será igual al área negativa. Además, según (7), cada elemento positivo de área $g(x')dx'$ puede compensarse por un elemento negativo $g(x'')dx''$, de modo que x'' esté

más alejada de μ que x' . Cuando se multiplican estos elementos de área por $(x - \mu)^2$, los elementos negativos quedan multiplicados por

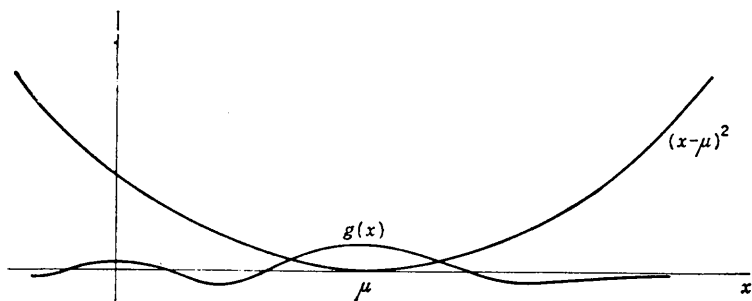


FIG. 5-2.

factores mayores que los correspondientes elementos positivos (véase Fig. 5-2); por tanto,

$$\int_{-\infty}^{\infty} (x - \mu)^2 g(x) dx < 0$$

a menos que $f_1(x)$ y $f_2(x)$ sean iguales. De aquí se deduce que $\sigma_1^2 < \sigma_2^2$. La recíproca no es cierta; esto es, si se sabe que $\sigma_1^2 < \sigma_2^2$ no puede deducirse de ello que las correspondientes funciones de densidad satisfagan a (7) para todos los valores de a , aunque sí puede verse que (7) debe ser cierta para determinados valores de a . Por consiguiente, la condición $\sigma_1^2 < \sigma_2^2$ no proporciona información precisa sobre la naturaleza de las distribuciones correspondientes, pero pone de manifiesto que $f_1(x)$ tiene más área cerca de la media que $f_2(x)$, al menos

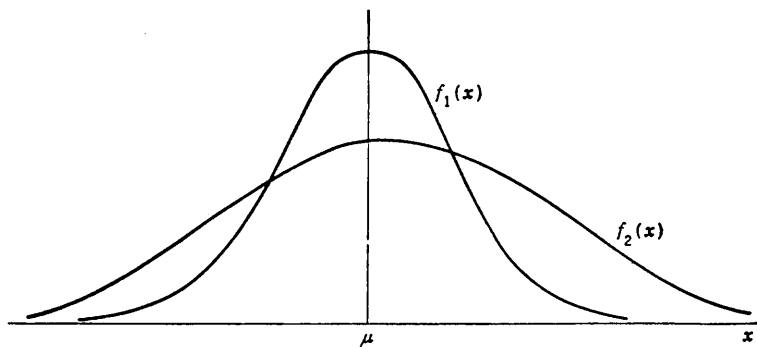


FIG. 5-3.

para ciertos intervalos próximos a la media. Las dos densidades de la figura 5-3 podrían tener varianzas aproximadamente iguales, y cabría alterar ligeramente cualquiera de ellas de modo que su varianza resultase mayor que la de la otra.

El tercer momento μ_3 respecto a la media se denomina a veces medida de asimetría o *deformación*. Puede demostrarse que distribuciones simétricas, como las representadas en las figuras 5-3 y 6-2, tienen $\mu_3=0$. Una curva cuya representación sea como la de $f_1(x)$ de la figura 5-4 se dice que es deformada a la izquierda y puede verse que tiene el tercer momento respecto a la media, negativo; otra cuyo aspecto sea como el de la $f_2(x)$, se dice que es deformada a la derecha y puede verse que tiene el tercer momento respecto a la media, positivo. Sin embargo, el conocimiento del tercer momento dice muy poco acerca del aspecto de la distribución, y si lo mencionamos es sobre todo para indicar esto. Así, la función de densidad $f_3(x)$ de la figura 5-4 tiene $\mu_3=0$ y dista mucho de ser simétrica. Mediante un pequeño cambio de forma podría conseguirse que su tercer momento fuese, a nuestro deseo, positivo o negativo.

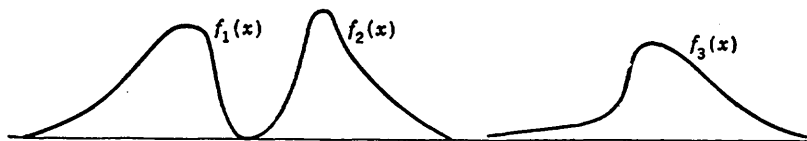


FIG. 5-4.

Así como un solo momento o algunos momentos proporcionan escasa información sobre la distribución correspondiente, el conjunto de todos los momentos (μ'_1 , μ'_2 , μ'_3 , ...) determina exactamente, en general, la distribución, razón por la cual tendremos ocasión de emplear los momentos en el desarrollo de la teoría.

En estadística aplicada, los dos primeros momentos tienen gran importancia, como veremos, pero los momentos de órdenes superiores rara vez son útiles. Lo corriente es que no se sepa cuál es la distribución con que se trabaja en un problema práctico dado y no suele importar demasiado el conocimiento de la forma efectiva de la distribución. Pero suele ser necesario conocer al menos la posición de la distribución y tener cierta idea de su dispersión. Estas características pueden estimarse examinando una muestra extraída de un conjunto de objetos que tengan dicha distribución. Este problema de estimación es probablemente el más importante de la estadística aplicada, y gran parte de este libro lo dedicaremos a su estudio.

Ejemplo 5-6.—Hállense la media y la varianza de la distribución hipergeométrica:

$$f(x) = \frac{\binom{m}{x} \binom{n}{k-x}}{\binom{m+n}{k}} \quad x=0, 1, 2, \dots, k \quad (8)$$

Este problema servirá de ilustración de una técnica que cabe utilizar para hallar los momentos de gran número de distribuciones discretas. Lo primero que se hace es utilizar la distribución para determinar una identidad entre los parámetros. Dado que $\sum f(x)=1$, se deduce que

$$\sum_{x=0}^k \binom{m}{x} \binom{n}{k-x} = \binom{m+n}{k} \quad (9)$$

para valores positivos cualesquiera de m , n , y k (en realidad, como hemos visto antes, el recorrido depende de los valores relativos de m , n y k ; pero podemos evitar ocuparnos de estos detalles, definiendo el coeficiente binomial

$$\binom{a}{b} = \frac{a!}{b!(a-b)!}$$

como igual a cero cuando b o $a-b$ sean negativos).

La media de la distribución es

$$\begin{aligned} \mu = E(x) &= \sum_{x=0}^k x f(x) \\ &= \frac{\sum_{x=0}^k x \binom{m}{x} \binom{n}{k-x}}{\binom{m+n}{k}} \end{aligned} \quad (10)$$

En esta expresión la x puede reducirse con otra x del denominador de $\binom{m}{x}$, con lo cual se obtiene

$$x \binom{m}{x} = m \binom{m-1}{x-1}$$

y resulta

$$\mu = \frac{\sum_{x=1}^k m \binom{m-1}{x-1} \binom{n}{k-x}}{\binom{m+n}{k}} \quad (11)$$

en donde la suma se extiende de 1 a k , porque el primer término de (10) se anula y puede omitirse. En realidad, puesto que por definición consideramos igual a cero todo coeficiente binomial cuyo índice inferior sea negativo, no habría inconveniente en seguir considerando como límites 0 y k . Ahora bien: si en la última expresión se sustituye $x-1$ por y y se sacan factores comunes del signo de sumación, se obtiene

$$\mu = \frac{m}{\binom{m+n}{k}} \sum_{y=0}^{k-1} \binom{m-1}{y} \binom{n}{k-1-y} \quad (12)$$

Esta suma puede calcularse mediante la identidad (9); basta reemplazar m por $m-1$ y k por $k-1$ en el segundo miembro de (9) para obtener

$$\begin{aligned} \mu &= \frac{m}{\binom{m+n}{k}} \binom{m-1+n}{k-1} \\ &= \frac{mk}{m+n} \end{aligned} \quad (13)$$

Para calcular la varianza necesitamos el segundo momento

$$\mu'_2 = \sum_{x=0}^k x^2 f(x)$$

Si sustituimos directamente el valor de $f(x)$, solo podremos dividir por una de las x , pero nos quedará otra x impidiéndonos utilizar la identidad que sirve para calcular la suma. El artificio que debe emplearse consiste en escribir x^2 en la forma

$$x(x-1) + x$$

con lo cual se obtiene

$$\mu'_2 = \sum x(x-1)f(x) + \sum xf(x) \quad (14)$$

Hemos calculado la segunda suma al obtener la media, y aplicando el mismo procedimiento a la primera, se tiene

$$\begin{aligned}
 E[x(x-1)] &= \frac{\sum_{x=0}^k x(x-1) \binom{m}{x} \binom{n}{k-x}}{\binom{m+n}{k}} \\
 &= \frac{\sum_{x=2}^k m(m-1) \binom{m-2}{x-2} \binom{n}{k-x}}{\binom{m+n}{k}} \\
 &= \frac{m(m-1)}{\binom{m+n}{k}} \sum_{y=0}^{k-2} \binom{m-2}{y} \binom{n}{k-2-y} \\
 &= \frac{m(m-1)}{\binom{m+n}{k}} \binom{m-2+n}{k-2} \\
 &= \frac{m(m-1)k(k-1)}{(m+n)(m+n-1)} \quad (15)
 \end{aligned}$$

Sumando (13) a este resultado, obtenemos μ'_2 , de acuerdo con (14); para hallar la varianza, basta restar el cuadrado de (13) de μ'_2 , en virtud de (5). La varianza es, por tanto,

$$\begin{aligned}
 \sigma^2 &= \frac{m(m-1)k(k-1)}{(m+n)(m+n-1)} + \frac{mk}{m+n} - \left(\frac{mk}{m+n} \right)^2 \\
 &= \frac{mnk(m+n-k)}{(m+n)^2(m+n-1)} \quad (16)
 \end{aligned}$$

El método general para calcular los momentos de orden superior resulta ahora evidente; para obtener el tercer momento, halláramos el valor esperado de

$$x(x-1)(x-2)$$

y puesto que este producto es igual a $x^3 - 3x^2 + 2x$, tenemos

$$\mu'_3 - 3\mu'_2 + 2\mu'_1 = E[x(x-1)(x-2)]$$

Una vez calculado el segundo miembro de esta expresión, podríamos despejar μ'_3 , ya que μ'_2 y μ'_1 han sido determinados anteriormente. Obtenido el tercer momento, podríamos obtener el cuarto, hallando el valor esperado de $x(x-1)(x-2)(x-3)$, y despejando μ'_4 en

$$\mu'_4 - 6\mu'_3 + 11\mu'_2 - 6\mu'_1 = E[x(x-1)(x-2)(x-3)]$$

El segundo miembro de esta expresión recibe el nombre de cuarto *momento factorial* de la distribución. El momento factorial r -ésimo es

$$E[x(x-1)(x-2) \dots (x-r+1)]$$

Ejemplo 5-7.—Hállense la media y la desviación estándar de la variante x con densidad $f(x)=2(1-x)$, $0 < x < 1$. El momento r -ésimo es

$$\begin{aligned}\mu'_r = E(x^r) &= \int_0^1 x^r 2(1-x) dx \\ &= 2 \int_0^1 (x^r - x^{r+1}) dx \\ &= \frac{2}{(r+1)(r+2)}\end{aligned}$$

La media es

$$\mu = \mu'_1 = \frac{2}{2 \times 3} = \frac{1}{3}$$

y la varianza

$$\sigma^2 = \mu'_2 - \mu^2 = \frac{2}{3 \times 4} - \frac{1}{9} = \frac{1}{18}$$

Por tanto,

$$\sigma = \sqrt{\frac{1}{18}} = \frac{\sqrt{2}}{6}$$

5-3. Funciones generatrices de momentos.—Los momentos de una función de densidad desempeñan un papel muy importante en estadística teórica y aplicada. En algunos casos, si se conocen todos los momentos, puede determinarse la función de densidad. Esto se estudiará brevemente en la sección 5-5. Puesto que los momentos de una función de densidad son tan importantes, sería muy útil hallar una función que diese una representación de todos los momentos; tal función se llama generatriz de momentos. Con mayor precisión, esta función se define a continuación.

Definición 5-3.—Sea x una variable aleatoria con densidad $f(x)$. Se llama *función generatriz de momentos* de x al valor esperado de e^{tx} si este valor existe para todo valor de t de determinado intervalo $-h^2 < t < h^2$. La *función generatriz de momentos* (f. g. m.) se representa por

$$m(t) = E(e^{tx}) = \int_{-\infty}^{\infty} e^{tx} f(x) dx \quad (1)$$

si la variable aleatoria es continua, y por

$$m(t) = E(e^{tx}) = \sum_x e^{tx} f(x)$$

si la variable aleatoria es discreta.

Se utiliza a veces la notación $m_x(t)$.

Si existe una función generatriz de momentos, $m(t)$ es indefinidamente derivable en un entorno del origen. Si derivamos r veces respecto a t la f. g. m., se obtiene

$$\frac{d^r}{dt^r} m(t) = \int_{-\infty}^{\infty} x^r e^{tx} f(x) dx \quad (2)$$

y haciendo $t=0$, hallamos

$$\frac{d^r}{dt^r} m(0) = E(x^r) = \mu'_r \quad (3)$$

donde el símbolo que aparece en el primer miembro debe interpretarse como la derivada r -ésima de $m(t)$ en el punto $t=0$. Por consiguiente, los momentos de una distribución pueden obtenerse derivando la función generatriz de momentos.

Si en la ecuación (1) sustituimos e^{xt} por su desarrollo en serie, resulta el desarrollo en serie de $m(t)$ en función de los momentos de $f(x)$; esto es,

$$\begin{aligned} m(t) &= E \left(1 + xt + \frac{1}{2!} (xt)^2 + \frac{1}{3!} (xt)^3 + \dots \right) \\ &= 1 + \mu'_1 t + \frac{1}{2!} \mu'_2 t^2 + \dots \\ &= \sum_{i=0}^{\infty} \frac{1}{i!} \mu'_i t^i \end{aligned} \quad (4)$$

luego vuelve a ser evidente que μ'_r se obtiene derivando $m(t)$ r veces y haciendo después $t=0$.

Como aclaración de esta técnica para el cálculo de momentos, hallemos la media y la varianza de la distribución de Poisson:

$$f(x) = \frac{e^{-a} a^x}{x!} \quad x = 0, 1, 2, \dots$$

Tenemos

$$\begin{aligned} m(t) &= E(e^{xt}) = \sum_{x=0}^{\infty} \frac{e^{xt} e^{-a} a^x}{x!} \\ &= e^{-a} \sum_{x=0}^{\infty} \frac{(ae^t)^x}{x!} \\ &= e^{-a} e^{ae^t} \end{aligned}$$

Las dos primeras derivadas son

$$\begin{aligned} m'(t) &= e^{-a} a e^t e^{ae^t} \\ m''(t) &= e^{-a} a e^t e^{ae^t} (1 + ae^t) \end{aligned}$$

de donde

$$\begin{aligned} \mu &= m'(0) = a \\ \mu_2' &= m''(0) = a(1 + a) \\ \sigma^2 &= a(1 + a) - a^2 = a \end{aligned}$$

La *función generatriz de los momentos factoriales* se define como $E(t^*)$, y se hallan los momentos factoriales a partir de esta función, de igual forma que los momentos ordinarios se obtienen a partir de $E(e^{xt})$, con la diferencia de que se hace $t=1$ en lugar de igualar a cero. Esta función simplifica algunas veces el problema del cálculo de momentos en distribuciones discretas. Sin embargo, este método no sirve en el ejemplo de la sección anterior, ya que la suma $\sum t^* f(x)$ no tiene una expresión sencilla. Para la distribución de Poisson:

$$E(t^*) = e^{a(t-1)}$$

de donde

$$\begin{aligned} E(x) &= \left. a e^{a(t-1)} \right|_{t=1} = a \\ E[x(x-1)] &= \left. a^2 e^{a(t-1)} \right|_{t=1} = a^2 \end{aligned}$$

que da los mismos momentos anteriores.

A veces tendremos ocasión de hablar de los momentos de una *función* de una variable aleatoria. Así, podemos necesitar los momentos de $h(x)$, sabiendo que la distribución de x es $f(x)$. El momento r -ésimo de $h(x)$ es

$$E[h(x)]^r = \int_{-\infty}^{\infty} [h(x)]^r f(x) dx \quad (5)$$

y una función generatriz de momentos es, evidentemente,

$$E(e^{th(x)}) = \int_{-\infty}^{\infty} e^{th(x)} f(x) dx \quad (6)$$

5-4. Momentos para distribuciones multivariantes.—Las ideas anteriores se generalizan fácilmente a distribuciones de dos o más variantes. Supongamos, p. ej., que tenemos tres variantes (x, y, z) con función de densidad $f(x, y, z)$. El momento r -ésimo de y , p. ej., es

$$E(y^r) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y^r f(x, y, z) dz dy dx \quad (1)$$

Además de los momentos de cada una de las variantes, existen varios *momentos mixtos* definidos en general por

$$E(x^q y^r z^s) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^q y^r z^s f(x, y, z) dz dy dx \quad (2)$$

en donde q, r y s son números naturales o cero. El momento mixto más importante es la *covarianza*, que es el momento mixto respecto a las medias del producto de dos variantes. Por tanto, la covarianza entre x y z es

$$\sigma_{xz} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [x - E(x)][z - E(z)] f(x, y, z) dz dy dx \quad (3)$$

y hay otras dos covarianzas σ_{xy} y σ_{yz} , definidas análogamente. La *correlación* entre dos variantes tales como x y z se representa por ρ_{xz} y se define por

$$\rho_{xz} = \frac{\sigma_{xz}}{\sigma_x \sigma_z} \quad (4)$$

en donde σ_x y σ_z son las desviaciones estándar de x y z .

Una *función generatriz de momentos mixta* de (x, y, z) se define por

$$m(t_1, t_2, t_3) = E(e^{t_1 x + t_2 y + t_3 z}) \quad (5)$$

si existe para todos los valores de t_1, t_2, t_3 tales que $-h^2 < t_i < h^2$ para algún h .

Es evidente que el momento r -ésimo de y puede obtenerse derivando r veces respecto a t_2 la función generatriz de momentos y haciendo después iguales a cero todas las t . Análogamente, el momento mixto (2) se obtendría derivando la función q veces respecto a t_1 , r veces respecto a t_2 , s veces respecto a t_3 , y haciendo a continuación iguales a cero todas las t .

5-5. El problema de los momentos.—Hemos visto que, en general, una función de densidad determina un conjunto de momentos $\mu'_1, \mu'_2, \dots$, siempre que estos existan. Uno de los problemas más importantes de la estadística teórica es este: Dado un conjunto de momentos, ¿cuál es la función de densidad de que provienen?, y ¿existe una única función de densidad que genere estos momentos particulares? Enunciaremos, sin demostrarlo, un teorema que puede utilizarse para responder a estas preguntas.

Teorema 5-7.—Sean x e y dos variables aleatorias, continuas o discretas, con funciones de densidad $f(x)$ y $g(y)$, respectivamente. Supongamos que existen las funciones generatrices de momentos de x e y y que ambas son iguales para todo t del intervalo $-h^2 < t < h^2$. Entonces, las dos funciones de densidad son iguales (salvo quizá en puntos en que no son continuas si x e y son variables aleatorias continuas).

En el caso de distribuciones multivariantes, es válido un teorema análogo.

Ejemplo 5-8.—Supongamos que una variable aleatoria x tiene una función generatriz de momentos $m_x(t) = 1/(1-t)$, que existe para todo t tal que $-1 < t < 1$. Vemos que la distribución $g(y) = e^{-y}$ ($0 < y < \infty$) admite la función generatriz de momentos

$$m_y(t) = \int_0^{\infty} e^{yt} e^{-y} dy = 1/(1-t)$$

Luego, por el teorema 5-7, x tiene como función de densidad e^{-x} , $0 < x < \infty$.

Para hallar la distribución de x en el ejemplo 5-8, hemos encontrado una distribución que tenía como f. g. m. $1/(1-t)$. Esto no siempre es fácil; sin embargo, puede utilizarse la teoría de la transformación de Laplace para deducir la función de densidad a partir del conocimiento de la f. g. m. No obstante, esta técnica excede de los límites de este libro.

5-6. Esperanzas condicionales.—En los siguientes capítulos, tendremos ocasión de hallar el valor esperado de variables aleatorias en distribuciones condicionales.

Definición 5-4.—Sean $f(x, y)$ la función de densidad conjunta de la variable aleatoria bidimensional (x, y) , $g(y|x)$ la función de densidad condicional de y , dado $x=x$, y $h(x)$ la función de densidad marginal de x . La esperanza condicional de y , dado $x=x$, designada por $E(y|x)$ se define como

$$E(y|x) = \int_{-\infty}^{\infty} yg(y|x) dy = \int_{-\infty}^{\infty} y \frac{f(x, y)}{h(x)} dy$$

Observemos que $E(y|x)$ es, en general, una función de x ; no es una variable aleatoria, ya que x es un valor particular de x .

Ejemplo 5-9.—Sea $f(x, y) = 8xy$, $0 < y < x < 1$, la función de densidad conjunta de la variable aleatoria bidimensional (x, y) . Hállese $E(y|x)$. La distribución marginal de x es

$$h(x) = \int_0^x 8xy dy = 4x^3 \quad 0 < x < 1$$

luego la función de densidad condicional de y , dado x , será

$$g(y|x) = \frac{f(x, y)}{h(x)} = \frac{8xy}{4x^3} = \frac{2y}{x^2} \quad 0 < y < x, \quad 0 < x < 1$$

y

$$E(y|x) = \int_0^x y \cdot g(y|x) dy = \int_0^x y \cdot \frac{2y}{x^2} dy = \frac{2}{3} x \quad 0 < x < 1$$

Como dijimos anteriormente, $E(y|x)$ es, en general, función de x . Designémosla por $u(x)$; es decir, $E(y|x) = u(x)$. Podemos calcular el valor esperado de la variable aleatoria $u(x)$, donde la variable aleatoria x tiene la función de densidad $h(x)$. Esto nos da

$$\begin{aligned} E[E(y|x)] &= E[u(x)] = \int_{-\infty}^{\infty} u(x)h(x) dx = \int_{-\infty}^{\infty} E(y|x)h(x) dx \\ &= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} yg(y|x) dy \right] h(x) dx = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} yg(y|x)h(x) dy dx \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} yf(x, y) dy dx = E(y) \end{aligned}$$

Queda así demostrado el siguiente teorema:

Teorema 5-8.—Sea (x, y) una variable aleatoria bidimensional; en tal supuesto, $E(y) = E[E(y|x)]$.

Enunciaremos, sin demostrarlo, otro teorema:

Teorema 5.9.—Sea $(y, x_1, x_2, \dots, x_k)$ una variable aleatoria $(k+1)$ -dimensional con función de densidad $f(y, x_1, \dots, x_k)$ y con función de densidad condicional $g(y|x_1, x_2, \dots, x_k)$. El valor esperado de una función u de y y de las x_i , dadas $x_1, x_2, \dots, x_k$, es

$$E[u(y, x_1, \dots, x_k)|x_1, x_2, \dots, x_k] = \int_{-\infty}^{\infty} u(y, x_1, \dots, x_k)g(y|x_1, x_2, \dots, x_k) dy$$

PROBLEMAS

1. Si se venden 3000 billetes de lotería, a un dólar cada uno, para el sorteo de un coche de 4000 dólares, ¿cuál es la ganancia esperada de una persona que compre un billete?

2. La distribución de una variable aleatoria x está dada por

x	1	2	3
$f(x)$	$1/2$	$1/3$	$1/6$

Hállese $E(x)$.

3. Se lanza una moneda hasta que aparece cara. ¿Cuál es el número esperado de tiradas?

4. La probabilidad de que un suceso ocurra es p , y la de que no ocurra $q=1-p$. En una sola prueba, ¿cuáles son la media y la varianza de x , siendo este el número de aciertos?

5. Si se realizan n pruebas independientes del suceso descrito en el problema 4, y x es el número total de aciertos, ¿cuáles son la media y la varianza de x ?

6. Hállese la media de la variante continua x cuya distribución es

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-(x-a)^2/2}, \quad -\infty < x < \infty$$

7. Hállese la media y la varianza de x si $f(x)=2, 0 < x < 1/2$.

8. Hállese la media y la varianza de $2x^2$ si $f(x)=2, 0 < x < 1/2$.

9. Hállese la media y la varianza de x si

$$f(x) = \frac{3}{(x+1)^4} \quad 0 < x < \infty$$

10. Demuéstrese que $E(xy)=E(x)E(y)$ cuando x e y tienen distribuciones independientes.

11. Demuéstrase que

$$\mu_r = \sum_{i=0}^r \binom{r}{i} (-1)^{r-i} \mu'_i (\mu'_1)^{r-i}$$

12. ¿Cuál es la mediana de x si $f(x)=2(1-x)$, $0 < x < 1$?

13. Hállese la función generatriz de momentos asociada a la distribución que tiene $f(x)=ae^{-ax}$, $x > 0$, y utilícese para obtener la media y la varianza de x .

14. Hállese la función generatriz de los momentos factoriales para la distribución binomial y utilícese para obtener el tercer momento.

15. Si x tiene como función de densidad $f(x)=x/2$, $0 < x < 2$, hállese el momento r -ésimo de x^2 . Demuéstrase entonces que $y=x^2$ tiene la distribución

$$g(y)=1/4 \quad 0 < y < 4$$

probando que y tiene los mismos momentos que x^2 .

16. Si $f(x, y)=a^2e^{-a(x+y)}$, $x > 0$, $y > 0$, hállese la función generatriz para los momentos de $u=x+y$, y dedúzcase la distribución de u a partir de la forma de esta función generatriz.

17. Demuéstrase que si una función de densidad $f(x)$ es simétrica respecto a un cierto punto b [esto es, $f(b+c)=f(b-c)$ para todo valor de c], este punto debe ser la media de x . Demuéstrase también en este caso que todos los momentos impares respecto a la media deben ser nulos.

18. Dada la función generatriz de momentos $m(t)$ para los momentos μ'_r respecto al origen, ¿cómo se obtendría la función generatriz de momentos para los momentos μ_r respecto a la media?

19. A menudo resulta útil caracterizar una distribución, no por los momentos μ'_r , sino por otro conjunto infinito de constantes γ_r , denominadas *cumulantes* de la distribución. Los cumulantes se definen por la función generatriz $c(t)=\log m(t)$, en donde $m(t)$ es la función generatriz para los μ'_r ; esto es,

$$\gamma_r = \frac{d^r c(t)}{dt^r}$$

calculada para $t=0$. Demostrar que $\gamma_1=\mu'_1$ y $\gamma_2=\sigma^2$.

20. Hállese el cumulante r -ésimo γ_r para la función de densidad

$$f(x)=ae^{-ax}, \quad x > 0.$$

21. Demuéstrase que si $M(t)$ engendra los momentos respecto a un punto arbitrario b , esto es,

$$M(t) = \int_{-\infty}^{\infty} e^{t(x-b)} f(x) dx$$

$C(t) = \log M(t)$ engendrará correctamente todos los cumulantes, excepto el primero. Por esto, se dice que los cumulantes de una distribución, posteriores al γ_1 , son invariantes respecto a las traslaciones de la variable aleatoria.

22. Si x tiene por cumulantes γ_r , demuéstrese que $y = kx$ tiene por cumulantes $k^r \gamma_r$.

23. Demuéstrese que la correlación entre dos variantes es cero si están distribuidas independientemente (la recíproca no es cierta, como lo demuestra el problema siguiente).

24. Supongamos que x tiene la función de densidad marginal

$$f_1(x) = 1, \quad -1/2 < x < 1/2$$

y sea la función de densidad condicional de y

$$\begin{aligned} f(y|x) &= 1 & x < y < x+1, & -1/2 < x < 0 \\ &= 1 & -x < y < 1-x, & 0 < x < 1/2 \\ &= 0 & \text{en otro caso.} \end{aligned}$$

Hállese la correlación entre x e y .

25. ¿Podría utilizarse la función $E[1/(1+tx)]$ para engendrar los momentos de una variante x ?

26. En el problema 16, hállese la función generatriz de momentos mixta de (x, y) . Derivando esta f. g. m., dedúzcase $E(y^2 x)$.

27. Si x e y son dos variables aleatorias independientes, demuéstrese que $E(y|x)$ no depende de x .

28. Sea (x, y) una variable aleatoria bidimensional con función de densidad conjunta

$$f(x, y) = 1/8(6-x-y) \quad 0 < x < 2, \quad 2 < y < 4$$

a) Hállese $E(y|x)$.

b) Obténgase $E(y^2|x)$.

c) Hállese $\text{var}(y|x)$ definida por $E(y^2|x) - [E(y|x)]^2$.

29. La distribución trinomial de dos variantes x e y es

$$f(x, y) = \frac{n!}{x!y!(n-x-y)!} p^x q^y (1-p-q)^{n-x-y} \quad \begin{matrix} x, y = 0, 1, \dots, n \\ x+y \leq n \end{matrix}$$

Obténgase la distribución marginal de y .

30. En el problema 29, hállese la distribución condicional de x , dado y , y obténgase su valor esperado.

31. En el problema 28, demuéstrese que

$$E[E(y|x)] = E(y)$$

32. Sea $f(x, y)$ la distribución conjunta de x e y , y sean $u(x)$ y $v(y)$ dos funciones de x e y , respectivamente. Pruébese que

$$E[u(x) \cdot v(y)|x] = u(x)E[v(y)|x]$$

33. En el problema 28, hállese $E(xy|x)$.

34. Si x e y son dos variables aleatorias y $E(y|x) = \mu$, donde μ no depende de x , demuéstrese que

$$\text{var}(y) = E[\text{var}(y|x)].$$

35. Sea la variable aleatoria x con función de densidad $f(x)$ y f. g. m. $m_x(t)$.

a) Si c es una constante, pruébese que la f. g. m. de $c+x$ es $e^{ct}m_x(t)$.

b) Si c es una constante, demuéstrese que la f. g. m. de cx es $m_x(ct)$, $c \neq 0$.

36. Sea $x_1, x_2, \dots, x_n$ una muestra aleatoria procedente de la distribución $f(x)$, y sea $m_x(t)$ la f. g. m. de x_i .

a) Demuéstrese que la f. g. m. de y es $[m_x(t)]^n$, donde

$$y = \sum_{i=1}^n x_i.$$

b) Pruébese que la f. g. m. de $\bar{x}$ es $[m_x(t/n)]^n$, donde

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i.$$

BIBLIOGRAFIA

1. BRUNK, H.: *An Introduction to Mathematical Statistics*, Ginn & Company, Boston, 1960.
2. FELLER, W.: *An Introduction to Probability Theory and Its Applications*, 2.^a ed., vol. I, John Wiley & Sons, Inc., Nueva York, 1957.
3. HOEL, P. G.: *Introduction to Mathematical Statistics*, 2.^a ed., John Wiley & Sons, Inc., Nueva York, 1954.
4. HOGG, R., y A. CRAIG: *Introduction to Mathematical Statistics*, The Macmillan Company, Nueva York, 1959.
5. KENDALL, M. G.: *The Advanced Theory of Statistics*, vol. I, Charles Griffin & Co., Ltd., Londres, 1948.

CAPITULO 6

DISTRIBUCIONES CONTINUAS ESPECIALES

6-1. Distribución uniforme.—La distribución más sencilla de variante continua es la que tiene densidad uniforme:

$$f(x) = \frac{1}{\beta - \alpha} \quad \alpha < x < \beta \quad (1)$$
$$= 0 \quad \text{en otro caso}$$

representada en la figura 6-1. La probabilidad de que una observación caiga en cualquier intervalo $\alpha < x < \beta$ es igual a $1/(\beta - \alpha)$ multiplicado por la longitud del intervalo. La distribución resulta particularmente útil en estadística teórica por la sencillez de su estudio desde el punto de vista matemático.

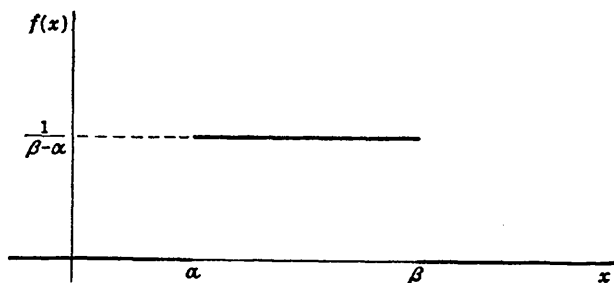


FIG. 6-1.

Podemos limitarnos a estudiar esta distribución al ocuparnos de ciertas propiedades de las distribuciones en general, mediante el siguiente teorema:

Teorema 6-1.—*Toda distribución de una variante continua x puede transformarse en la de densidad uniforme*

$$f(y) = 1 \quad 0 < y < 1 \quad (2)$$

haciendo $y = G(x)$, en donde $G(x)$ es la distribución acumulativa de x . Dejamos al cuidado del lector la demostración.

Por medio de este teorema se pueden demostrar muchas propiedades de las distribuciones continuas en general, estableciéndolas simplemente para la distribución uniforme sobre el intervalo unidad.

6-2. La distribución normal.—Muchas de las técnicas utilizadas en estadística aplicada se basan en la distribución normal, por lo que gran parte de lo que resta de este libro se dedicará al estudio de tal distribución.

Definición 6-1.—Se dice que la variable aleatoria x se distribuye normalmente si su función de densidad está dada por

$$n(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2} \quad -\infty < x < \infty \quad (1)$$

La función está representada en la figura 6-2 para varios valores de σ . Un cambio en el valor de μ se limita a trasladar la curva a la derecha o a la izquierda sin alterar su forma. En realidad, la fun-

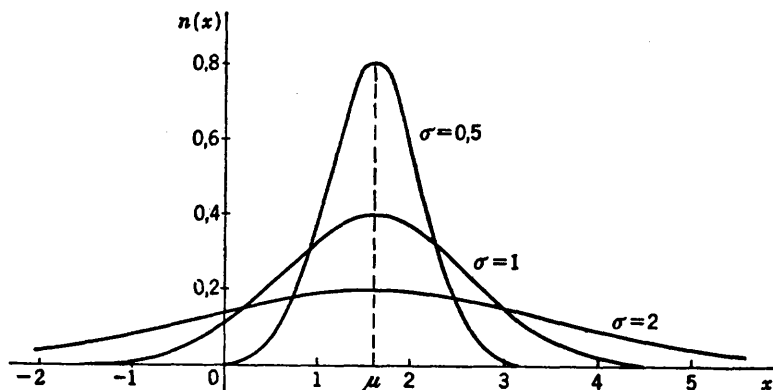


FIG. 6-2.

ción anterior representa una familia de distribuciones con dos parámetros, μ y σ^2 . Hemos utilizado los símbolos μ y σ^2 para representar los parámetros, porque, según veremos, estos son precisamente la media y la varianza, respectivamente, de la distribución.

Puesto que $n(x)$ es una función de densidad habrá de verificarse

$$\int_{-\infty}^{\infty} n(x) dx = 1$$

pero hemos de comprobar que verdaderamente ocurre así. La comprobación resulta algo laboriosa debido a que esta función particular no puede integrarse de modo directo. Supongamos que representamos por A el área limitada por la curva; tendremos

$$A = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{-(x-\mu)^2/2\sigma^2} dx$$

y haciendo la sustitución

$$y = \frac{x - \mu}{\sigma}$$

hallamos

$$A = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}y^2} dy$$

Queremos demostrar que $A=1$; el procedimiento más sencillo consiste en ver que A^2 es igual a 1 y deducir después que $A=1$ puesto que $f(x)$ es positiva. Podemos hacer

$$\begin{aligned} A^2 &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}y^2} dy \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}z^2} dz \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(y^2+z^2)} dy dz \end{aligned}$$

escribiendo el producto de las dos integrales como una integral doble. En esta integral hacemos el cambio de variables cartesianas a polares, mediante la sustitución

$$y = r \sin \theta$$

$$z = r \cos \theta$$

con lo que la integral se transforma en

$$\begin{aligned} A^2 &= \frac{1}{2\pi} \int_0^{\infty} \int_0^{2\pi} r e^{-\frac{1}{2}r^2} d\theta dr \\ &= \int_0^{\infty} r e^{-\frac{1}{2}r^2} dr \\ &= 1 \end{aligned}$$

Puesto que la integral de $n(x)$ no tiene una expresión funcional sencilla, solo podemos presentar formalmente la distribución acumulativa:

$$N(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^x e^{-[(t-\mu)^2/2\sigma^2]} dt \quad (2)$$

y haciendo

$$y = \frac{t - \mu}{\sigma}$$

hallamos

$$N(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(x-\mu)/\sigma} e^{-\frac{1}{2}y^2} dy \quad (3)$$

y dado un valor determinado de $(x-\mu)/\sigma$, la integral puede calcularse por métodos numéricos. En la tabla II puede encontrarse esta función tabulada. Puesto que la distribución es simétrica respecto a μ , esto es, debido a que

$$n(\mu - a) = n(\mu + a),$$

se deduce que $N(x)$ para valores negativos de $(x-\mu)/\sigma$ es igual a $1 - N(x')$, en donde $(x' - \mu)/\sigma = -(x - \mu)/\sigma$. La representación gráfica de $N(x)$ está dada en la figura 6-3.

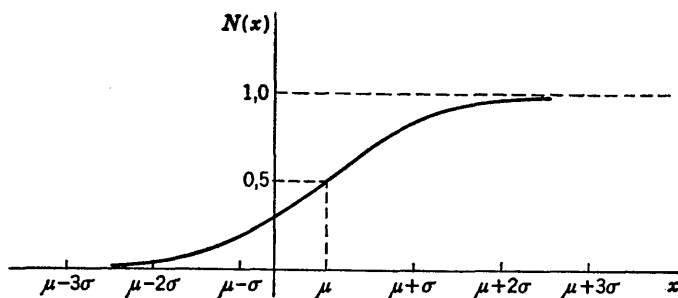


FIG. 6-3.

Para ilustrar el uso de la tabla, hallaremos $P(-1 < x < 4)$, cuando x tiene la función de densidad

$$n(x) = \frac{1}{4\sqrt{2\pi}} e^{-[(x-2)^2/32]} \quad (4)$$

Observemos que

$$\mu = 2 \quad \sigma = 4$$

y, por tanto, que los valores de $(x-\mu)/\sigma$ correspondientes a -1 y 4 son

$$\frac{-1-2}{4} = -\frac{3}{4} \quad \frac{4-2}{4} = \frac{1}{2}$$

de donde,

$$\begin{aligned} P(-1 < x < 4) &= N(4) - N(-1) \\ &= 0.6915 - (1 - 0.7734) \\ &= 0.4649 \end{aligned}$$

Resulta muy conveniente que $N(x)$ sea de tal forma que no necesite tabulación para las diversas combinaciones de valores de μ y σ . La transformación $y=(x-\mu)/\sigma$ hace que todas las distribuciones normales tomen la misma forma, que se denomina *forma estandarizada o normalizada*. Desde ahora reservaremos las letras n y N para indicar la función de densidad normal y su forma acumulativa. Ocurre a menudo que quieren ponerse de manifiesto los parámetros, y entonces se representan las funciones por $n(x; \mu, \sigma^2)$ y $N(x; \mu, \sigma^2)$, separando los parámetros de la variante mediante un punto y coma. Con esta notación, la distribución (4) vendría representada por $n(x; 2, 16)$. La distribución normal estándar es, entonces

$$n(x; 0, 1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \quad (5)$$

y su forma acumulativa

$$N(x; 0, 1) = \int_{-\infty}^x n(t; 0, 1) dt \quad (6)$$

Calcularemos ahora los momentos de $n(x; \mu, \sigma^2)$, empezando por hallar la función generatriz de estos. El cálculo es el siguiente:

$$\begin{aligned} m(t) &= E(e^{tx}) = e^{t\mu} E(e^{t(x-\mu)}) \\ &= e^{t\mu} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{t(x-\mu)} e^{-(1/2\sigma^2)(x-\mu)^2} dx \\ &= e^{t\mu} \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{-(1/2\sigma^2)[(x-\mu)^2 - 2\sigma^2 t(x-\mu)]} dx \end{aligned}$$

Completando el cuadrado interior al paréntesis, se obtiene

$$\begin{aligned} (x-\mu)^2 - 2\sigma^2 t(x-\mu) &= (x-\mu)^2 - 2\sigma^2 t(x-\mu) + \sigma^4 t^2 - \sigma^4 t^2 \\ &= (x-\mu - \sigma^2 t)^2 - \sigma^4 t^2 \end{aligned}$$

y tenemos

$$m(t) = e^{t\mu} e^{\sigma^2 t^2/2} \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{-(x-\mu-\sigma^2 t)^2/2\sigma^2} dx$$

La integral, con el factor $1/\sqrt{2\pi}\sigma$, es necesariamente igual a uno, por ser el área limitada por una distribución normal de media $\mu + \sigma^2 t$ y varianza σ^2 . Por tanto, tenemos el siguiente teorema:

Teorema 6-2.—Si x tiene la distribución $n(x; \mu, \sigma^2)$, es decir, la distribución normal, la función generatriz de momentos de x es

$$m(t) = e^{t\mu + (\sigma^2 t^2)/2} \quad (7)$$

Derivando dos veces esta función y haciendo $t=0$ en los resultados, hallamos

$$\mu'_1 = \mu$$

$$\mu'_2 = \sigma^2 + \mu^2$$

$$\text{Varianza}(x) = \mu'_2 - (\mu'_1)^2 = \sigma^2$$

lo que justifica el uso de los símbolos correspondientes a los momentos para los parámetros.

6-3. La distribución gamma.—Otra distribución que desempeña un importante papel en estadística es la distribución gamma que definimos a continuación.

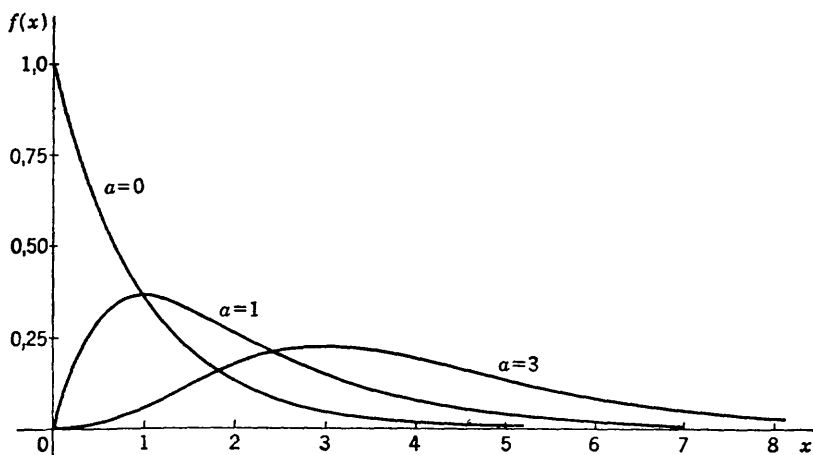


FIG. 6-4.

Definición 6-2.—Se dice que una variable aleatoria x se distribuye según una distribución gamma si su función de densidad es

$$f(x; \alpha, \beta) = \frac{1}{\alpha! \beta^{\alpha+1}} x^{\alpha} e^{-x/\beta} \quad 0 < x < \infty \quad (1)$$

$$= 0 \quad \text{en otro caso}$$

Es una familia de distribuciones con dos parámetros, α y β ; β ha de ser positivo, y α , mayor que -1 . La función se representa en la fi-

gura 6-4 para $\beta=1$ y diversos valores de α . Un cambio en el valor de β hace variar solo la escala sobre ambos ejes, lo que resulta evidente si examinamos la forma de la función.

Para demostrar que la función representa una densidad (que tiene área igual a la unidad) calcularemos la integral

$$A = \int_0^{\infty} \frac{1}{\beta^{\alpha+1}} x^{\alpha} e^{-x/\beta} dx$$

que se convierte en

$$A = A(\alpha) = \int_0^{\infty} y^{\alpha} e^{-y} dy$$

sustituyendo x/β por y ; se deduce, por tanto, que A es función solo de α . Si $\alpha > 0$, podemos integrar inmediatamente por partes, y se obtiene:

$$\begin{aligned} A(\alpha) &= -y^{\alpha} e^{-y} \Big|_0^{\infty} + \int_0^{\infty} \alpha y^{\alpha-1} e^{-y} dy \\ &= \alpha \int_0^{\infty} y^{\alpha-1} e^{-y} dy \end{aligned}$$

de donde se deduce que

$$A(\alpha) = \alpha A(\alpha - 1) \quad (2)$$

Si α es un número natural, aplicando la fórmula de recurrencia (2) sucesivamente, se obtiene

$$A(\alpha) = \alpha(\alpha - 1)(\alpha - 2) \dots (2)(1) A(0)$$

y puesto que

$$A(0) = \int_0^{\infty} e^{-y} dy = 1$$

resulta

$$A(\alpha) = \alpha!$$

cuando α es entero. La función $A(\alpha)$ se representa a menudo por $\Gamma(\alpha + 1)$ en las obras de matemáticas, pero aquí emplearemos el símbolo $\alpha!$ aunque α no sea entero.

En muchas aplicaciones de esta distribución, α es entero o múltiplo de $1/2$. Por tanto, para nuestro propósito bastará con calcular $(1/2)!$, con lo que nos será posible obtener $\alpha!$ para cualquier valor

de α que necesitamos:

$$\begin{aligned} (1/2)! &= 1/2(-1/2)! \\ &= 1/2 \int_0^{\infty} y^{-1/2} e^{-y} dy \end{aligned}$$

y si hacemos $y = z^2/2$,

$$\begin{aligned} (1/2)! &= 1/2 \int_0^{\infty} \sqrt{2} e^{-(z^2/2)} dz \\ &= \sqrt{\pi} \int_0^{\infty} \frac{1}{\sqrt{2\pi}} e^{-(z^2/2)} dz \\ &= \frac{\sqrt{\pi}}{2} \end{aligned}$$

ya que la integral es la mitad del área limitada por la función de densidad normal, y vale, por tanto, $1/2$. Conocido este número, podemos calcular $\alpha!$ para cualquier múltiplo de $1/2$ utilizando la relación (2); p. ej.,

$$\begin{aligned} (5/2)! &= 5/2(3/2)! = 5/2 \times 3/2(1/2)! \\ &= \frac{15\sqrt{\pi}}{8} \end{aligned}$$

La distribución acumulativa es

$$F(x) = \int_0^x \frac{1}{\alpha! \beta^{\alpha+1}} t^{\alpha} e^{-t/\beta} dt \quad x > 0 \quad (3)$$

que, por supuesto, es igual a cero cuando $x \leq 0$. Debe calcularse por métodos numéricos, a menos que α sea un número natural, en cuyo caso la función puede hallarse por integración sucesiva por partes, y se obtiene

$$\begin{aligned} F(x) &= 1 - \left[1 + \frac{x}{\beta} + \frac{1}{2!} \left(\frac{x}{\beta} \right)^2 + \frac{1}{3!} \left(\frac{x}{\beta} \right)^3 + \dots + \frac{1}{\alpha!} \left(\frac{x}{\beta} \right)^{\alpha} \right] e^{-x/\beta} \\ &= 0 \end{aligned} \quad \begin{array}{l} x > 0 \\ x \leq 0 \end{array} \quad (4)$$

Pero en todo caso suele ser más sencillo utilizar tablas de la función para resolver problemas concretos. La función F recibe el nombre de *función gamma incompleta* y ha sido extensamente tabulada.

La función generatriz de momentos de esta distribución es

$$\begin{aligned} m(t) &= \int_0^{\infty} e^{tx} \frac{1}{\alpha! \beta^{\alpha+1}} x^{\alpha} e^{-x/\beta} dx \\ &= \int_0^{\infty} e^{\beta t y} \frac{1}{\alpha!} y^{\alpha} e^{-y} dy \end{aligned}$$

después de sustituir x/β por y . Esta puede escribirse también en la forma

$$\begin{aligned} m(t) &= \frac{1}{\alpha!} \int_0^{\infty} y^{\alpha} e^{-y(1-\beta t)} dy \\ &= \frac{1}{(1-\beta t)^{\alpha+1}} \int_0^{\infty} \frac{(1-\beta t)^{\alpha+1}}{\alpha!} y^{\alpha} e^{-y(1-\beta t)} dy \\ &= \frac{1}{(1-\beta t)^{\alpha+1}} \end{aligned} \quad (5)$$

suponiendo que $t < 1/\beta$, ya que la última integral representa el área limitada por la distribución gamma de parámetros α y $\beta' = 1/(1-\beta t)$, y es, por tanto, igual a la unidad. Derivando $m(t)$ dos veces y haciendo $t=0$ en los resultados, tenemos

$$\mu = \beta(\alpha + 1) \quad (6)$$

$$\mu'_2 = \beta^2(\alpha + 1)(\alpha + 2) \quad (7)$$

$$\sigma^2 = \beta^2(\alpha + 1) \quad (8)$$

Hemos demostrado el siguiente teorema:

Teorema 6-3.—*Si la variable aleatoria x tiene una distribución gamma, la función generatriz de momentos de x es*

$$m(t) = (1 - \beta t)^{-(\alpha+1)} \quad t < \frac{1}{\beta}$$

6-4. La distribución beta.—Otra distribución útil en estadística es la distribución beta.

Definición 6-3.—*Se dice que una variable aleatoria x tiene una distribución beta si su función de densidad está dada por*

$$\begin{aligned} f(x; \alpha, \beta) &= \frac{(\alpha + \beta + 1)!}{\alpha! \beta!} x^{\alpha} (1-x)^{\beta} & 0 < x < 1 \\ &= 0 & \text{en otro caso} \end{aligned} \quad (1)$$

Esta función constituye una familia de distribuciones con dos parámetros, de la que se representan algunos ejemplos en la figura 6-5. Los parámetros α y β deben ser ambos mayores que -1 . Se reduce a la distribución uniforme sobre el intervalo unitario cuando $\alpha = \beta = 0$.

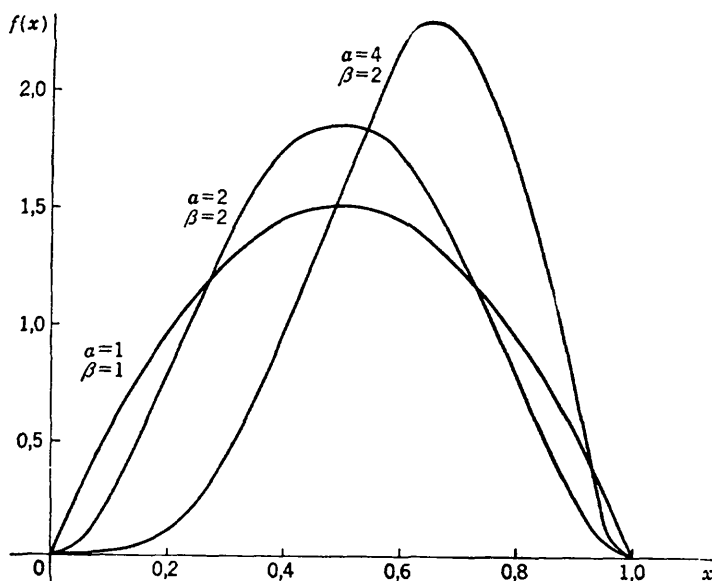


FIG. 6-5.

Para ver que el área limitada por $f(x)$ es igual a la unidad, calculemos la integral

$$A(\alpha, \beta) = \int_0^1 x^\alpha (1-x)^\beta dx \quad (2)$$

Claro es que A es función de α y β ; vamos a probar que es igual al recíproco del multiplicador constante que aparece en (1). Volviendo a referirnos a la distribución gamma, podemos escribir

$$\begin{aligned} \alpha! \beta! &= \left(\int_0^\infty x^\alpha e^{-x} dx \right) \left(\int_0^\infty y^\beta e^{-y} dy \right) \\ &= \int_0^\infty \int_0^\infty x^\alpha y^\beta e^{-(x+y)} dx dy \end{aligned}$$

y en esta última integral cambiaremos la variable x por la u , mediante la sustitución

$$u = \frac{x}{x+y}$$

o bien

$$x = \frac{uy}{1-u} \quad dx = \frac{y \, du}{(1-u)^2}$$

Ya que evidentemente u tiene el recorrido de 0 a 1, la integral es ahora

$$\alpha! \beta! = \int_0^\infty \int_0^1 \left(\frac{uy}{1-u} \right)^\alpha y^\beta e^{-y/(1-u)} \frac{y}{(1-u)^2} du dy$$

En esta integral cambiamos y por v , mediante la sustitución

$$y = (1-u)v \quad dy = (1-u) dv$$

obteniendo

$$\begin{aligned} \alpha! \beta! &= \int_0^\infty \int_0^1 u^\alpha (1-u)^\beta v^{\alpha+\beta+1} e^{-v} du dv \\ &= \left(\int_0^\infty v^{\alpha+\beta+1} e^{-v} dv \right) \left[\int_0^1 u^\alpha (1-u)^\beta du \right] \\ &= (\alpha + \beta + 1)! \int_0^1 u^\alpha (1-u)^\beta du \end{aligned}$$

lo que demuestra que $A(\alpha, \beta)$ tiene el valor dicho. $A(\alpha-1, \beta-1)$ recibe el nombre de *función beta* de α y β , y suele designarse por $B(\alpha, \beta)$.

La distribución acumulativa, a menudo denominada *función beta incompleta*, es

$$\begin{aligned} F(x) &= 0 & x &\leq 0 \\ &= \int_0^x \frac{(\alpha + \beta + 1)!}{\alpha! \beta!} t^\alpha (1-t)^\beta dt & 0 < x < 1 \\ &= 1 & x &\geq 1 \end{aligned} \quad (3)$$

y ha sido también tabulada.

La función generatriz de momentos de esta distribución no tiene forma sencilla, pero los momentos pueden hallarse directa-

mente con facilidad:

$$\begin{aligned}\mu'_r = E(x^r) &= \frac{(\alpha + \beta + 1)!}{\alpha! \beta!} \int_0^1 x^{r+\alpha} (1-x)^\beta dx \\ &= \frac{(\alpha + \beta + 1)! (\alpha + r)!}{(\alpha + \beta + r + 1)! \alpha!} \int_0^1 \frac{(\alpha + \beta + r + 1)!}{(\alpha + r)! \beta!} x^{r+\alpha} (1-x)^\beta dx \\ &= \frac{(\alpha + \beta + 1)! (\alpha + r)!}{(\alpha + \beta + r + 1)! \alpha!}\end{aligned}\quad (4)$$

ya que la integral debe ser igual a la unidad.

6-5. Otras distribuciones.—Una distribución que nos será útil a efectos didácticos es la de Cauchy:

$$f(x) = \frac{1}{\pi} \frac{1}{1 + (x - \mu)^2} \quad -\infty < x < \infty \quad (1)$$

que solo tiene media en sentido restringido y carece de momentos de orden superior. La distribución acumulativa es

$$\begin{aligned}F(x) &= \frac{1}{\pi} \int_{-\infty}^x \frac{dt}{1 + (t - \mu)^2} \\ &= \frac{1}{2} + \frac{1}{\pi} \arctan(x - \mu)\end{aligned}$$

Logarítmico normal.—Sea x una variable aleatoria, y sea $y = \log_e x$. Si y es una variable normal, se dice que x tiene una distribución logarítmico normal. La función de densidad es

$$\begin{aligned}f(x) &= \frac{1}{x\beta\sqrt{2\pi}} e^{-(1/2\beta^2)(\log x - \log \alpha)^2} & x > 0 \\ &= 0 & \text{en otro caso}\end{aligned}$$

En esta densidad hay dos parámetros α y β , ambos mayores que cero.

6-6. Funciones de densidad completas.—En esta sección estudiaremos una propiedad de las funciones de densidad que será útil al desarrollar la teoría en capítulos posteriores. En lo que sigue designaremos por $f(x; \theta)$ una función de densidad, donde θ es el parámetro. Supondremos que θ está en el intervalo $\alpha_0 < \theta < \alpha_1$;

es decir, cada θ del intervalo define una densidad diferente $f(x; \theta)$. Queda así definido un conjunto de densidades al tomar θ todos los valores del intervalo $\alpha_0 < \theta < \alpha_1$; conjunto al que frecuentemente designaremos como una familia de funciones de densidad. Supongamos, además, que $f(x; \theta)$ es mayor que cero para todo x del intervalo $a < x < b$, e igual a cero en otro caso. También supondremos que a y b no dependen de θ . Más concretamente:

1. $f(x; \theta) > 0$ para $a < x < b$ y para $\alpha_0 < \theta < \alpha_1$. Si $f(x; \theta)$ es una densidad discreta, $a < x < b$ representará los valores enteros del intervalo para el cual

$$f(x; \theta) > 0 \quad (1)$$

2. $f(x; \theta) = 0$ en otro caso; es decir, cuando x no está en el intervalo $a < x < b$.

3. a y b no dependen de θ .

Antes de pasar a definir una familia completa de densidades, daremos dos ejemplos para aclarar las hipótesis hechas en 1).

Ejemplo 6-1.—Sea la variable aleatoria x con distribución normal de media θ y varianza 1. La densidad está dada por

$$f(x; \theta) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-\theta)^2} \quad -\infty < x < \infty$$

donde θ , en este problema, puede tomar los valores del intervalo $-\infty < \theta < \infty$. Si hacemos $a = -\infty$, $b = \infty$ y $\alpha_0 = -\infty$, $\alpha_1 = \infty$, $f(x; \theta) > 0$ si $a < x < b$ y $\alpha_0 < \theta < \alpha_1$; por tanto, esta densidad satisface las hipótesis de 1).

Ejemplo 6-2.—Sea la variable aleatoria x con distribución de Poisson de parámetro θ . La densidad está dada por

$$f(x; \theta) = \begin{cases} \frac{\theta^x e^{-\theta}}{x!} & x = 0, 1, 2, \dots \\ = 0 & \text{en otro caso} \end{cases}$$

donde θ toma, p. ej., los valores del intervalo $0 < \theta < 20$.

Si hacemos $a = 0$, $b = \infty$ y $\alpha_0 = 0$, $\alpha_1 = 20$, vemos que esta densidad discreta satisface las hipótesis hechas en 1).

Realmente, la hipótesis de que a y b son independientes de θ no es esencial para definir una función de densidad completa, pero simplificará las cosas y es suficiente para la mayoría de las densidades que se considerarán en este libro.

Analizaremos el concepto de *completa* con más detalle antes de dar la definición. Sea u cualquier función de x que *no dependa* de θ , que *no* sea idénticamente igual a cero para cada x del inter-

valo $a < x < b$, y que sea continua en el intervalo. Representamos por u_0 la función cero en el intervalo $a < x < b$; es decir, $u_0(x) = 0$ para todo x del intervalo $a < x < b$. Ahora bien, $E[u_0(x)] = 0$ para todo θ del intervalo $\alpha_0 < \theta < \alpha_1$ puesto que

$$E[u_0(x)] = \int_a^b 0 \cdot f(x; \theta) dx = 0$$

Si u_0 es la *única* función continua de x para $a < x < b$, cuyo valor esperado es 0 para cada θ de $\alpha_0 < \theta < \alpha_1$, se dice que $f(x; \theta)$ es una familia *completa* de densidades.

Esto puede establecerse así: Supongamos que no existe una función continua u de x en $a < x < b$ para la cual $E[u(x)] = 0$ para todo θ del intervalo $\alpha_0 < \theta < \alpha_1$, excepto la función cero u_0 . Entonces $f(x; \theta)$ es *completa* (una familia completa de densidades). La definición siguiente resume lo anterior.

Definición 6-4.—Sea $f(x; \theta)$ una familia de densidades y apliquemos las hipótesis hechas en 1). Si no existe una función u , continua en el intervalo $a < x < b$, para la cual $E[u(x)] = 0$ para todo θ del intervalo $\alpha_0 < \theta < \alpha_1$, a excepción de la función cero u_0 , entonces $f(x; \theta)$ se define como una familia completa de densidades. Si existe al menos una función continua v tal que $E[v(x)] = 0$ para cada θ del intervalo $\alpha_0 < \theta < \alpha_1$ y v no es la función cero, $f(x; \theta)$ no es completa.

La demostración de que una determinada función de densidad es completa requiere, en general, el empleo de recursos matemáticos que superan los límites de esta obra. Para la mayor parte de las funciones de densidad que manejamos en este libro nos limitaremos simplemente a decir si son o no completas.

Algunas veces puede demostrarse que una función de densidad *no* es completa encontrando una función v no cero para la cual $E[v(x)] = 0$ para todo θ del intervalo $\alpha_0 < \theta < \alpha_1$.

Ejemplo 6-3.—Vamos a ver aquí una función de densidad *no* completa. Sea x una variable aleatoria que se distribuye normalmente con media cero y varianza σ^2 ; es decir,

$$f(x; \sigma^2) = \frac{1}{\sigma \sqrt{2\pi}} e^{-(x^2)/2\sigma^2} \quad -\infty < x < \infty$$

donde σ^2 toma sus valores en el intervalo $0 < \sigma^2 < \infty$. Se ve fácilmente que $E(x) = 0$ para cada σ^2 del intervalo $0 < \sigma^2 < \infty$. Pero, evidentemente, $v(x) = x$ no es la función cero y, por tanto, $f(x; \sigma^2)$ *no* es completa.

He aquí algunas funciones de densidad que sí son completas:

$$f(x; \mu) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-\mu)^2} \quad -\infty < x < \infty \quad (2)$$

$$\alpha_0 < \mu < \alpha_1$$

donde α_0 y α_1 son $-\infty$ y $+\infty$, respectivamente, o números reales cualesquiera tales que $-\infty < \alpha_0 < \alpha_1 < \infty$.

$$f(x; \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \quad -\infty < x < \infty \quad (3)$$

$$\alpha_0 < \mu < \alpha_1$$

$$\beta_0 < \sigma^2 < \beta_1$$

donde $-\infty \leq \alpha_0 < \alpha_1 \leq \infty$, $0 < \beta_0 < \beta_1 \leq \infty$.

$$f(x; p) = p^x(1-p)^{1-x} \quad x=0, 1; \alpha_0 \leq p \leq \alpha_1 \quad (4)$$

$$0 \leq \alpha_0 < \alpha_1 \leq 1$$

$$f(x; \lambda) = \frac{\lambda^x e^{-\lambda}}{x!} \quad x=0, 1, 2, \dots; \alpha_0 < \lambda < \alpha_1 \quad (5)$$

con $0 < \alpha_0 < \alpha_1 \leq \infty$.

Daremos ahora un ejemplo de cómo puede utilizarse la propiedad de que una densidad sea completa. Sea x una variable aleatoria con función de densidad

$$f(x; \mu) = \frac{1}{\sqrt{2\pi}} e^{-(x-\mu)^2/2} \quad -\infty < x < \infty$$

$$-\infty < \mu < \infty$$

Evidentemente, $E(x^2) = \mu^2 + 1$. Planteemos ahora el problema siguiente: Además de x^2 , ¿existe otra función continua de x cuyo valor esperado sea también $\mu^2 + 1$? Si tal función existe, designémosla por $v(x)$; es decir,

$$E[v(x)] = \mu^2 + 1$$

Luego

$$E[v(x)] - E[(x^2)] = (\mu^2 + 1) - (\mu^2 + 1) = 0$$

o

$$E[v(x) - x^2] = 0 \quad (6)$$

para todo μ del intervalo $-\infty < \mu < \infty$. Hagamos $v(x) - x^2 = u(x)$. Así (6) puede escribirse $E[u(x)] = 0$ para todo μ del intervalo $-\infty < \mu < \infty$. Ahora bien, puesto que la densidad es completa

[véase (2)], esto implica que $u(x) \equiv 0$ para todo x del intervalo $0 < x < \infty$, o, en otras palabras, $v(x) \equiv x^2$. Luego x^2 es la única función continua cuyo valor esperado es $\mu^2 + 1$. Hemos resaltado el hecho de que $u(x)$, $v(x)$, etc. son continuas en el intervalo $a < x < b$. Ello no es esencial para la definición de densidad completa, pero sí será suficiente para cuanto necesitamos en este libro. La definición 6-4 se extiende a las densidades de más de una variable aleatoria si se reemplaza x por $x, y, \dots$, y para más de un parámetro θ .

PROBLEMAS

1. Hállese y represéntese la forma acumulativa de la distribución uniforme.

2. ¿Qué transformación cambia la variante x en otra cuya distribución sea la uniforme sobre el intervalo unitario, si

$$f(x) = \frac{(x-1)}{2}$$

$1 < x < 3$ y cero en otro caso? ¿Qué intervalo de la nueva variante es el que corresponde al $1,1 < x < 2,9$?

3. Represéntense $n(x; 0, 0,25)$, $n(x; 1, 0,25)$ y $n(x; 1, 9)$ sobre el mismo gráfico. ¿Cuál sería el aspecto de la distribución para valores pequeños de σ ? (Utilícese la tabla I.)

4. Si la distribución de x es la normal, con media unidad y $\sigma=0,6$, hállese $P(x > 0)$ y $P(0,2 < x < 1,8)$.

5. Hállese un número k tal que para una variante con distribución normal se verifique $P(\mu - k\sigma < x < \mu + k\sigma) = 0,95$. ¿Cuál sería el valor de k para $P=0,90$? ¿Y para $P=0,99$? ¿Para qué valor de k se verifica $P(x > \mu - k\sigma) = 0,95$?

6. Hállese la función generatriz $E(e^{t(x-\mu)})$ para los momentos respecto a la media de una distribución normal.

7. Hállese μ_r en función de σ para una distribución normal, para r par y para r impar. Desarróllese en serie infinita la función generatriz anterior.

8. ¿Qué multiplicador constante cambia la función e^{-x^2+x} en una función de densidad? ¿Cuáles son la media y la varianza de la distribución resultante?

9. Calcúlese $\int_1^2 e^{-x^2} dx$.

10. Calcúlese $\int_0^\infty x^{3/2} e^{-x/2} dx$.

11. Represéntese la función de densidad gamma para $\alpha=1$, $\beta=1$; $\alpha=1$, $\beta=2$; $\alpha=2$, $\beta=1$; $\alpha=4$, $\beta=1$.

12. Hállese el tercer momento μ'_3 de la distribución gamma.
13. Si en la distribución gamma se hace $\beta=2$ y $\alpha=(n-2)/2$, la distribución resultante se denomina distribución ji cuadrado con n grados de libertad. Hállense su función generatriz de momentos y su media y varianza.
14. Hállese k de manera que $P(x > k) = 0,05$ para la distribución ji cuadrado con 2 grados de libertad.
15. Hállese el momento r -ésimo de la distribución gamma sin utilizar la función generatriz de momentos.
16. Hállese el momento r -ésimo de la distribución gamma utilizando la función generatriz.
17. Represéntese la función de densidad beta para $\alpha=0, \beta=0$; $\alpha=1, \beta=1$; $\alpha=3, \beta=3$; $\alpha=2, \beta=3$; $\alpha=3, \beta=2$. ¿Qué aspecto tendría la función para valores grandes de α y β ?
18. Hállense la media y la varianza de la distribución beta.
19. Pruébese que la distribución beta es simétrica respecto del punto $x=1/2$ cuando $\alpha=\beta$.
20. Hállese la media de la distribución de Cauchy si

$$\frac{1}{\pi} \int_{-\infty}^{\infty} \frac{x dx}{1+(x-\mu)^2}$$

se define como

$$\lim_{A \rightarrow \infty} \int_{\mu-A}^{\mu+A} \frac{1}{\pi} \frac{x dx}{1+(x-\mu)^2}$$

Demuéstrese que esta distribución carece de momentos de orden superior.

21. Sean x e y dos variables aleatorias tales que $g(y|x)$ es una distribución normal con media μ y varianza σ^2/x , donde μ y σ^2 no dependen de x . Supongamos también que $x > 0$.

a) Escribbase la función de densidad $g(y|x)$.

b) Sea la distribución marginal de x una distribución gamma con $\alpha=m/2-1, \beta=2$; es decir,

$$h(x) = 1/[\Gamma(m/2)2^{m/2}]x^{(m-2)/2}e^{-x/2} \quad x > 0$$

Hállese la función de densidad conjunta de x e y .

22. a) Obténgase la distribución marginal de y en el problema 21.

b) Hállese $\text{var}(y)$ a partir del resultado encontrado en a).

23. Utilizando el problema 34 del capítulo 5, hállese $\text{var}(y)$ sin emplear la distribución marginal de y .

24. Compárese la distribución de Cauchy con la normal para $\sigma=2$, representándolas en un mismo gráfico con media igual a cero. Obsérvese que la varianza no es un buen criterio para comparar dos distribuciones, a menos que tengan la misma expresión funcional.

25. ¿Cuáles son los cumulantes de la distribución normal?
26. Supongamos que x tiene la distribución gamma, con parámetros $\alpha=10$, $\beta=1$. ¿Cuántos momentos tiene $y=1/x$?
27. Si x tiene la distribución gamma, hállese la función generatriz de momentos de $y=\log x$.
28. Una variante x tiene por función de densidad

$$f(x) = 2 \sqrt{\frac{2}{\pi}} x e^{-x^4/2} \quad x > 0$$

Hállense su media y su varianza.

29. Una variante tiene por momentos $\mu'_r = r!$. Hállese su función generatriz de momentos y dedúzcase de esta su distribución.
30. Una variante x tiene la distribución uniforme sobre el intervalo unitario; ¿qué función de x tendrá la distribución gamma con $\alpha=0$, $\beta=1$?
31. Una variante x tiene la distribución beta con $\alpha=0$, $\beta=1$. ¿Qué función de x tiene la distribución gamma con $\alpha=0$, $\beta=1$?
32. Una variante tiene por momentos

$$\mu'_r = \frac{r!}{(r/2)!}$$

cuando r es par y $\mu'_r = 0$ si r es impar. Dedúzcase la distribución de esta variante a partir de su función generatriz de momentos.

33. Demuéstrese cómo puede calcularse la distribución de Poisson acumulativa

$$\sum_{y=0}^n \frac{e^{-m} m^y}{y!}$$

mediante tablas de la función gamma incompleta $F(x; \alpha, \beta)$.

34. Si $\log x$ tiene una distribución normal con $\mu=1$, $\sigma^2=4$, hállese
- $$P(1/2 < x < 2)$$
- ($\log 2=0,693$).

35. Una variante x tiene por función de densidad

$$f(x) = 2 \sqrt{\frac{2}{\pi}} x e^{-x^4/2} \quad x > 0$$

Hállese $P(x < 4)$.

36. Determinense la media y la varianza de la distribución normal, derivando la identidad

$$\int_{-\infty}^{\infty} n(x; \mu, \sigma^2) dx = 1$$

respecto a μ y respecto a σ^2 .

37. Se dice que una variante x ha sido estandarizada en su escala si se divide por su desviación estándar. Demuéstrese que los cumulantes de x/σ son iguales a $\gamma_r/\gamma_r^{r/2}$, siendo γ_r el cumulante r -ésimo de x .

38. Demuéstrese que la distribución gamma se aproxima a la normal cuando α es grande, comparando los cumulantes de ambas distribuciones con escala estandarizada.

39. Una variante x tiene una distribución normal con media μ y varianza σ^2 . Demuéstrese que la media de la distribución condicional de x , siendo

$$a < x < b$$

es

$$\mu + \frac{n(a) - n(b)}{N(b) - N(a)} \sigma^2$$

40. Una variante x tiene como función de densidad $f(x)$. ¿Cómo podría determinarse una función $u=u(x)$ tal que la distribución de u sea $g(u)$?

41. La variable aleatoria bidimensional (x, y) tiene la función de densidad

$$f(x, y) = p^{x+y}(1-p)^{2-x-y} \quad x=0, 1; y=0, 1; 0 < p < 1$$

a) Demuéstrese que $E(x)=E(y)$.

b) Utilícese la parte a) para demostrar que la densidad no es completa, encontrando una función u no nula tal que $E[u(x, y)]=0$ para todo p tal que $0 < p < 1$.

42. La variable aleatoria bidimensional (x, y) tiene la función de densidad

$$f(x, y) = \frac{1}{2\pi} \exp \left[-\left(\frac{1}{2}\right)(x-\alpha)^2 - \left(\frac{1}{2}\right)(y-\alpha)^2 \right] \quad \begin{array}{l} -\infty < x < \infty \\ -\infty < y < \infty \\ 0 < \alpha < \infty \end{array}$$

Pruébese que esta densidad no es completa.

43. La variable aleatoria discreta x tiene la función de densidad

$$f(x) = p^x(1-p)^{1-x} \quad x=0, 1; 0 < p < 1$$

a) Sea u cualquier función de x (no dependiente de p) salvo la función cero; es decir, $u(1)$ y $u(0)$ no son ambas iguales a cero. Hállese $E[u(x)]$.

b) Pruébese, utilizando la parte a), que $E[u(x)]$ no puede ser idénticamente nula para todo p del intervalo $0 < p < 1$, al menos que u sea la función cero.

c) Utilícense las partes a) y b) para deducir que $f(x)$ es una densidad completa.

BIBLIOGRAFIA

1. ANDERSON, T. W.: *An Introduction to Multivariate Statistical Analysis*, John Wiley & Sons, Inc., Nueva York, 1958.
2. BRUNK, H.: *An Introduction to Mathematical Statistics*, Ginn & Company, Boston, 1960.
3. FELLER, W.: *An Introduction to Probability Theory and Its Applications*, 2.^a ed., vol. I, John Wiley & Sons, Inc., Nueva York, 1957.
4. HOEL, P. G.: *Introduction to Mathematical Statistics*, 2.^a ed., John Wiley & Sons, Inc., Nueva York, 1954.
5. HOGG, R., y A. CRAIG: *Introduction to Mathematical Statistics*, The Macmillan Company, Nueva York, 1959.
6. KENDALL, M. G.: *Advanced Theory of Statistics*, vol. I, Charles Griffin & Co., Ltd., Londres, 1948.

CAPITULO 7

MUESTREO

7-1. Inferencia inductiva.—Hasta ahora nos hemos ocupado de algunos aspectos de la teoría de la probabilidad. El estudio del muestreo nos lleva a la teoría de la estadística propiamente dicha, y pasamos ahora a tratar brevemente de una rama importante de la teoría de la estadística y de sus relaciones con el muestreo.

El progreso científico va unido a la experimentación: el investigador realiza un experimento y obtiene varios datos; sobre la base de estos datos se sacan ciertas conclusiones, y estas conclusiones suelen ir más allá de los materiales y operaciones del experimento particular realizado. En otras palabras, puede ocurrir que el científico generalice ciertas conclusiones del experimento particular a todas las clases de experimentos semejantes. Tal tipo de extensión de lo particular a lo general se denomina *inferencia inductiva*, y es un procedimiento para hallar nuevo conocimiento científico.

Es bien sabido que la inferencia inductiva constituye un proceso arriesgado. En efecto, es un teorema de lógica que toda inferencia inductiva exacta es imposible. Una generalización perfectamente válida no puede hacerse; sin embargo, sí cabe hacer inferencias inseguras, y el grado de incertidumbre es susceptible de medición si el experimento se ha realizado de acuerdo con determinados principios. Una de las misiones de la estadística consiste en conseguir técnicas para efectuar inferencias inductivas y para medir el grado de incertidumbre de tales inferencias. La medida de la incertidumbre viene expresada en probabilidad, y esta es la razón por la cual hemos dedicado tanta extensión a la teoría de la probabilidad.

Antes de seguir, diremos unas palabras sobre otra clase de inferencia: la *inferencia deductiva*. Mientras que las conclusiones sacadas por inferencia inductiva son solo probables, las deducidas por inferencia deductiva son concluyentes. Para dar un ejemplo de inferencia deductiva, consideremos las dos afirmaciones siguientes:

1. Uno de los ángulos interiores de todo triángulo rectángulo es igual a 90° .

2. El triángulo *A* es un triángulo rectángulo.

Si aceptamos estas afirmaciones, llegamos necesariamente a la conclusión:

3. Uno de los ángulos del triángulo rectángulo A es igual a 90° .

Este es un ejemplo de inferencia deductiva, la cual puede describirse como un método de deducir información (afirmación 3) a partir de hechos aceptados (afirmaciones 1 y 2). La afirmación 1 se llama premisa mayor; la afirmación 2, premisa menor, y la afirmación 3, conclusión. Como otro ejemplo, consideremos el siguiente:

Premisa mayor: 1. Todos los graduados de West Point tienen más de 18 años.

Premisa menor: 2. Juan es un graduado de West Point.

Conclusión: 3. Juan tiene más de 18 años.

En la inferencia deductiva la conclusión es verdadera si lo son las premisas. El ejemplo puede representarse gráficamente como en la figura 7-1.

Los graduados de West Point forman un subconjunto de todas las personas que tienen más de dieciocho años de edad, y Juan es un elemento del subconjunto de los graduados de West Point; por tanto, Juan es también un elemento del conjunto de personas cuya edad es superior a los dieciocho años.

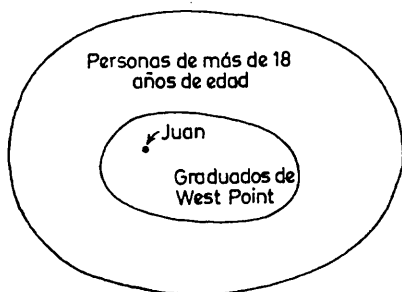


FIG. 7-1.

Aunque la inferencia deductiva es extremadamente importante, muchos de los nuevos conocimientos del mundo real se han adquirido por el proceso de inferencia inductiva. En las ciencias matemáticas, p. ej., la inferencia *deductiva* se utiliza para *demonstrar* teoremas, mientras que en las ciencias empíricas se emplea la inferencia *inductiva* para hallar nuevos conocimientos.

Ilustremos la inferencia inductiva con un ejemplo sencillo. Imaginemos un almacén que contiene, p. ej., 10 millones de semillas, de las cuales sabemos que producen flores blancas o rojas. La información que deseamos es: *¿cuántas de estos 10 millones de semillas (o qué porcentaje) producirán flores blancas?* La única manera de estar *seguros* de dar una respuesta correcta a esta pregunta es plantar todas las semillas y observar el número de las que producen flores blancas. Sin embargo, esto no es posible, pues deseamos vender las semillas; aunque no quisiéramos vender las semillas, preferiríamos obtener una respuesta sin invertir tanto esfuerzo. Naturalmen-

te, sin plantar todas las semillas y observar el color de la flor producida por cada una no podemos conocer *con certeza* el número de semillas que producen flores blancas. Otra idea que se nos ocurre es: *¿Podemos plantar unas pocas semillas y, basándonos en los colores de estas pocas flores, hacer una afirmación sobre el número de semillas de los 10 millones que producirán flores blancas?* La respuesta es que no es posible hacer una predicción exacta acerca de cuántas flores blancas producirán las semillas, pero sí cabe hacer una afirmación probabilística si seleccionamos esas pocas semillas de cierta manera. Esto es inferencia inductiva: Seleccionamos unas pocas de los 10 millones de semillas, las plantamos y observamos el número de las que producen flores blancas y, basándonos en estas pocas semillas, hacemos una predicción sobre cuántas de los 10 millones producirán flores blancas; a partir del conocimiento del color de unas pocas generalizamos al total de los 10 millones. No podemos estar seguros de nuestra respuesta, pero sí cabe tener confianza en ella en el sentido de la relación entre frecuencia y probabilidad.

7.2. Poblaciones y muestras.—En la sección anterior hemos visto que un problema central en el descubrimiento de nuevos conocimientos acerca del mundo real consiste en observar unos pocos de los elementos bajo discusión y, sobre la base de estos, hacer una afirmación referente a la totalidad de los mismos. Analicemos esta forma de proceder con más detalle.

Definición 7-1. Población objetivo.—*La totalidad de los elementos en discusión y acerca de los cuales se desea información, se denominará población objetivo.*

En el ejemplo de la sección anterior, los 10 millones de semillas del almacén forman la población objetivo. Esta puede estar constituida por todas las cabezas de ganado vacuno de Wisconsin en una fecha dada, o por los precios de la carne en Nueva York en una fecha determinada, o por la sucesión hipotética de caras y cruces obtenidas al lanzar una moneda un número infinito de veces, o por el conjunto hipotético de un número infinito de medidas de la velocidad de la luz, etc. Lo importante es que la población objetivo sea susceptible de ser definida con absoluta precisión; por lo demás, puede ser real o hipotética.

Desde el punto de vista de la estadística el problema de la inferencia inductiva se trata del siguiente modo: el objeto de un experimento es averiguar algo sobre una determinada población. Es imposible o impracticable examinar la población completa, pero puede examinarse una parte o muestra de la misma, y sobre la base de esta investigación limitada hacer inferencias relativas a la población total.

Surge inmediatamente el problema de cómo debe seleccionarse la muestra de la población. Dijimos en la sección anterior que se podrían hacer afirmaciones probabilísticas acerca de la población si la muestra se seleccionaba de cierta manera. El caso de una *muestra aleatoria simple*, llamada a veces *muestra aleatoria*, es de particular importancia. Su definición se da a continuación.

Definición 7-2. Muestra aleatoria.—Sean $x_1, x_2, \dots, x_n$ variables aleatorias cuya distribución conjunta es

$$g(x_1, x_2, \dots, x_n) = f(x_1) f(x_2) \dots f(x_n)$$

donde la función de densidad de cada x_i es $f(x)$. En tal supuesto se dice que $x_1, x_2, \dots, x_n$ es una muestra aleatoria de tamaño n de la población con densidad $f(x)$.

Frecuentemente no es posible seleccionar una muestra aleatoria de la población objetivo, sino que se toma de una población relacionada con ella. Para distinguir las dos poblaciones damos la siguiente definición:

Definición 7-3. Población muestreada.—Sea $x_1, x_2, \dots, x_n$ una muestra aleatoria de una población con densidad $f(x)$; esta población recibe el nombre de población muestreada.

Pueden hacerse afirmaciones probabilísticas válidas, acerca de poblaciones muestreadas, sobre la base de muestras aleatorias, pero las afirmaciones relativas a las poblaciones objetivo, basadas en la relación entre probabilidad y frecuencia relativa, no son válidas, salvo en el caso de que la población objetivo sea también la población muestreada. Daremos algunos ejemplos que aclaren la distinción entre población muestreada y población objetivo.

Ejemplo 7-1.—Supongamos que un sociólogo desea estudiar las costumbres religiosas de los varones de veinte años de los Estados Unidos. Para hacer su estudio, extrae una muestra de varones de veinte años de una gran ciudad. En este caso la población objetivo está formada por los varones de veinte años de los Estados Unidos, y la población muestreada es la de los varones de veinte años de la ciudad en que muestreó. El sociólogo puede sacar conclusiones probabilísticas válidas para su población muestreada, pero tiene que utilizar su juicio personal para extrapolar a la población objetivo, y la fiabilidad de la extrapolación no puede medirse en términos de probabilidad.

Ejemplo 7-2.—Un investigador está estudiando el rendimiento de cierta variedad de trigo en el estado de Colorado. Tiene a su disposición cinco granjas diseminadas por dicho estado, en las cuales planta el trigo y observa el rendimiento. La población muestreada está formada por los rendimientos de estas cinco granjas,

mientras que la población objetivo la forman los rendimientos de esta variedad de trigo en todas las granjas de dicho estado.

En este libro se estudiará el problema de seleccionar (extraer) una muestra de una población muestreada con densidad $f(x)$ y, sobre la base de estas observaciones muestrales, se harán afirmaciones probabilísticas acerca de $f(x)$. Así desarrollaremos las ideas de muestreo.

7-3. Distribuciones muestrales.—Supongamos una variante x con distribución $f(x)$ en determinada población. Y supongamos que se extrae aleatoriamente una muestra de dos valores, x_1 y x_2 , de x . Este par de números (x_1, x_2) determina un punto en el plano, y la población de todos los posibles pares de números que pueden obtenerse forma una población bivalente. Lo que nos interesa es hallar la distribución de esta población bivalente basándonos en la distribución original $f(x)$.

La función de densidad conjunta de x_1 y x_2 deberá ser una cierta función $g(x_1, x_2)$ tal que cualesquiera que sean a_1, a_2, b_1, b_2 , tengamos

$$P(a_1 < x_1 < b_1, a_2 < x_2 < b_2) = \int_{a_1}^{b_1} \int_{a_2}^{b_2} g(x_1, x_2) dx_2 dx_1 \quad (1)$$

Ahora bien: al decir muestra aleatoria entendemos que el valor de la primera observación no tiene efecto alguno sobre el valor de la segunda. O, dicho con otras palabras, para una muestra aleatoria, x_1 y x_2 son independientes en sentido probabilístico. Cuando las dos variantes de la distribución bivalente son independientes en sentido probabilístico, hemos visto que la distribución conjunta es el producto de las distribuciones marginales. En este ejemplo, las distribuciones marginales son simplemente $f(x_1)$ y $f(x_2)$, de modo que tenemos por definición de aleatoriedad

$$g(x_1, x_2) = f(x_1)f(x_2) \quad (2)$$

o, lo que es lo mismo,

$$P(a_1 < x_1 < b_1, a_2 < x_2 < b_2) = P(a_1 < x_1 < b_1)P(a_2 < x_2 < b_2) \quad (3)$$

para todo a_i y b_i .

Como ejemplo sencillo, supongamos que x solo tiene dos valores, cero y uno, con probabilidades q y p , respectivamente; esto es, x es una variante discreta con la distribución binomial

$$f(x) = \binom{1}{x} p^x q^{1-x} \quad x=0, 1 \quad (4)$$

y puesto que $\binom{1}{0} = \binom{1}{1} = 1$, podemos escribir

$$f(x) = p^x q^{1-x}$$

La función de densidad conjunta para una muestra de dos valores de x es

$$g(x_1, x_2) = p^{x_1+x_2} q^{2-x_1-x_2} \quad x_1 = 0, 1; \quad x_2 = 0, 1 \quad (5)$$

que está definida en los cuatro puntos (0, 0), (0, 1), (1, 0), (1, 1) del plano x_1, x_2 . Debe observarse que esta función no es la que se habría obtenido extrayendo dos elementos de una población binomial y contando el número y de aciertos. En ese caso, la densidad sería

$$h(y) = \binom{2}{y} p^y q^{2-y} \quad y = 0, 1, 2, \quad (6)$$

que difiere de (5) en que es la distribución de una sola variante $y = x_1 + x_2$. La ecuación (5) da la distribución conjunta de las dos variables aleatorias x_1 y x_2 .

Debe observarse que $g(x_1, x_2)$ proporciona la distribución de la muestra en el *orden de extracción*. Así, en (5), $g(0, 1) = pq$ y no $2pq$; $g(0, 1)$ se refiere a la probabilidad de obtener primero un cero y después un uno, y, en general, (1) representa la probabilidad de que la primera observación obtenida caiga en el intervalo (a_1, b_1) y la segunda en el (a_2, b_2) . El suceso opuesto no satisface las condiciones especificadas, salvo, naturalmente, en el caso de que los dos intervalos coincidan.

Razonando de la misma manera que antes, hallamos que la distribución conjunta para una muestra aleatoria de tamaño n , $x_1, x_2, \dots, x_n$, de una población cuya distribución sea $f(x)$, es

$$g(x_1, x_2, \dots, x_n) = f(x_1)f(x_2) \dots f(x_n) \quad (7)$$

que vuelve a darnos la distribución de la muestra en el orden de extracción.

Nuestra definición de muestreo aleatorio elimina automáticamente el muestreo sin reemplazamiento de una población finita. Si, p. ej., extraemos 2 bolas de una urna que contenga 2 bolas blancas y 3 bolas negras, el resultado de la primera extracción afecta seguramente a la probabilidad del resultado de la segunda extracción. Las dos extracciones no son independientes en el sentido probabilístico. En este caso debe adoptarse otra definición de muestreo aleatorio (problemas 26 y 32). Por consiguiente, tanto en este capítulo como en los siguientes nos ocuparemos del muestreo en poblaciones con-

tinuas (en donde no surge la cuestión de si el muestreo es con o sin reemplazamiento) y del muestreo con reemplazamiento en poblaciones finitas.

Algunas veces emplearemos la expresión «población $f(x)$ » para indicar «una población con densidad $f(x)$ ».

7-4. Momentos muestrales.—Uno de los problemas centrales de la estadística es el siguiente: Se desea estudiar una población que tiene la distribución $f(x; \theta)$, de la que se conoce la forma, pero que contiene un parámetro desconocido θ (si θ es conocido, la función de densidad está completamente especificada). El procedimiento es tomar una muestra aleatoria $x_1, x_2, \dots, x_n$ de tamaño n de esta distribución y formar una función $u(x_1, x_2, \dots, x_n)$ que «represente» o «estime» el parámetro desconocido θ . El problema consiste en determinar qué función será «la mejor» para estimar θ . Este problema será formulado con mayor detalle en el próximo capítulo. En esta sección examinaremos ciertas funciones de las muestras aleatorias, llamadas momentos muestrales. En primer lugar, definiremos lo que se entiende por estadístico.

Definición 7-4. Estadístico.—*Un estadístico es una función de variables aleatorias observables que no contiene parámetros desconocidos.*

Así, por ejemplo, si la variable aleatoria x tiene la distribución $n(x; \mu, \sigma^2)$, donde μ y σ^2 son desconocidas, $x - \mu$ no es un estadístico, ni tampoco lo es x/σ ; pero x , $x + 3$, $x^2 + \log x^2$ sí son estadísticos.

En la formulación anterior, uno de los problemas fundamentales de la estadística es hallar un estadístico (función de las variables aleatorias $x_1, x_2, \dots, x_n$) apropiado para «representar» θ .

A continuación definiremos y analizaremos ciertos e importantes estadísticos: los momentos muestrales.

Definición 7-5. Momentos muestrales.—*Sea $x_1, x_2, \dots, x_n$ una muestra aleatoria de la distribución $f(x)$. El momento muestral r -ésimo respecto a cero es*

$$m'_r = \frac{1}{n} \sum_{i=1}^n x_i^r$$

En particular, si $r=1$, tenemos la media muestral, que se representa corrientemente por $\bar{x}$; es decir,

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

En el capítulo 5 definíamos el momento r -ésimo de una densidad $f(x)$ como igual a $E(x^r)$. Veremos ahora que los momentos muestrales reflejan los momentos de la población en el sentido de que los valores esperados de los momentos muestrales coinciden con los momentos de la población.

Teorema 7-1.—Sea $x_1, x_2, \dots, x_n$ una muestra aleatoria de una población con densidad $f(x)$. El valor esperado del momento muestral r -ésimo es igual al momento r -ésimo de la población; es decir, $E[m'_r] = \mu'_r$.

Demostración.—Observemos, en primer lugar, que

$$m'_r = \frac{1}{n} \sum x_i^r$$

y que m'_r es una variable aleatoria, puesto que es función de las variables aleatorias x_i . En virtud del teorema 5-6, tenemos

$$E[m'_r] = E \left[\frac{1}{n} \sum x_i^r \right] = \frac{1}{n} E \left[\sum x_i^r \right] = \frac{1}{n} \sum E[x_i^r]$$

Pero, por la definición de momentos de una población $E(x_i^r) = \mu'_r$; por tanto,

$$E[m'_r] = \frac{1}{n} \sum_{i=1}^n \mu'_r = \mu'_r$$

Como caso particular, para $r=1$ se obtiene el siguiente corolario.

Corolario 7-1-1.—Sea $x_1, x_2, \dots, x_n$ una muestra aleatoria de una población con densidad $f(x)$, y sea μ la media de la población. Se tiene entonces que $E(\bar{x}) = \mu$.

Consideraremos a continuación la varianza de la media $\bar{x}$ de una muestra aleatoria.

Teorema 7-2.—Sea $x_1, x_2, \dots, x_n$ una muestra aleatoria de tamaño n de la densidad $f(x)$, y sea

$$\bar{x} = \frac{1}{n} \sum x_i.$$

Supongamos también que la densidad $f(x)$ tiene varianza finita σ^2 . En tal supuesto, la varianza de $\bar{x}$, que representamos por $\sigma_{\bar{x}}^2$, es σ^2/n .

Demostración.—Por definición,

$$\sigma_{\bar{x}}^2 = E[\bar{x} - E(\bar{x})]^2 \quad (1)$$

o sea

$$\begin{aligned}
 \sigma_{\bar{x}}^2 &= E(\bar{x} - \mu)^2 \\
 &= E\left(\frac{1}{n} \sum x_i - \mu\right)^2 \\
 &= E\left[\frac{1}{n} \sum (x_i - \mu)\right]^2 \\
 &= \frac{1}{n^2} E\left[\sum (x_i - \mu)\right]^2
 \end{aligned} \tag{2}$$

Elevando al cuadrado la suma, obtenemos n términos de la forma $(x_i - \mu)^2$ y $\binom{n}{2}$ términos de la forma $2(x_i - \mu)(x_j - \mu)$, siendo $i < j$.

El valor esperado de $(x_i - \mu)^2$ solo depende de la distribución marginal de x_i , ya que en la integral

$$\int \int \dots \int (x_i - \mu)^2 \prod_j [f(x_j) dx_j]$$

todos los factores en que no interviene x_i se hacen iguales a la unidad y nos queda

$$\int (x_i - \mu)^2 f(x_i) dx_i = \sigma^2 \tag{3}$$

en donde σ^2 es la varianza poblacional. Análogamente

$$\begin{aligned}
 E[(x_i - \mu)(x_j - \mu)] &= \int \int (x_i - \mu)(x_j - \mu)f(x_i)f(x_j) dx_i dx_j \\
 &= \int (x_i - \mu)f(x_i) dx_i \int (x_j - \mu)f(x_j) dx_j \\
 &= 0
 \end{aligned} \tag{4}$$

La ecuación (2) es entonces

$$\begin{aligned}
 \sigma_{\bar{x}}^2 &= \frac{1}{n^2} \sum_{i=1}^n E(x_i - \mu)^2 \\
 &= \frac{1}{n^2} \sum \sigma^2 \\
 &= \frac{\sigma^2}{n}
 \end{aligned} \tag{5}$$

Por tanto, la varianza de la media muestral es igual a la varianza poblacional dividida por el tamaño de la muestra.

Este hecho tiene mucha importancia en estadística aplicada. Significa que cualquiera que sea la distribución poblacional (con tal que tenga varianza finita), la distribución de la media muestral se concentra más y más alrededor de la media poblacional al aumentar el tamaño de la muestra. Se deduce que cuanto mayor sea la muestra, más ciertos podremos estar de que la media muestral será una buena estimación de la media poblacional. En esto consiste esencialmente la ley de los grandes números. Obtendremos un enunciado más preciso de la misma en la sección siguiente.

7.5. Ley de los grandes números.—Sea $f(x; \theta)$ la densidad de una variable aleatoria x . Hemos examinado el hecho de que una forma de tener información sobre la función de densidad consiste en observar una muestra aleatoria y hacer una inferencia desde la muestra a la población. Si se conociese θ , la función de densidad estaría completamente determinada y la inferencia sería innecesaria. Por tanto, parece que sería deseable que la muestra aleatoria nos facilitase alguna información sobre el parámetro desconocido θ . Este problema será examinado con detalle en el próximo capítulo. En esta sección estudiaremos otro problema relacionado con el primero.

Designemos por μ el valor esperado $E(x)$ en la densidad $f(x)$; se trata de estimar μ . En sentido amplio, $E(x)$ es la media de un número infinito de valores de la variable aleatoria x . En un problema del mundo real, solo podemos observar un número finito de valores de la variable aleatoria x . Una cuestión crucial es entonces la siguiente: *¿Pueden hacerse inferencias fiables acerca de $E(x)$, «media» de un número infinito de valores de x , utilizando únicamente un número finito de valores de x (es decir, una muestra aleatoria de tamaño n)?* La respuesta es afirmativa; es posible hacer inferencias fiables sobre $E(x)$ utilizando solo una muestra finita, y lo demostraremos probando lo que se llama la ley débil de los grandes números. Esta ley afirma lo siguiente: Se puede determinar un n tal que, si se toma una muestra aleatoria de tamaño n o mayor, de una población con densidad $f(x)$ [con $E(x) = \mu$], la probabilidad de que la media muestral difiera de μ en menos de una cantidad determinada, arbitrariamente pequeña, llega a hacerse tan próxima a 1 como se desee. Con mayor precisión: la ley débil de los grandes números dice que para dos números pequeños determinados ϵ y δ , donde $\epsilon > 0$ y $0 < \delta < 1$, existe un entero n tal que si se obtiene de $f(x)$ una muestra aleatoria de tamaño igual o mayor que n y se calcula la media, designada por $\bar{x}_n$, la probabilidad de que $\bar{x}_n$ se desvíe de μ en menos de ϵ (es decir, esté arbitrariamente cer-

ca de μ) es mayor que $1 - \delta$ (o sea tan próxima a 1 como se desee). En símbolos, esto se escribe así: Para cualesquiera $\epsilon > 0$ y δ entre 0 y 1, existe un entero n tal que, para todo $m \geq n$,

$$P(|\bar{x}_m - \mu| < \epsilon) > 1 - \delta$$

Antes de demostrar la ley débil de los grandes números, establecemos la desigualdad de Tchebysheff.

Teorema 7-3.—Representemos por $f(x)$ una densidad con media μ y varianza finita σ^2 . Sea a cualquier número positivo y $\bar{x}_n$ la media de una muestra aleatoria de $f(x)$, de tamaño n . Entonces,

$$P\left(-\frac{a\sigma}{\sqrt{n}} \leq \bar{x}_n - \mu \leq \frac{a\sigma}{\sqrt{n}}\right) \geq 1 - \frac{1}{a^2}$$

Demostración.—En la demostración reemplazaremos $\bar{x}_n$ por $\bar{x}$.

Supongamos que la distribución de la media muestral viene dada por $g(\bar{x})$, en donde $\bar{x}$ es la media de una muestra de tamaño n , pro-

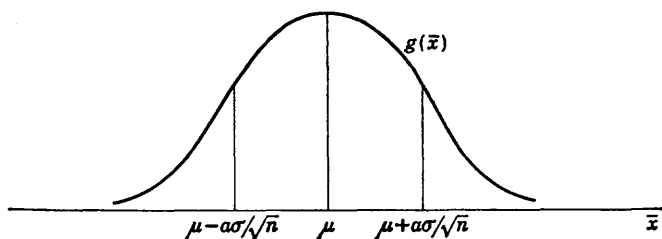


FIG. 7-2.

cedente de una población con distribución $f(x)$. Hemos visto que la media y la varianza de $\bar{x}$ son μ y σ^2/n , siendo μ y σ^2 la media y la varianza de $f(x)$. De la definición de varianza se deduce que

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n} = \int_{-\infty}^{\infty} (\bar{x} - \mu)^2 g(\bar{x}) d\bar{x} \quad (1)$$

Dividamos el campo de integración en tres partes, como se indica en la figura 7-2:

$$\begin{aligned} \frac{\sigma^2}{n} = \int_{-\infty}^{\mu - (a\sigma/\sqrt{n})} (\bar{x} - \mu)^2 g(\bar{x}) d\bar{x} &+ \int_{\mu - (a\sigma/\sqrt{n})}^{\mu + (a\sigma/\sqrt{n})} (\bar{x} - \mu)^2 g(\bar{x}) d\bar{x} \\ &+ \int_{\mu + (a\sigma/\sqrt{n})}^{\infty} (\bar{x} - \mu)^2 g(\bar{x}) d\bar{x} \quad (2) \end{aligned}$$

en donde a es un número positivo arbitrariamente elegido. Vamos a obtener una desigualdad modificando el segundo miembro de la ecuación (2). Prescindiremos de la segunda integral, y como esta es positiva, dicho segundo miembro disminuirá. Además, en la primera integral sustituiremos el factor $(\bar{x} - \mu)^2$ por $a^2\sigma^2/n$. Esto reducirá evidentemente el valor de la integral, puesto que en el campo de integración

$$|\bar{x} - \mu| \geq \frac{a\sigma}{\sqrt{n}}$$

La misma sustitución reducirá también la tercera integral, y tendremos

$$\frac{\sigma^2}{n} \geq \frac{a^2\sigma^2}{n} \int_{-\infty}^{\mu - (a\sigma/\sqrt{n})} g(\bar{x}) d\bar{x} + \frac{a^2\sigma^2}{n} \int_{\mu + (a\sigma/\sqrt{n})}^{\infty} g(\bar{x}) d\bar{x} \quad (3)$$

o, lo que es lo mismo,

$$\frac{1}{a^2} \geq P \left(|\bar{x} - \mu| \geq \frac{a\sigma}{\sqrt{n}} \right) \quad (4)$$

puesto que las dos integrales de (3) dan exactamente la probabilidad de que $\bar{x}$ esté fuera del intervalo

$$\left[\mu - \left(\frac{a\sigma}{\sqrt{n}} \right) \quad \text{a} \quad \mu + \left(\frac{a\sigma}{\sqrt{n}} \right) \right]$$

Transformando (4) se completa la demostración.

Consideremos ahora la ley débil de los grandes números.

Teorema 7-4.—*Designemos por $f(x)$ una densidad con media μ y varianza finita σ^2 , y sea $\bar{x}_n$ la media de una muestra aleatoria de tamaño n de $f(x)$. Sean ϵ y δ dos números pequeños dados y tales que $\epsilon > 0$, $0 < \delta < 1$. Si n es un entero mayor que $\sigma^2/\epsilon^2\delta$,*

$$P(-\epsilon < \bar{x}_n - \mu < \epsilon) \geq 1 - \delta \quad (5)$$

Demostración.—En el teorema 7-3 tomamos el número positivo a de manera que $1/a^2 = \delta$ o, en otras palabras, elegimos a tal que $a = 1/\sqrt{\delta}$. A continuación tomamos n tal que $a\sigma/\sqrt{n} < \epsilon$; o sea, $n > \sigma^2/\delta\epsilon^2$. Sustituyendo en la desigualdad probabilística del teorema 7-3, se deduce la inecuación (5).

Como ejemplo, supongamos una distribución de media desconocida y con varianza igual a 1. ¿Qué tamaño habrá de tener la muestra para que la probabilidad de que la media muestral diste menos

de 0,5 de la media poblacional sea al menos de 0,95? Tenemos $\sigma^2=1$, $\epsilon=0,5$, $\delta=0,05$. Luego

$$n > \frac{\sigma^2}{\delta \epsilon^2} = \frac{1}{0,05(0,5)^2} = 80$$

Queda demostrado que, mediante una muestra aleatoria, resulta posible hacer inferencias inductivas a poblaciones, y que la fiabilidad de la inferencia puede medirse en términos de probabilidad.

7-6. El teorema central del límite.—El teorema central del límite es una de las proposiciones más importantes de la estadística. Justifica el esfuerzo realizado en el estudio de la función de densidad normal; es también uno de los teoremas más notables de toda la matemática.

Teorema 7-5.—Representemos por $f(x)$ una densidad con media μ y varianza finita σ^2 , y sea $\bar{x}_n$ la media de una muestra aleatoria de tamaño n de $f(x)$. Definamos la variable aleatoria y_n por

$$y_n = \frac{\bar{x}_n - \mu}{\sigma} \sqrt{n}$$

La distribución de y_n se aproxima a la normal de media 0 y varianza 1 cuando n crece indefinidamente.

Lo extraordinario en este teorema es que nada se dice acerca de la forma de la función de densidad dada. Cualquiera que sea esta distribución, con tal que tenga varianza finita, la media muestral tendrá aproximadamente, para muestras grandes, la distribución normal. La condición de que la varianza sea finita no es una restricción excesiva en lo que se refiere a la estadística aplicada, ya que, prácticamente, siempre será finito el recorrido de la variante, en cuyo caso la varianza es necesariamente finita.

La importancia de este teorema, en lo que concierne a las aplicaciones prácticas, se debe al hecho de que la media $\bar{x}_n$ de una muestra aleatoria procedente de *cualquier* distribución con varianza finita σ^2 y media μ , se distribuye aproximadamente como una variante normal de media μ y varianza σ^2/n .

No nos es posible demostrar este teorema, porque ello exige el empleo de técnicas matemáticas superiores. No obstante, para facilitar su comprensión, vamos a esbozar su demostración para el caso más restringido en que la distribución posee función generatriz de momentos. Se trata esencialmente de ver que la función generatriz de momentos para la media muestral tiende a la función generatriz

de momentos de la distribución normal. Empezaremos por obtener la función generatriz de momentos para

$$y = \frac{x' - \mu'}{\sigma'}$$

en donde x' tiene la distribución normal. La función generatriz es

$$m_1(t) = \int_{-\infty}^{\infty} e^{ty} n(x'; \mu', \sigma'^2) dx' \quad (1)$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma'} e^{t(x' - \mu')/\sigma'} e^{-\frac{1}{2}(x' - \mu')^2/\sigma'^2} dx' \quad (2)$$

y, lo mismo que en la sección 6-2, hallamos

$$m_1(t) = e^{\frac{1}{2}t^2} \quad (3)$$

Supongamos ahora que x tiene una distribución cualquiera $f(x)$ con media μ y varianza σ^2 , que tiene función generatriz de momentos. La función generatriz de momentos de $(x - \mu)/\sigma$, se define así:

$$m_2(t) = \int_{-\infty}^{\infty} e^{t(x - \mu)/\sigma} f(x) dx \quad (4)$$

La media $\bar{x}$ de una muestra de tamaño n tendrá una distribución $g(\bar{x})$, que, como hemos visto, tiene por media μ y por varianza σ^2/n . La función generatriz de momentos de

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \quad (5)$$

$m_3(t)$, se define así:

$$m_3(t) = \int_{-\infty}^{\infty} \exp\left(t \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}\right) g(\bar{x}) d\bar{x} \quad (6)$$

Queremos demostrar que $m_3(t)$ se aproxima a $m_1(t)$ al crecer el tamaño n de la muestra.

Podemos determinar $m_3(t)$ en función de $m_2(t)$; $m_3(t)$ es el valor esperado

$$E \left[\exp \left(t \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \right) \right] = E \left[\exp \left(\frac{t}{n} \sum \frac{x_i - \mu}{\sigma/\sqrt{n}} \right) \right]$$

y como sabemos que la distribución conjunta de $x_1, x_2, \dots, x_n$ es

$\prod_{i=1}^n f(x_i)$, escribiremos

$$\begin{aligned} m_3(t) &= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} e^{\frac{t}{\sqrt{n}} \sum_{i=1}^n x_i} \prod_{i=1}^n f(x_i) dx_i \\ &= \prod_{i=1}^n \left[\int_{-\infty}^{\infty} e^{\frac{t}{\sqrt{n}} \frac{x_i - \mu}{\sigma}} f(x_i) dx_i \right] \end{aligned} \quad (7)$$

y en virtud de (4), cada factor de este producto es $m_2(t/\sqrt{n})$; por tanto,

$$m_3(t) = \left[m_2\left(\frac{t}{\sqrt{n}}\right) \right]^n \quad (8)$$

La derivada r -ésima de $m_2(t/\sqrt{n})$ en $t=0$ da, evidentemente, el momento r -ésimo respecto a la media, dividido por $(\sigma\sqrt{n})^r$. Podemos escribir:

$$m_2\left(\frac{t}{\sqrt{n}}\right) = 1 + \frac{\mu_1}{\sigma} \frac{t}{\sqrt{n}} + \frac{1}{2!} \frac{\mu_2}{\sigma^2} \left(\frac{t}{\sqrt{n}}\right)^2 + \frac{1}{3!} \frac{\mu_3}{\sigma^3} \left(\frac{t}{\sqrt{n}}\right)^3 + \dots \quad (9)$$

y, dado que $\mu_1=0$, $\mu_2=\sigma^2$, se tendrá

$$m_2\left(\frac{t}{\sqrt{n}}\right) = 1 + \frac{1}{n} \left(\frac{1}{2} t^2 + \frac{1}{3! \sqrt{n}} \frac{\mu_3}{\sigma^3} t^3 + \frac{1}{4! n} \frac{\mu_4}{\sigma^4} t^4 + \dots \right) \quad (10)$$

Si recordamos que la definición de e^u es

$$e^u = \lim_{n \rightarrow \infty} \left(1 + \frac{u}{n} \right)^n$$

vemos que $m_3(t)$ toma precisamente esta forma cuando n tiende a infinito, en donde u representa el paréntesis de (10), y al tender n a infinito, todos los términos de u se anulan a excepción del primero, de modo que

$$\lim_{n \rightarrow \infty} m_3(t) = e^{t^2/2} \quad (11)$$

Por tanto, en el límite, z posee la misma función generatriz de momentos que y , y, por un teorema análogo al 5-7, tiene la misma

distribución. Así, pues, en el límite, la media muestral tendrá la distribución normal, cualquiera que sea la distribución $f(x)$, siempre que $f(x)$ tenga función generatriz de momentos, o, más generalmente, siempre que $f(x)$ tenga momento de segundo orden.

El grado de aproximación depende, por supuesto, del tamaño de la muestra y de la función particular $f(x)$. En la figura 7-3 puede verse la tendencia a la distribución normal para el caso particular $f(x)=e^{-x}$, $x > 0$. Las curvas de trazo continuo representan las distribuciones efectivas, y las de trazo discontinuo las correspon-

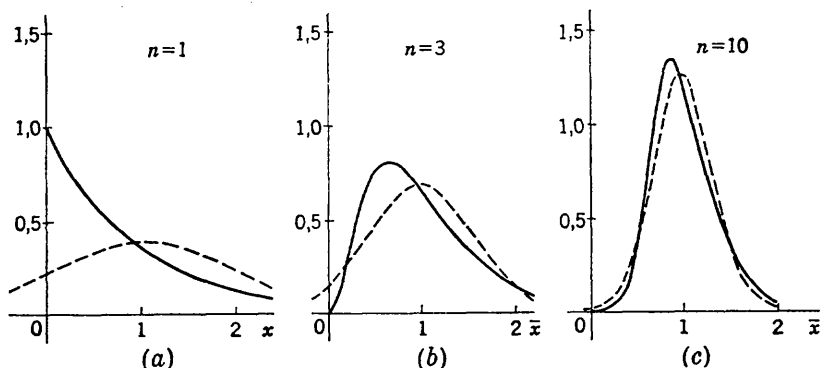


FIG. 7-3.

dientes aproximaciones normales: (a) da la distribución original, que corresponde a muestras de tamaño 1; (b) representa la distribución de las medias muestrales para $n=3$; (c) da la distribución de las medias muestrales para $n=10$. Las curvas exageran algo la aproximación a la distribución normal, ya que no muestran lo que ocurre en las ramas extremas de la distribución. Ordinariamente, las distribuciones de medias muestrales se aproximan a la normal con bastante rapidez al aumentar el tamaño de la muestra, considerando la región próxima a la media, pero más lentamente para los puntos que se alejan de esta. En general, cuanto mayor sea la distancia de un punto a la media, más lentamente se aproxima la distribución dada a la normal.

El teorema central del límite se aplica tanto a distribuciones discretas como a distribuciones continuas. Las funciones generatrices de momentos que hemos utilizado en esta sección podrían haber correspondido a distribuciones discretas y el razonamiento hubiera sido el mismo, sin más que sustituir las integrales por sumas. En la sección siguiente investigaremos la naturaleza de esta aproximación para un caso particular de distribución discreta.

7-7. Aproximación normal a la distribución binomial.—Consideremos la función

$$f(x) = p^x q^{1-x} \quad x = 0, 1 \quad (1)$$

en la cual

$$\mu = p \quad \sigma^2 = pq \quad (2)$$

y supongamos una muestra $x_1, x_2, \dots, x_n$, de tamaño n , de esta distribución. La muestra será simplemente una sucesión de ceros y unos, designando por 1 un acierto y por 0 un fallo. Y

$$\bar{x} = \frac{1}{n} \sum x_i$$

es la proporción de aciertos en la muestra. Hemos visto que la media y la varianza de $\bar{x}$ son

$$E(\bar{x}) = \mu = p \quad (3)$$

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n} = \frac{pq}{n} \quad (4)$$

La distribución de $\bar{x}$ es discreta; en realidad, $\bar{x}$ solo puede tomar los valores

$$0, \frac{1}{n}, \frac{2}{n}, \dots, \frac{j}{n}, \dots, 1$$

y sabemos que la distribución de $j = n\bar{x}$ es

$$\binom{n}{j} p^j q^{n-j} \quad j = 0, 1, 2, \dots, n \quad (5)$$

Así, pues, la distribución de $\bar{x}$ es

$$h(\bar{x}) = \binom{n}{n\bar{x}} p^{n\bar{x}} q^{n(1-\bar{x})} \quad \bar{x} = 0, \frac{1}{n}, \frac{2}{n}, \dots, 1 \quad (6)$$

En la figura 7-4 puede verse cómo esta distribución discreta se aproxima a una continua.

Supongamos contruidos rectángulos de altura $h(\bar{x})$ y anchura $1/n$, siendo j/n ($j=0, 1, 2, \dots, n$) los puntos medios de las bases. La parte superior de estos rectángulos forma una quebrada que podemos representar por $g(\bar{x})$. Puesto que $\sum h(\bar{x}) = 1$, el área limitada por $g(\bar{x})$ será $1/n$. Es evidente que

$$P\left(\frac{a}{n} \leq \bar{x} \leq \frac{b}{n}\right) = n \int_{(a-\frac{1}{2})/n}^{(b+\frac{1}{2})/n} g(\bar{x}) d\bar{x} \quad (7)$$

para números enteros cualesquiera a y b ($b > a$) del recorrido de j , ya que la integral es sencillamente el área limitada por la parte

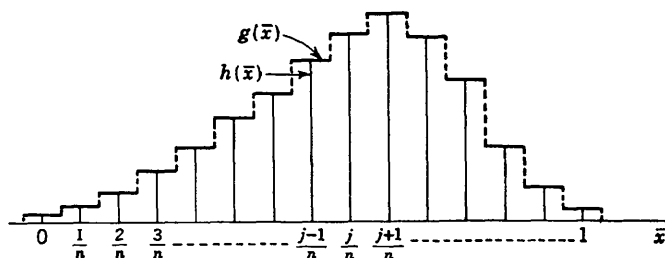


FIG. 7-4.

superior de los rectángulos construidos sobre los puntos a hasta b y, por tanto, vale

$$\sum_{\bar{x}=a/n}^{b/n} h(\bar{x}) \frac{1}{n} = \frac{1}{n} \sum_{j=a}^b \binom{n}{j} p^j q^{n-j} \quad (8)$$

Al aumentar n , disminuye la anchura de los rectángulos y los escalones de la función $ng(\bar{x})$ se aproximan entre sí, tomando un aspecto parecido al de la función de la figura 7-5. La aproximación normal a la distribución binomial puede considerarse como la forma límite de esta quebrada cuando n tiende a infinito.

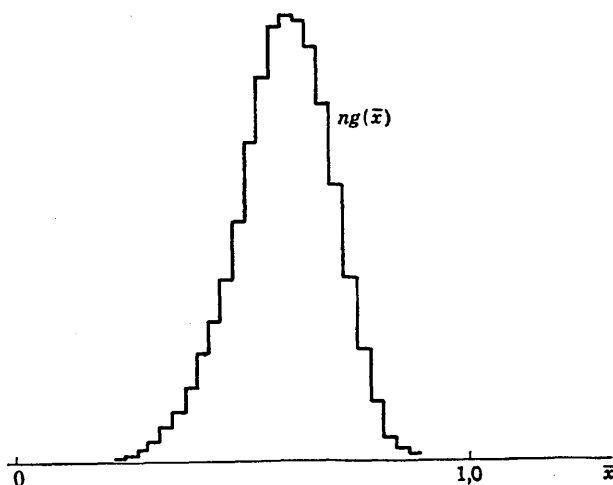


FIG. 7-5.

Esta aproximación normal tiene interés especial por proporcionar un método para calcular fácilmente el valor aproximado de sumas de la distribución binomial. Supongamos, p. ej., el lanzamiento de un dado bien construido, en el que se consideran como aciertos la aparición de un 1 o de un 2. En este caso $p=1/3$, $q=2/3$. Para una muestra de 300 pruebas, la distribución del número total de aciertos, j , viene dada por

$$f(j) = \binom{300}{j} \left(\frac{1}{3}\right)^j \left(\frac{2}{3}\right)^{300-j} \quad j=0, 1, \dots, 300$$

Supongamos que quisiéramos conocer la probabilidad de que el número de aciertos no se desvíe de 100 en más de 15; tendríamos que sumar $f(j)$ sobre el recorrido de 85 a 115, lo que da lugar a un cálculo bastante tedioso. Es posible obtener una aproximación de esta suma basándose en que

$$P(85 \leq j \leq 115) = P\left(\frac{85}{300} \leq \frac{j}{300} \leq \frac{115}{300}\right)$$

y puesto que $\bar{x} = j/300$ tiene una distribución aproximadamente normal con media $1/3$ y varianza $1/3 \times 2/3 \times 1/300$, tenemos

$$\begin{aligned} P(85/300 \leq \bar{x} \leq 115/300) &\cong \int_{85/300}^{115/300} n(\bar{x}; 1/3, 2/2700) d\bar{x} \\ &\cong \int_{85/300}^{115/300} \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{2/2700}} \exp\left[-\frac{1/2(\bar{x} - 1/3)^2}{2/2700}\right] d\bar{x} \end{aligned}$$

y haciendo $t = (\bar{x} - 1/3)/\sqrt{2/2700}$, tenemos

$$P\left(\frac{85}{300} \leq \bar{x} \leq \frac{115}{300}\right) \cong \int_{-1,84}^{1,84} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$

ya que

$$\frac{85/300 - 1/3}{\sqrt{2/2700}} \cong -1,84 \quad \frac{115/300 - 1/3}{\sqrt{2/2700}} \cong 1,84$$

Utilizando tablas de la distribución normal, obtenemos

$$P(85 \leq j \leq 115) \cong 0,934 \quad (9)$$

La aproximación podría todavía mejorarse poniendo $85 - 1/2$ y $115 + 1/2$ al calcular los límites de la integral, según se indica en (7).

En general, para la distribución binomial resulta ahora evidente que

$$P(a \leq j \leq b) = \sum_{j=a}^b \binom{n}{j} p^j q^{n-j} \quad (10)$$

$$\cong \int_{a'}^{b'} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt \quad (11)$$

en donde

$$a' = \frac{[(a - 1/2)/n] - p}{\sqrt{pq/n}} \quad b' = \frac{[(b + 1/2)/n] - p}{\sqrt{pq/n}} \quad (12)$$

Un análisis más detallado mostraría que el error de esta aproximación es inferior a

$$\frac{0,15}{\sqrt{npq}} \quad (13)$$

supuesto $npq > 25$. Así, en el ejemplo anterior el error máximo viene medido por

$$\frac{0,15}{\sqrt{300 \times 1/3 \times 2/3}} = 0,018$$

de modo que la aproximación (9) no llega a tener dos cifras decimales exactas, si se tiene en cuenta (13). Uspensky (en su *Introduction to Mathematical Probability*, Cap. VII, McGraw-Hill Book Company, Inc., Nueva York, 1937) proporciona algunas aproximaciones de mayor precisión.

7-8. Papel de la distribución normal en estadística.—En los capítulos siguientes veremos que la distribución normal desempeña un papel predominante. Claro es que el teorema central del límite nos asegura ya que así habría de ocurrir, pero es que además existen otras razones casi de la misma importancia.

En primer lugar, muchas de las poblaciones que se encuentran en investigaciones realizadas en muy diversos campos parecen tener una distribución bastante aproximada a la normal. Se ha dicho muchas veces que este fenómeno es perfectamente razonable si se tiene en cuenta el teorema central del límite. Podemos considerar como ilustración los disparos contra un cierto blanco: la trayectoria del proyectil está afectada por muchísimos factores, cada uno de los cuales ejerce un efecto pequeño; la desviación efec-

tiva es el resultado de todos estos factores. Supongamos que el resultado o efecto de cada factor es una observación procedente de una población determinada; el efecto total será esencialmente la media de un conjunto de observaciones procedentes de un conjunto de poblaciones. Puesto que las desviaciones efectivamente observadas son de la naturaleza de las medias, es de esperar que su distribución sea aproximadamente normal. No queremos indicar con esto que la mayor parte de las distribuciones encontradas en la práctica sean normales, porque no es este el caso en absoluto, pero sí que es muy frecuente encontrar distribuciones aproximadas a la normal.

Otra consideración a favor de la distribución normal es el hecho de que las distribuciones en el muestreo basadas en una distribución madre que sea normal, resultan bastante cómodas desde el punto de vista analítico. Al hacer inferencias sobre poblaciones a partir de muestras procedentes de las mismas, es necesario conocer las distribuciones de varias funciones de las observaciones muestrales. El problema matemático de obtención de estas distribuciones suele ser más sencillo cuando las muestras proceden de una población normal que en otro caso.

Al aplicar métodos estadísticos basados en la distribución normal, el experimentador deberá conocer, al menos aproximadamente, la forma general de distribución que siguen sus datos. Si esta es normal, podrá usar directamente los métodos que hemos mencionado; en caso contrario, transformará los datos de modo que las observaciones transformadas sigan la distribución normal. Cuando el experimentador desconozca la forma de distribución poblacional, deberá usar otros métodos de análisis más generales y que suelen resultar menos eficaces, llamados métodos *no paramétricos*. En el capítulo final del libro nos ocuparemos de algunos de estos métodos.

PROBLEMAS

1. Cítese un ejemplo en el que la población objetivo y la población muestreada coincidan.
2. Dése un ejemplo en el que la población objetivo y la población muestreada no sean la misma.
3. Una compañía fabrica transistores en tres plantas diferentes, A, B y C, cuyos métodos de fabricación son muy similares. Se decide inspeccionar los transistores fabricados en la A, por ser esta la mayor y disponerse en ella de estadísticos. Para inspeccionar la producción semanal, se seleccionan al azar 100 transistores y se comprueban sus defectos. Defínase la población muestreada y la población objetivo.

4. En el problema 3 se decide utilizar los resultados obtenidos en la planta A para sacar conclusiones acerca de las plantas B y C. Defínase la población objetivo.

5. Sea (x_1, x_2) una variable aleatoria bidimensional con distribución conjunta $p^{x_1+x_2}q^{2-x_1-x_2}$, $x_1=0, 1$; $x_2=0, 1$.

a) Defínense los puntos (x_1, x_2) del espacio bidimensional que toma la variable aleatoria.

b) Hágase $x_1=y-x_2$ y escríbanse los valores que puede tomar la variable aleatoria (y, x_2) .

6. En el problema 5, hállese la distribución conjunta de y y x_2 , dando esta en forma de tabla de doble entrada.

7. Hállese la distribución marginal de y , a partir de los resultados del problema anterior.

8. ¿Cuál es la probabilidad de que las dos observaciones de una muestra de tamaño dos, de una población con distribución rectangular sobre el intervalo unitario, no difieran en más de un medio?

9. ¿Cuál es la probabilidad de que la media de una muestra de tamaño dos, de una distribución rectangular (sobre el intervalo unitario), esté comprendida entre $1/4$ y $3/4$?

10. Se extraen bolas, con reemplazamiento, de una urna que contiene una bola blanca y dos negras. Representemos por $x=0$ una bola blanca, y por $x=1$ una bola negra. Para muestras $x_1, x_2, \dots, x_9$, de tamaño nueve, ¿cuál es la distribución conjunta de las observaciones? ¿Y la distribución de la suma de las observaciones?

11. Con referencia al problema 10, hállese los valores esperados de la media muestral y de la varianza muestral.

12. Para muestras de tamaño dos de una población con varianza σ^2 , demuéstrese que el valor esperado de la varianza muestral es $\sigma^2/2$.

13. Generalícese el resultado del problema 12 a muestras de tamaño n .

14. ¿Cuál es el valor de y que hace mínimo $\sum_1^n (x_i - y)^2$?

15. Si

$$\bar{x} = (1/n) \sum_1^n x_i,$$

demuéstrese que

$$\sum_1^n (x_i - \mu)^2 = \sum_1^n (x_i - \bar{x})^2 + n(\bar{x} - \mu)^2$$

Utilizando este resultado y el del problema 14, explíquese por qué el valor esperado de la varianza muestral no coincide con la varianza poblacional.

16. Hállese $E(m_3)$ para muestras de tamaño dos de una población con momento de tercer orden finito, donde

$$m_3 = \left(\frac{1}{n} \right) \sum_{i=1}^n (x_i - \bar{x})^3.$$

17. Demuéstrese que $E[(1/n)\sum(x_i - \mu)^r] = \mu_r$ para muestras de tamaño n de una población con media μ y momento r -ésimo μ_r .

18. Utilícese la desigualdad de Tchebysheff para hallar cuántas veces debe lanzarse una moneda para que la probabilidad de que $\bar{x}$ esté entre 0,4 y 0,6 sea, al menos, 0,90 (suponiendo que la moneda esté bien construida).

19. ¿Cómo podría determinarse con más precisión el número necesario de tiradas en el problema 18?; esto es, ¿cómo podría hacerse que la probabilidad se aproximara mucho a 0,90? ¿Cuál sería el número de tiradas?

20. Si una población tiene $\sigma=2$, y $\bar{x}$ es la media de muestras de tamaño 100, hállese los límites entre los cuales estará situado $\bar{x} - \mu$ con probabilidad 0,90. Utilícense la desigualdad de Tchebysheff y el teorema central del límite; ¿por qué difieren los dos resultados?

21. Supongamos que $\bar{x}_1$ y $\bar{x}_2$ son medias de dos muestras de tamaño n de una población con varianza σ^2 . Determinése n de modo que la probabilidad de que las dos medias muestrales difieran en un valor superior a σ sea, aproximadamente, 0,01. (Considérese la variante $y = \bar{x}_1 - \bar{x}_2$.)

22. Supongamos que las lámparas fabricadas mediante un determinado proceso tienen una vida media de 2000 horas con desviación estándar de 250 horas. Supongamos también que se considera aconsejable sustituir el proceso si la vida media puede aumentarse al menos en un 10%. Un ingeniero desea poner a prueba un nuevo proceso, admitiendo que la desviación estándar de la distribución de vidas de lámparas es aproximadamente la misma que para el proceso considerado al principio. ¿Qué tamaño de muestra deberá examinar si quiere que la probabilidad de no adoptar el nuevo proceso sea 0,01, aproximadamente, cuando con él se obtienen en efecto lámparas con una vida media de 2250 horas?

23. Un investigador quiere estimar la media de una población utilizando una muestra suficientemente grande para que la probabilidad de que la media muestral no difiera de la media poblacional en más del 25% de la desviación estándar sea 0,95. ¿Qué tamaño deberá adoptar para la muestra?

24. Un instituto de opinión pública quiere obtener una muestra de votantes de un cierto Estado, suficientemente grande para que la probabilidad de que la proporción obtenida a favor de un cierto candidato resulte inferior al 50%, siendo la verdadera del 52%, sea solo 0,01. ¿Qué tamaño deberá adoptar para la muestra?

25. Se sabe que una droga de uso establecido, utilizada en el tratamiento de infecciones, resulta eficaz en el 80% de los casos. Se ha obser-

vado que una nueva droga ha resultado eficaz en 85 de los primeros 100 casos en que se ha utilizado. ¿Ha quedado bien establecida la superioridad de la nueva droga? (Si la nueva droga fuese tan eficaz como la antigua, ¿cuál sería la probabilidad de obtener 85 ó más éxitos en una muestra de 100?)

26. Una urna contiene 5 fichas numeradas del 1 al 5. Se dice que una muestra sin reemplazamiento de esta población finita es aleatoria si todos los pares posibles formados con las cinco fichas tienen igual probabilidad de extracción. ¿Cuál es el valor esperado de la media muestral? ¿Cuál es la varianza de la media muestral?

27. Suponiendo que las dos fichas del problema 26 se extrajeran con reemplazamiento, ¿cuál sería la varianza de la media muestral? ¿Por qué podría conjeturarse que esta varianza será superior a la anterior?

28. Si una distribución dada por $f(x)$ tiene por función generatriz de momentos $m(t)$, demuéstrese que la media de las muestras de tamaño n tiene como función generatriz de momentos $[m(t/n)]^n$.

29. Utilícese el resultado del problema 28 para ver que la media y la varianza de la media muestral son μ y σ^2/n .

30. Hállese el tercer momento respecto a la media de la media muestral, para muestras de tamaño n de una población binomial. Demuéstrese que este se acerca a cero al crecer n (según debe ocurrir si es válida la aproximación normal).

31. Supongamos que la distribución de la vida de una pieza de una cierta máquina viene dada por $0,01e^{-0,01t}$, en donde t está medido en días. La máquina viene provista de un repuesto. ¿Cuál es la distribución de la vida combinada de la pieza y de su repuesto?

32. Generalícese el problema 26 considerando N fichas y muestras de tamaño n . La varianza de la media muestral es

$$\frac{\sigma^2}{n} \frac{N-n}{N-1}$$

siendo σ^2 la varianza poblacional,

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N \left(i - \frac{N+1}{2} \right)^2$$

BIBLIOGRAFIA

1. BROSS, I. D. J.: *Design for Decision*, The Macmillan Company, Nueva York, 1953.
2. BRUNK, H.: *An Introduction to Mathematical Statistics*, Ginn & Company, Boston, 1960.
3. COCHRAN, W.: *Sampling Techniques*, John Wiley & Sons, Inc., Nueva York, 1953.

4. DIXON, W. J., y F. J. MASSEY, Jr.: *Introduction to Statistical Analysis*, McGraw-Hill Book Company, Inc., Nueva York, 1957.
5. FELLER, W.: *An Introduction to Probability Theory and Its Applications*, 2.^a ed., vol. I, John Wiley & Sons, Inc., Nueva York, 1957.
6. FRASER, D. A. S.: *Statistics—An Introduction*, John Wiley & Sons, Inc., Nueva York, 1958.
7. HOGG, R., y A. CRAIG: *Introduction to Mathematical Statistics*, The Macmillan Company, Nueva York, 1959.
8. NEYMAN, J.: *First Course in Probability and Statistics*, Holt, Rinehart y Winston, Nueva York, 1950.
9. WALLIS, W. A., y H. V. ROBERTS: *Statistics—A New Approach*, Free Press, Glencoe, Ill., 1956.
10. WILKS, S. S.: *Elementary Statistical Analysis*, Princeton University Press, Princeton, N. J., 1948.
11. YOUTEN, W. J.: *Statistical Methods for Chemists*, John Wiley & Sons, Inc., Nueva York, 1951.

CAPITULO 8

ESTIMACION PUNTUAL

8-1. Teoría de la decisión.—Estudiaremos en este capítulo uno de los más importantes problemas de la estadística, la estimación puntual, pero en primer lugar expondremos el problema general de la decisión. Imaginemos que un estadístico o un científico tiene que seleccionar una acción entre cierto número de acciones posibles. Supongamos también que la acción apropiada depende de un parámetro desconocido θ , que determina la distribución $f(x; \theta)$ de la población que va a ser muestreada. Si θ fuera conocido, se conocería la función de densidad y, por tanto, la acción apropiada.

El método a emplear consiste en seleccionar una muestra aleatoria $x_1, x_2, \dots, x_n$ de $f(x; \theta)$ y tomar una decisión sobre la acción basándose en los valores de la muestra. Se procede de la siguiente forma:

1. Se define el conjunto de *todos* los valores posibles que puede tomar θ en el problema. Tal conjunto de valores de θ es el espacio paramétrico y se designa por Ω . (En este libro Ω será un intervalo o un número finito de puntos.)

2. Se define el conjunto de *todas* las acciones posibles que pueden adoptarse en el problema particular. Este conjunto se llama espacio de *acción* (o espacio de decisión) y se designa por A . A veces, estas acciones se denominan *acciones finales*.

3. Se selecciona una función d de la muestra aleatoria; p. ej., sea

$$a = d(x_1, x_2, \dots, x_n)$$

donde a pertenece al espacio de acción A . Es decir, planeamos o decidimos tomar la acción a , donde $a = d(x_1, x_2, \dots, x_n)$, si observamos $x_1, x_2, \dots, x_n$. La función d se llama *función de decisión* o estrategia. Daremos algunos ejemplos que ayuden a aclarar estas ideas.

Ejemplo 8-1.—Sea $f(x; \theta)$ una densidad normal con media μ y varianza 1. Por tanto, θ es el parámetro μ , y supongamos que μ puede ser cualquier número real. Entonces, el espacio paramétrico Ω es el eje real; es decir,

$$\Omega = \{\mu: -\infty < \mu < \infty\}$$

Imaginemos que existe una acción distinta para cada μ . Podemos

representar una acción por una *estimación* $\hat{\mu}$ de μ , por lo que el espacio de acción A es también el eje real, o sea

$$A = \{\hat{\mu} : -\infty < \hat{\mu} < \infty\}$$

Como función de decisión supongamos que utilizamos d_1 que asigna $\hat{\mu} = \bar{x}$. Esto quiere decir que se selecciona una muestra aleatoria $x_1, x_2, \dots, x_n$ de una distribución normal y se toma la acción $\hat{\mu} = \bar{x} = d_1(x_1, x_2, \dots, x_n)$ como estimación de μ . Podrían utilizarse otras funciones de decisión. Así, p. ej., cabría emplear d_2 , definida por

$$d_2(x_1, x_2, \dots, x_n) = \sum_{i=1}^n x_i^2$$

o d_3 , definida por

$$d_3(x_1, x_2, \dots, x_n) = \sum_{i=1}^n x_i^5$$

o muchas otras funciones. Naturalmente, desearemos utilizar una función de decisión que dé «buenas» estimaciones, cuestión que veremos más adelante. Estos tipos de problemas de decisión se llaman problemas de estimación puntual y se analizarán con detalle en este capítulo.

Ejemplo 8-2.—Sea $f(x; \mu)$ la distribución normal con media μ y varianza 1. Supongamos que queremos tomar una de las dos acciones posibles siguientes: la acción 1, designada por a_1 , será la afirmación de que $\mu < 0$; la acción 2, representada por a_2 , será la afirmación de que $\mu \geq 0$. El espacio paramétrico es el mismo que en el ejemplo 8-1; es decir,

$$\Omega = \{\mu : -\infty < \mu < \infty\}$$

pero el espacio de acción A está formado solo por dos elementos:

$$A = \{a : a = a_1, a_2\}$$

Supongamos que nuestra función de decisión d es tal que se toma la acción a_1 si $\bar{x} < 0$ y la acción a_2 si $\bar{x} \geq 0$. Así, se observa una muestra aleatoria $x_1, x_2, \dots, x_n$, se calcula $\bar{x}$, y se hace la afirmación $\mu < 0$ si $\bar{x} < 0$ y la afirmación $\mu \geq 0$ si $\bar{x} \geq 0$. Subrayemos de nuevo la posibilidad de utilizar otras funciones de decisión. Expondremos más adelante criterios que dan «buenas» funciones de decisión para este problema. Estos tipos de problemas de decisión se denominan problemas de docimasia de hipótesis y se estudiarán en el capítulo 12.

Ejemplo 8-3.—Un florista tiene un almacén lleno de semillas de una cierta clase de plantas. Sabe que no todas las semillas se desarrollarán cuando se planten y desea hacer una estimación de la proporción de semillas que germinarán. El florista realiza un experimento mediante el que selecciona aleatoriamente 1000 semillas del almacén, las pone en una estufa de germinación y, basándose en el resultado del experimento, toma una decisión sobre el valor de p , fracción de la población (semillas del almacén) que germinará. El espacio paramétrico es ahora

$$\Omega = \{p: 0 \leq p \leq 1\}$$

y el espacio de acción,

$$A = \{\hat{p}: 0 \leq \hat{p} \leq 1\}$$

donde $\hat{p}$ es la estimación de p . Sea

$$f(x; p) = p^x(1-p)^{1-x}, \quad x=0, 1,$$

la función de densidad de la población de semillas. Sea $x_1, x_2, \dots, x_{1000}$ una muestra aleatoria de tamaño 1000 de esta población, tal que $x_i=1$ si la semilla i -ésima germina. Una función de decisión que podría utilizarse en este problema es la definida por $\hat{p} = \bar{x}$ = la fracción de las 1000 semillas que germina.

Puesto que en cada problema particular existen muchas funciones diferentes que cabe utilizar como funciones de decisión, se hace necesario desarrollar una teoría que permita evaluar las funciones de decisión y seleccionar las adecuadas. Para evaluar una función de decisión, es necesario primero evaluar las consecuencias de las acciones finales. Para ello introducimos una *función de pérdida*, representada por $l(a; \theta)$. Esta es una función real no negativa que refleja la pérdida al tomar la acción a cuando es θ el parámetro. Esta pérdida es cero si a es la acción óptima para θ . En un problema estadístico en el que no se conoce θ , no es posible determinar la mejor acción. En tal caso aplicamos una estrategia d que conduce a la acción

$$a = d(x_1, x_2, \dots, x_n)$$

y a la pérdida correspondiente

$$l(a; \theta) = l[d(x_1, x_2, \dots, x_n); \theta]$$

que depende de los valores muestrales; es decir, aun cuando las funciones l y d están determinadas, en la práctica la pérdida depende de las observaciones particulares x_i obtenidas en la muestra aleatoria.

Ya que la pérdida es aleatoria, definiremos el *riesgo*, esto es, el valor esperado de la función de pérdida. El riesgo para la función de decisión d cuando el parámetro es θ se representa por $R(d; \theta)$ y depende de la función de decisión d , de la pérdida l y del valor del parámetro θ . No depende, en cambio, de la muestra aleatoria particular elegida. Por tanto, tenemos

$$\begin{aligned} R(d; \theta) &= E[l(a; \theta)] \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} l[d(x_1, x_2, \dots, x_n); \theta] f(x_1; \theta) f(x_2; \theta) \\ &\quad \dots f(x_n; \theta) dx_1 \dots dx_n \end{aligned}$$

Entonces una «buena» función de decisión es toda aquella que hace mínimo, de alguna manera, el riesgo para cada valor de θ del espacio paramétrico Ω .

Una de las dificultades del empleo de la teoría de la decisión en problemas de aplicación estriba en la determinación de una función de pérdida realista. Es imposible, p. ej., determinar con exactitud la pérdida en que se incurre al tomar una decisión errónea para sostener una hipótesis científica o al fallar en aconsejar una nueva variedad de trigo si esta es realmente superior a la que corrientemente se planta. En algunos problemas de juegos la función de pérdida podría reflejar la pérdida numérica real, pero aun entonces es discutible que la esperanza matemática de la pérdida sea una medida apropiada de las pérdidas aleatorias cuando el problema estadístico se afronta solo una vez. Estas dificultades pueden resolverse parcialmente como sigue: Primero, la experiencia con problemas estadísticos demuestra que los «buenos» procedimientos son insensibles a pequeños cambios de la función de pérdida, especialmente si se dispone de datos considerables. Por tanto, no son absolutamente necesarios valores precisos de la pérdida. Segundo, se puede medir la pérdida aleatoria tomando esperanzas cuando estas pérdidas están medidas en términos de una *función de utilidad*. Para el análisis de este concepto, remitimos al lector a la bibliografía.

Los tipos de problemas que estudiaremos son los siguientes, donde $f(x; \theta)$ es una función de densidad con parámetro θ (el símbolo θ puede representar más de un parámetro):

1. Estimación puntual de θ , cuestión que trataremos en este capítulo.
2. Docimasia de hipótesis sobre θ ; de ello nos ocuparemos en el capítulo 12.

A continuación damos algunos ejemplos de funciones de pérdida y de riesgo para ciertos problemas particulares; después precisaremos estos conceptos en definiciones formales.

Ejemplo 8-4.—Consideremos la distribución dada en el ejemplo 8-1. Hallemos una estimación de la media μ . Supongamos que la función de pérdida en este problema particular está dada por

$$l(a; \mu) = (a - \mu)^2$$

Si utilizamos la función de decisión d , definida por

$$a = \bar{x} = d(x_1, x_2, \dots, x_n),$$

la pérdida vendrá dada por

$$l(\bar{x}; \mu) = (\bar{x} - \mu)^2$$

En este caso, el riesgo es

$$R(d; \mu) = E[(\bar{x} - \mu)^2] = \frac{1}{n}$$

Otras funciones de pérdida apropiadas serían:

$$l(a; \mu) = |a - \mu|$$

o

$$l(a; \mu) = (a - \mu)^4$$

u otras muchas.

Ejemplo 8-5.—Sea x una variable aleatoria distribuida según $n(x; 0, \sigma^2)$, donde σ^2 puede tomar cualquier valor mayor que cero; es decir, $\Omega = \{\sigma^2: \sigma^2 > 0\}$. Se trata en este caso de hallar una estimación de σ^2 . El espacio de acción A es $A = \{\hat{\sigma}^2: \hat{\sigma}^2 > 0\}$. Supongamos que $x_1, x_2, \dots, x_n$ es una muestra aleatoria procedente de $n(x; 0, \sigma^2)$ y que se utiliza la función de decisión definida por

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2$$

Una típica función de pérdida sería $l(\hat{\sigma}^2; \sigma^2) = (\hat{\sigma}^2 - \sigma^2)^2$ o bien cabría emplear $(\log \hat{\sigma}^2 / \sigma^2)^2$, o muchas otras.

Más adelante expondremos una función de pérdida que resulta adecuada en muchos problemas de estimación puntual, pero ahora vamos a precisar las definiciones de funciones de pérdida y riesgo.

Definición 8-1. Funciones de pérdida.—Sea θ un parámetro cuyos valores pertenecen al espacio paramétrico Ω , y a una acción final que puede tomar valores del espacio de acción A . Entonces l es una función de pérdida si es una función real de a y θ que satisface las dos condiciones siguientes:

a) $l(a; \theta) \geq 0$ para todo a de A y todo θ de Ω .

b) Para cada θ de Ω existe al menos un a de A para el cual $l(a; \theta) = 0$.

Aquellos valores de a para los cuales la pérdida es cero constituyen la decisión o acción correcta cuando el parámetro es θ . Las dos condiciones anteriores establecen que la función de pérdida tiene que ser no negativa; la pérdida es igual a cero si se toma la decisión correcta.

Definición 8-2. Función de riesgo.—La función de riesgo (llamada algunas veces simplemente «el riesgo») para la función de decisión d y la función de pérdida l es

$$\begin{aligned} R(d; \theta) &= E[l(a; \theta)] \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} l[d(x_1, x_2, \dots, x_n); \theta] f(x_1; \theta) \\ &\quad \dots f(x_n; \theta) dx_1 \dots dx_n \end{aligned}$$

si $a = d(x_1, x_2, \dots, x_n)$ es la acción resultante al aplicar la estrategia d a la muestra aleatoria $x_1, x_2, \dots, x_n$ de una población con densidad $f(x; \theta)$.

Evidentemente, el riesgo depende de la función de decisión d que elijamos para el problema, de la función de pérdida l y del parámetro desconocido θ .

8-2. Estimación puntual.—Si la acción representada por $\hat{\theta}$ consiste en operar con $\hat{\theta}$ como si fuese θ , será $l(\hat{\theta}; \theta) = 0$ para $\hat{\theta} = \theta$. Supongamos que hemos dado una función de pérdida l tal que $l(\hat{\theta}; \theta) = 0$ si y solamente si $\hat{\theta} = \theta$. Esta función de pérdida define un problema de *estimación puntual*. La función de decisión d , donde

$$\hat{\theta} = d(x_1, x_2, \dots, x_n),$$

se llama entonces un *estimador* de θ . Al estadístico le interesa hallar un estimador con riesgo relativamente pequeño. En muchos problemas de decisión, pueden darse razones convincentes a favor del empleo de una función de pérdida de la forma

$$l(\hat{\theta}; \theta) = c(\theta) (\hat{\theta} - \theta)^2 \quad (1)$$

donde $c(\theta) > 0$ para todos los valores de θ . Esta función implica que la pérdida es mayor cuanto más alejado esté $\hat{\theta}$ del valor verdadero θ . Puede resultar muy difícil encontrar un $c(\theta)$ que haga realista esta pérdida, pero $c(\theta)$ desempeña un papel secundario en la determinación del mérito relativo de dos funciones de decisión.

Un argumento importante en favor de una función de pérdida de la forma (1) es que muchas funciones de θ cuyo valor mínimo es cero en $\theta = \theta$ pueden sustituirse, con aproximación aceptable, por funciones cuadráticas para θ próximo a θ . La experiencia ha demostrado que una función de pérdida de la forma dada en (1) resulta adecuada en muchos problemas de aplicación. En el resto de este capítulo supondremos funciones de pérdida de esta forma o con $c(\theta)=1$, en cuyo caso (1) recibe el nombre de función de pérdida de error cuadrático.

El problema que abordamos ahora es el de hallar una función de decisión apropiada d de la muestra aleatoria observada, $x_1, x_2, \dots, x_n$, procedente de la distribución $f(x; \theta)$, que haga mínima la función de riesgo

$$R(d; \theta) = E[c(\theta) (\hat{\theta} - \theta)^2] = c(\theta) E[(\hat{\theta} - \theta)^2] \quad (2)$$

En lo que queda de este capítulo emplearemos el término «estimadores» en lugar del de funciones de decisión, y la palabra «estimación» en vez del término «acción». Si es θ el parámetro que va a ser estimado, utilizaremos el símbolo $\hat{\theta}$ para representar la estimación (acción). Sin embargo, siempre que no haya peligro de ambigüedad, también denominaremos estimador a $\hat{\theta}$. Cuando llamemos estimador a $\hat{\theta}$ querremos decir la función de decisión d tal que $\hat{\theta} = d(x_1, x_2, \dots, x_n)$.

Supongamos que deseamos elegir entre dos estimadores

$$\hat{\theta}_1 = d_1(x_1, x_2, \dots, x_n)$$

y

$$\hat{\theta}_2 = d_2(x_1, x_2, \dots, x_n).$$

Queremos saber cómo se compara $R(d_1; \theta)$ con $R(d_2; \theta)$ para valores concretos de θ . Si

$$R(d_1; \theta) > R(d_2; \theta)$$

podemos también escribir

$$E[l(\hat{\theta}_1; \theta)] > E[l(\hat{\theta}_2; \theta)].$$

Si utilizamos la función de pérdida (1), tendremos

$$c(\theta) E[(\hat{\theta}_1 - \theta)^2] > c(\theta) E[(\hat{\theta}_2 - \theta)^2]$$

o bien, puesto que $c(\theta)$ es positivo,

$$E[(\hat{\theta}_1 - \theta)^2] > E[(\hat{\theta}_2 - \theta)^2].$$

Hemos demostrado así que, si se utiliza una función de pérdida de la forma (1), $\hat{\theta}_1$ tiene mayor riesgo que $\hat{\theta}_2$ si, y solamente si,

$$E[(\hat{\theta}_1 - \theta)^2] > E[(\hat{\theta}_2 - \theta)^2].$$

Para comparar dos estimadores

$$\hat{\theta}_1 = d_1(x_1, x_2, \dots, x_n)$$

y

$$\hat{\theta}_2 = d_2(x_1, x_2, \dots, x_n)$$

cuando la función de pérdida es de la forma dada en (1), la *eficiencia relativa* de $\hat{\theta}_1$ a $\hat{\theta}_2$, o más exactamente de d_1 a d_2 , es igual a

$$r(d_1, d_2) = \frac{R(d_2; \theta)}{R(d_1; \theta)}$$

Si $r(d_1, d_2)$ es mayor que 1 (para ciertos valores de θ ; p. ej., θ en una región C), parece razonable considerar que $\hat{\theta}_1$ es un estimador «mejor» que $\hat{\theta}_2$ (para θ en C). En resumen, si la función de pérdida que utilizamos es de la forma dada en (1), el problema de hallar un estimador con riesgo mínimo es equivalente al de hallar un estimador $\hat{\theta}$ que haga mínima la cantidad $E[(\hat{\theta} - \theta)^2]$, llamada *error cuadrático medio*.

Desgraciadamente, para la mayoría de las funciones de densidad $f(x; \theta)$, no existe un estimador que haga mínimo el error cuadrático medio para todos los valores de θ , sino que un estimador puede dar un error cuadrático medio mínimo para algunos valores de θ , y otro hacer mínimo dicho error para otros valores distintos de θ . Puesto que θ es desconocido, esta circunstancia complica el problema. Con todo esto queremos poner de manifiesto que, excepto en casos triviales, no existe un estimador cuyo error cuadrático medio sea mínimo para todos los valores de θ en Ω . Parece, por tanto, que es inevitable cierta arbitrariedad en la búsqueda de estimadores de riesgo mínimo. No obstante, utilizaremos como guía el error cuadrático medio. Así, p. ej., si $\hat{\theta}_1$, $\hat{\theta}_2$ y $\hat{\theta}_3$ son estimadores diferentes de θ , con densidades $g_1(\hat{\theta}; \theta)$, $g_2(\hat{\theta}; \theta)$ y $g_3(\hat{\theta}; \theta)$, tal como se representa en la figura 8-1, hablando en términos vulgares, parece que $\hat{\theta}_2$ es «mejor» estimador que $\hat{\theta}_1$ o $\hat{\theta}_3$, y $\hat{\theta}_3$ es «mejor» que $\hat{\theta}_1$.

Más adelante se citan otros criterios distintos del error cuadrático medio que parecen valiosos para aplicarlos a los estimadores. Examinaremos cada uno de ellos para mostrar en qué forma son útiles. En un problema de aplicación, no existe posiblemente un estimador con error cuadrático medio mínimo para todos los valores

de θ , de modo que el experimentador deberá seleccionar los estimadores que posean aquellas propiedades de las relacionadas a continuación que parezcan deseables para su situación particular.

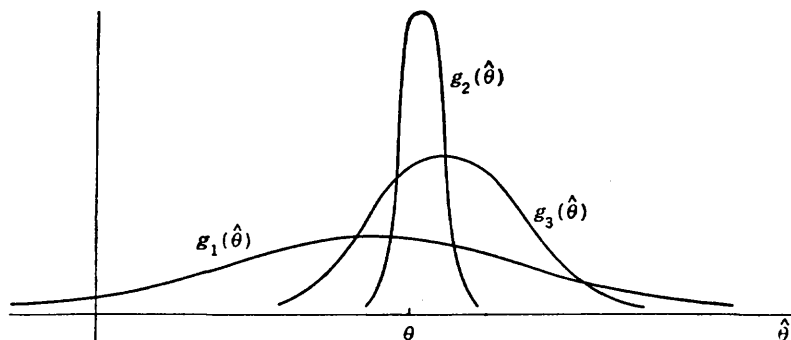


FIG. 8-1.

Relación de propiedades de un estimador

1. Insesgado.
2. Consistente o conciliable (consistencia simple y consistencia de error cuadrático).
3. Asintóticamente eficiente y óptimo asintóticamente normal.
4. Insesgado de varianza mínima.

Estas propiedades se analizarán oportunamente, pero veremos primero una importante propiedad de las funciones de densidad, la *suficiencia*.

8-3. Estadísticos suficientes; caso de un solo parámetro.—

En esta sección consideraremos una densidad $f(x; \theta)$, con un único parámetro θ perteneciente al conjunto Ω , y emplearemos como estimador de θ un estadístico o estadígrafo $\hat{\theta} = d(x_1, x_2, \dots, x_n)$ función de una muestra aleatoria de n valores $x_1, x_2, \dots, x_n$. Ahora bien, $\hat{\theta}$ es una variable aleatoria y, puesto que hemos comenzado con n variables aleatorias $x_1, x_2, \dots, x_n$ que han quedado reducidas a una única θ , mediante el empleo de $d(x_1, x_2, \dots, x_n)$, estamos interesados en ver si en este proceso de condensación se ha perdido «información». Así, p. ej., una posible elección de $d(x_1, x_2, \dots, x_n)$ es x_1 y tendríamos $\hat{\theta} = x_1$. En este caso es evidente que no se utiliza toda la «información» de la muestra.

En muchos casos, el estadístico $\hat{\theta}$ puede proporcionar toda la «información» que contiene la muestra acerca del parámetro θ . En-

tonces preferimos trabajar con $\hat{\theta}$ en lugar de hacerlo con n variables aleatorias $x_1, x_2, \dots, x_n$, por la sencilla razón de que es más fácil de manejar una sola variable aleatoria que n . Debemos examinar qué entendemos al decir que un estadístico contiene toda la «información» acerca de un parámetro incluida en los n valores muestrales.

La única «información» sobre θ en la densidad $f(x; \theta)$ está contenida en los valores muestrales $x_1, x_2, \dots, x_n$. No es posible obtener información acerca de θ por muestreo de una densidad $g(x)$ que no contenga a θ o a parámetros relacionados con θ . En otras palabras, si deseamos hallar la estatura media de los habitantes de Nueva York, no cabe muestrear peces del lago Ontario o extraer una muestra de una tabla de números aleatorios y esperar que con ello consigamos la información deseada.

En el caso de que exista un estadístico $\hat{\theta}$ que contenga toda la «información» sobre θ incluida en la muestra, $\hat{\theta}$ se llamará un estadístico suficiente. Puntualizaremos ahora su definición.

Definición 8-3. Estadístico suficiente.—Sea $x_1, x_2, \dots, x_n$ una muestra aleatoria de la densidad $f(x; \theta)$, y $\hat{\theta} = d(x_1, x_2, \dots, x_n)$ un estadístico (función solo de las x_i). Sea $\theta^* = d^*(x_1, x_2, \dots, x_n)$ otro estadístico, que no es función de $\hat{\theta}$. Si, para cada uno de los estadísticos θ^* , la distribución condicional de θ^* , dado $\hat{\theta}$, no depende de θ , se dice que $\hat{\theta}$ es un estadístico suficiente para θ ; es decir, si $p(\theta^*|\hat{\theta})$ no contiene a θ , $\hat{\theta}$ es un estadístico suficiente para θ .

Sea $\hat{\theta}_1$ cualquier estadístico no función del estadístico suficiente $\hat{\theta}$. Por la definición de estadístico suficiente, la distribución condicional de $\hat{\theta}_1$, dado $\hat{\theta}$, no depende de θ . Por tanto, $\hat{\theta}_1$ no puede facilitar ninguna información sobre θ que no nos haya dado ya el estadístico suficiente. El concepto de estadístico suficiente fue estudiado por primera vez en una serie de trabajos de R. A. Fisher.

J. Neyman ha desarrollado un criterio relativamente fácil para examinar si un estadístico $\hat{\theta}$ es suficiente. En el teorema siguiente, cuya demostración se omite, se da este criterio.

Teorema 8-1.—Sea $x_1, x_2, \dots, x_n$ una muestra aleatoria de tamaño n de la densidad $f(x; \theta)$, $a < x < b$, donde a y b no dependen de θ , y sea

$$g(x_1, x_2, \dots, x_n; \theta) = f(x_1; \theta)f(x_2; \theta) \dots f(x_n; \theta)$$

la distribución conjunta de estas n variables aleatorias. Si esta función de densidad se descompone de la siguiente forma en producto de factores

$$g(x_1, x_2, \dots, x_n; \theta) = h(\theta; \theta)k(x_1, x_2, \dots, x_n)$$

donde $k(x_1, x_2, \dots, x_n)$ no depende del parámetro θ , $\hat{\theta}$ es en tal caso un estadístico suficiente para θ .

Este teorema proporciona un método relativamente fácil para juzgar si un estadístico es suficiente. Se habrá observado que nos hemos abstenido de emplear la denominación estimador suficiente, si bien se ha utilizado el término estadístico suficiente. Como dijimos anteriormente, se pretende solo reducir n variables aleatorias a una sola para facilitar su manejo; sin embargo, puesto que hemos demostrado que toda la «información» que está en la muestra se halla contenida en el estadístico suficiente, parece natural que deseemos emplear una función del estadístico suficiente como estimador de θ . Si utilizamos como estimador una función del estadístico suficiente, cabe preguntarse si esta función contiene también toda la «información» incluida en la muestra. La respuesta la da el siguiente teorema, que nos limitamos a enunciar.

Teorema 8-2.—Sea $\hat{\theta} = d(x_1, x_2, \dots, x_n)$ un estadístico suficiente para θ basado en una muestra aleatoria de tamaño n de la densidad $f(x; \theta)$. Si u es una función de θ que admite función inversa uniforme, entonces $\bar{\theta} = u(\hat{\theta})$ es también un estadístico suficiente para θ , y $\hat{\theta}$ es un estadístico suficiente para $u(\theta)$.

Daremos ahora algunos ejemplos con el fin de aclarar estos teoremas y definiciones.

Ejemplo 8-6.—Sea $x_1, x_2, \dots, x_n$ una muestra aleatoria de la distribución normal $n(x; \mu, 1)$, $-\infty < x < \infty$. La distribución conjunta de $x_1, x_2, \dots, x_n$ está dada por

$$\begin{aligned} g(x_1, x_2, \dots, x_n; \mu) &= n(x_1; \mu, 1) \cdot n(x_2; \mu, 1) \dots n(x_n; \mu, 1) \\ &= \frac{1}{(2\pi)^{\frac{1}{2}}} e^{-\frac{1}{2}(x_1 - \mu)^2} \cdot \frac{1}{(2\pi)^{\frac{1}{2}}} e^{-\frac{1}{2}(x_2 - \mu)^2} \dots \frac{1}{(2\pi)^{\frac{1}{2}}} e^{-\frac{1}{2}(x_n - \mu)^2} \\ &= \frac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2} \sum_{i=1}^n (x_i - \mu)^2} \quad (1) \end{aligned}$$

Pero

$$\begin{aligned} \sum (x_i - \mu)^2 &= \sum [(x_i - \bar{x}) - (\mu - \bar{x})]^2 \\ &= \sum (x_i - \bar{x})^2 + \sum (\bar{x} - \mu)^2 \\ &= \sum (x_i - \bar{x})^2 + n(\bar{x} - \mu)^2 \end{aligned}$$

Sustituyendo en (1):

$$\begin{aligned} g(x_1, x_2, \dots, x_n; \mu) &= \frac{1}{(2\pi)^{n/2}} \exp \left[-\frac{1}{2} n(\bar{x} - \mu)^2 \right] \\ &\quad \cdot \exp \left[-\frac{1}{2} \sum (x_i - \bar{x})^2 \right] \end{aligned}$$

Si hacemos $\hat{\theta} = \bar{x}$, se podrá utilizar el teorema 8-1 y se deduce que $\bar{x}$ es un estadístico suficiente para μ . Por el teorema 8-2, las cantidades $\bar{x} + 8$, $6\bar{x}$, $n\bar{x}$, $\Sigma(x_i - 4)$, etc., son también estadísticos suficientes para μ , y $\bar{x}$ es un estadístico suficiente para 16μ , $3\mu + 5$, etc.

Ejemplo 8-7.—Sea $x_1, x_2, \dots, x_n$ una muestra aleatoria procedente de $n(x; 0, \sigma^2)$. La distribución conjunta es

$$g(x_1, x_2, \dots, x_n; \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\frac{1}{2\sigma^2} \sum_i x_i^2}$$

Si hacemos

$$h(\hat{\sigma}^2; \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\frac{\hat{\sigma}^2}{2\sigma^2}}$$

y $k(x_1, x_2, \dots, x_n) = 1$, por el teorema 8-1, el estadístico $\hat{\sigma}^2 = \Sigma x_i^2$ es un estadístico suficiente para σ^2 . Por el teorema 8-2, cada una de las cantidades $\Sigma x_i^2/n$, $\Sigma x_i^2 + 8$, $1/(2 + \Sigma x_i^2)$, etc. es así mismo un estadístico suficiente para σ^2 , y $\hat{\sigma}^2$ es un estadístico suficiente para $3\sigma^2$, $\sigma^2 + 18$, etc.

8-4. Estadísticos suficientes; más de un parámetro.—En esta sección supondremos que $f(x; \theta_1, \theta_2, \dots, \theta_k)$ es una función de densidad con k parámetros desconocidos, y $x_1, x_2, \dots, x_n$ una muestra aleatoria de tamaño n . De nuevo estamos interesados en condensar las n variables aleatorias en menos de n estadísticos. En este caso, la función de densidad contiene k parámetros desconocidos y no es posible condensar las n variables aleatorias en un *único* estadístico y retener toda la «información». Sin embargo, sí es posible condensarlas en m estadísticos, donde $m < n$, sin pérdida de «información». Estos estadísticos forman lo que se llama un *conjunto* de estadísticos suficientes; la definición se da a continuación.

Definición 8-4. Estadísticos suficientes; caso multiparamétrico.—Sea $x_1, x_2, \dots, x_n$ una muestra aleatoria de tamaño n de la densidad

$$f(x; \theta_1, \theta_2, \dots, \theta_k)$$

y sean $\hat{\theta}_1 = d_1(x_1, x_2, \dots, x_n)$,

$$\hat{\theta}_2 = d_2(x_1, x_2, \dots, x_n), \dots, \hat{\theta}_m = d_m(x_1, x_2, \dots, x_n)$$

m estadísticos. Si la distribución condicional de $x_1, x_2, \dots, x_n$, dados $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_m$, es independiente de $\theta_1, \theta_2, \dots, \theta_k$, entonces $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_m$ constituye un conjunto de m estadísticos suficientes para los parámetros $\theta_1, \theta_2, \dots, \theta_k$.

El entero m puede ser menor, mayor o igual que k ; es decir, el número de estadísticos suficientes de un conjunto de estadísticos suficientes puede ser menor, mayor o igual que el número de parámetros de la función de densidad. En general, en este libro estudiaremos distribuciones en las que $m=k$. Obsérvese que siempre existe un conjunto de n estadísticos suficientes, ya que por el teorema 8-3, que enunciaremos más adelante, las n variables aleatorias de la muestra $x_1, x_2, \dots, x_n$ forman un conjunto suficiente. Naturalmente, en cualquier distribución existirán muchos conjuntos de estadísticos suficientes, pero siempre nos interesará el conjunto «más pequeño», en el sentido de que este conjunto sea una función de todos los otros conjuntos de estadísticos suficientes. Tal conjunto se dice que es un *conjunto mínimo* de estadísticos suficientes.

Como en el caso de un parámetro, también en el caso multiparamétrico disponemos de un criterio que ayuda a comprobar la suficiencia de un conjunto de estadísticos. Nos limitamos a enunciarlo sin dar la demostración.

Teorema 8-3. Sea $f(x; \theta_1, \theta_2, \dots, \theta_k)$, $a < x < b$, una función de densidad, donde a y b no dependen de las θ . Si la función de densidad conjunta de una muestra aleatoria de esta distribución puede descomponerse en la forma

$$g(x_1, \dots, x_n; \theta_1, \theta_2, \dots, \theta_k) = h(\theta_1, \dots, \theta_m; \theta_1, \dots, \theta_k) q(x_1, \dots, x_n)$$

donde $q(x_1, \dots, x_n)$ no contiene las θ_i , entonces $\hat{\theta}_1, \dots, \hat{\theta}_m$ es un conjunto de m estadísticos suficientes.

Teorema 8-4.—Sea $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_m$ un conjunto de estadísticos suficientes. Supongamos que una transformación biunívoca sobre estos m estadísticos define $\hat{\alpha}_1, \dots, \hat{\alpha}_m$. En tal hipótesis, el conjunto $\hat{\alpha}_1, \hat{\alpha}_2, \dots, \hat{\alpha}_m$ forma también un conjunto suficiente.

Ilustraremos los teoremas anteriores con un ejemplo.

Ejemplo 8-8.—Sea $x_1, x_2, \dots, x_n$ una muestra aleatoria de $n(x; \mu, \sigma^2)$, donde $-\infty < \mu < \infty$, $0 < \sigma^2 < \infty$. La función de densidad conjunta es

$$\begin{aligned} g(x_1, \dots, x_n; \mu, \sigma^2) &= n(x_1; \mu, \sigma^2) \dots n(x_n; \mu, \sigma^2) \\ &= \frac{1}{(\sigma\sqrt{2\pi})^n} \exp \left[-\frac{1}{2\sigma^2} \sum (x_i - \mu)^2 \right] \\ &= \frac{1}{(\sigma\sqrt{2\pi})^n} \exp \left\{ -\frac{1}{2\sigma^2} \left[\sum (x_i - \bar{x})^2 + n(\bar{x} - \mu)^2 \right] \right\} \\ &= \frac{1}{(\sigma\sqrt{2\pi})^n} \exp \left\{ -\frac{1}{2\sigma^2} [n-1]\hat{\sigma}^2 + n(\hat{\mu} - \mu)^2 \right\} \cdot 1 \\ &= h(\hat{\sigma}^2, \hat{\mu}; \sigma^2, \mu) \cdot q(x_1, \dots, x_n) \end{aligned}$$

donde

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2 \quad \hat{\mu} = \bar{x} \quad q(x_1, \dots, x_n) = 1$$

Por el teorema 8-3, $\hat{\sigma}^2$ y $\hat{\mu}$ son un conjunto de dos estadísticos suficientes. Por el teorema 8-4, $\sum(x_i - \bar{x})^2$ y $\sum x_i$ forman también un conjunto de estadísticos suficientes, como lo es el conjunto $\sum x_i^2$ y $\sum x_i$.

A continuación examinaremos las propiedades de los estimadores enunciadas en la sección 8-2.

8-5. Estimador insesgado.—El error cuadrático medio puede escribirse así

$$\begin{aligned} E[(\hat{\theta} - \theta)^2] &= E[\{\hat{\theta} - E(\hat{\theta})\} - \{\theta - E(\hat{\theta})\}]^2 \\ &= E[\hat{\theta} - E(\hat{\theta})]^2 + [\theta - E(\hat{\theta})]^2 \\ &= \text{var}(\hat{\theta}) + [\theta - E(\hat{\theta})]^2 \end{aligned} \quad (1)$$

Por tanto, el error cuadrático medio es la suma de dos cantidades no negativas. El término $\theta - E(\hat{\theta})$ se llama sesgo del estimador y puede ser positivo, negativo o cero.

Si es posible hallar un estimador con sesgo próximo a cero y tal que $\text{var}(\hat{\theta})$ sea pequeño, el error cuadrático medio será pequeño. Parece, por tanto, deseable disponer de un estimador cuyo sesgo sea cero. Definiremos a continuación el estimador insesgado.

Definición 8-5. Estimador insesgado.—Un estimador $\hat{\theta}$ se llama estimador insesgado de θ si el valor esperado de $\hat{\theta}$ es igual a θ ; es decir, si $E(\hat{\theta}) = \theta$ para todos los valores de θ en Ω .

Ejemplo 8-9.—Sea $x_1, x_2, \dots, x_n$ una muestra aleatoria de una densidad cuya media es μ . En tal caso $E(\bar{x}) = \mu$; así mismo $E(x_i) = \mu$ y, por tanto, $\bar{x}$ y x_i son estimadores insesgados de μ . Sea

$$y = \sum_{i=1}^n a_i x_i$$

donde las a_i son constantes tales que $\sum a_i = 1$. Entonces

$$\begin{aligned} E(y) &= E(\sum a_i x_i) \\ &= \sum a_i E(x_i) \\ &= \sum a_i \mu \\ &= \mu \end{aligned}$$

luego y es también un estimador insesgado de μ .

8-6. Estimador consistente.—Parece que un «buen» estimador sería aquel para el cual el riesgo disminuyese cuando crece el tamaño de la muestra. Es decir, el estimador es mejor cuando se basa en una muestra de veinte observaciones, p. ej., que si se basa solo en dos. Para ser más explícitos, supongamos que $\hat{\theta}_1 = d_1(x_1)$ es un estimador de θ basado en una muestra de tamaño 1 de $f(x; \theta)$; sea $\hat{\theta}_2 = d_2(x_1, x_2)$ un estimador de θ que se basa en una muestra de tamaño 2; y, en general, representemos por $\hat{\theta}_n = d_n(x_1, x_2, \dots, x_n)$ un estimador de θ basado en una muestra de tamaño n . Es decir, sea $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_n, \dots$ una sucesión de estimadores de θ , que a veces escribiremos en la forma abreviada $\{\hat{\theta}_n\}$ o $\{d_n\}$, indistintamente. La condición sugerida anteriormente es que

$$\lim_{n \rightarrow \infty} R(\theta; d_n) = 0 \quad \text{para todo } \theta \text{ de } \Omega.$$

En el caso de pérdida medida por el error cuadrático, la definición formal es la siguiente:

Definición 8-6 a. Consistencia en error cuadrático.—Sea $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_n, \dots$ una sucesión de estimadores de θ (con más precisión, sea $d_1, d_2, \dots, d_n, \dots$ una sucesión de estimadores para θ). Esta sucesión es un estimador consistente o conciliable en error cuadrático de θ si

$$\lim_{n \rightarrow \infty} E[(\hat{\theta}_n - \theta)^2] = 0 \quad \text{para todo } \theta \text{ de } \Omega$$

Puesto que

$$R(\theta; d_n) = E[(\hat{\theta}_n - \theta)^2] = \text{var}(\hat{\theta}_n) + [\theta - E(\hat{\theta}_n)]^2$$

la consistencia en error cuadrático implica que tanto el sesgo como la varianza de $\hat{\theta}_n$ tienden a cero. Como en la deducción de la ley de los grandes números, la aplicación de la desigualdad de Tchebysheff demostraría que si el sesgo y la varianza de $\hat{\theta}_n$ tienden a cero, queda satisfecha la siguiente definición de *consistencia* (simple).

Definición 8-6 b. Consistencia.—Sea $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_n, \dots$ una sucesión de estimadores de θ . Esta sucesión es un estimador consistente simple de θ si para todo $\epsilon > 0$ se verifica que

$$\lim_{n \rightarrow \infty} P(\theta - \epsilon < \hat{\theta}_n < \theta + \epsilon) = 1 \quad \text{para todo } \theta \text{ de } \Omega$$

La condición de consistencia (en lo sucesivo emplearemos el término *consistencia* para designar la *consistencia simple*), que es más débil que la *consistencia en error cuadrático*, establece que para

muestras grandes $\hat{\theta}_n$ tiende a aproximarse a θ . (Algunas veces diremos que $\hat{\theta}$ es consistente para indicar que lo es $\{\hat{\theta}_n\}$.)

Ya que el error cuadrático no es necesariamente una buena aproximación de la pérdida cuando $\hat{\theta}$ se halla alejado de θ , puede avanzarse que la consistencia simple es frecuentemente más fundamental que la consistencia en error cuadrático.

8-7. Estimadores asintóticamente eficientes.—Hemos dicho anteriormente que en general no existen estimadores que hagan mínimo el error cuadrático medio para todo θ . Esto se debe al hecho de que el estimador $\hat{\theta} = \theta_0$, que se forma prescindiendo de los datos y conjeturando que θ es igual a θ_0 , es mejor que cualquier otro estimador si θ es realmente igual a θ_0 , y es bastante bueno cuando θ está próximo a θ_0 . La existencia de estimadores consistentes implica que el recorrido de valores de θ para los cuales nuestra conjetura es «mejor» que un estimador consistente disminuye a medida que crece el tamaño de la muestra. Esto sugiere la posibilidad de que existan estimadores que sean eficientes en un sentido límite apropiado cuando $n \rightarrow \infty$.

Para el caso de pérdida medida por el error cuadrático, damos la siguiente definición.

Definición 8-7 a. Estimadores asintóticamente eficientes en error cuadrático.—Sea $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_n, \dots$ un estimador consistente en error cuadrático de θ . Se dice que esta sucesión es un estimador asintóticamente eficiente en error cuadrático (en lo sucesivo utilizaremos generalmente el término estimador eficiente) si no existe otro estimador consistente en error cuadrático $\theta_1^*, \theta_2^*, \dots, \theta_n^*, \dots$ para el cual

$$\lim_{n \rightarrow \infty} \frac{E[(\hat{\theta}_n - \theta)^2]}{E[(\theta_n^* - \theta)^2]} > 1$$

para todo θ de algún intervalo abierto.

El límite superior $c = \overline{\lim}_{n \rightarrow \infty} c_n$ es el mayor número (puede ser $+\infty$) tal que para todo $\epsilon > 0$, $c_n \geq c - \epsilon$ para infinitos n .

Es interesante observar que, en muchos problemas, existen estimadores que satisfacen la definición anterior. Por razones técnicas relacionadas con las que nos llevaron a la definición de consistencia (simple), encontraremos algunas veces más apropiada otra definición que implica la distribución límite de $\sqrt{n}(\hat{\theta}_n - \theta)$. A este fin, fijaremos nuestra atención en estimadores $\hat{\theta}_n$ para los cuales $\sqrt{n}(\hat{\theta}_n - \theta)$ tiene una distribución límite normal, con media 0 y

varianza $\sigma^2(\theta)$. El símbolo $\sigma^2(\theta)$ indica que la varianza depende de θ .

Definición 8-7 b. Estimadores óptimos asintóticamente normales (estimadores OAN).—*La sucesión de estimadores $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_n, \dots$ es un estimador óptimo asintóticamente normal (OAN) de θ si se satisfacen las tres condiciones siguientes:*

a) *La distribución de $\sqrt{n}(\hat{\theta}_n - \theta)$ tiende a la distribución normal con media 0 y varianza $\sigma^2(\theta)$ cuando n tiende a infinito.*

b) *Para todo $\epsilon > 0$,*

$$\lim_{n \rightarrow \infty} P\{|\hat{\theta}_n - \theta| > \epsilon\} = 0 \quad \text{para todo } \theta \text{ de } \Omega$$

c) *No existe otra sucesión de estimadores consistentes $\theta_1^*, \theta_2^*, \dots, \theta_n^*, \dots$ para la cual la distribución de $\sqrt{n}(\theta_n^* - \theta)$ tienda a la distribución normal con media cero y varianza $\sigma^{*2}(\theta)$, y tal que*

$$\frac{\sigma^2(\theta)}{\sigma^{*2}(\theta)} > 1$$

para todo θ de algún intervalo abierto.

La utilidad de esta definición se deriva en parte de los teoremas que prueban la existencia de estimadores OAN y del hecho de que muchos estimadores razonables se distribuyen de una forma aproximadamente normal.

Puede demostrarse que, para muestras extraídas de una densidad normal con media μ y varianza σ^2 , la sucesión

$$\hat{\theta}_n = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}_n \quad \text{para } n=1, 2, \dots$$

es un estimador eficiente y OAN de μ . Naturalmente, la distribución límite de $\sqrt{n}(\bar{x}_n - \mu)$ es normal con media cero y varianza σ^2 , y ningún otro estimador tiene varianza límite más pequeña cualquiera que sea el intervalo de valores de μ . Sin embargo, existen otros muchos estimadores para este problema que son también estimadores eficientes y OAN de μ ; es decir, estimadores con la misma distribución normal en el límite. Así, p. ej.,

$$y_n = \frac{1}{n+1} \sum_{i=1}^n x_i \quad n=1, 2, \dots$$

es un estimador eficiente y OAN de μ . Los estimadores eficientes y óptimos asintóticamente normales son necesariamente consistentes.

8-8. Estimadores insesgados de varianza mínima.—Puesto que raramente existen estimadores con error cuadrático medio mínimo, una forma razonable de proceder consiste en restringir la clase de funciones de estimación y buscar estimadores con error cuadrático medio mínimo en la clase restringida. Así, p. ej., podemos considerar solo estimadores insesgados y, dentro de la clase de estimadores insesgados, ver si es posible encontrar un estimador con error cuadrático medio mínimo. Según (8-5-1), el error cuadrático medio puede escribirse así

$$\begin{aligned} E\{(\hat{\theta} - \theta)^2\} &= E\{[\hat{\theta} - E(\hat{\theta})]^2 + [\theta - E(\hat{\theta})]^2\} \\ &= \text{var}(\hat{\theta}) + [\theta - E(\hat{\theta})]^2 \end{aligned} \quad (1)$$

Si únicamente consideramos estimadores insesgados, $E(\hat{\theta}) = \theta$ y vemos que el error cuadrático medio de un estimador insesgado es igual a la varianza del estimador; por tanto, (1) puede escribirse

$$E\{(\hat{\theta} - \theta)^2\} = \text{var}(\hat{\theta})$$

si $\hat{\theta}$ es insesgado. Precisaremos la definición.

Definición 8-8. Estimadores insesgados de varianza mínima.

Sea $x_1, x_2, \dots, x_n$ una muestra aleatoria de $f(x; \theta)$, y designemos por $\hat{\theta} = d(x_1, x_2, \dots, x_n)$ un estimador de θ tal que:

- a) $E(\hat{\theta}) = \theta$; es decir, $\hat{\theta}$ es insesgado.
- b) $\text{var}(\hat{\theta})$ es menor que la varianza de cualquier otro estimador que satisfaga a); es decir, $\text{var}(\hat{\theta})$ es menor que la varianza de cualquier otro estimador insesgado.

En tales condiciones, $\hat{\theta}$ es el estimador insesgado de varianza mínima de θ .

Si en vez de θ , estamos interesados en estimar una función de θ , p. ej., $u(\theta)$, buscamos un estimador insesgado de $u(\theta)$ con varianza mínima. Pueden encontrarse estimadores insesgados de varianza mínima para muchos de los parámetros que desempeñan un papel importante en estadística aplicada.

Los estimadores consistentes, OAN y eficientes gozan de propiedades óptimas cuando el tamaño de la muestra es grande, pero la propiedad de insesgado de varianza mínima es óptima para cualquier tamaño de la muestra. Aunque los estimadores insesgados de varianza mínima no son tan deseables como los de error cuadrático medio mínimo, existen frecuentemente, mientras que es

rara la existencia de estimadores de error cuadrático medio mínimo.

Demostraremos que los conceptos de estadístico suficiente y de funciones de densidad completas nos ayudarán mucho a encontrar estimadores incesgados de varianza mínima. Hablando en términos generales, un estimador incesgado que sea función del estadístico suficiente tiene varianza más pequeña que un estimador incesgado no basado en un estadístico suficiente. Representemos por $f(x; \theta)$ una densidad y supongamos que queremos estimar $u(\theta)$. Sea $t=t(x_1, x_2, \dots, x_n)$ un estimador incesgado de $u(\theta)$, y $\hat{\theta}=d(x_1, \dots, x_n)$, un estadístico suficiente para θ . Puede demostrarse que existe una función $v(\hat{\theta})$ de $\hat{\theta}$ tal que $E[v(\hat{\theta})]=u(\theta)$; es decir, existe una función del estadístico suficiente $\hat{\theta}$ que es así mismo un estimador incesgado de $u(\theta)$. También la varianza de $v(\hat{\theta})$ es menor o igual que la varianza de t . Por tanto, en nuestra búsqueda de estimadores incesgados de varianza mínima, solo necesitamos considerar aquellos estimadores que sean funciones de estadísticos suficientes. Formalizaremos estas ideas en el teorema siguiente.

Teorema 8-5.—Sea $x_1, x_2, \dots, x_n$ una muestra aleatoria de la densidad $f(x; \theta)$, y $\hat{\theta}=d(x_1, x_2, \dots, x_n)$, un estadístico suficiente para θ . Sea, además, el estadístico $t=t(x_1, x_2, \dots, x_n)$ un estimador incesgado de $u(\theta)$ y supongamos que t no es función de $\hat{\theta}$. Designemos por $v(\hat{\theta})$ la esperanza condicional de t , dado $\hat{\theta}$; es decir, sea $E(t|\hat{\theta})=v(\hat{\theta})$. Entonces:

a) $E[v(\hat{\theta})]=u(\theta)$; es decir, si la esperanza de t es igual a $u(\theta)$, la esperanza de $v(\hat{\theta})$ es también igual a $u(\theta)$.

b) $\text{var}[v(\hat{\theta})] < \text{var}(t)$.

c) $v(\hat{\theta})$ no contiene a θ y, por tanto, es un estadístico y puede calcularse a partir de los valores muestrales observados.

Demostración.—Se tiene (Cap. 5) que $E(t|\hat{\theta})$ es función de $\hat{\theta}$, la cual se ha designado por $v(\hat{\theta})$. En primer lugar, demostraremos la parte c) del teorema. Representemos por $g(t, \hat{\theta}; \theta)$ la función de densidad conjunta de t y $\hat{\theta}$, y por $p(t|\hat{\theta})$, la distribución condicional de t , dado $\hat{\theta}$. Por la definición de estadístico suficiente, $p(t|\hat{\theta})$ no depende de θ ; por tanto, el valor esperado de t , dado $\hat{\theta}$, que es $v(\hat{\theta})$, no puede depender de θ . Demostraremos ahora la parte a). Puesto que t es un estimador incesgado de $u(\theta)$, se tiene

$$E(t) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} t \cdot g(t, \hat{\theta}; \theta) dt d\hat{\theta} = u(\theta) \quad (2)$$

Pero podemos escribir

$$g(t, \hat{\theta}; \theta) = p(t|\hat{\theta})h(\hat{\theta}; \theta) \quad (3)$$

donde $h(\theta; \theta)$ es la densidad marginal de $\hat{\theta}$. Sustituyendo (3) en (2)

$$\begin{aligned} u(\theta) = E(\mathbf{t}) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} t \cdot g(t, \theta; \theta) dt d\theta = \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} t \cdot p(t|\theta) h(\theta; \theta) dt d\theta = \\ &= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} t p(t|\theta) dt \right] h(\theta; \theta) d\theta \end{aligned} \quad (4)$$

Pero la integral entre corchetes, es decir,

$$\int_{-\infty}^{\infty} t p(t|\theta) dt = v(\theta) \quad (5)$$

es, por definición, igual a $E(\mathbf{t}|\theta)$ y es función solo de θ ; no depende de θ puesto que $p(t|\theta)$ no contiene a θ . Sustituyendo (5) en (4), se obtiene

$$u(\theta) = E(\mathbf{t}) = \int_{-\infty}^{\infty} v(\theta) h(\theta; \theta) d\theta = E[v(\hat{\theta})]$$

y la parte a) queda demostrada. Para probar la parte b) tenemos

$$\begin{aligned} \text{var}(\mathbf{t}) &= E[\mathbf{t} - u(\theta)]^2 = E\{[\mathbf{t} - v(\hat{\theta})] + [v(\hat{\theta}) - u(\theta)]\}^2 \\ &= E[v(\hat{\theta}) - u(\theta)]^2 + E[\mathbf{t} - v(\hat{\theta})]^2 + 2E\{[\mathbf{t} - v(\hat{\theta})][v(\hat{\theta}) - u(\theta)]\} \end{aligned} \quad (6)$$

Demostraremos ahora que el tercer término de (6) es igual a cero. Resulta

$$\begin{aligned} E\{[\mathbf{t} - v(\hat{\theta})][v(\hat{\theta}) - u(\theta)]\} &= \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [t - v(\theta)][v(\theta) - u(\theta)] g(t, \theta, \theta) dt d\theta \\ &= \int_{-\infty}^{\infty} \left\{ \int_{-\infty}^{\infty} [t - v(\theta)] p(t|\theta) dt \right\} [v(\theta) - u(\theta)] h(\theta; \theta) d\theta \end{aligned}$$

Pero, por (5), el término entre llaves es igual a 0. Queda demostrado que (6) se reduce a

$$\text{var}(\mathbf{t}) = \text{var}[v(\hat{\theta})] + E[\mathbf{t} - v(\hat{\theta})]^2$$

o bien

$$\text{var}(\mathbf{t}) > \text{var}[v(\hat{\theta})]$$

puesto que

$$E[t - v(\hat{\theta})]^2 > 0$$

lo que prueba la parte b).

Las ideas del teorema 8-5 fueron examinadas por vez primera por C. R. Rao y D. Blackwell.

Este teorema representa una gran ayuda en nuestra búsqueda de estimadores ineseados de varianza mínima, puesto que esencialmente nos dice que solo necesitamos considerar los estimadores ineseados que sean funciones de estadísticos suficientes. Sin embargo, el teorema no proporciona la solución definitiva, ya que es posible que existan muchos estimadores basados en estadísticos suficientes que sean ineseados. El teorema no dice cuál de estos tiene varianza mínima. Sin embargo, el teorema 8-6 puede utilizarse a menudo. Este teorema asegura que si la densidad del estadístico suficiente es *completa*, existe únicamente *un* estimador ineseado de $u(\theta)$ basado en el estadístico suficiente. Si solo hay un estimador ineseado basado en el estadístico suficiente, él es *el* estimador ineseado de varianza mínima.

Teorema 8-6.—Sea $x_1, x_2, \dots, x_n$ una muestra aleatoria de $f(x; \theta)$, y $\hat{\theta} = d(x_1, x_2, \dots, x_n)$, un estadístico suficiente. Supongamos que la función de densidad de $\hat{\theta}$ es completa. Si existe una función $v(\hat{\theta})$ de $\hat{\theta}$ tal que $E[v(\hat{\theta})] = u(\theta)$, $v(\hat{\theta})$ es entonces el estimador ineseado de varianza mínima de $u(\theta)$.

Ejemplo 8-10.—Sea $x_1, x_2, \dots, x_n$ una muestra aleatoria de $n(x; \mu, 1)$; $\bar{x}$ es un estadístico suficiente para μ y la función de densidad de $\bar{x}$ es completa. Supóngase que deseamos hallar estimadores ineseados de varianza mínima para a) μ ; b) μ^2 ; c) 3μ ; d) $\mu^2 + 6\mu$. Por el teorema 8-6, si podemos hallar en cada caso a), b), c), d) un estimador ineseado que sea función del estadístico suficiente de densidad completa $\bar{x}$, este será el estimador ineseado de varianza mínima. El lector comprobará que los siguientes estimadores son ineseados: a) $\bar{x}$; b) $\bar{x}^2 - 1/n$; c) $3\bar{x}$; d) $\bar{x}^2 + 6\bar{x} - 1/n$.

En el caso de densidades con más de un parámetro, son válidos los teoremas sobre estimadores ineseados de varianza mínima análogos a los anteriores.

Ejemplo 8-11.—Sea $x_1, x_2, \dots, x_n$ una muestra aleatoria de $n(x; \mu, \sigma^2)$; $\bar{x}$ y $\sum x_i^2$ forman un conjunto de estadísticos suficientes y su densidad conjunta es completa. Supongamos que queremos hallar estimadores ineseados de varianza mínima para: a) μ ; b) σ^2 ; c) $\mu + \sigma^2$; d) $6\mu + 8\sigma^2$; por la generalización del teorema 8-6 al caso multiparamétrico, bastará hallar estimadores ineseados, basados en los estadísticos suficientes, de la función de los parámetros en a),

b), c) y d), los cuales son los estimadores *insesgados de varianza mínima*. El lector comprobará que los siguientes son estimadores insesgados para a), b), c) y d):

$$a) \quad \bar{x};$$

$$b) \quad \frac{1}{n-1} \left(\sum x_i^2 - n\bar{x}^2 \right);$$

$$c) \quad \bar{x} + \frac{1}{n-1} \sum x_i^2 - \frac{n}{n-1} \bar{x}^2;$$

$$d) \quad 6\bar{x} + \frac{8}{n-1} \sum x_i^2 - \frac{8n}{n-1} \bar{x}^2.$$

8-9. Principio de máxima verosimilitud.—Aunque un experimentador decide sobre qué propiedades desea que posea un estimador, tiene que enfrentarse con el problema de cómo obtener dicho estimador. Existen varios métodos que proporcionan estimadores con algunas de las distintas propiedades consideradas en secciones anteriores. Pero aquí, los únicos que estudiaremos son: 1) el *método de máxima verosimilitud*; 2) el *método de los momentos*; 3) el *método de Bayes*. En un capítulo posterior nos ocuparemos de otro método que se utiliza en ciertos tipos de problemas de estimación: el *método de los mínimos cuadrados*. En esta sección vamos a exponer el principio de máxima verosimilitud. Como introducción consideraremos un problema muy sencillo de estimación. Supongamos que una urna contiene cierto número de bolas negras y blancas, y admitamos que se sabe que la razón de los números de ambos colores es de tres a uno, pero no se conoce si las más numerosas son las bolas blancas o las negras. Esto es, la probabilidad de sacar una bola negra es $1/4$ ó $3/4$. Si se extraen n bolas de la urna, con reemplazamiento, la distribución del número de bolas negras x es la binomial:

$$f(x; p) = \binom{n}{x} p^x q^{n-x} \quad x=0, 1, 2, \dots, n \quad (1)$$

$$p = 1/4, 3/4$$

donde $q=1-p$ y p es la probabilidad de extraer una bola negra.

Extraeremos una muestra de tres bolas con reemplazamiento, tratando de estimar el parámetro desconocido p de la distribución. El problema de estimación resulta especialmente sencillo en este caso, ya que solo tenemos que elegir entre los números 0,25 y 0,75. Vamos a anticipar el resultado de la extracción de la muestra. Damos a

continuación los números que pueden obtenerse y sus probabilidades en las dos posibilidades consideradas:

x	0	1	2	3
$f\left(x; \frac{3}{4}\right)$	$\frac{1}{64}$	$\frac{9}{64}$	$\frac{27}{64}$	$\frac{27}{64}$
$f\left(x; \frac{1}{4}\right)$	$\frac{27}{64}$	$\frac{27}{64}$	$\frac{9}{64}$	$\frac{1}{64}$

En este ejemplo, si obtenemos $x=0$ en una muestra de tamaño 3, preferiremos la estimación 0,25 a la 0,75 para p , por ser la probabilidad $27/64$ superior a $1/64$; esto es, porque es más probable obtener una muestra con $x=0$ si la población tiene $p=1/4$, que si tiene $p=3/4$. En general, tomaremos como estimación de p 0,25 si $x=0$ ó 1, y 0,75 si $x=2$ ó 3. El estimador puede definirse así:

$$\begin{aligned}\hat{p}(x) &= 0,25 & x=0, 1 \\ &= 0,75 & x=2, 3\end{aligned}\quad (2)$$

Por tanto, se toma como estimación para cada x el valor de p , o sea $\hat{p}$, tal que

$$f(x; \hat{p}) > f(x; p')$$

siendo p' el otro valor posible de p .

Con mayor generalidad, si fuesen posibles varios valores de p , podríamos proceder razonablemente del mismo modo. Así, si obtenemos $x=6$ en una muestra de 25 de una población binomial, sustituiremos todos los valores posibles de p en la expresión

$$f(6; p) = \binom{25}{6} p^6 (1-p)^{19} \quad 0 \leq p \leq 1 \quad (3)$$

eligiendo como estimación el valor de p que hace máxima $f(6; p)$. Para los valores posibles de p , nuestra estimación sería $6/25$. La posición de su valor máximo se halla igualando a cero su derivada respecto a p , y resolviendo la ecuación en p resultante. Así se obtiene

$$\frac{d}{dp} f(6; p) = \binom{25}{6} p^5 (1-p)^{18} [6(1-p) - 19p] \quad (4)$$

que igualada a cero y resuelta respecto a p proporciona como raíces: $p=0$; 1; $6/25$. Las dos primeras dan un mínimo, y la estimación es, por tanto, $\hat{p}=6/25$. Esta estimación cumple la propiedad

$$f(6; \hat{p}) > f(6; p') \quad (5)$$

en donde p' es cualquier otro valor de p del intervalo $0 \leq p \leq 1$.

Definición 8-9. Función de verosimilitud.—La función de verosimilitud de n variables aleatorias $x_1, x_2, \dots, x_n$ es la densidad conjunta de las n variables $g(x_1, x_2, \dots, x_n; \theta)$, la cual se supone también función de θ . En particular, si $x_1, x_2, \dots, x_n$ es una muestra aleatoria de la densidad $f(x; \theta)$, la función de verosimilitud es $g(x_1, x_2, \dots, x_n; \theta) = f(x_1; \theta) f(x_2; \theta) \dots f(x_n; \theta)$.

La función de verosimilitud $g(x_1, x_2, \dots, x_n; \theta)$ da la verosimilitud relativa de que las variables aleatorias tomen un valor particular $x_1, x_2, \dots, x_n$. Supongamos por un momento que se conoce θ , y designemos por θ_0 su valor. El conjunto de valores particulares de las variables aleatorias «más verosímiles» es el $x'_1, x'_2, \dots, x'_n$ tal que $g(x_1, x_2, \dots, x_n; \theta_0)$ sea máximo. Así, p. ej., si por sencillez admitimos que $n=1$ y que x_1 tiene la distribución normal con media 6 y varianza 1, el valor de la variable aleatoria «más verosímil que ocurra» es $x_1=6$. Entendemos por valor de x_1 «más verosímil que ocurra» el x'_1 tal que $n(x'_1; 6, 1) > n(x_1; 6, 1)$. Supongamos ahora que la densidad conjunta de n variables aleatorias es $g(x_1, x_2, \dots, x_n; \theta)$, en donde θ es desconocido. Representemos por $x'_1, x'_2, \dots, x'_n$ los valores particulares observados. Deseamos saber de qué densidad es «más verosímil» que proceda este conjunto particular de valores. Expresado de otra forma: cuando θ toma los diferentes valores en Ω , queda definida una familia de densidades. Queremos conocer qué densidad (qué valor de θ) da la mayor verosimilitud de obtener el conjunto $x'_1, x'_2, \dots, x'_n$. En otras palabras, queremos hallar el valor de θ en Ω , representado por $\hat{\theta}$, que hace máxima la función de verosimilitud $g(x_1, x_2, \dots, x_n; \theta)$. El valor $\hat{\theta}$ que hace máxima la función de verosimilitud es, en general, una función de $x_1, x_2, \dots, x_n$; es decir, $\hat{\theta} = d(x_1, x_2, \dots, x_n)$. Cuando así sucede, se dice que $\hat{\theta} = d(x_1, x_2, \dots, x_n)$ es el estimador máximo-verosímil de θ . (Suponemos que existe el máximo de la función de verosimilitud.) Precisamos a continuación la definición de estimador máximo-verosímil.

Definición 8-10. Estimador máximo-verosímil.—Sea

$$L(\theta) = g(x_1, x_2, \dots, x_n; \theta)$$

la función de verosimilitud para las variables aleatorias $x_1, x_2, \dots, x_n$. Si $\hat{\theta}$ [donde $\hat{\theta} = d(x_1, x_2, \dots, x_n)$] es el valor de θ en Ω que hace má-

xima $L(\theta)$, $\hat{\theta}$, o más exactamente, $d(x_1, x_2, \dots, x_n)$, es entonces el estimador máximo-verosímil de θ .

Los casos más importantes que consideraremos son aquellos en los que $x_1, x_2, \dots, x_n$ es una muestra aleatoria de una densidad $f(x; \theta)$, y, por tanto, la función de verosimilitud es

$$L(\theta) = f(x_1; \theta)f(x_2; \theta) \dots f(x_n; \theta)$$

Muchas funciones de verosimilitud satisfacen condiciones de regularidad, de modo que el estimador máximo-verosímil es la solución de la ecuación

$$\frac{dL(\theta)}{d\theta} = 0$$

$L(\theta)$ y $\log [L(\theta)]$ toman su máximo para el mismo valor de θ , y algunas veces resulta más fácil hallar el máximo del logaritmo de la verosimilitud.

Si la función de verosimilitud contiene k parámetros, es decir, si

$$L(\theta_1, \theta_2, \dots, \theta_k) = \prod_{i=1}^n f(x_i; \theta_1, \theta_2, \dots, \theta_k)$$

los estimadores máximo-verosímiles de los parámetros $\theta_1, \theta_2, \dots, \theta_k$ son las variables aleatorias $\hat{\theta}_1 = d_1(x_1, x_2, \dots, x_n)$; $\hat{\theta}_2 = d_2(x_1, x_2, \dots, x_n)$; $\dots$; $\hat{\theta}_k = d_k(x_1, x_2, \dots, x_n)$, donde $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_k$ son los valores en Ω que hacen máxima a $L(\theta_1, \theta_2, \dots, \theta_k)$.

Si se satisfacen ciertas condiciones de regularidad, el punto en que la verosimilitud es máxima es una solución del sistema de k ecuaciones

$$\frac{\partial L(\theta_1, \dots, \theta_k)}{\partial \theta_1} = 0$$

$$\frac{\partial L(\theta_1, \dots, \theta_k)}{\partial \theta_2} = 0$$

$$\dots \dots \dots$$

$$\frac{\partial L(\theta_1, \dots, \theta_k)}{\partial \theta_k} = 0$$

También en este caso puede ser más fácil trabajar con el logaritmo de la verosimilitud.

En la sección siguiente aclararemos estas definiciones con algunos ejemplos.

8-10. Algunos estimadores máximo-verosímiles.—En esta sección obtendremos estimadores máximo-verosímiles de los parámetros de algunas de las distribuciones más corrientes. Frecuentemente, la función de verosimilitud satisface condiciones de regularidad, de modo que el valor máximo se obtiene igualando a cero las derivadas de la función de verosimilitud y resolviendo el sistema de ecuaciones resultantes respecto a los parámetros.

Ejemplo 8-12.—Supongamos que se extrae una muestra aleatoria de tamaño n de la distribución binomial puntual

$$f(x; p) = p^x q^{1-x} \quad x=0, 1; \quad 0 \leq p \leq 1 \quad (1)$$

Los valores muestrales, $x_1, x_2, \dots, x_n$, serán una sucesión de ceros y unos, y la verosimilitud es

$$L(p) = \prod_{i=1}^n p^{x_i} q^{1-x_i} = p^{\sum x_i} q^{n - \sum x_i} \quad (2)$$

y haciendo

$$y = \sum x_i \quad (3)$$

tenemos:

$$L^* = \log L(p) = y \log p + (n - y) \log q \quad (4)$$

$$\frac{dL^*}{dp} = \frac{y}{p} - \frac{n - y}{q} \quad (5)$$

recordando que $q = 1 - p$. Si se iguala esta expresión a cero y se despeja p , obtenemos el estimador

$$\hat{p} = \frac{y}{n} = \frac{1}{n} \sum x_i = \bar{x} \quad (6)$$

que es, naturalmente, la expresión que cabía esperar para el estimador de este parámetro.

Demostraremos que este estimador es suficiente y, puesto que es también insesgado y completo, será el único estimador insesgado de varianza mínima de p . Necesitamos demostrar que la distribución condicional de las x_i , dada $\bar{x}$, es independiente de p . Puesto que la distribución marginal de $n\bar{x} = y$ viene dada por

$$\binom{n}{y} p^y q^{n-y} \quad (7)$$

la distribución condicional de las x_i , dado y , se logra dividiendo (2) por (7), con lo que se obtiene

$$g(x_1, x_2, \dots, x_n | \hat{p}) = \frac{1}{\binom{n}{n\hat{p}}} \quad x_i = 0, 1; \quad \sum x_i = n\hat{p}, \quad (8)$$

distribución que es independiente del parámetro p .

Ejemplo 8-13.—Una muestra aleatoria de tamaño n , procedente de la distribución normal, tiene la densidad

El logaritmo de la verosimilitud es

$$L^* = -\frac{n}{2} \log 2\pi - \frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum (x_i - \mu)^2 \quad (10)$$

Para hallar la posición del máximo, calculamos

$$\frac{\partial L^*}{\partial \mu} = \frac{1}{\sigma^2} \sum (x_i - \mu) \quad (11)$$

$$\frac{\partial L^*}{\partial \sigma^2} = -\frac{n}{2} \frac{1}{\sigma^2} + \frac{1}{2\sigma^4} \sum (x_i - \mu)^2 \quad (12)$$

e igualando a cero estas derivadas y resolviendo las ecuaciones que resultan respecto a μ y a σ^2 , hallamos los estimadores

$$\hat{\mu} = \frac{1}{n} \sum x_i = \bar{x} \quad (13)$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum (x_i - \bar{x})^2 \quad (14)$$

que son los momentos muestrales correspondientes a μ y σ^2 . El estimador $\hat{\mu}$ es insesgado, pero $\hat{\sigma}^2$ no lo es, ya que

$$E(\hat{\sigma}^2) = \frac{n-1}{n} \sigma^2 \quad (15)$$

Este par de estimadores es suficiente para los parámetros; la distribución muestral para valores dados de $\hat{\mu}$ y $\hat{\sigma}^2$ no comprende μ ni σ^2 . En este caso observemos que es posible estimar μ sin estimar σ^2 , pero no lo es estimar σ^2 sin empezar por estimar μ .

Ejemplo 8-14.—Sea la variable aleatoria x , con densidad uniforme dada por

$$f(x) = \frac{1}{\beta - \alpha} \quad \alpha < x < \beta \quad (16)$$

$$= 0 \quad \text{en otro caso}$$

donde α y β son dos números tales que $\alpha < \beta$. La función de verosimilitud para una muestra aleatoria de tamaño n es

$$L(\alpha, \beta) = g(x_1, \dots, x_n; \alpha, \beta) = \frac{1}{(\beta - \alpha)^n} \quad \alpha < x_i < \beta \quad (17)$$

$$= 0 \quad \text{en otro caso}$$

Si las derivadas de esta expresión respecto a α y β se igualan a cero e intentamos resolver el sistema para α y β , hallamos que al menos uno de los parámetros α , β debe ser infinito, lo cual es un resultado absurdo. La dificultad estriba en que la función de verosimilitud no tiene su tangente horizontal en el valor máximo, de modo que para localizar este hay que recurrir a otros medios. Por (17) resulta evidente que la verosimilitud alcanza su máximo cuando $\beta - \alpha$ es tan pequeño como sea posible. Dada una muestra de n observaciones $x_1, x_2, \dots, x_n$, supongamos que x' designa la menor y x'' la mayor de ellas. Claro es que α no puede ser superior a x' ni β inferior a x'' ; por tanto, el menor valor posible de $\beta - \alpha$ es $x'' - x'$. Los estimadores máximo-verosímiles son, evidentemente,

$$\hat{\alpha} = x' \quad (18)$$

$$\hat{\beta} = x''$$

resultado bastante curioso, ya que es independiente de las observaciones intermedias.

Estos tres ejemplos son suficientes para aclarar la aplicación del método de máxima verosimilitud. El último de ellos muestra que no es necesario basarse en el proceso de derivación para hallar el máximo. La función $L(\theta)$ puede representarse por la curva de la figura 8-2, en donde el máximo efectivo corresponde a θ , mientras que, por derivación, se encontraría para el máximo θ' . Debe recordarse así mismo que la ecuación $\partial L / \partial \theta = 0$ se emplea para obtener tanto máximos como mínimos y, por tanto, debe evitarse utilizar una raíz de la ecuación que en realidad corresponda a un mínimo.

Nada hemos dicho acerca de la estimación de un parámetro que aparezca en forma factorial en la expresión funcional de la distribución. Esto puede hacerse para cualquier problema dado mediante

tablas de las derivadas de la función factorial. No obstante, este problema surge tan raramente que no merece que nos ocupemos ahora de él. Estos parámetros (n en la distribución binomial, α en

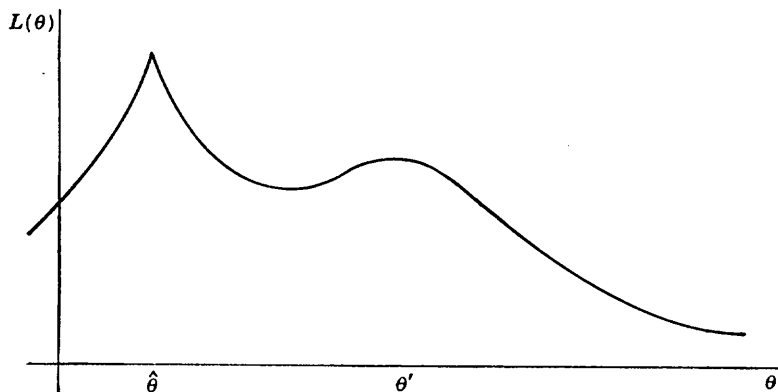


FIG. 8-2.

la distribución gamma, y α y β en la distribución beta) suelen determinarse mediante el tamaño muestral y aquí consideramos solo el caso en que se conoce el tamaño de la muestra.

8-11. Propiedades de los estimadores máximo-verosímiles.

En muchos de los problemas que tienen importancia en estadística aplicada, es bastante fácil obtener estimadores máximo-verosímiles, y frecuentemente estos estimadores poseen algunas de las propiedades óptimas consideradas en las secciones precedentes. Bajo condiciones de regularidad bastante generales de la función de densidad $f(x; \theta)$, los estimadores máximo-verosímiles son: 1) estimadores (asintóticamente) eficientes y óptimos asintóticamente normales; 2) estimadores consistentes (simples) y consistentes en error cuadrático; 3) función de los estadísticos suficientes mínimos. Además de estas propiedades, los estimadores máximo-verosímiles poseen una propiedad, llamada invarianza, que es la siguiente: Sea $\hat{\theta} = d(x_1, x_2, \dots, x_n)$ el estimador máximo-verosímil de θ en la densidad $f(x; \theta)$. Si $u(\theta)$ es una función de θ con función inversa uniforme, el estimador máximo-verosímil de $u(\theta)$ es $u(\hat{\theta})$. Así, p. ej., en la distribución normal, el estimador máximo-verosímil de σ^2 es

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

Por la propiedad de invarianza de los estimadores máximo-verosímiles, el estimador máximo-verosímil de σ es

$$\hat{\sigma} = \sqrt{\frac{1}{n} \sum (x_i - \bar{x})^2}$$

Análogamente, el estimador máximo-verosímil de, p. ej., $\log \sigma^2$ es

$$\log \left[\frac{1}{n} \sum (x_i - \bar{x})^2 \right]$$

etcétera.

Un estimador máximo-verosímil no siempre es insesgado, pero mucha veces una pequeña modificación puede hacer que lo sea. Así, p. ej.,

$$\hat{\sigma}^2 = \frac{1}{n} \sum (x_i - \bar{x})^2$$

es un estimador sesgado de σ^2 , pero $[n/(n-1)]\hat{\sigma}^2$ es insesgado. Para una discusión más general de la máxima verosimilitud, remitimos al lector a las referencias dadas en la bibliografía.

8-12. Estimación por el método de los momentos.—Sea $f(x; \theta_1, \dots, \theta_k)$ una función de densidad con k parámetros, y designemos por $\mu'_1, \mu'_2, \dots, \mu'_k$ los primeros k momentos respecto al origen; es decir,

$$\mu'_t = \int_{-\infty}^{\infty} x^t f(x; \theta_1, \dots, \theta_k) dx \quad t = 1, 2, \dots, k$$

En general, μ'_t será función de los k parámetros $\theta_1, \dots, \theta_k$, lo que representaremos escribiendo $\mu'_t = \mu'_t(\theta_1, \theta_2, \dots, \theta_k)$. Sea $x_1, x_2, \dots, x_n$ una muestra aleatoria de tamaño n de la densidad $f(x; \theta_1, \theta_2, \dots, \theta_k)$. A partir de esta muestra formamos los k primeros momentos muestrales $m'_1, m'_2, \dots, m'_k$, donde

$$m'_t = \frac{1}{n} \sum_{i=1}^n x_i^t$$

Sea $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_k$ la solución en función de θ_i de las k ecuaciones

$$m'_t = \mu'_t \quad t = 1, 2, \dots, k$$

Estas soluciones son los estimadores obtenidos por el método de los momentos. Así, p. ej., supongamos que la distribución de x es

$n(x; \mu, \sigma^2)$. En el capítulo 6 habíamos encontrado que

$$\mu'_1 = \mu; \quad \mu'_2 = \sigma^2 + \mu^2$$

Sea ahora $x_1, x_2, \dots, x_n$ una muestra aleatoria de tamaño n de esta distribución. En tal caso los momentos muestrales son

$$m'_1 = \frac{1}{n} \sum x_i$$

$$m'_2 = \frac{1}{n} \sum x_i^2$$

Igualando los momentos de la población y de la muestra, tenemos:

$$\hat{\mu} = \bar{x} \quad \hat{\sigma}^2 + \hat{\mu}^2 = \frac{1}{n} \sum x_i^2$$

o bien,

$$\hat{\mu} = \bar{x} \quad \hat{\sigma}^2 = \frac{1}{n} \left(\sum x_i^2 - n\bar{x}^2 \right)$$

En esta distribución los estimadores obtenidos por el método de los momentos coinciden con los estimadores máximo-verosímiles.

Se puede demostrar que, en condiciones bastante generales, los estimadores deducidos por el método de los momentos son: 1) estimadores consistentes (simples) y consistentes en error cuadrático, y 2) asintóticamente normales, aunque, en general, no asintóticamente eficientes ni óptimos asintóticamente normales.

8-13. Estimadores de Bayes.—En las secciones anteriores hemos considerado una densidad $f(x; \theta)$ de una variable aleatoria x , en donde θ es un parámetro desconocido que puede tomar valores en el espacio Ω . En algunas situaciones reales, representadas por la densidad $f(x; \theta)$, existe frecuentemente información adicional sobre θ (hasta ahora la única hipótesis que hemos hecho sobre θ es que puede tomar valores en Ω). Así, p. ej., el experimentador puede tener la evidencia de que θ se comporta como una variable aleatoria, para la cual es capaz de postular una función de densidad realista. Supongamos, p. ej., que se va a examinar una máquina que estampa piezas de automóvil para ver qué fracción p de piezas defectuosas se está obteniendo. En cierto día, se examinan 10 piezas producidas por la máquina, representándose las observaciones por $x_1, x_2, \dots, x_{10}$, donde $x_i = 1$ si la i -ésima pieza es defectuosa y $x_i = 0$

si no lo es. Las observaciones pueden considerarse como una muestra aleatoria de tamaño 10 de la distribución binomial puntual

$$f(x; p) = p^x(1-p)^{1-x} \quad x=0, 1; \quad 0 \leq p \leq 1$$

la cual indica que la probabilidad de que una pieza dada sea defectuosa es igual al número desconocido p . La densidad conjunta de las 10 variables aleatorias $x_1, x_2, \dots, x_{10}$ es

$$g(x_1, x_2, \dots, x_{10}; p) = p^{\sum x_i} (1-p)^{10-\sum x_i} \quad x_i=0, 1; \quad 0 \leq p \leq 1$$

El estimador máximo-verosímil de p , como se vio en secciones anteriores, es $\hat{p} = \bar{x}$. El método de los momentos proporciona el mismo estimador.

Imaginemos, sin embargo, que el experimentador dispone de información adicional sobre p ; supongamos que ha observado que el valor de p varía y le parece que el cambio puede representarse como una variable aleatoria con densidad

$$h(p) = 6p(1-p) \quad 0 \leq p \leq 1$$

¿Cómo podrá utilizarse esta información *adicional* sobre p para estimar p ? En muchos problemas quizá no sea realista suponer que p se comporta como una variable aleatoria; en otros, aunque parezca razonable suponer que p se comporta como una variable aleatoria, puede ser desconocida la función de densidad de p . Sin embargo, en algunos problemas resulta posible hacer hipótesis realistas; examinaremos tal situación en esta sección. Hasta aquí hemos empleado la notación $f(x; \theta)$ para representar la densidad de una variable aleatoria x para cada valor de θ en Ω . Cuando queramos indicar que el parámetro es también una variable aleatoria, denotaremos la densidad de x por $f(x|\theta)$, en lugar de por $f(x; \theta)$.

Sea $x_1, x_2, \dots, x_n$ una muestra aleatoria de tamaño n de la densidad $f(x|\theta)$, e imaginemos que deseamos estimar la θ que determina la densidad de la que procede la muestra aleatoria. Supongamos que la densidad marginal de θ es $p(\theta)$, y la pérdida, $l(\theta; \theta)$. Recordemos que, aunque estamos suponiendo que θ es una variable aleatoria, deseamos estimar un valor particular de θ : el valor θ que determina la densidad $f(x|\theta)$ de la cual fue seleccionada la muestra aleatoria. En otras palabras, al variar θ , quedan determinadas diferentes densidades, y la muestra aleatoria se tomó de una de estas densidades. Queremos estimar el valor de θ que determina tal densidad. El riesgo es $E[l(\hat{\theta}; \theta)] = R(d; \theta)$. Puesto que θ es una

variable aleatoria, interesará determinar la función d que hace mínimo el riesgo *esperado*. El riesgo esperado puede escribirse

$$B(d) = E[R(d, \theta)] = \int_{-\infty}^{\infty} R(d, \theta) p(\theta) d\theta \quad (1)$$

$$= \int_{-\infty}^{\infty} \left\{ \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} l[d(x_1, \dots, x_n); \theta] g(x_1, \dots, x_n | \theta) dx_1 \dots dx_n \right\} p(\theta) d\theta \quad (2)$$

Un «buen» estimador será una función d de las x_i que haga mínimo a $B(d)$; esta función recibe el nombre de *estimador de Bayes*. Si en (2) intercambiamos el orden de integración de las variables x y θ , tendremos:

$$B(d) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \left\{ \int_{-\infty}^{\infty} l[d(x_1, \dots, x_n); \theta] g(x_1, \dots, x_n | \theta) p(\theta) d\theta \right\} dx_1 \dots dx_n \quad (3)$$

Ahora bien: $B(d)$ se hará mínimo si es posible hallar una función d de las x_i que haga mínima la cantidad encerrada entre llaves en (3) para todo conjunto de las x . Esto es, queremos hallar la $d(x_1, \dots, x_n)$ que hace mínimo a

$$\int_{-\infty}^{\infty} l[d(x_1, \dots, x_n); \theta] g(x_1, \dots, x_n | \theta) p(\theta) d\theta \quad (4)$$

La cantidad $g(x_1, \dots, x_n | \theta) \cdot p(\theta)$ en (4) es la distribución conjunta de $x_1, \dots, x_n, \theta$, y la designaremos por $q(x_1, \dots, x_n, \theta)$. La distribución marginal de las x está dada por

$$k(x_1, \dots, x_n) = \int_{-\infty}^{\infty} q(x_1, \dots, x_n, \theta) d\theta = \int_{-\infty}^{\infty} g(x_1, \dots, x_n | \theta) \cdot p(\theta) d\theta \quad (5)$$

y la distribución condicional de θ , dadas $x_1, x_2, \dots, x_n$, es

$$h(\theta | x_1, \dots, x_n) = \frac{q(x_1, \dots, x_n, \theta)}{k(x_1, \dots, x_n)} = \frac{g(x_1, \dots, x_n | \theta) p(\theta)}{k(x_1, \dots, x_n)} \quad (6)$$

y se llama densidad *a posteriori*. Así, podemos escribir (4) en la forma

$$k(x_1, \dots, x_n) \int_{-\infty}^{\infty} l[d(x_1, \dots, x_n); \theta] h(\theta | x_1, \dots, x_n) d\theta \quad (7)$$

Luego un estimador de Bayes es el *valor de θ* que, para cada muestra posible $x_1, x_2, \dots, x_n$, hace mínima la cantidad

$$v(\theta; x_1, \dots, x_n) = \int_{-\infty}^{\infty} l(\theta; \theta) h(\theta | x_1, \dots, x_n) d\theta \quad (8)$$

La función v representa el *riesgo a posteriori* para estimar θ , dado que $x_1 = x_1, \dots, x_n = x_n$. Todo lo anterior queda resumido en la siguiente definición.

Definición 8-11.—Sean $x_1, \dots, x_n$ una muestra aleatoria de la densidad $f(x|\theta)$, $p(\theta)$ la densidad de θ y $g(x_1, \dots, x_n|\theta)$ la densidad condicional de las x , dada θ . Además, sea $h(\theta|x_1, \dots, x_n)$ la densidad condicional de θ , dadas las x , y $l(\theta; \theta)$, la pérdida. El estimador de Bayes de θ es una función, definida por $\hat{\theta} = d(x_1, \dots, x_n)$, que hace mínimo $B(d)$ dado en (1).

De las fórmulas (2) a (8) anteriores, se obtiene el teorema siguiente:

Teorema 8-7.—El valor de θ en función de las x , que hace mínimo el riesgo a posteriori $v[\theta; x_1, \dots, x_n]$ de la ecuación (8), es el estimador de Bayes de θ para las densidades y función de pérdida dadas en la definición 8-11.

Si las densidades de la definición 8-11 fuesen discretas en lugar de continuas, todas las integrales utilizadas en esta sección se reemplazarían por signos de suma.

Ejemplo 8-15.—Sea $x_1, \dots, x_n$ una muestra aleatoria de la densidad

$$f(x|\theta) = \theta^x (1-\theta)^{1-x}; \quad x=0, 1; \quad 0 \leq \theta \leq 1.$$

La densidad condicional de las x , dada θ , es

$$g(x_1, \dots, x_n|\theta) = \theta^{\sum x_i} (1-\theta)^{n-\sum x_i}$$

Supongamos que la pérdida es el error cuadrático; es decir, $l(\hat{\theta}; \theta) = (\hat{\theta} - \theta)^2$. Supongamos además que la densidad de θ es uniforme, de modo que $p(\theta) = 1$, $0 \leq \theta \leq 1$. Entonces

$$q(x_1, \dots, x_n, \theta) = \theta^{\sum x_i} (1-\theta)^{n-\sum x_i} \cdot 1$$

y

$$k(x_1, \dots, x_n) = \int_0^1 \theta^{\sum x_i} (1-\theta)^{n-\sum x_i} d\theta$$

que (Sec. 6-4) es igual a

$$\frac{[(\sum x_i)!][(n - \sum x_i)!]}{(n+1)!}$$

de modo que

$$h(\theta|x_1, \dots, x_n) = (n+1)! \frac{\theta^{\sum x_i} (1-\theta)^{n-\sum x_i}}{[(\sum x_i)!] [(n-\sum x_i)!]} \quad 0 \leq \theta \leq 1$$

Por (8), queremos hallar una función d , definida por $\hat{\theta} = d(x_1, \dots, x_n)$, que haga mínimo a

$$v(\theta; x_1, \dots, x_n) = \int_0^1 (\theta - \theta)^2 \frac{(n+1)! \theta^{\sum x_i} (1-\theta)^{n-\sum x_i}}{[(\sum x_i)!] [(n-\sum x_i)!]} d\theta$$

Por la sección 6-4, tenemos

$$\begin{aligned} v(\theta; x_1, \dots, x_n) &= \theta^2 - 2\theta \frac{[(n+1)!] [(\sum x_i + 1)!]}{[(n+2)!] [(\sum x_i)!]} \\ &\quad + \frac{[(n+1)!] [(\sum x_i + 2)!]}{[(n+3)!] [(\sum x_i)!]} \\ &= \theta^2 - 2\theta \frac{\sum x_i + 1}{n+2} + \frac{(\sum x_i + 2)(\sum x_i + 1)}{(n+2)(n+3)} \end{aligned}$$

El valor de θ , en función de las x , que hace mínimo a $v(\theta; x_1, \dots, x_n)$ se obtiene igualando a cero la derivada de v respecto a θ . La solución es

$$\theta = \frac{\sum x_i + 1}{n+2}$$

y, por tanto, el estimador de Bayes para este problema será

$$\hat{\theta} = \frac{\sum x_i + 1}{n+2}$$

Ejemplo 8-16.—Sea $x_1, \dots, x_n$ una muestra aleatoria de la densidad normal con varianza 1:

$$f(x|\mu) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-\mu)^2}$$

Entonces

$$g(x_1, \dots, x_n|\mu) = \frac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2}\sum (x_i - \mu)^2} = \frac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2}(\sum x_i^2 - 2\mu \sum x_i + n\mu^2)}$$

Supongamos que μ es una variable aleatoria cuya densidad es

$$p(\mu) = \frac{1}{\sqrt{2\pi}} e^{-\mu^2/2} \quad -\infty < \mu < \infty$$

Entonces

$$q(x_1, \dots, x_n, \mu) = \frac{1}{(2\pi)^{(n+1)/2}} \exp \left\{ -\frac{1}{2} \left[\sum x_i^2 + (n+1)\mu^2 - 2\mu n\bar{x} \right] \right\}$$

y

$$k(x_1, \dots, x_n) = \frac{1}{(2\pi)^{(n+1)/2}} \exp \left(-\frac{1}{2} \sum x_i^2 \right) \int_{-\infty}^{\infty} \exp \left\{ -\frac{1}{2} [(n+1)\mu^2 - 2\mu n\bar{x}] \right\} d\mu$$

Completando el cuadrado en el exponente del integrando, este se transforma en

$$\frac{1}{(2\pi)^{n/2}} \exp \left[-\frac{1}{2} \left(\sum x_i^2 - \frac{n^2 \bar{x}^2}{n+1} \right) \right] \left\{ \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{\infty} \exp \left[-\frac{1}{2} (n+1) \left(\mu - \frac{n\bar{x}}{n+1} \right)^2 \right] d\mu \right\}$$

Por la sección 6-2, la cantidad entre llaves vale $(n+1)^{-1/2}$ y, por tanto,

$$k(x_1, \dots, x_n) = \frac{1}{(n+1)^{1/2} (2\pi)^{n/2}} \exp \left[-\frac{1}{2} \left(\sum x_i^2 - \frac{n^2 \bar{x}^2}{n+1} \right) \right]$$

y

$$\begin{aligned} h(\mu | x_1, \dots, x_n) &= \frac{(2\pi)^{-(n+1)/2} \exp \left\{ -\frac{1}{2} \left[\sum x_i^2 + (n+1)\mu^2 - 2n\bar{x}\mu \right] \right\}}{(2\pi)^{-(n/2)} (n+1)^{-1/2} \exp \left[-\frac{1}{2} \left(\sum x_i^2 - \frac{n^2 \bar{x}^2}{n+1} \right) \right]} \\ &= \frac{(n+1)^{1/2}}{(2\pi)^{1/2}} \exp \left\{ -\frac{1}{2} (n+1) \left[\mu^2 - \frac{2n\bar{x}\mu}{n+1} + \frac{n^2 \bar{x}^2}{(n+1)^2} \right] \right\} \\ &= \frac{(n+1)^{1/2}}{(2\pi)^{1/2}} \exp \left\{ -\frac{1}{2} (n+1) \left[\mu - \frac{n\bar{x}}{n+1} \right]^2 \right\} \end{aligned}$$

Así, la distribución condicional de μ , dadas $x_1, \dots, x_n$, es normal con media $\bar{x}n/(n+1)$ y varianza $(n+1)^{-1}$. Supongamos que la función de pérdida que queremos considerar es el error cuadrático

$$l(\hat{\mu}; \mu) = (\hat{\mu} - \mu)^2.$$

Entonces el estimador de Bayes será el valor de $\hat{\mu}$, función de las x , que hace mínimo a $v(\hat{\mu}; x_1, \dots, x_n)$, donde

$$\begin{aligned} v(\hat{\mu}; x_1, \dots, x_n) &= \int_{-\infty}^{\infty} (\hat{\mu} - \mu)^2 \cdot h(\mu | x_1, \dots, x_n) d\mu \\ &= \frac{(n+1)^{\frac{1}{2}}}{(2\pi)^{\frac{1}{2}}} \int_{-\infty}^{\infty} (\hat{\mu} - \mu)^2 \\ &\quad \exp \left\{ -\frac{1}{2} (n+1) \left[\mu - \frac{n\bar{x}}{n+1} \right]^2 \right\} d\mu \\ &= \hat{\mu}^2 - \frac{2\hat{\mu}\bar{x}n}{n+1} + \frac{1}{n+1} + \frac{\bar{x}^2 n^2}{(n+1)^2} \end{aligned}$$

Para hallar el valor de $\hat{\mu}$ que hace mínimo a $v(\hat{\mu}; x_1, \dots, x_n)$ resolvemos la ecuación

$$\frac{\partial[v(\hat{\mu}; x_1, \dots, x_n)]}{\partial \hat{\mu}} = 2\hat{\mu} - \frac{2\bar{x}n}{n+1} = 0$$

lo que da $\hat{\mu} = \Sigma x_i / (n+1)$. Luego el estimador de Bayes de μ es

$$\hat{\mu} = \frac{\Sigma x_i}{n+1}$$

Bajo condiciones muy generales, se demuestra que el estimador de Bayes correspondiente a una distribución de probabilidad *a priori* arbitraria, con densidad positiva, es: 1) consistente; 2) (asintóticamente) eficiente y OAN; 3) función del estadístico suficiente mínimo. Además, puede demostrarse que el estimador de Bayes difiere del estimador máximo-verosímil en una cantidad pequeña, comparada con $1/\sqrt{n}$.

PROBLEMAS

1. Sea x_1, x_2 una muestra aleatoria de tamaño 2 de una distribución normal con media μ desconocida y varianza 1, pudiendo variar μ entre $-\infty$ y $+\infty$. Queremos hallar una estimación puntual de μ . Consideraremos tres estimadores

$$\hat{\mu}_1 = d_1(x_1, x_2) = \frac{2}{3}x_1 + \frac{1}{3}x_2$$

$$\hat{\mu}_2 = d_2(x_1, x_2) = \frac{1}{4}x_1 + \frac{3}{4}x_2$$

$$\hat{\mu}_3 = d_3(x_1, x_2) = \frac{1}{2}x_1 + \frac{1}{2}x_2$$

y la función de pérdida

$$l(\hat{\mu}; \mu) = 3\mu^2(\hat{\mu} - \mu)^2$$

Hallar $R(d_1; \mu)$.

2. En el problema 1, defínase el espacio de acción y el espacio paramétrico.

3. En el problema 1, hállese $R(d_2; \mu)$ y $R(d_3; \mu)$.

4. ¿Cómo se comparan d_1 , d_2 o d_3 en cuanto a riesgo mínimo para cada valor de μ ?

5. Demuéstrese que los tres estimadores del problema 1 son insesgados.

6. En el problema 1, hállese las eficiencias relativas: a) d_1 respecto a d_2 ; b) d_1 respecto a d_3 ; c) d_2 respecto a d_3 .

7. Hállese el estimador máximo-verosímil de μ , la media poblacional, dada una muestra de tamaño n de una población con $f(x)=1/\beta$, $0 < x < \beta$. Estímese β por el método de los momentos.

8. La muestra 1,3; 0,6; 1,7; 2,2; 0,3; 1,1 procede de una población con función de densidad $f(x)=1/\beta$, $0 < x < \beta$. ¿Cuáles son las estimaciones máximo-verosímiles de la media y de la varianza de la población?

9. ¿Cuál es el estimador máximo-verosímil de α , si la función de densidad es $f(x)=(\alpha+1)x^\alpha$, $0 < x < 1$? Estímese α por el método de los momentos.

10. Supuesto conocida α , hállese el estimador máximo-verosímil de β para una muestra aleatoria de tamaño n de la distribución gamma. ¿Es un estadístico suficiente? ¿Es insesgado?

11. Para una muestra aleatoria de tamaño n , hállese el estimador máximo-verosímil del parámetro de la distribución de Poisson. ¿Es un estadístico suficiente? ¿Es insesgado?

12. Para una muestra aleatoria de tamaño n , hállese el estimador máximo-verosímil de la varianza de una población normal, suponiendo conocida la media. ¿Es un estadístico suficiente? ¿Es insesgado?

13. Para una muestra aleatoria de tamaño 1, hállese el estimador máximo-verosímil de la varianza de la distribución gamma, suponiendo α conocida.

14. Si x se distribuye según $f(x)=1/\beta$, $0 < x < \beta$ y se consideran muestras de una sola observación x , puesto que $E(x)=\beta/2$, el estimador de β basado en el método de los momentos será $\hat{\beta}_1=2x$. Por otra parte, el estimador máximo-verosímil de β es $\hat{\beta}_2=x$. ¿Puede elegirse entre estos dos estimadores, basándose en la eficiencia relativa?

15. Sea $x_1, \dots, x_n$ una muestra aleatoria de tamaño n de una densidad normal con media μ y varianza σ^2 , en donde $-\infty < \mu < \infty$ y $0 < \sigma^2 < \infty$. Si

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2$$

puede demostrarse que

$$\text{var}(\hat{\sigma}^2) = \frac{2\sigma^2}{n-1}$$

Pruebase que $\hat{\sigma}^2$ es un estimador consistente en error cuadrático de σ^2 .

16. Sea $x_1, \dots, x_n$ una muestra aleatoria de la densidad $f(x; \theta)$, discreta o continua, y sea $\{\hat{\theta}_n\}$ una sucesión de estimadores que forman un estimador consistente en error cuadrático de θ . Utilícese la desigualdad de Tchebysheff para demostrar que $\{\hat{\theta}_n\}$ es también un estimador consistente (simple) de θ .

17. Si la distribución de x es normal con media μ y varianza σ^2 , hállese para muestras de tamaño k el estimador máximo-verosímil del punto A tal que

$$\int_A^\infty n(x; \mu, \sigma^2) dx = 0,05.$$

Obtégase el estimador insesgado de varianza mínima de A .

18. En el capítulo 10 se prueba que la media de una muestra procedente de una población normal tiene una distribución exactamente normal. Utilícese este resultado para demostrar que la media muestral es un estimador suficiente de la media poblacional, cuando σ es conocida.

19. En investigaciones genéticas, las muestras proceden frecuentemente de una distribución binomial

$$f(x) = \binom{m}{x} p^x q^{m-x},$$

con la excepción de que las observaciones de $x=0$ son imposibles, de modo que, en realidad, las muestras proceden de la distribución condicional

$$g(x|x > 0) = \binom{m}{x} \frac{p^x q^{m-x}}{1 - q^m} \quad x=1, 2, \dots, m$$

Hállese el estimador máximo-verosímil de p , en el caso $m=2$, para muestras de tamaño n . ¿Es insesgado?

20. Hállese el estimador de α para la función de densidad

$$f(x; \alpha) = \frac{2}{\alpha^2} (\alpha - x) \quad 0 < x < \alpha$$

y muestras de tamaño 2. ¿Es un estadístico suficiente? Estímese α por el método de los momentos.

21. Con referencia al problema 20, ¿cuál es el estimador máximo-verosímil de la media poblacional?

22. Una urna contiene bolas negras y blancas. Se extrae una muestra de tamaño n , con reemplazamiento. ¿Cuál es el estimador máximo-verosímil de la razón R de bolas negras a bolas blancas en la urna?

23. Con referencia al problema 22, supóngase que se extraen las bolas una a una, con reemplazamiento, hasta que aparece una bola negra. Sea x el número necesario de extracciones (sin contar la última). Esta operación se repite n veces, obteniendo una muestra $x_1, x_2, \dots, x_n$. ¿Cuál es el estimador máximo-verosímil de R , basado en esta muestra?

24. Sea $x_1, x_2, \dots, x_n$ una muestra aleatoria de tamaño n de una densidad normal con media μ y varianza σ^2 . Los estimadores $\bar{x} = \bar{\mu}$ y $\hat{\sigma}^2 = (n-1)^{-1} \sum (x_i - \bar{x})^2$ son suficientes y completos. Hállese el estimador insesgado de varianza mínima de: a) $6\mu + 4\sigma^2$; b) $\mu^2 - 5\sigma^2$.

25. Sea $x_1, x_2, \dots, x_n$ una muestra aleatoria de tamaño n de la densidad binomial puntual $p^x(1-p)^{1-x}$, $x=0, 1$. El estimador $\hat{p} = \bar{x}$ es suficiente y completo. Hállese el estimador insesgado de varianza mínima de: a) $3p$; b) $5p-1$.

26. Supongamos que se eligen al azar n piezas cilíndricas entre las producidas por cierta máquina, y que se miden sus diámetros y longitudes. Se observa que en n_{11} ambas medidas son inferiores a los límites de tolerancia; en n_{12} las longitudes son satisfactorias, pero no los diámetros; en n_{21} , los diámetros, pero no las longitudes; y en n_{22} ninguna de las dos medidas es satisfactoria. Se verifica que $\sum n_{ij} = n$. Cada pieza puede considerarse procedente de una población polinomial, con densidad

$$p_{11}^{x_{11}} p_{12}^{x_{12}} p_{21}^{x_{21}} (1-p_{11}-p_{12}-p_{21})^{x_{22}} \quad x_{ij}=0, 1; \quad \sum x_{ij}=1$$

con tres parámetros. ¿Cuáles son las estimaciones máximo-verosímiles de los parámetros si $n_{11}=90$, $n_{12}=6$, $n_{21}=3$, $n_{22}=1$?

27. Con referencia al problema anterior, supongamos que no hay motivo para creer que exista relación alguna entre diámetros defectuosos y longitudes defectuosas. En este caso, la distribución de las x_{ij} puede expresarse en función de dos parámetros: p_1 , probabilidad de que la longitud sea satisfactoria, y q_1 , probabilidad de que el diámetro sea satisfactorio. La distribución de las x_{ij} es ahora

$$(p_1 q_1)^{x_{11}} [p_1(1-q_1)]^{x_{12}} [(1-p_1)q_1]^{x_{21}} [(1-p_1)(1-q_1)]^{x_{22}} \quad x_{ij}=0, 1; \quad \sum x_{ij}=1$$

¿Cuáles son las estimaciones máximo-verosímiles de estos parámetros? Las probabilidades correspondientes a las cuatro clases en este modelo, ¿son diferentes de las obtenidas en el problema anterior?

28. Una muestra de tamaño n_1 procede de una población normal de media μ_1 y varianza σ_1^2 . Si se toma una segunda muestra de tamaño n_2 de una población normal de media μ_2 y varianza σ_2^2 , ¿cuál es el estimador máximo-verosímil de $\alpha = \mu_1 - \mu_2$? Suponiendo fijo el tamaño $n = n_1 + n_2$ de la muestra total, ¿cómo deberían dividirse las n observaciones de ambas poblaciones para que fuese mínima la varianza de $\hat{\alpha}$?

29. Se toman muestras de tamaño n de cada una de cuatro poblaciones normales con igual varianza σ^2 . Las medias de estas cuatro poblaciones son $a+b+c$, $a+b-c$, $a-b+c$, $a-b-c$. ¿Cuáles son los estimadores de máxima verosimilitud de a , b , c y σ^2 ? (Las observaciones muestrales pueden designarse por x_{ij} ; $i=1, 2, 3, 4$, y $j=1, 2, \dots, n$.)

30. Las observaciones $x_1, x_2, \dots, x_n$ proceden de poblaciones normales de igual media μ , pero diferentes varianzas $\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2$. ¿Es posible estimar todos los parámetros? Si se suponen conocidas todas las σ_i^2 , ¿cuál es el estimador máximo-verosímil de μ ?

31. ¿Es $\hat{\sigma}$, raíz cuadrada de la expresión del segundo miembro de la ecuación (8-10-14), una estimación insesgada de σ ?

32. Sea μ el verdadero valor del cociente intelectual de cierto estudiante. Para medir su C. I. realiza un *test* y se sabe que las puntuaciones obtenidas se distribuyen normalmente, con media μ y desviación estándar 5. El estudiante realiza el *test* y alcanza la puntuación 130. ¿Cuál es el estimador máximo-verosímil de μ ?

33. En el problema 32, supongamos que se sabe que el verdadero cociente intelectual de los estudiantes de cierto grupo de edad se distribuye normalmente, con media 100 y varianza 225; es decir, suponemos que μ se distribuye normalmente con media 100 y varianza 225. Así, $f(x|\mu)$ en el problema 32 es $n(x; \mu, 25)$ y $p(\mu)$ es $n(\mu; 100, 225)$. Hállense $q(x, \mu)$ y $k(x)$ de la sección 8-13.

34. En el problema 33, demuéstrese que la densidad condicional $h(\mu|x)$ es normal, con media $0,9x+10$ y varianza $45/2$.

35. Utilizando la pérdida, $l(\hat{\mu}; \mu) = (\hat{\mu} - \mu)^2$ en el problema 34, hállese el estimador de Bayes de μ , cociente intelectual del estudiante, si la puntuación del *test* ha sido $x=130$. Obsérvese que no es el mismo que el estimador máximo-verosímil del problema 32.

36. La fracción defectuosa de la producción de un día de cierto producto es θ . Sea x una observación de uno de los ítems de la producción de un día dado. La distribución de x es

$$f(x|\theta) = \theta^x(1-\theta)^{1-x} \quad x=0, 1; 0 \leq \theta \leq 1$$

donde $x=1$ se identifica con un ítem defectuoso y $x=0$ con uno no defectuoso. Aunque la proporción de defectuosos permanece constante en un día dado, se observa que θ varía de un día a otro, comportándose como una variable aleatoria con función de densidad

$$p(\theta) = 6\theta(1-\theta) \quad 0 \leq \theta \leq 1$$

Consideremos la pérdida $l(\hat{\theta}; \theta) = 2(\hat{\theta} - \theta)^2$, donde $\hat{\theta}$ es el estimador de θ . Si no se dispone de observaciones x , el valor de $\hat{\theta}$ que hace mínimo a

$$E[l(\hat{\theta}; \theta)] = \int_0^1 2(\hat{\theta} - \theta)^2 p(\theta) d\theta$$

podría utilizarse para estimar θ . Hállese este valor de $\hat{\theta}$.

37. Si se dispone de una observación x en el problema 36, hállese $q(x; \theta)$ y $k(x)$ de la sección 8-13.

38. En el problema 37, hállese $h(\theta|x)$ de la sección 8-13.

39. En el problema 37, hállese $v(\hat{\theta}; x)$ de la sección 8-13.

40. En el problema 37, hállese el estimador de Bayes de θ .

41. Si

$$f(x|\theta) = \frac{2x}{\theta^2} \quad 0 < x < \theta$$

y

$$p(\theta)=1 \quad 0 < \theta < 1$$

hállese el estimador de Bayes de θ , utilizando la pérdida $l(\hat{\theta}; \theta) = \theta^2(\hat{\theta} - \theta)^2$.

42. Demuéstrase que para la pérdida $(\hat{\theta} - \theta)^2$ el estimador de Bayes está dado por θ^* , donde

$$\theta^* = E(\theta|x) = \int_{-\infty}^{\infty} \theta h(\theta|x) d\theta$$

43. Sea $x_1, x_2, \dots, x_n$ una muestra aleatoria de tamaño n de la distribución de Poisson

$$f(x|\lambda) = \frac{\lambda^x e^{-\lambda}}{x!} \quad x=0, 1, \dots$$

Supongamos que λ tiene la densidad

$$p(\lambda) = e^{-\lambda} \quad 0 < \lambda < \infty$$

Hállese la densidad *a posteriori*, $h(\lambda|x_1, \dots, x_n)$.

44. En el problema 43, hállese $E(\lambda|x_1, \dots, x_n)$ y demuéstrase que esta es el estimador de Bayes para la pérdida

$$l(\hat{\lambda}; \lambda) = (\hat{\lambda} - \lambda)^2.$$

BIBLIOGRAFIA

1. BLACKWELL, D.: «Conditional expectation and unbiased sequential estimation», *Annals of Mathematical Statistics*, vol. 18 (1947), páginas 105-110.
2. CHERNOFF, Herman: «Remarks on a rational selection of a decision function», Cowles Commission Discussion Paper 326, Statistics, Enero 10, 1949. No publicado.
3. CHERNOFF, H., y L. E. MOSES: *Elementary Decision Theory*, John Wiley & Sons, Inc., Nueva York, 1959.
4. CRAMÉR, H.: *Métodos matemáticos de estadística*, Aguilar, 4.^a ed., Madrid, 1967.
5. DOOB, J.: «Statistical estimation», *Transactions of the American Mathematical Society*, vol. 39 (1936), págs. 410-421.
6. FISHER, R. A.: «On the mathematical foundations of theoretical statistics», *Philosophical Transactions of the Royal Society of London*, series A, vol. 222 (1922).
7. FISHER, R. A.: *Statistical Methods for Research Workers*, Oliver & Boyd, Ltd., Edimburgo y Londres, 1925.
8. FISHER, R. A.: «Theory of statistical estimation», *Proceedings of the Cambridge Philosophical Society*, vol. 22 (1925).
9. FISHER, R. A.: *The Design of Experiments*, Oliver & Boyd, Ltd., Edimburgo y Londres, 1935.
10. LEHMAN, E. L.: *Testing Statistical Hypotheses*, John Wiley & Sons, Inc., Nueva York, 1959.

11. LEHMANN, E. L., y H. SCHEFFÉ: «Completeness, similar regions and unbiased estimation», *Sankhyā*, vol. 10 (1950), págs. 305-340.
12. NEYMAN, Jerzy: «Contributions to the theory of the χ^2 test», *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability*, Universidad de California, Berkeley, Calif., 1949, págs. 239-273.
13. PITMAN, E. J. G.: «The 'closest' estimate of statistical parameters», *Proceedings of the Cambridge Philosophical Society*, vol. 33 (1937), páginas 212-222.
14. RAO, C. R.: «Information and accuracy attainable in the estimation of statistical parameters», *Bulletin of the Calcutta Mathematical Society*, vol. 37 (1945), pág. 81.
15. RAO, C. R.: «Minimum variance and the estimation of several parameters», *Proceedings of the Cambridge Philosophical Society*, vol. 43 (1947), pág. 280.
16. RAO, C. R.: «Sufficient statistics and minimum variance estimates», *Proceeding of the Cambridge Philosophical Society*, vol. 45 (1948), pág. 213.
17. THRALL, Robert M., Clyde H. COOMBS y Robert L. DAVIS: *Decision Processes*, John Wiley & Sons, Inc., Nueva York, 1954.
18. WALD, Abraham: *On the Principles of Statistical Inference*, Universidad de Notre Dame, Notre Dame, Ind., 1942.
19. WALD, Abraham: *Statistical Decision Functions*, John Wiley & Sons, Inc., Nueva York, 1950.
20. WEISS, Lionel: *Statistical Decision Theory*, McGraw-Hill Book Company Inc., Nueva York, 1961.
21. WOLFOWITZ, J.: «On Wald's proof of the consistency of the maximum-likelihood estimates», *Annals of Mathematical Statistics*, vol. 20 (1949), págs. 601-602.

CAPITULO 9

DISTRIBUCION NORMAL MULTIVARIANTE

9-1. La distribución normal bivalente.—Una de las densidades multivariantes más importantes es la normal multivariante, que constituye una generalización de la distribución normal de una sola variante. En esta sección estudiaremos un caso especial, el de la distribución normal bivalente. El estudio del caso de más de dos variantes resulta muy prolijo si no se utilizan matrices y vectores. En consecuencia, en la sección 9-2 se darán los elementos necesarios de teoría matricial, y el resto del capítulo se dedicará al análisis de la distribución normal multivariante utilizando matrices. Los lectores no familiarizados con las matrices y que no quieran gastar tiempo en el estudio de la sección 9-2, pueden estudiar la normal bivalente en esta sección y pasar directamente al capítulo 10.

Definición 9-1. La distribución normal bivalente.—Sea la variable aleatoria bidimensional (x, y) con densidad conjunta

$$f(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)} \left[\left(\frac{x-\mu_x}{\sigma_x} \right)^2 - 2\rho \frac{x-\mu_x}{\sigma_x} \frac{y-\mu_y}{\sigma_y} + \left(\frac{y-\mu_y}{\sigma_y} \right)^2 \right]} \quad (1)$$

$-\infty < x < \infty$, $-\infty < y < \infty$, donde σ_x , σ_y , μ_x , μ_y , ρ , son constantes tales que $-1 < \rho < 1$; $0 < \sigma_y$; $0 < \sigma_x$; $-\infty < \mu_x < \infty$; $-\infty < \mu_y < \infty$. En estas condiciones se dice que la variable aleatoria tiene una distribución normal bivalente.

La densidad (1) puede representarse por una superficie de forma de campana $z=f(x, y)$, tal como la dibujada en la figura 9-1. Todo plano paralelo al x, y , que corte a la superficie, lo hace según una curva elíptica, y todo plano perpendicular al x, y la corta según una curva normal. La probabilidad de que un punto (x, y) tomado al azar esté situado en una región R determinada del plano x, y se obtiene integrando $f(x, y)$ sobre tal región

$$P[(x, y) \text{ esté en } R] = \int_R f(x, y) dy dx \quad (2)$$

Esta función podría, p. ej., representar la distribución de impactos sobre un plano vertical (Cap. 4), en donde x y y son las desvia-

ciones horizontales y verticales respecto a dos rectas centrales. La distribución se aproxima, efectivamente, tanto a esta como a muchas otras poblaciones bivariantes que se encuentran en la práctica.

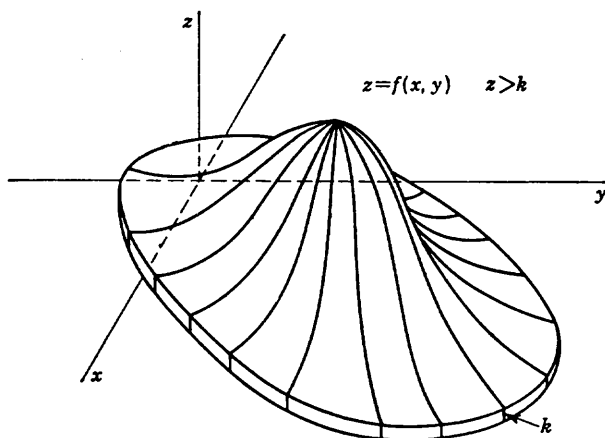


FIG. 9-1.

Debemos comenzar por demostrar que la función representa, en efecto, una distribución, viendo que la integral extendida a todo el plano es igual a uno; esto es,

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dy dx = 1 \quad (3)$$

La densidad es, por supuesto, positiva. Para simplificar la integral, hacemos la sustitución

$$u = \frac{x - \mu_x}{\sigma_x} \quad (4)$$

$$v = \frac{y - \mu_y}{\sigma_y}$$

con lo cual tenemos

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-[1/2(1-\rho^2)](u^2-2\rho uv+v^2)} dv du$$

Al completar el cuadrado en la variante u del exponente, resulta

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-[1/2(1-\rho^2)][(u-\rho v)^2+(1-\rho^2)v^2]} dv du$$

y haciendo la sustitución

$$w = \frac{u - \rho v}{\sqrt{1 - \rho^2}} \quad dw = \frac{du}{\sqrt{1 - \rho^2}}$$

la integral puede escribirse como producto de dos integrales simples

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-(w^2/2)} dw \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-(v^2/2)} dv \quad (5)$$

cada una de las cuales es igual a la unidad, como hemos visto al estudiar la distribución normal univariante. Con esto queda comprobada la ecuación (3).

Para obtener los momentos de x e y , hallaremos la función generatriz de momentos mixtos; esto es

$$m(t_1, t_2) = E(e^{t_1 x + t_2 y}) \quad (6)$$

$$= \iint e^{t_1 x + t_2 y} f(x, y) dy dx \quad (7)$$

Si ahora sustituimos x e y en función de u y v , obtenemos

$$m(t_1, t_2) = e^{t_1 \mu_x + t_2 \mu_y} \iint e^{t_1 \sigma_x u + t_2 \sigma_y v} \frac{1}{2\pi \sqrt{1 - \rho^2}} e^{-[1/2(1 - \rho^2)](u^2 - 2\rho uv + v^2)} dv du \quad (8)$$

Los exponentes combinados que aparecen en el integrando pueden escribirse

$$-\frac{1}{2(1 - \rho^2)} [u^2 - 2\rho uv + v^2 - 2(1 - \rho^2)t_1 \sigma_x u - 2(1 - \rho^2)t_2 \sigma_y v]$$

y completando el cuadrado, primero en u y luego en v , la expresión se transforma en

$$-\frac{1}{2(1 - \rho^2)} \{ [u - \rho v - (1 - \rho^2)t_1 \sigma_x]^2 + (1 - \rho^2)(v - \rho t_1 \sigma_x - t_2 \sigma_y)^2 - (1 - \rho^2)(t_1^2 \sigma_x^2 + 2\rho t_1 t_2 \sigma_x \sigma_y + t_2^2 \sigma_y^2) \}$$

que, haciendo la sustitución

$$w = \frac{u - \rho v - (1 - \rho^2)t_1 \sigma_x}{\sqrt{1 - \rho^2}}$$

$$z = v - \rho t_1 \sigma_x - t_2 \sigma_y$$

toma la forma

$$-1/2w^2 - 1/2z^2 + 1/2(t_1^2\sigma_x^2 + 2\rho t_1 t_2 \sigma_x \sigma_y + t_2^2\sigma_y^2)$$

con lo cual la integral de (8), se escribe

$$\begin{aligned} m(t_1, t_2) &= e^{t_1\mu_x + t_2\mu_y} e^{\frac{1}{2}(t_1^2\sigma_x^2 + 2\rho t_1 t_2 \sigma_x \sigma_y + t_2^2\sigma_y^2)} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{2\pi} e^{-(w^2/2) - (z^2/2)} dw dz \\ &= e^{t_1\mu_x + t_2\mu_y + \frac{1}{2}(t_1^2\sigma_x^2 + 2\rho t_1 t_2 \sigma_x \sigma_y + t_2^2\sigma_y^2)} \end{aligned} \quad (9)$$

puesto que, evidentemente, la integral vale 1.

Resulta así el siguiente teorema.

Teorema 9-1.—*La función generatriz de momentos de la distribución normal bivalente es*

$$m(t_1, t_2) = e^{t_1\mu_x + t_2\mu_y + \frac{1}{2}(t_1^2\sigma_x^2 + 2\rho t_1 t_2 \sigma_x \sigma_y + t_2^2\sigma_y^2)}$$

Los momentos pueden obtenerse calculando las derivadas de $m(t_1, t_2)$ en $t_1=0, t_2=0$. Así, pues,

$$E(x) = \left. \frac{\partial m}{\partial t_1} \right|_{t_1, t_2=0} = \mu_x \quad (10)$$

$$E(x^2) = \left. \frac{\partial^2 m}{\partial t_1^2} \right|_{t_1, t_2=0} = \mu_x^2 + \sigma_x^2 \quad (11)$$

y, por tanto, la varianza de x es

$$E(x - \mu_x)^2 = E(x^2) - \mu_x^2 = \sigma_x^2 \quad (12)$$

Análogamente, derivando respecto a t_2 , se hallan la media y la varianza de y , que son μ_y y σ_y^2 . Se obtienen también los momentos mixtos

$$E(x^r y^s)$$

derivando $m(t_1, t_2)$ r veces respecto a t_1 y s veces respecto a t_2 , y haciendo a continuación t_1 y t_2 iguales a cero. La *covarianza* de x e y es

$$\begin{aligned} E[(x - \mu_x)(y - \mu_y)] &= E(xy - x\mu_y - y\mu_x + \mu_x\mu_y) \\ &= E(xy) - \mu_x\mu_y \\ &= \rho\sigma_x\sigma_y \end{aligned} \quad (13)$$

El parámetro ρ recibe el nombre de *correlación* entre x e y . Se observará que cuando la correlación es cero, $f(x, y)$ en (1) es el pro-

ducto de dos distribuciones normales univariantes; por tanto, en este caso ($\rho=0$), x e y serán independientes en sentido probabilístico.

La densidad marginal de una de las variables, p. ej., x , es, por definición,

$$f_1(x) = \int_{-\infty}^{\infty} f(x, y) dy; \quad (14)$$

sustituyendo otra vez

$$v = \frac{y - \mu_y}{\sigma_y}$$

y completando el cuadrado en v , se tiene

$$f_1(x) = \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma_x\sqrt{1-\rho^2}} e^{-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - \frac{1}{2(1-\rho^2)}\left(v - \rho\frac{x-\mu_x}{\sigma_x}\right)^2} dv$$

Entonces, la sustitución

$$w = \frac{v - \rho[(x - \mu_x)/\sigma_x]}{\sqrt{1-\rho^2}} \quad dw = \frac{dv}{\sqrt{1-\rho^2}}$$

nos da inmediatamente:

$$f_1(x) = \frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2} \quad (15)$$

que es la función de densidad normal univariante. Análogamente se halla la función de densidad marginal de y ,

$$f_2(y) = \frac{1}{\sqrt{2\pi}\sigma_y} e^{-\frac{1}{2}\left(\frac{y-\mu_y}{\sigma_y}\right)^2} \quad (16)$$

Tenemos el siguiente teorema.

Teorema 9-2.—Sea la variable aleatoria bidimensional (x, y) , con la densidad bivalente dada en la definición 9.1. La distribución marginal de x es normal con media μ_x y varianza σ_x^2 . Así mismo, la distribución marginal de y es normal con media μ_y y varianza σ_y^2 .

Una vez obtenidas las distribuciones marginales, es posible determinar las distribuciones condicionales. Así, la función de den-

sidad condicional de x para valores fijos de y es

$$f(x|y) = \frac{f(x, y)}{f_y(y)}$$

que, una vez sustituidos los valores de las funciones en el segundo miembro, toma la forma

$$f(x|y) = \frac{1}{\sqrt{2\pi} \sigma_x \sqrt{1 - \rho^2}} e^{-\frac{1}{2\sigma_x^2(1-\rho^2)} \left[x - \mu_x - \frac{\rho\sigma_x}{\sigma_y} (y - \mu_y) \right]^2} \quad (17)$$

que es una función de densidad normal univariante, con media $\mu_x + (\rho\sigma_x/\sigma_y)(y - \mu_y)$ y varianza $\sigma_x^2(1 - \rho^2)$. La distribución condicional de y puede obtenerse reemplazando x por y en (17), con lo cual resulta

$$f(y|x) = \frac{1}{\sqrt{2\pi} \sigma_y \sqrt{1 - \rho^2}} e^{-\frac{1}{2\sigma_y^2(1-\rho^2)} \left[y - \mu_y - \frac{\rho\sigma_y}{\sigma_x} (x - \mu_x) \right]^2} \quad (18)$$

De este modo, se deduce el siguiente teorema.

Teorema 9-3.—Sea la variable aleatoria bidimensional (x, y) , con la densidad bivalente dada en la definición 9-1. La distribución condicional de x , dado $y=y$, es normal con media $\mu_x + (\rho\sigma_x/\sigma_y)(y - \mu_y)$ y varianza $\sigma_x^2(1 - \rho^2)$. Análogamente, la distribución condicional de y , dado $x=x$, es normal con media $\mu_y + (\rho\sigma_y/\sigma_x)(x - \mu_x)$ y varianza $\sigma_y^2(1 - \rho^2)$.

El valor medio de una variante en una distribución condicional recibe el nombre de *regresión* cuando se considera como función de las variantes fijadas en la distribución condicional. Así, la regresión para x en (17) es $\mu_x + (\rho\sigma_x/\sigma_y)(y - \mu_y)$, que en este caso es una función lineal de y . Para distribuciones bivariantes en general, la media de x en la distribución condicional de x , dado $y=y$, será cierta función, $g(y)$, y la ecuación

$$x = g(y)$$

representada en el plano x, y da la *curva de regresión* para x . Se trata de una curva que da la situación de la media de x para los diversos valores de y en la densidad condicional de x dado y .

Para la distribución normal bivalente, la curva de regresión es la recta obtenida representando

$$x = \mu_x + \frac{\rho\sigma_x}{\sigma_y} (y - \mu_y) \quad (19)$$

como se indica en la figura 9-2. La función de densidad condicional de x , $f(x|y)$, se ha representado también en esta figura para dos valores particulares, y_0 y y_1 , de y .

La distribución normal bivalente acumulativa

$$F(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(s, t) dt ds$$

puede reducirse a una forma que solo contiene el parámetro ρ , haciendo la sustitución (4).

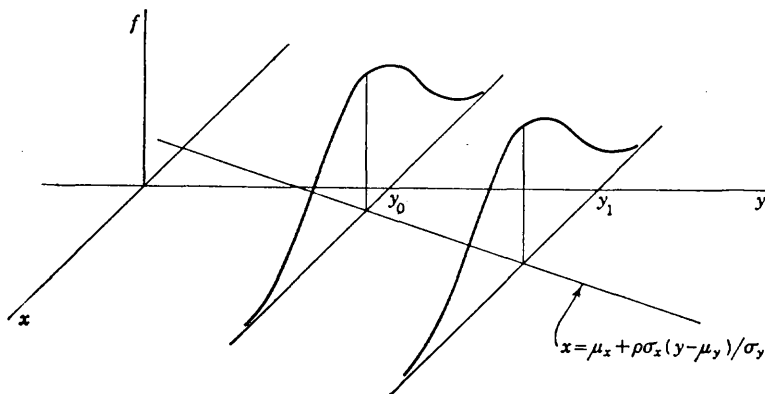


FIG. 9-2.

9-2. Matrices y determinantes.—Daremos en esta sección algunos teoremas sobre matrices y determinantes que serán necesarios para desarrollar la distribución normal p -variante general. No daremos las demostraciones de los teoremas, pero suponemos que el lector está familiarizado con los conceptos elementales del álgebra de matrices. En todo este capítulo utilizaremos letras mayúsculas V , R , R_{11} , R_{12} , R_{21} , R_{22} , V_{11} , V_{12} , V_{21} y A para representar matrices, y las letras Y , μ , U_2 , Y_2^* , Y_1^* , U_1 , X^* para representar vectores. Esta sección y las restantes del capítulo pueden omitirse sin interrumpir la continuidad del libro, excepto para unas pocas cuestiones que están señaladas con un asterisco. En el capítulo 14, y en parte del 13, se utilizarán extensamente matrices y vectores.

Definición 9-2.—Una matriz V con p filas y q columnas es una disposición rectangular de elementos σ_{ij} . Los elementos σ_{ij} pueden ser números, funciones, variables aleatorias, etc. La cantidad σ_{ij} representa el elemento situado en la intersección de la i -ésima fila con la j -ésima columna de V .

Así, p. ej., sea

$$V = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{pmatrix} \quad \text{e} \quad Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{pmatrix} \quad (1)$$

V es una matriz $p \times p$ cuyo elemento ij -ésimo es σ_{ij} ; a veces escribiremos $V = (\sigma_{ij})$. Y es una matriz $p \times 1$, cuyo elemento i -ésimo es y_i . (Si una matriz tiene una única fila o una única columna, se dice que es un vector.)

Emplearemos el símbolo $\mathbf{0}$ para la matriz nula, en la que todos sus elementos son cero. I representará la matriz identidad; esto es, una matriz cuadrada tal que $\sigma_{ii} = 1$ y $\sigma_{ij} = 0$ si $i \neq j$. Con el símbolo V' se indicará la traspuesta de la matriz V , obtenida cambiando filas por columnas. V^{-1} representará la inversa de la matriz V ; es decir, una matriz cuadrada tal que $VV^{-1} = V^{-1}V = I$. $|V|$ designará el determinante de la matriz cuadrada V . Si $V = V'$, la matriz se llama simétrica.

Haremos un extenso empleo de las formas cuadráticas.

Definición 9-3.—Sea Y un vector $p \times 1$ con elementos y_i , y V una matriz simétrica $p \times p$ con elementos σ_{ij} . Entonces $Y'VY$ es una forma cuadrática en los elementos y_i , y se dice que V es la matriz de la forma cuadrática, en donde

$$Y'VY = \sum_{i=1}^p \sum_{j=1}^p y_i y_j \sigma_{ij} \quad (2)$$

Para una matriz general V , la forma cuadrática puede ser positiva, negativa o nula. Sin embargo, estaremos especialmente interesados en una clase de matrices para las cuales la forma cuadrática es siempre positiva, excepto cuando Y es el vector nulo.

Definición 9-4.—Si la forma cuadrática $Y'VY$ es positiva para todo vector Y no nulo, se dice que $Y'VY$ es una forma cuadrática definida positiva, y V se denomina matriz (simétrica) definida positiva.

Para determinar si una forma cuadrática es definida positiva, utilizaremos el teorema siguiente.

Teorema 9-4.—Sea V una matriz simétrica $p \times p$ representada por (1). Una condición necesaria y suficiente para que V sea una matriz simétrica definida positiva es que los p determinantes si-

guientes sean positivos:

$$\sigma_{11}; \begin{vmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{vmatrix}; \begin{vmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{vmatrix}; \dots; \begin{vmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2p} \\ \vdots & \vdots & & \vdots \\ \sigma_{p1} & \sigma_{p2} & \dots & \sigma_{pp} \end{vmatrix} \quad (3)$$

Por ejemplo, si

$$V = \begin{pmatrix} 3 & 2 & 0 \\ 2 & 2 & 1 \\ 0 & 1 & 2 \end{pmatrix}$$

entonces

$$\sigma_{11} = 3 > 0 \quad \begin{vmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{vmatrix} = \begin{vmatrix} 3 & 2 \\ 2 & 2 \end{vmatrix} = 2 > 0 \quad \text{y} \quad |V| = 1 > 0$$

luego V es definida positiva.

También, si $Y' = (y_1, y_2, y_3)$

$$\begin{aligned} Y'VY &= (y_1, y_2, y_3) \begin{pmatrix} 3 & 2 & 0 \\ 2 & 2 & 1 \\ 0 & 1 & 2 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} \\ &= 3y_1^2 + 2y_2^2 + 2y_3^2 + 4y_1y_2 + 2y_2y_3 \end{aligned}$$

Puede demostrarse que

$$V^{-1} = \begin{pmatrix} 3 & -4 & 2 \\ -4 & 6 & -3 \\ 2 & -3 & 2 \end{pmatrix}$$

Es con frecuencia útil descomponer una matriz en submatrices tales como

$$V = \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix} \quad (4)$$

donde V es de orden $p \times p$, V_{11} es $k \times k$, V_{12} es $k \times (p-k)$, V_{21} es $(p-k) \times k$ y V_{22} es $(p-k) \times (p-k)$. Obsérvese también que si V es simétrica, $V'_{12} = V_{21}$. Si V es simétrica definida positiva, también V_{11} es simétrica definida positiva, lo mismo que V_{22} , y, por tanto, existen V_{11}^{-1} y V_{22}^{-1} . En general, V_{12} no tiene inversa. Sea R otra matriz $p \times p$ que descompondremos en submatrices del mismo orden que las de V . Al multiplicar, se tiene

$$VR = \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix} \begin{pmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{pmatrix} = \begin{pmatrix} V_{11}R_{11} + V_{12}R_{21} & V_{11}R_{12} + V_{12}R_{22} \\ V_{21}R_{11} + V_{22}R_{21} & V_{21}R_{12} + V_{22}R_{22} \end{pmatrix} \quad (5)$$

Supongamos que R es la inversa de V ; por tanto, $R^{-1} = V$ y $VR = I$. Si V es simétrica definida positiva, también R es simétrica definida positiva y existen R_{11}^{-1} y R_{22}^{-1} , pero, en general, R_{12} carece de inversa. Particionemos ahora la matriz identidad $p \times p$ de tal forma que $VR = I$ se transforme en

$$\begin{pmatrix} V_{11}R_{11} + V_{12}R_{21} & V_{11}R_{12} + V_{12}R_{22} \\ V_{21}R_{11} + V_{22}R_{21} & V_{21}R_{12} + V_{22}R_{22} \end{pmatrix} = \begin{pmatrix} I & \mathbf{0} \\ \mathbf{0} & I \end{pmatrix}$$

de donde

$$V_{11}R_{11} + V_{12}R_{21} = I \quad \text{y} \quad V_{11}R_{12} + V_{12}R_{22} = \mathbf{0}$$

de modo que

$$V_{12} = -V_{11}R_{12}R_{22}^{-1};$$

sustituyendo, se obtiene

$$V_{11}R_{11} - V_{11}R_{12}R_{22}^{-1}R_{21} = I$$

Multiplicando por V_{11}^{-1} resulta:

$$V_{11}^{-1} = R_{11} - R_{12}R_{22}^{-1}R_{21}$$

Utilizando otras ecuaciones, se llega al siguiente teorema.

Teorema 9-5.—Sea $V^{-1} = R$ y particionemos las dos matrices simétricas definidas positivas como en (5); se obtienen así las siguientes ecuaciones matriciales

$$\begin{array}{ll} a) \quad V_{11}^{-1} = R_{11} - R_{12}R_{22}^{-1}R_{21} & d) \quad R_{22}^{-1} = V_{22} - V_{21}V_{11}^{-1}V_{12} \\ b) \quad V_{22}^{-1} = R_{22} - R_{21}R_{11}^{-1}R_{12} & e) \quad V_{11}^{-1}V_{12} = -R_{12}R_{22}^{-1} \\ c) \quad R_{11}^{-1} = V_{11} - V_{12}V_{22}^{-1}V_{21} & f) \quad V_{22}^{-1}V_{21} = -R_{21}R_{11}^{-1} \end{array} \quad (6)$$

Si V es simétrica definida positiva, el determinante de V puede escribirse así

$$\begin{aligned} |V| &= |V_{22}| |V_{11} - V_{12}V_{22}^{-1}V_{21}| \\ &= |V_{11}| |V_{22} - V_{21}V_{11}^{-1}V_{12}| \end{aligned}$$

También

$$\begin{aligned} |R| &= |R_{11}| |R_{22} - R_{21}R_{11}^{-1}R_{12}| \\ &= |R_{22}| |R_{11} - R_{12}R_{22}^{-1}R_{21}| \end{aligned} \quad (7)$$

Particionemos la matriz del ejemplo anterior de manera que

$$V_{11} = 3$$

Entonces,

$$V_{12} = (2, 0) \quad V_{21} = \begin{pmatrix} 2 \\ 0 \end{pmatrix} \quad V_{22} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$$

Sea $R = V^{-1}$ y descompongámosla de manera que

$$R_{11} = 3 \\ R_{12} = (-4, 2) \quad R_{21} = \begin{pmatrix} -4 \\ 2 \end{pmatrix} \quad R_{22} = \begin{pmatrix} 6 & -3 \\ -3 & 2 \end{pmatrix}$$

Puesto que $RV = I$, la parte a) del teorema 9-5 se demuestra así:

$$\begin{aligned} R_{11} - R_{12}R_{22}^{-1}R_{21} &= 3 - (-4, 2) \begin{pmatrix} 6 & -3 \\ -3 & 2 \end{pmatrix}^{-1} \begin{pmatrix} -4 \\ 2 \end{pmatrix} \\ &= 3 - (-4, 2) \begin{pmatrix} 2/3 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} -4 \\ 2 \end{pmatrix} = 3 - 8/3 = 1/3 \end{aligned}$$

Pero $V_{11} = 3$, luego $V_{11}^{-1} = 1/3$, lo que prueba la parte a) del teorema 9-5.

También $|R| = 1$, pero

$$R_{11} - R_{12}R_{22}^{-1}R_{21} = 1/3$$

y

$$|R_{22}| = 12 - 9 = 3$$

luego

$$|R_{22}| \cdot |R_{11} - R_{12}R_{22}^{-1}R_{21}| = 3 \cdot 1/3 = 1$$

Se deja al cuidado del lector la demostración de las restantes relaciones del teorema 9-5.

Otro resultado importante de los determinantes es que si $V^{-1} = R$, $|R| = 1/|V|$.

*** 9-3. Distribución normal multivariante.**—Sea $(y_1, y_2, \dots, y_p)$ una variable aleatoria p -dimensional que designaremos como elementos de un vector Y por

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{pmatrix} \quad (1)$$

Llamaremos a Y vector $p \times 1$ aleatorio.

* Véase explicación de Sec. 9-2.

Definición 9-5.—El vector aleatorio Y se distribuye según la normal p -variante si la densidad conjunta de $y_1, y_2, \dots, y_p$ es

$$f(Y) = f(y_1, y_2, \dots, y_p) = \frac{|R|^{\frac{1}{2}}}{(2\pi)^{p/2}} e^{-\frac{1}{2}(Y-\mu)'R(Y-\mu)} \quad -\infty < y_i < \infty \quad (2)$$

$$i = 1, 2, \dots, p$$

donde

a) R es una matriz simétrica definida positiva, cuyos elementos r_{ij} son constantes (no variables aleatorias).

b) μ es un vector $p \times 1$, cuyos elementos μ_i son constantes.

Observemos que si $p=1$, $R=r_{11}$ que, por la parte a), tiene que ser positivo. Si hacemos $r_{11}=1/\sigma^2$, es evidente que la densidad (2) es la distribución normal definida en el capítulo 6. La cantidad $Q=(Y-\mu)'R(Y-\mu)$ se llama forma cuadrática de la normal p -variante. Es una forma cuadrática en los elementos $y_i - \mu_i$, y puede escribirse así:

$$Q = \sum_{j=1}^p \sum_{i=1}^p (y_i - \mu_i)(y_j - \mu_j)r_{ij} \quad (3)$$

Para probar que (2) satisface las propiedades que la califican como una densidad, hay que demostrar: 1) que $f(Y) \geq 0$, lo que es evidente puesto que R , determinante de una matriz definida positiva, es positivo; 2) que la integral de $f(Y)$ es igual a 1. No demostraremos la condición (2), pero una consecuencia es que

$$\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} e^{-\frac{1}{2}(Y-\mu)'R(Y-\mu)} dy_1, dy_2, \dots, dy_p = \frac{(2\pi)^{p/2}}{|R|^{\frac{1}{2}}} \quad (4)$$

lo que proporciona el siguiente teorema:

Teorema 9-6.—La integral múltiple

$$\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} e^{-\frac{1}{2}(Y-\mu)'R(Y-\mu)} dy_1 \dots dy_p \quad (5)$$

es igual a $(2\pi)^{p/2}|R|^{-1/2}$ y no depende del vector μ .

Para hallar la distribución marginal de una de las variables aleatorias de Y , p. ej., y_1 , tenemos

$$g(y_1) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} |R|^{\frac{1}{2}} (2\pi)^{-p/2} e^{-\frac{1}{2}(Y-\mu)'R(Y-\mu)} dy_2 dy_3 \dots dy_p$$

Si particionamos los vectores y la matriz del exponente, se obtiene

$$(Y - \mu)'R(Y - \mu) = \begin{pmatrix} y_1 - \mu_1 \\ Y_2 - U_2 \end{pmatrix}' \begin{pmatrix} r_{11} & R_{12} \\ R_{21} & R_{22} \end{pmatrix} \begin{pmatrix} y_1 - \mu_1 \\ Y_2 - U_2 \end{pmatrix} \quad (6)$$

donde Y_2 y U_2 son vectores $(p-1) \times 1$, R_{12} es $1 \times (p-1)$ y R_{22} es $(p-1) \times (p-1)$. Multiplicando resulta

$$\begin{aligned} (Y - \mu)'R(Y - \mu) &= (y_1 - \mu_1)r_{11}(y_1 - \mu_1) + (y_1 - \mu_1)R_{12}(Y_2 - U_2) \\ &\quad + (Y_2 - U_2)'R_{21}(y_1 - \mu_1) + (Y_2 - U_2)'R_{22}(Y_2 - U_2) \end{aligned} \quad (7)$$

Puesto que R es simétrica definida positiva, se deduce que $R'_{12} = R_{21}$, que existe R_{22}^{-1} y es simétrica, y que $r_{11} > 0$. Podemos escribir (7) así

$$\begin{aligned} (Y - \mu)'R(Y - \mu) &= (y_1 - \mu_1)(r_{11} - R_{12}R_{22}^{-1}R_{21})(y_1 - \mu_1) \\ &\quad + [(Y_2 - U_2) + R_{22}^{-1}R_{21}(y_1 - \mu_1)]'R_{22}[(Y_2 - U_2) + R_{22}^{-1}R_{21}(y_1 - \mu_1)] \end{aligned} \quad (8)$$

La ecuación (8) puede comprobarse por multiplicación, y comparando término a término con (7). Mediante (8), la densidad marginal de y_1 se escribe

$$g(y_1) = |R|^{\frac{1}{2}} (2\pi)^{-p/2} e^{-(y_1 - \mu_1)^2 / 2\sigma^2} \cdot F \quad (9)$$

donde F es la integral múltiple

$$F = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} e^{-\frac{1}{2}(Y_2 - h)'R_{22}(Y_2 - h)} dy_2 \dots dy_p$$

donde

$$1/\sigma^2 = r_{11} - R_{12}R_{22}^{-1}R_{21} \quad \text{y} \quad h = U_2 - R_{22}^{-1}R_{21}(y_1 - \mu_1).$$

Por (4), se tiene

$$F = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} e^{-\frac{1}{2}(Y_2 - h)'R_{22}(Y_2 - h)} dy_2 dy_3 \dots dy_p = (2\pi)^{(p-1)/2} |R_{22}|^{-\frac{1}{2}}$$

y, por tanto,

$$g(y_1) = \frac{|R|^{\frac{1}{2}}}{|R_{22}|^{\frac{1}{2}} (2\pi)^{\frac{1}{2}}} e^{-(y_1 - \mu_1)^2 / 2\sigma^2}$$

Por la ecuación (9-2-7),

$$|R| = |R_{22}| \cdot |r_{11} - R_{12}R_{22}^{-1}R_{21}|$$

y, en consecuencia,

$$g(y_1) = \frac{1}{\sigma \sqrt{2\pi}} e^{-(y_1 - \mu_1)^2 / 2\sigma^2}$$

que, por la sección 6-2, indica que y_1 es una variable normal con media μ_1 y varianza $\sigma^2 = (r_{11} - R_{12}R_{22}^{-1}R_{21})^{-1}$. Hemos utilizado la variable y_1 , pero puede hacerse una prueba similar para cualquier variable y_i . Queda así demostrado el siguiente teorema.

Teorema 9-7.—*Supongamos que $\mathbf{Y}$ se distribuye según una normal p -variante con densidad dada por (2). La densidad marginal de y_1 es normal, con media μ_1 y varianza $(r_{11} - R_{12}R_{22}^{-1}R_{21})^{-1}$.*

En lugar de la densidad marginal de una sola variable aleatoria y_1 , puede obtenerse la densidad conjunta de k variables aleatorias de $\mathbf{Y}$ reemplazando en la demostración anterior y_1 por el vector $k \times 1$ $\mathbf{Y}_1$, μ_1 por $\mathbf{U}_1$ y r_{11} por R_{11} .

Corolario 9-7-1.—*Supongamos que el vector aleatorio $p \times 1$ $\mathbf{Y}$ es normal, con densidad dada por (2). Sean $\mathbf{Y}_1$ el vector formado por las k primeras componentes de $\mathbf{Y}$, $\mathbf{U}_1$ el formado con las k primeras componentes de $\boldsymbol{\mu}$, y particionemos R como sigue*

$$R = \begin{pmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{pmatrix}$$

donde R_{11} es $k \times k$. Entonces $\mathbf{Y}_1$ se distribuye según una normal k -variante con densidad

$$f(\mathbf{Y}_1) = \frac{|R_{11} - R_{12}R_{22}^{-1}R_{21}|^{\frac{1}{2}}}{(2\pi)^{k/2}} e^{-\frac{1}{2}(\mathbf{Y}_1 - \mathbf{U}_1)'(R_{11} - R_{12}R_{22}^{-1}R_{21})(\mathbf{Y}_1 - \mathbf{U}_1)}$$

Más adelante examinaremos los momentos de la normal p -variante y, en particular, nos interesarán los momentos primeros y segundos; no obstante, definiremos antes el valor esperado de una matriz.

Definición 9-6.—*El valor esperado de una matriz o vector $\mathbf{A}$, que denotaremos con el símbolo $E(\mathbf{A})$, se definirá como el valor esperado de cada elemento de $\mathbf{A}$.*

Así, p. ej., si

$$\mathbf{A} = \begin{pmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \end{pmatrix}$$

entonces

$$E(\mathbf{A}) = \begin{pmatrix} E(y_{11}) & E(y_{12}) \\ E(y_{21}) & E(y_{22}) \end{pmatrix} \quad (10)$$

Los momentos primeros de la normal p -variante, $E(\mathbf{y}_i)$, son los momentos de las respectivas distribuciones marginales. Escribiremos

$$E(\mathbf{Y}) = \begin{pmatrix} E(y_1) \\ E(y_2) \\ \vdots \\ E(y_p) \end{pmatrix} \quad (11)$$

lo que nos permite examinar simultáneamente los momentos primeros de cada componente.

Teorema 9-8.—Supongamos que $\mathbf{Y}$, vector aleatorio $p \times 1$, se distribuye según la normal p -variante; entonces $E(\mathbf{Y}) = \boldsymbol{\mu}$.

Demostración.—En virtud de (11), debemos demostrar que $E(y_i) = \mu_i$. La definición de valor esperado de la variable aleatoria y_i da (de nuevo utilizamos y_1 en lugar de y_i , pero la demostración es general)

$$\begin{aligned} E(y_1) &= \frac{|R|^{\frac{1}{2}}}{(2\pi)^{p/2}} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} y_1 e^{-\frac{1}{2}(\mathbf{Y}-\boldsymbol{\mu})'R(\mathbf{Y}-\boldsymbol{\mu})} dy_1 dy_2 \dots dy_p \\ &= \int_{-\infty}^{\infty} y_1 \left[\frac{|R|^{\frac{1}{2}}}{(2\pi)^{p/2}} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} e^{-\frac{1}{2}(\mathbf{Y}-\boldsymbol{\mu})'R(\mathbf{Y}-\boldsymbol{\mu})} dy_2 \dots dy_p \right] dy_1 \end{aligned}$$

Pero por el teorema 9-7 el término entre corchetes es la densidad marginal de y_1 ; luego

$$E(y_1) = \int_{-\infty}^{\infty} \frac{y_1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2\sigma^2} (y_1 - \mu_1)^2} dy_1 = \mu_1$$

con lo que queda demostrado.

La varianza de la variable aleatoria y_i es, por definición, $E[y_i - E(y_i)]^2$, o sea $E(y_i - \mu_i)^2$, y la covarianza de las dos variables aleatorias y_i e y_j es $E(y_i - \mu_i)(y_j - \mu_j)$, $i \neq j$. En una normal p -variante habrá p varianzas, una para cada variable aleatoria y_i , y $p(p-1)/2$ covarianzas. Definiremos una matriz $p \times p$ que tenga, como elemento i, j -ésimo la covarianza de y_i e y_j , si $i \neq j$, y como elemento i -ésimo de su diagonal la varianza de y_i . Así,

$$V = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2p} \\ \vdots & \vdots & & \vdots \\ \sigma_{p1} & \sigma_{p2} & \dots & \sigma_{pp} \end{pmatrix} \quad (12)$$

donde $\sigma_{ij} = \sigma_{ji} = E(\mathbf{y}_i - \mu_i)(\mathbf{y}_j - \mu_j)$. Esta matriz puede escribirse también así

$$V = E(\mathbf{Y} - \mu)(\mathbf{Y} - \mu)' \quad (13)$$

V recibe el nombre de matriz de covarianzas del vector $\mathbf{Y}$, existiendo una notable relación en la normal p -variante entre la matriz R y la matriz V .

Teorema 9-9.—*En la normal p -variante, la matriz R es la inversa de la matriz de covarianzas V ; es decir, $V^{-1} = R$. Naturalmente, también $R^{-1} = V$.*

Omitimos la demostración de este teorema, pero es un resultado muy importante. En lo sucesivo escribiremos la función de densidad normal p -variante en la forma

$$\frac{1}{|V|^{\frac{1}{2}}(2\pi)^{p/2}} e^{-\frac{1}{2}(\mathbf{Y}-\mu)'V^{-1}(\mathbf{Y}-\mu)}$$

Si el elemento ij -ésimo ($i \neq j$) de la matriz de covarianzas es 0, las dos variables aleatorias y_i e y_j son independientes. No demostraremos esta proposición, que nos será útil más adelante. Evidentemente, el elemento ij -ésimo es 0 si, y solamente si, la correlación entre las dos variables y_i e y_j es 0, puesto que, por definición, la correlación es

$$\rho_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}}$$

Teorema 9-10.—*Supongamos que $\mathbf{Y}$ tiene la distribución normal p -variante con media μ y matriz de covarianzas V , y sea $\alpha_1, \dots, \alpha_p$ un conjunto de constantes. En tal supuesto,*

$$z = \sum_{i=1}^p \alpha_i y_i$$

se distribuye según la normal univariante, con media $\sum \alpha_i \mu_i$ y varianza

$$\sum \alpha_i^2 \sigma_i^2 + \sum_i \sum_{j \neq i} \alpha_i \alpha_j \sigma_{ij}$$

Ejemplo 9-1.—Como ejemplo para ilustrar los teoremas y definiciones anteriores, supongamos que el vector 3×1 , $\mathbf{Y}$, tiene una densidad normal p -variante, en la que

$$\mu = \begin{pmatrix} 3 \\ -1 \\ 0 \end{pmatrix} \quad R = \begin{pmatrix} 2 & 0 & 3 \\ 0 & 1 & 0 \\ 3 & 0 & 5 \end{pmatrix}$$

La matriz de covarianzas es

$$R^{-1} = V = \begin{pmatrix} 5 & 0 & -3 \\ 0 & 1 & 0 \\ -3 & 0 & 2 \end{pmatrix}$$

La media de y_2 , p. ej., es -1 , y su varianza, 1. La covarianza de y_1 e y_3 es -3 , e y_1 e y_2 son independientes. La media de

$$z = y_1 - 3y_2 + 2y_3$$

es

$$1(3) - 3(-1) + 2(0) = 6.$$

La varianza de z es

$$(1)^2(5) + (-3)^2(1) + (2)^2(2) + 2[(1)(-3)(0) + (1)(2)(-3) + (-3)(2)(0)] = 10$$

En virtud del teorema 9-9 y de (12), la varianza de y_1 es $\sigma_{11} = 5$. Pero por el teorema 9-7, la varianza de y_1 es $(r_{11} - R_{12}R_{22}^{-1}R_{21})^{-1}$. Según esto,

$$r_{11} - R_{12}R_{22}^{-1}R_{21} = 2 - (0, 3) \begin{pmatrix} 1 & 0 \\ 0 & 1/5 \end{pmatrix} \begin{pmatrix} 0 \\ 3 \end{pmatrix} = 2 - 9/5 = 1/5$$

y, por tanto, $\sigma_{11} = 5$, igual que antes.

Particionemos el vector $p \times 1$, Y , el vector $p \times 1$, μ , la matriz V y la matriz R en la forma siguiente:

$$Y = \begin{pmatrix} Y_1^* \\ Y_2^* \end{pmatrix} \quad \mu = \begin{pmatrix} U_1 \\ U_2 \end{pmatrix} \quad R = \begin{pmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{pmatrix} \quad V = \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix} \quad (14)$$

donde

$$Y_1^* = \begin{pmatrix} y_1 \\ \vdots \\ y_k \end{pmatrix} \quad U_1 = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_k \end{pmatrix}$$

R_{11} y V_{11} son ambas de orden $k \times k$, y los órdenes de las restantes matrices están determinados. La distribución condicional de $y_1, y_2, \dots, y_k$, dadas $y_{k+1}, \dots, y_p$, es, por definición,

$$h(y_1, \dots, y_k | y_{k+1}, \dots, y_p) = \frac{f(y_1, \dots, y_p)}{g(y_{k+1}, \dots, y_p)}$$

Algunas veces escribiremos esta densidad así:

$$h(Y_1^* | Y_2^*) = \frac{f(Y)}{g(Y_2^*)} \quad (15)$$

Por el corolario 9-7-1, Y_2^* es una normal $(p-k)$ -variante, con media U_2 y covarianza V_{22} . Luego (15) se transforma en

$$\begin{aligned} h(Y_1^* | Y_2^*) &= \frac{|V|^{-1/2} (2\pi)^{-p/2} e^{-\frac{1}{2}(Y-\mu)'V^{-1}(Y-\mu)}}{|V_{22}|^{-1/2} (2\pi)^{-(p-k)/2} e^{-\frac{1}{2}(Y_2^*-U_2)'V_{22}^{-1}(Y_2^*-U_2)}} \\ &= |V_{11} - V_{12}V_{22}^{-1}V_{21}|^{-1/2} (2\pi)^{-k/2} e^{-\frac{1}{2}[(Y-\mu)'V^{-1}(Y-\mu) - (Y_2^*-U_2)'V_{22}^{-1}(Y_2^*-U_2)]} \end{aligned}$$

Utilizando (14), la forma cuadrática del exponente puede escribirse

$$\begin{aligned} &(Y-\mu)'R(Y-\mu) - (Y_2^*-U_2)'V_{22}^{-1}(Y_2^*-U_2) \\ &= \begin{pmatrix} Y_1^*-U_1 \\ Y_2^*-U_2 \end{pmatrix}' \begin{pmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{pmatrix} \begin{pmatrix} Y_1^*-U_1 \\ Y_2^*-U_2 \end{pmatrix} \\ &\quad - (Y_2^*-U_2)'(R_{22} - R_{21}R_{11}^{-1}R_{12})(Y_2^*-U_2) \\ &= [Y_1^*-U_1 - V_{12}V_{22}^{-1}(Y_2^*-U_2)]'R_{11}[Y_1^*-U_1 - V_{12}V_{22}^{-1}(Y_2^*-U_2)] \\ &= (Y_1^*-U_1^*)(R_{11}^{-1})^{-1}(Y_1^*-U_1^*) \quad (16) \end{aligned}$$

donde

$$U_1^* = U_1 + V_{12}V_{22}^{-1}(Y_2^*-U_2).$$

Queda así probado el teorema siguiente:

Teorema 9-11.—Supongamos que Y , vector $p \times 1$, se distribuye según la normal p -variante, con media μ y matriz de covarianzas V , y particionemos estos como en (14). La distribución condicional de Y_1^* , dado Y_2^* , es la normal k -variante con media

$$U_1^* = U_1 + V_{12}V_{22}^{-1}(Y_2^*-U_2)$$

y matriz de covarianzas

$$R_{11}^{-1} = V_{11} - V_{12}V_{22}^{-1}V_{21}$$

A veces designaremos R_{11}^{-1} por $V_{11.2}$. Obsérvese que la matriz de covarianzas de Y_1^* , dado Y_2^* , no depende del valor que tenga Y_2^* . El elemento ii -ésimo de $V_{11.2}$ se llama varianza de y_i ($i \leq k$), dados $y_{k+1}, \dots, y_p$, y se representa por $\sigma_{ii \cdot (k+1) \dots p}$. Análogamente, el elemento ij -ésimo ($i \neq j$) de $V_{11.2}$ se representará por $\sigma_{ij \cdot (k+1) \dots p}$, donde

los subíndices que siguen al punto son los que figuran en Y_2^* . La correlación parcial de y_i e y_j , ($i, j < k$), dados $y_{k+1}, \dots, y_p$, se define en la forma siguiente:

$$\rho_{ij.(k+1)\dots p} = \frac{\sigma_{ij.(k+1)\dots p}}{\sqrt{\sigma_{ii.(k+1)\dots p} \sigma_{jj.(k+1)\dots p}}} \quad (17)$$

Daremos algunas reglas que serán útiles para hallar las distintas distribuciones marginales y condicionales.

μ y V dados:

Regla 1.—Para hallar la distribución marginal de $Y_1^* = (y_1, y_2, \dots, y_k)$, tachemos las $p-k$ últimas filas y columnas de V y los $p-k$ últimos elementos de μ . El vector U_1 y la matriz V_{11} resultantes son el vector media y la matriz de covarianzas de la variable aleatoria $(y_1, y_2, \dots, y_k)$. Si las variables no son las k primeras, se permutan los elementos de Y hasta que las variables aleatorias deseadas coincidan con las k primeras, y se procede como antes.

Regla 2.—Para hallar la densidad condicional de Y_1^* , dado Y_2^* , se calculan $V_{12}V_{22}^{-1}$ y $V_{12}V_{22}^{-1}V_{21}$. Luego se calculan

$$U_1^* = U_1 + V_{12}V_{22}^{-1}(Y_2^* - U_2)$$

y

$$V_{11.2} = V_{11} - V_{12}V_{22}^{-1}V_{21}.$$

La distribución condicional de Y_1^* , dado Y_2^* , es la normal k -variante con vector media U_1^* y matriz de covarianzas $V_{11.2}$.

Ejemplo 9-2.—Supongamos que el vector Y (4×1) tiene vector media μ y matriz de covarianzas V , donde

$$\mu = \begin{pmatrix} 1 \\ 0 \\ -2 \\ 3 \end{pmatrix} \quad V = \begin{pmatrix} 2 & 0 & 1 & 0 \\ 0 & 5 & 0 & -3 \\ 1 & 0 & 1 & 0 \\ 0 & -3 & 0 & 2 \end{pmatrix}$$

Para hallar la densidad marginal de $Y_1^* = (y_1, y_2)$, tenemos

$$U_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad V_{11} = \begin{pmatrix} 2 & 0 \\ 0 & 5 \end{pmatrix}$$

Luego Y_1^* se distribuye según la normal bivalente con media U_1 y covarianza V_{11} . Obsérvese que y_1 e y_2 son independientes, puesto que el elemento de V situado en la primera fila y segunda columna es 0. Naturalmente, también sucede lo mismo en V_{11} . Para hallar la distribución condicional de y_1, y_4 , dadas y_2, y_3 , permutamos

primero y_2 , y_3 e y_4 para tener un nuevo vector $Y' = (y_1, y_4, y_2, y_3)$ que tiene media y covarianza

$$\mu = \begin{pmatrix} 1 \\ 3 \\ 0 \\ -2 \end{pmatrix} \quad V = \begin{pmatrix} 2 & 0 & 0 & 1 \\ 0 & 2 & -3 & 0 \\ 0 & -3 & 5 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix} \quad (18)$$

y

$$U_1 = \begin{pmatrix} 1 \\ 3 \end{pmatrix} \quad U_2 = \begin{pmatrix} 0 \\ -2 \end{pmatrix} \quad V_{11} = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} \\ V_{22}^{-1} = \begin{pmatrix} 1/5 & 0 \\ 0 & 1 \end{pmatrix} \quad V_{12} = \begin{pmatrix} 0 & 1 \\ -3 & 0 \end{pmatrix}$$

y

$$V_{11.2} = V_{11} - V_{12}V_{22}^{-1}V_{21} = \begin{pmatrix} 1 & 0 \\ 0 & 1/5 \end{pmatrix} \\ U_1^* = U_1 + V_{12}V_{22}^{-1}(Y_2^* - U_2) = \begin{pmatrix} y_3 + 3 \\ -3/5y_2 + 3 \end{pmatrix}$$

La densidad condicional de (y_1, y_4) , dadas (y_2, y_3) , es la normal bivalente con media U_1^* y covarianza $V_{11.2}$.

μ y R dados:

Regla 3.—Para hallar las distribuciones marginales, se halla $V = R^{-1}$ y se procede según la regla 1, o se halla V_{11} por la igualdad $V_{11} = (R_{11} - R_{12}R_{22}^{-1}R_{21})^{-1}$, según resulte más fácil.

Regla 4.—Para hallar las distribuciones condicionales, se puede utilizar la ecuación $V_{12}V_{22}^{-1} = -R_{11}^{-1}R_{12}$ para calcular U_1^* . La matriz de covarianzas es R_{11}^{-1} .

Ejemplo 9-3.—En el ejemplo anterior, hállese la distribución condicional de y_1 , y_4 , dadas y_2 , y_3 , cuando se conocen μ y R . Tenemos:

$$\mu = \begin{pmatrix} 1 \\ 3 \\ 0 \\ -2 \end{pmatrix} \quad R = \begin{pmatrix} 1 & 0 & 0 & -1 \\ 0 & 5 & 3 & 0 \\ 0 & 3 & 2 & 0 \\ -1 & 0 & 0 & 2 \end{pmatrix} \quad (19) \\ R_{11} = \begin{pmatrix} 1 & 0 \\ 0 & 5 \end{pmatrix}$$

y, por tanto,

$$V_{11.2} = R_{11}^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & 1/5 \end{pmatrix}$$

y

$$U_1^* = U_1 - R_{11}^{-1} R_{12} (Y_2^* - U_2) = \begin{pmatrix} y_3 + 3 \\ -3/5 y_2 + 3 \end{pmatrix}$$

igual que cuando se emplea la regla 2. Obsérvese en $V_{11.2}$ que $\sigma_{11.23} = 1$, $\sigma_{44.23} = 1/5$, $\sigma_{14.23} = 0$ y $\rho_{14.23} = 0$. También, en la V dada en (18), $\sigma_{11} = 2$ y $\sigma_{44} = 2$ y, por tanto, $\sigma_{11.23} < \sigma_{11}$ y $\sigma_{44.23} < \sigma_{44}$. Este resultado se establece mediante el teorema siguiente.

Teorema 9-12.—*En la distribución condicional de $y_1, \dots, y_k$, dadas $y_{k+1}, \dots, y_p$, las varianzas condicionales son menores o iguales que las varianzas marginales; es decir, $\sigma_{ii.(k+1) \dots p} \leq \sigma_{ii}$, donde $i = 1, 2, \dots, k$.*

La función generatriz de momentos de la normal p -variante es

$$\begin{aligned} m(t_1, t_2, \dots, t_p) &= E(e^{t_1 y_1 + t_2 y_2 + \dots + t_p y_p}) \\ &= |V|^{-1/2} (2\pi)^{-p/2} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} e^{\sum t_i y_i} e^{-\frac{1}{2}(Y-\mu)' V^{-1}(Y-\mu)} \\ &\quad dy_1 dy_2 \dots dy_p \quad (20) \end{aligned}$$

El exponente puede escribirse [siendo el vector $1 \times p$ $T' = (t_1, \dots, t_p)$]:

$$-1/2 \{ [Y - (\mu + VT)]' V^{-1} [Y - (\mu + VT)] - 2T'\mu - T'VT \}$$

Luego (20) se escribirá

$$m(T) = e^{T'\mu + \frac{1}{2}T'VT} \left[|V|^{-1/2} (2\pi)^{-p/2} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} e^{-\frac{1}{2}(Y-\mu-VT)' V^{-1}(Y-\mu-VT)} dy_1 \dots dy_p \right] = e^{T'\mu + \frac{1}{2}T'VT}$$

Teorema 9-13.—*Si el vector Y se distribuye normalmente con media μ y matriz de covarianzas V , la f. g. m. de Y es*

$$m(T) = m(t_1, \dots, t_p) = e^{T'\mu + \frac{1}{2}T'VT} = e^{\sum_i \mu_i t_i + \frac{1}{2} \sum_i \sum_j t_i t_j \sigma_{ij}}$$

Para las demostraciones y ampliación de las cuestiones tratadas en esta sección, el lector puede consultar la obra 2 de la bibliografía reseñada al final del capítulo.

PROBLEMAS

1. Demuéstrese que las curvas de nivel de la distribución normal bi-variante [o sea, las curvas para las cuales se verifica $f(x, y) = \text{constante}$] son elipses.

2. Pruébese que cualquier plano perpendicular al plano x, y corta a la superficie normal según una curva de forma normal.

3. Si la forma cuadrática de una densidad normal bivalente es

$$1/3[6(x+1)^2 - 2(x+1)(y-2) + (y-2)^2],$$

¿cuáles son las medias, varianzas y covarianzas de las variantes?

4. ¿Cuál es la función generatriz de momentos de la distribución definida en el problema 3?

5. ¿Cuál es la función generatriz de momentos para los momentos respecto a la media en una distribución normal bivalente?

*6. Hállese la inversa de la matriz
$$\begin{pmatrix} 3 & 1 & 0 & 0 \\ 1 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 4 \end{pmatrix}$$

*7. Obténganse las varianzas y covarianzas de las variantes normales en cuya distribución interviene la forma cuadrática

$$2y_1^2 + y_2^2 + 4y_3^2 - y_1y_2 - 2y_1y_3$$

*8. ¿Cuál es la función de densidad marginal de y_2 en el problema anterior?

*9. ¿Cuál es, en el problema 7, la densidad condicional de y_1 , dados y_2, y_3 ?

*10. Si la matriz del problema 6 es la matriz R de una distribución normal de (y_1, y_2, y_3, y_4) , demuéstrese que la distribución condicional de y_1 e y_2 es la misma que la distribución marginal de y_1 e y_2 , y, por tanto, que el par (y_1, y_2) se distribuye independientemente del par (y_3, y_4) .

*11. Supongamos que el vector Y (4×1) se distribuye normalmente, con media μ y matriz de covarianzas V , donde

$$\mu = \begin{pmatrix} 6 \\ -1 \\ 0 \\ 3 \end{pmatrix} \quad y \quad V = \begin{pmatrix} 2 & 0 & 3 & 0 \\ 0 & 5 & 0 & 2 \\ 3 & 0 & 5 & 0 \\ 0 & 2 & 5 & 1 \end{pmatrix}$$

Hállense

a) $\sigma_{11,34}$

b) $\sigma_{12,34}$

c) $\rho_{12,34}$

d) σ_{11}

e) ρ_{12}

f) La media en la distribución condicional de y_3 , dado y_4 .

*12. Demuéstrese que el determinante de orden k

$$\begin{vmatrix} a & b & b & \dots & b \\ b & a & b & \dots & b \\ \vdots & \vdots & \vdots & & \vdots \\ \vdots & \vdots & \vdots & & \vdots \\ b & b & b & \dots & a \end{vmatrix}$$

* En los problemas señalados con asterisco se utilizan conceptos dados en las secciones 9-2 y 9-3.

en el cual los elementos de la diagonal principal son iguales a a y los demás iguales a b , tiene por valor

$$(a-b)^{k-1}[a+(k-1)b]$$

Antes de desarrollar el determinante, se resta la segunda fila de la primera, la tercera de la segunda, etc.; después, se suma la primera columna a la segunda, la segunda a la tercera, etc.

13. Supongamos que (x, y) tiene una densidad normal bivalente con medias μ_x y μ_y , y que la matriz de covarianzas es

$$V = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}$$

Hállense R y la forma cuadrática del exponente; pruébese que este último es igual al exponente de la ecuación (9-1-1).

14. En el problema anterior, demuéstrese que $|V|^{-\frac{1}{2}}$ es igual a 2π veces el coeficiente de la ecuación (9-1-1); es decir, $1/\sigma_x\sigma_y(1-\rho^2)^{\frac{1}{2}}$.

15. Dada la muestra aleatoria $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$, procedente de la normal bivalente del problema 13, hállense los estimadores máximo-verosímiles de $\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho$.

16. Dada una muestra, (2,5; 7,0), (4,0; 9,0), (0,4; 1,7), (1,2; 2,0), (0,3; 0,0), (1,5; 3,7), de una población normal bivalente, hállese la estimación máximo-verosímil de la función de regresión para la distribución condicional de y_2 , dada y_1 . Representense las observaciones muestrales y la función de regresión.

*17. Consideremos una densidad multivalente cualquiera $f(y_1, y_2, \dots, y_k)$. Se utilizarán los símbolos siguientes:

$$\text{Medias:} \quad \mu_i = E(y_i).$$

$$\text{Varianzas:} \quad \sigma_{ii} = E[(y_i - \mu_i)^2].$$

$$\text{Covarianzas:} \quad \sigma_{ij} = E[(y_i - \mu_i)(y_j - \mu_j)].$$

$$\text{Correlaciones:} \quad \rho_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}}$$

¿Cuáles son la media y la varianza de la función lineal $x = \sum \alpha_i y_i$ de las y , si $\alpha_i = 1/n$?

*18. Sea α' el vector $\alpha' = (\alpha_1, \alpha_2, \dots, \alpha_p)$. En el problema 17,

$$x = \sum_{i=1}^p \alpha_i y_i$$

se convierte en $x = \alpha'Y$. Demuéstrese que $E(x) = \alpha'\mu$ y que $\text{var}(x) = \alpha'V\alpha$, donde $V = (\sigma_{ij})$ y $\mu = (\mu_i)$.

*19. Con referencia al problema 17, ¿cuál es la correlación entre dos funciones lineales

$$x = \sum \alpha_i y_i = \alpha'Y \quad \text{y} \quad z = \sum b_i y_i = b'Y,$$

donde $x \neq kz$, $b' = (b_1, \dots, b_p)$ y k es una constante cualquiera?

*20. Supongamos que Y se distribuye según una normal p -variante, con media O y matriz de covarianzas I . Demuéstrese que la densidad de Y es la misma que la densidad conjunta de una muestra aleatoria $y_1, \dots, y_p$ de una densidad normal con media O y varianza 1.

*21. Supongamos que el vector Y ($p \times 1$) se distribuye según una normal p -variante con media O y matriz de covarianzas I . Sea A una matriz $q \times p$ de rango $q \leq p$. Pruébese que el vector $q \times 1$, $X^* = AY$, se distribuye según una normal q -variante, con media O y covarianzas AA' . NOTA: Hállese la f. g. m. de X^* y utilícense los teoremas 9-13 y 5-7.

*22. Demuéstrese que el teorema 9-10 es un caso especial del teorema del problema 21.

*23. Supongamos que el vector $p \times 1$, Y , se distribuye según una normal p -variante con media μ y covarianza V . Sea A una matriz $q \times p$ de rango $q \leq p$. Pruébese que el vector $q \times 1$, $X^* = AY$, se distribuye según una normal q -variante con media $A\mu$ y covarianza AVA' .

*24. Una matriz $p \times p$ ortogonal, P , es tal que $P'P = I$; es decir, $P' = P^{-1}$. En el problema 21, si $p = q$ y $A = P$, demuéstrese que X^* e Y tienen la misma distribución.

*25. Sean dos vectores $p \times 1$, α y b , definidos como en los problemas 18 y 19. Supongamos que Y se distribuye según una normal p -variante con media μ y covarianza I . Pruébese que una condición necesaria y suficiente para que $x = \alpha'Y$ y $z = b'Y$ sean independientes es $\alpha'b = 0$.

26. Supongamos que $y_1, y_2, \dots, y_{2k}$ representan puntuaciones relativas a $2k$ preguntas de un *test* de aptitud, y admitamos que la distribución de las puntuaciones es normal, con la misma media y varianza (μ, σ^2) para cada puntuación, y tales que la correlación entre cualquier par de preguntas sea $\rho > 0$. Si

$$y_1 = \sum_{i=1}^k y_{2i-1} \quad \text{e} \quad y_2 = \sum_{i=1}^k y_{2i}$$

son las puntuaciones totales relativas a las preguntas pares e impares, hállese la correlación entre y_1 e y_2 , y demuéstrese que puede aproximarse a la unidad tanto como queramos sin más que alargar suficientemente el *test*.

*27. Supongamos que el vector $p \times 1$, Y , se distribuye según una normal p -variante con media μ y covarianza V , y sean Y_1^* y U_1 particiones $k \times 1$ de Y y μ , respectivamente, y V_{11} una partición $k \times k$ de V . Demuéstrese que los elementos del vector $k \times 1$, $X^* = Y_1^* - U_1 - V_{12}V_{22}^{-1}(Y_2^* - U_2)$, son in-

dependientes de los elementos del vector $(p-k) \times k$, Y_2^* . NOTA: Obsérvese que

$$E(Y_1^* - U_1)(Y_2^* - U_2)' = V_{12}.$$

***28.** Representemos por $y_{1,23}, \dots, y_r$ la desviación de y_1 respecto a su función de regresión en la distribución condicional de y_1 , dados $y_2, y_3, \dots, y_r$. Demuéstrese, para una distribución normal trivariante, que $y_1, y_{2,1}, y_{3,21}$ tienen distribuciones normales independientes.

***29.** Generalícese el resultado del problema anterior a k variantes.

***30.** Supongamos que el vector Y ($p \times 1$) se distribuye según una normal p -variante con media μ y covarianza V . Particionemos Y , μ y V como en la ecuación (9-3-14), con $k=1$. En la distribución condicional de y_1 , dado Y_2^* , la función de regresión es

$$\mu_1 + V_{12}V_{22}^{-1}(Y_2^* - U_2).$$

Sea

$$z = V_{12}V_{22}^{-1}(Y_2^* - U_2).$$

La correlación entre z e y_1 se llama coeficiente de correlación múltiple de y_1 sobre $y_2, \dots, y_p$ y se representa por $R_{1,23 \dots p}$. Demuéstrese que

$$E(z) = 0 \quad \text{y} \quad E(z^2) = E(y_1).$$

***31.** En el problema 30, y para $p=3$, pruébese que

$$\sigma_{11,23} = \sigma_{11}(1 - R_{1,23}^2)$$

***32.** Con referencia al problema 30, demuéstrese que

$$\sigma_{11,23 \dots k} = \sigma_{11}(1 - R_{1,23 \dots k}^2)$$

***33.** Para la distribución normal trivariante, demuéstrese que

$$\rho_{12,3} = \frac{\rho_{12} - \rho_{13}\rho_{23}}{[(1 - \rho_{13}^2)(1 - \rho_{23}^2)]^{1/2}}$$

BIBLIOGRAFIA

1. ANDERSON, T. W.: *An Introduction to Multivariate Statistical Analysis*, John Wiley & Sons, Inc., Nueva York, 1958.
2. GRAYBILL, F. A.: *An Introduction to Linear Statistical Models*, vol. I, McGraw-Hill Book Company, Inc., Nueva York, 1961.
3. HALD, A.: *Statistical Theory with Engineering Applications*, John Wiley & Sons, Inc., Nueva York, 1952.
4. RAO, C. R.: *Advanced Statistical Methods in Biometric Research*, John Wiley & Sons, Inc., Nueva York, 1952.
5. ROY, S. N.: *Some Aspects of Multivariate Analysis*, John Wiley & Sons, Inc., Nueva York, 1957.
6. WILKS, S. S.: *Mathematical Statistics*, Princeton University Press, Princeton, N. J., 1947.

CAPITULO 10

DISTRIBUCIONES EN EL MUESTREO

10-1. Distribuciones de funciones de variables aleatorias. Para proseguir el estudio del problema de la estimación, es necesario conocer las distribuciones de los estimadores. En esta sección estudiaremos métodos para la obtención de estas distribuciones, y en las restantes secciones de este capítulo utilizaremos tales métodos para obtener ciertas distribuciones de interés especial.

Una variante x puede transformarse mediante cierta función u para definir una nueva variante u . La densidad de u , representada por $g(u)$, vendrá determinada por la transformación $u(x)$, junto con la función $f(x)$, función de densidad o de cuantía de x .

Si x es una variable discreta, la distribución de una función $u(x)$ queda determinada directamente por las leyes de probabilidad. Si x toma los valores $0, 1, 2, \dots, r$, p. ej., con probabilidades $f(0), f(1), \dots, f(r)$, los valores posibles de u , esto es $u_0, u_1, u_2, \dots, u_s$ se obtienen sustituyendo en $u(x)$ los valores sucesivos de x . Puede ocurrir que varios valores de x den lugar a un mismo valor de u . La probabilidad de que u tome un valor dado, u_i , es

$$g(u_i) = \Sigma' f(x) \quad (1)$$

donde la suma Σ' se extiende a todos los valores de x tales que $u(x) = u_i$.

Ejemplo 10-1.—Supongamos que x toma los valores $0, 1, 2, 3, 4, 5$, con probabilidades $p_0, p_1, p_2, p_3, p_4, p_5$; la función de densidad de $u = (x - 2)^2$ es

$$\begin{aligned} g(0) &= p_2 \\ g(1) &= p_1 + p_3 \\ g(4) &= p_0 + p_4 \\ g(9) &= p_5 \end{aligned}$$

y los valores posibles de u son $0, 1, 4, 9$. Análogamente, si u es una función de varias variantes discretas $x_1, x_2, \dots, x_k$, cuya distribución conjunta viene dada por $f(x_1, x_2, \dots, x_k)$, la probabilidad de que $u(x_1, x_2, \dots, x_k)$ tome uno de sus valores particulares u_i es

$$g(u_i) = \Sigma f(x_1, x_2, \dots, x_k) \quad (2)$$

en donde Σ se extiende a todos los valores de x tales que se verifique

$$u(x_1, x_2, \dots, x_k) = u_i$$

El método básico y con frecuencia más sencillo de hallar distribuciones de funciones de variables aleatorias continuas es el expuesto en el capítulo 4.

Supongamos que x tiene una densidad $f(x)$ para $a < x < b$ y 0 en otro caso, y deseamos hallar la densidad de $y = u(x)$, donde $u'(x)$ es continua y positiva para $a \leq x \leq b$. En tal caso la densidad de y será 0 si y no pertenece al intervalo $u(a) < y < u(b)$. Sea $x = v(y)$ la solución respecto a x de $y = u(x)$. Representemos por $g(y)$ la densidad de y . Para obtener $g(y)$, hallaremos la distribución acumulativa $G(y)$ de y y derivaremos en los puntos donde existe derivada. Si $u'(x)$ fuese negativa y continua, la demostración sería análoga:

$$\begin{aligned} G(y) &= P(y \leq y) = P[u(x) \leq y] = P[x \leq v(y)] = \int_a^{v(y)} f(x) dx \\ &= F[v(y)] - F(a) \quad \text{si} \quad u(a) < y < u(b) \\ G(y) &= 0 \quad \text{si} \quad y \leq u(a) \\ G(y) &= 1 \quad \text{si} \quad y \geq u(b) \end{aligned} \quad (3)$$

Luego

$$\begin{aligned} g(y) &= G'(y) = \frac{d}{dy} \{F[v(y)] - F(a)\} \quad u(a) < y < u(b) \\ &= 0 \quad \text{en otro caso} \end{aligned} \quad (4)$$

Ejemplo 10-2.—Supongamos que x tiene la densidad $f(x) = 2x$, $0 < x < 1$, y cero en otro caso. Se trata de hallar la densidad de $y = x^3 - 2$. En el intervalo $0 \leq x \leq 1$, $y'(x)$ es continua y positiva, $x = (y+2)^{1/3}$. Ahora bien:

$$\begin{aligned} G(y) &= P(y \leq y) = P(x^3 - 2 \leq y) = P(x^3 \leq y + 2) \\ &= P(x \leq \sqrt[3]{y+2}) = \int_0^{\sqrt[3]{y+2}} 2x dx = (y+2)^{2/3}, \quad -2 < y < -1 \\ G(y) &= 0 \quad \text{para } y \leq -2 \\ G(y) &= 1 \quad \text{para } y \geq -1 \end{aligned} \quad (5)$$

Luego

$$\begin{aligned} g(y) &= (2/3)(y+2)^{-1/3} \quad -2 < y < -1 \\ &= 0 \quad \text{en otro caso} \end{aligned} \quad (6)$$

Será instructivo ver otra forma de tratar este problema de hallar las distribuciones de funciones de variantes continuas.

En primer lugar, consideraremos funciones de una sola variable aleatoria x .

Sea $f(x)$ la densidad de una variable aleatoria continua x , tal que

$$f(x) > 0 \quad \text{para} \quad a < x < b$$

y

$$f(x) = 0 \quad \text{en otro caso}$$

Sea u una función de x dada por $u=u(x)$ y supongamos que $u(x)$ tiene derivada continua en todos los puntos x del intervalo

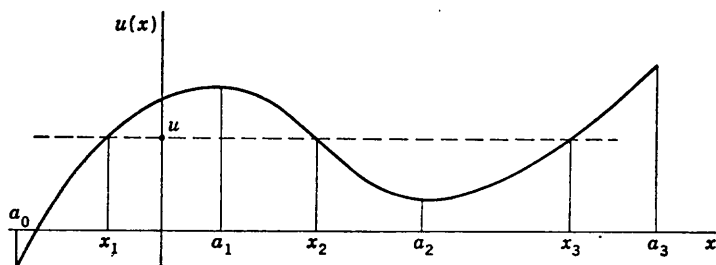


FIG. 10-1.

$a \leq x \leq b$, y que la derivada es 0 solo en los puntos $a_1, a_2, \dots, a_{n-1}$, en donde $a = a_0 < a_1 < \dots < a_{n-1} < a_n = b$. Dividamos el intervalo $a_0 < x < a_n$ en n intervalos disjuntos, $I_1, I_2, \dots, I_n$ designados por

$$a_0 < x < a_1; \quad a_1 < x < a_2; \quad \dots; \quad a_{n-1} < x < a_n$$

tales que u es monótona creciente o monótona decreciente en el intervalo i -ésimo para cada i (véase Fig. 10-1), y sea $x=x_i(u)$ la solución de $u=u(x)$ para x en el intervalo i -ésimo, en donde la inversa es $u=u_i(x)$. Podemos escribir

$$u(x) = \left\{ \begin{array}{llll} u_1(x), & a_0 < x < a_1; & \text{inversa} & x=x_1(u), \quad u \text{ en } A_1 \\ u_2(x), & a_1 < x < a_2; & \text{inversa} & x=x_2(u), \quad u \text{ en } A_2 \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ u_n(x), & a_{n-1} < x < a_n; & \text{inversa} & x=x_n(u), \quad u \text{ en } A_n \end{array} \right\}$$

donde

$$A_i = \{u: u_i(a_{i-1}) < u < u_i(a_i)\}$$

si u es función monótona creciente de x en el i -ésimo intervalo y

$$A_i = \{u : u_i(a_i) < u < u_i(a_{i+1})\}$$

si u es monótona decreciente en el mismo intervalo. Por los supuestos anteriores, existe en A_i para u la derivada $d[x_i(u)]/du$.

Si u es una variable aleatoria, la densidad de u en el punto $u = u$ es

$$g(u) = 0 \quad (7)$$

si no hay puntos x en I_i , cualquiera que sea i , tales que $u = u(x)$; y

$$g(u) = \sum f[x_i(u)] \frac{d[x_i(u)]}{du} \quad (8)$$

en donde la suma se extiende a aquellos valores de i para los cuales $u_i(x) = u$ para algún valor de x del i -ésimo intervalo I_i . Definimos $d[x_i(u)]/du$ como igual al valor absoluto de $d[x_i(u)]/du$.

Ejemplo 10-3.—Consideremos la variable aleatoria x con densidad

$$f(x) = \frac{2}{9}(x+1) \quad -1 < x < 2 \\ = 0 \quad \text{en otro caso}$$

(Véase Fig. 10-2.)

Se desea hallar la densidad de u , siendo $u = x^2$. Obtenemos $a = a_0 = -1$; $a_1 = 0$; $a_2 = b = 2$.

$$u(x) = \begin{cases} u = x^2; & -1 < x < 0: & \text{inversa } x = -\sqrt{u}; & 0 < u < 1 \\ u = x^2; & 0 < x < 2: & \text{inversa } x = +\sqrt{u}; & 0 < u < 4 \end{cases}$$

$$\frac{dx}{du} = \frac{1}{2\sqrt{u}}$$

Para u en el intervalo $0 < u < 1$, a partir de (8) se obtiene:

$$g(u) = f(-\sqrt{u}) \frac{1}{2\sqrt{u}} + f(\sqrt{u}) \frac{1}{2\sqrt{u}} \quad 0 < u < 1$$

Para u en el intervalo $1 < u < 4$, también a partir de (8), resulta:

$$g(u) = f(\sqrt{u}) \frac{1}{2\sqrt{u}} \quad 1 < u < 4$$

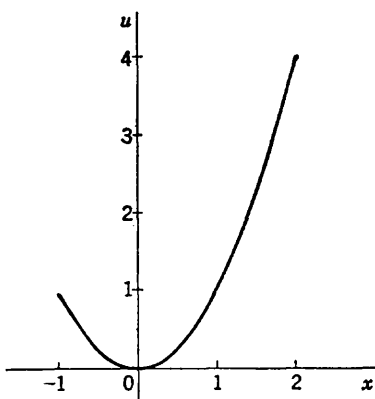


FIG. 10-2.

y

$$g(u)=0 \quad \text{en otro caso}$$

Esto da

$$g(u)=\frac{2}{9\sqrt{u}} \quad 0 < u < 1$$

$$g(u)=\frac{\sqrt{u}+1}{9\sqrt{u}} \quad 1 < u < 4$$

$$g(u)=0 \quad \text{en otro caso}$$

Si u es una función de varias variables aleatorias, puede obtenerse la distribución de u como distribución marginal. Supongamos que $x_1, x_2, \dots, x_k$ tienen la función de densidad $f(x_1, x_2, \dots, x_k)$ y que se desea obtener la función de densidad de $u(x_1, x_2, \dots, x_k)$. Se elimina una de las x , p. ej., x_1 , en función de u , y se resuelve la ecuación

$$u(x_1, x_2, \dots, x_k) = u$$

respecto a x_1 , obteniendo así una función $x_1(u, x_2, x_3, \dots, x_k)$, o varias funciones de este tipo $x_{1i}(u, x_2, \dots, x_k)$ si u no es función monótona de x_1 . Haciendo uso de un razonamiento análogo al utilizado para obtener (8), se deduce la función de densidad

$$g(u, x_2, \dots, x_k) = \sum_i f[x_{1i}(u, x_2, \dots, x_k), x_2, \dots, x_k] \frac{\partial x_{1i}}{\partial u} \quad (9)$$

a partir de la cual puede hallarse la función de densidad de u , integrando g respecto a $x_2, x_3, \dots, x_k$.

El procedimiento anteriormente descrito puede generalizarse para determinar la distribución conjunta de varias funciones $u_1(x_1, \dots, x_k), u_2(x_1, \dots, x_k), \dots, u_r(x_1, \dots, x_k)$ ($r \leq k$), de k variables aleatorias. Se hace

$$\begin{aligned} u_1(x_1, \dots, x_k) &= u_1 \\ u_2(x_1, \dots, x_k) &= u_2 \\ &\dots\dots\dots \\ u_r(x_1, \dots, x_k) &= u_r \end{aligned} \quad (10)$$

y se resuelve el sistema de ecuaciones resultante respecto $x_1, x_2, \dots, x_r$, para obtener r funciones $x_j^*(u_1, u_2, \dots, u_r, x_{r+1}, \dots, x_k)$, o si la solución no es única, se tendrán varios conjuntos de r funciones.

La función de densidad conjunta de las u y de las restantes x es, como puede verse,

$$g(u_1, \dots, u_r, x_{r+1}, \dots, x_k) = \sum f(x_1^*, \dots, x_r^*, x_{r+1}, \dots, x_k) \left| \frac{\partial x_i}{\partial u_j} \right|_+ \quad (11)$$

donde la suma se extiende a todos los conjuntos de soluciones. de (10).

La cantidad $\left| \frac{\partial x_i}{\partial u_j} \right|$ es el valor absoluto del jacobiano, o sea, del determinante

$$\begin{vmatrix} \frac{\partial x_1}{\partial u_1} & \frac{\partial x_1}{\partial u_2} & \dots & \frac{\partial x_1}{\partial u_k} \\ \frac{\partial x_2}{\partial u_1} & \frac{\partial x_2}{\partial u_2} & \dots & \frac{\partial x_2}{\partial u_k} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial x_k}{\partial u_1} & \frac{\partial x_k}{\partial u_2} & \dots & \frac{\partial x_k}{\partial u_k} \end{vmatrix}$$

Prescindimos de la demostración de (11); en esencia coincide con el proceso que se emplea para deducir las fórmulas de transformación de variables en integrales múltiples, el cual puede consultarse en cualquier libro de cálculo.

Uso de las funciones generatrices de momentos.—Hay otro método para determinar distribuciones de funciones de variables aleatorias que, como veremos, resulta particularmente útil. Si es $u(x_1, x_2, \dots, x_k)$ una función de las variables aleatorias x_i cuya distribución viene dada por $f(x_1, x_2, \dots, x_k)$, la función generatriz de momentos de u , si existe, es

$$\begin{aligned} m(t) &= E(e^{tu}) \\ &= \int \dots \int e^{tu(x_1, x_2, \dots, x_k)} f(x_1, x_2, \dots, x_k) \Pi dx_i \end{aligned} \quad (12)$$

Si se reconoce que la función de t que resulta es la función generatriz de momentos de alguna distribución conocida, se deduce que u tiene precisamente dicha distribución, en virtud del teorema 5-7.

Se trata de un método muy eficaz en relación con ciertas técnicas de matemáticas superiores (teoría de las transformaciones de Laplace) que permiten determinar la distribución asociada con mu-

chas funciones generatrices de momentos. El método puede también generalizarse para determinar la distribución conjunta de varias funciones de variables aleatorias.

10-2. Distribución de la media muestral para densidades normales.—Sea $x_1, x_2, \dots, x_n$ una muestra aleatoria de tamaño n , procedente de la densidad normal de media μ y de varianza σ^2 . Para obtener la densidad de

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i,$$

hallaremos la f.g.m. de $\bar{x}$ y aplicaremos el teorema 5-7. La f.g.m. de x_i es, por el teorema 6-2,

$$m_{x_i}(t) = e^{\mu t + (\sigma^2 t^2)/2} \quad i = 1, 2, \dots, n$$

Para hallar la f.g.m. de $\bar{x}$ utilizaremos los resultados de los problemas 35 y 36 (Cap. 5). Puesto que

$$\bar{x} = x_1/n + x_2/n + \dots + x_n/n,$$

necesitamos la f.g.m. de x_i/n . Por el problema 5-35 y lo visto anteriormente,

$$m_{x_i/n}(t) = m_{x_i}\left(\frac{t}{n}\right) = e^{\mu t/n + \sigma^2 t^2/2n^2}$$

Por el problema 5-36, la f.g.m. de $\bar{x}$ es

$$m_{\bar{x}}(t) = (e^{\mu t/n + \sigma^2 t^2/2n^2})^n = e^{\mu t + \sigma^2 t^2/2n}$$

pero esta es la f.g.m. de una densidad normal con media μ y varianza σ^2/n ; por tanto, por el teorema 5-7, queda probado el siguiente teorema:

Teorema 10-1.—*Si $x_1, \dots, x_n$ es una muestra aleatoria de una distribución normal, con media μ y varianza σ^2 , la media muestral $\bar{x}$ es una variable normal con media μ y varianza σ^2/n .*

10-3. Distribución ji cuadrado.—Vamos a obtener la distribución de

$$u = \sum_{i=1}^k \left(\frac{x_i - \mu_i}{\sigma_i} \right)^2 \quad (1)$$

en donde las x_i tienen distribuciones normales independientes, de medias μ_i y varianzas σ_i^2 . En la distribución conjunta de las x_i volvemos a hacer la transformación de variantes

$$y_i = \frac{x_i - \mu_i}{\sigma_i}$$

a fin de simplificar las ecuaciones; u se reduce ahora a Σy_i^2 . Utilizaremos el método de las funciones generatrices de momentos para obtener la distribución deseada.

La función generatriz de momentos de u es

$$m(t) = \left(\frac{1}{2\pi}\right)^{k/2} \int \int \dots \int e^{t \Sigma y_i^2} e^{-\frac{1}{2} \Sigma y_i^2} \prod_i dy_i \quad (2)$$

donde la integral múltiple se escribe como producto de k integrales de la forma

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(1-2t)y_i^2} dy_i \quad (3)$$

La integral (3) tiene por valor $1/\sqrt{1-2t}$, ya que el producto de la misma por $\sqrt{1-2t}$ representa el área limitada por la curva normal cuya varianza es $1/(1-2t)$. Se deduce que

$$m(t) = \left(\frac{1}{1-2t}\right)^{k/2} \quad t < \frac{1}{2} \quad (4)$$

La función generatriz de momentos tiene la forma correspondiente a la de la distribución gamma (Sec. 6-3), con $\alpha = (k/2) - 1$ y $\beta = 2$. Se deduce de ello que la función de densidad de u es

$$f(u) = \frac{1}{[(k/2) - 1]!} \frac{1}{2^{k/2}} u^{(k/2)-1} e^{-\frac{1}{2}u} \quad u > 0 \quad (5)$$

Este caso particular de la distribución gamma suele denominarse distribución ji cuadrado, con k *grados de libertad*. La variante u se designa con el símbolo χ^2 (cuadrado de la letra griega ji)

$$\chi^2 = \sum_i^k \left(\frac{x_i - \mu_i}{\sigma_i}\right)^2 \quad (6)$$

lo que justifica el nombre dado a la distribución. La frase *grados de libertad* se refiere al número de cuadrados independientes en

la suma que figura en (6); no obstante, cabe también considerarlo como designación del parámetro k en la función de densidad (5).

Teorema 10-2.—*Sea $y_1, \dots, y_n$ una muestra aleatoria de una distribución normal con media 0 y varianza 1. En tal supuesto,*

$$u = \sum_{i=1}^n y_i^2$$

tiene una distribución ji-cuadrado con n grados de libertad.

Observemos que (5) da en esencia la distribución del estimador máximo-verosímil de σ^2 en poblaciones normales, supuesto conocido μ . Si se consideran muestras de tamaño n de una población normal, con media conocida μ , el estimador máximo-verosímil de σ^2 es

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2 = \frac{\sigma^2}{n} u$$

en donde

$$u = \sum_{i=1}^n [(x_i - \mu)/\sigma]^2$$

tiene la distribución ji cuadrado con n grados de libertad. La función de densidad del estimador es, por tanto,

$$f(\hat{\sigma}^2) = \frac{1}{[(n/2) - 1]!} \left(\frac{n}{2\sigma^2} \right)^{n/2} (\hat{\sigma}^2)^{(n/2)-1} e^{-n\hat{\sigma}^2/2\sigma^2} \quad \hat{\sigma}^2 > 0 \quad (7)$$

ya que

$$\frac{du}{d\hat{\sigma}^2} = \frac{n}{\sigma^2}$$

Esta es una función de densidad gamma con parámetros $\alpha = (n/2) - 1$ y $\beta = 2\sigma^2/n$.

La distribución ji cuadrado está tabulada parcialmente en la tabla III.

*** 10-4. Independencia de la media y varianza muestrales en densidades normales.**—Generalmente, se desconoce el valor de la media de la población, por lo cual interesa más el estimador $(1/n)\sum(x_i - \bar{x})^2$ de σ^2 que el $(1/n)\sum(x_i - \mu)^2$, considerado en la sección anterior. Obtendremos ahora la distribución de este estimador,

* Si el lector omitió las secciones 9-2 y 9-3, puede estudiar el teorema 10-3 de esta sección sin detenerse en las demostraciones.

viendo al mismo tiempo si es independiente de la correspondiente a la media muestral.

Sean

$$y_i = \frac{x_i - \mu}{\sigma} \quad (1)$$

$$u = n\bar{y}^2 = \frac{1}{n} \left(\sum y_i \right)^2 \quad (2)$$

$$v = \sum_1^n (y_i - \bar{y})^2 \quad (3)$$

y hallemos la función generatriz de momentos conjunta de u y v ; esto es

$$m(t_1, t_2) = E(e^{t_1 u + t_2 v}) \quad (4)$$

$$\begin{aligned} &= \int \int \dots \int \left(\frac{1}{2\pi} \right)^{n/2} e^{(t_1/n)(\sum y_i)^2 + t_2 \sum (y_i - \bar{y})^2 - \frac{1}{2} \sum y_i^2} \prod_1^n dy_i \\ &= \int \int \dots \int \left(\frac{1}{2\pi} \right)^{n/2} e^{-\frac{1}{2} [\sum y_i^2 - (2t_1/n)(\sum y_i)^2 - 2t_2(\sum y_i - \bar{y})^2]} \prod_1^n dy_i \end{aligned} \quad (5)$$

La forma cuadrática puede escribirse

$$\begin{aligned} \sum y_i^2 - \frac{2t_1}{n} \left(\sum y_i \right)^2 - 2t_2 \sum (y_i - \bar{y})^2 \\ = \sum y_i^2 - \frac{2t_1}{n} \left(\sum y_i \right)^2 - 2t_2 \sum y_i^2 + 2nt_2 \bar{y}^2 \end{aligned} \quad (6)$$

$$\begin{aligned} &= (1 - 2t_2) \sum y_i^2 - \frac{2(t_1 - t_2)}{n} \left(\sum y_i \right)^2 \\ &= \sum r_{ij} y_i y_j \end{aligned} \quad (7)$$

donde

$$\begin{aligned} r_{ii} &= 1 - 2t_2 - \frac{2(t_1 - t_2)}{n} = a \\ r_{ij} &= -\frac{2(t_1 - t_2)}{n} = b \quad i \neq j \end{aligned}$$

Un determinante de orden n , cuyos elementos de la diagonal principal sean iguales a a y los restantes a b , tiene por valor

$$(a - b)^{n-1} [a + (n - 1)b]$$

Por tanto,

$$|R| = \left[1 - 2t_2 - \frac{2(t_1 - t_2)}{n} + \frac{2(t_1 - t_2)}{n} \right]^{n-1} \left[1 - 2t_2 - \frac{2(t_1 - t_2)}{n} - \frac{n-1}{n} 2(t_1 - t_2) \right] \\ = (1 - 2t_2)^{n-1} (1 - 2t_1) \quad (8)$$

De la distribución normal multivariante se deduce que

$$\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \left(\frac{1}{2\pi} \right)^{n/2} e^{-\frac{1}{2} \sum_{i,j} r_{ij} y_i y_j} \prod dy_i = \frac{1}{\sqrt{|R|}} \quad (9)$$

de donde resulta que la integral en (5) vale

$$m(t_1, t_2) = \left(\frac{1}{1 - 2t_1} \right) \cdot \left(\frac{1}{1 - 2t_2} \right)^{(n-1)/2} \quad t_1 < 1/2, \quad t_2 < 1/2 \quad (10)$$

El hecho de que la función generatriz de momentos mixtos se descomponga en dos factores, uno de ellos función de t_1 únicamente y el otro solo función de t_2 , quiere decir que u y v se distribuyen independientemente. No demostraremos esto de manera rigurosa, limitándonos a indicar que un razonamiento análogo al efectuado en la sección 5-5 permitiría demostrar que si dos distribuciones multivariantes tienen la misma función de momentos mixtos, ambas son iguales. Supongamos la función de densidad $f(u, v)$, cuya función generatriz de momentos mixtos sea (10). Dadas las distribuciones marginales $f_1(u)$ y $f_2(v)$, construimos la función bivalente

$$g(u, v) = f_1(u)f_2(v) \quad (11)$$

que, evidentemente, es una función de densidad. Además, su función generatriz de momentos será

$$m(t_1, 0)m(0, t_2) \quad (12)$$

en donde

$$m(t_1, t_2) = \iint e^{t_1 u + t_2 v} f(u, v) du dv \quad (13)$$

Puesto que (12) y (13) son idénticas, según (10), se deduce que $g(u, v)$ y $f(u, v)$ son una misma función de densidad y, por tanto, que $f(u, v)$ es igual al producto de sus funciones de densidad marginal.

Los dos factores de la ecuación (10) tienen la forma de la función generatriz de momentos de una distribución ji cuadrado; por tanto, se deduce que u y v poseen distribuciones independientes ji cuadrado, la primera con un grado de libertad y la segunda con $n-1$. El hecho de que $u = n\bar{y}^2$ se distribuya como una ji cuadrado con un grado de libertad, está de acuerdo con los resultados de las secciones 10-2 y 10-3. En efecto, hemos visto que $\bar{y}$ tiene distribución normal con media 0 y varianza $1/n$ y, según resulta de la sección 10-3 para $k=1$, se deduce que

$$u = \frac{(\bar{y} - 0)^2}{1/n} = n\bar{y}^2 = n \left(\frac{\bar{x} - \mu}{\sigma} \right)^2 \quad (14)$$

admitirá una distribución ji cuadrado con solo un grado de libertad. La cantidad

$$v = \sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{\sigma} \right)^2 \quad (15)$$

tiene la distribución dada por la ecuación (10-3-5), pero sustituyendo k por $n-1$ en vez de por n , como ocurriría si las desviaciones se midieran respecto de la media poblacional. Suele decirse que, al tomar la suma de cuadrados de las desviaciones respecto de la media muestral en vez de la media poblacional, se pierde un grado de libertad, o bien que se utiliza un grado de libertad al estimar la media. Aunque en la ecuación (15) v es la suma de n cuadrados, estos no son todos funcionalmente independientes. La relación $\sum y_i = n\bar{y}$ permite calcular cualquiera de las desviaciones $y_i - \bar{y}$, dadas las otras $n-1$.

Teorema 10-3.—Sea $y_1, y_2, \dots, y_n$ una muestra aleatoria de una distribución normal con media 0 y varianza 1. En tal caso:

- $\sum_{i=1}^n (y_i - \bar{y})^2$ es una variante ji cuadrado con $n-1$ g. de l.
- $\bar{y}$ tiene una distribución normal con media 0 y varianza $1/n$.
- $\bar{y}$ y $\sum_{i=1}^n (y_i - \bar{y})^2$ son independientes.

En función de v , de acuerdo con (15), el estimador

$$\hat{\sigma}^2 = \frac{1}{n} \sum (x_i - \bar{x})^2 \quad (16)$$

puede escribirse

$$\hat{\sigma}^2 = \frac{\sigma^2 v}{n}$$

La densidad correspondiente a este estimador es, por tanto,

$$f(\hat{\sigma}^2) = \frac{1}{[(n-3)/2]!} \left(\frac{n}{2\sigma^2} \right)^{(n-1)/2} (\hat{\sigma}^2)^{(n-3)/2} \exp \left(-\frac{n\hat{\sigma}^2}{2\sigma^2} \right) \quad \hat{\sigma}^2 > 0 \quad (17)$$

Los resultados de esta sección son aplicables solo a poblaciones normales. Puede demostrarse que para ninguna otra distribución se verifica que: 1) la media y la varianza muestrales tienen distribuciones independientes; 2) la media muestral tiene exactamente una distribución normal.

10-5. La distribución «F».—Una distribución, cuyo considerable interés práctico veremos más adelante, es la del cociente de dos variables independientes con distribución ji cuadrado. Supongamos que u y v tienen distribuciones independientes ji cuadrado, con m y n grados de libertad, respectivamente. Su función de densidad conjunta es, según (10-3-5),

$$f(u, v) = \frac{1}{[(m-2)/2]! [(n-2)/2]! 2^{(m+n)/2}} u^{(m-2)/2} v^{(n-2)/2} e^{-\frac{1}{2}(u+v)} \quad (1)$$

$u > 0, v > 0$

Halleemos la distribución de la cantidad

$$F = \frac{u/m}{v/n} = \frac{nu}{mv} \quad (2)$$

que corrientemente se designa con el nombre de *razón de varianza*. Obtendremos la función de densidad de F , eliminando u en función de F en (1), e integrando el resultado respecto a v . Puesto que

$$\frac{\partial u}{\partial F} = \frac{mv}{n} \quad (3)$$

y como $\frac{\partial u}{\partial F}$ es continua y positiva, la función de densidad conjunta de F y v es

$$g(F, v) = \frac{1}{[(m-2)/2]! [(n-2)/2]! 2^{(m+n)/2}} v^{(n-2)/2} \left(\frac{mvF}{n} \right)^{(m-2)/2} e^{-\frac{1}{2}[v+(mvF/n)]} \frac{mv}{n} \quad (4)$$

Para integrar respecto a v , habrá que calcular la integral

$$\int_0^{\infty} v^{(m+n-2)/2} e^{-\frac{1}{2}[1+(mF/n)]v} dv \quad (5)$$

que comprende los factores de (4) donde interviene v . Observemos que, prescindiendo de ciertas constantes, el integrando es la integral de una función de densidad gamma. En efecto, multiplicando esta integral por

$$\frac{\{1/2[1+(mF/n)]\}^{(m+n)/2}}{[(m+n-2)/2]!} \quad (6)$$

se obtiene exactamente el área limitada por la función de densidad gamma, con

$$\alpha = \frac{m+n-2}{2}$$

y

$$\beta = \frac{1}{1/2[1+(mF/n)]}$$

y el resultado sería igual a la unidad. Por tanto, el valor de (5) es el recíproco de la expresión (6). Luego la función de densidad de F es

$$\begin{aligned} h(F) &= \int_0^{\infty} g(F, v) dv \\ &= \frac{1}{\left(\frac{m-2}{2}\right)! \left(\frac{n-2}{2}\right)! 2^{\frac{m+n}{2}}} \\ &\quad \left(\frac{m}{n}\right)^{\frac{m}{2}} F^{\frac{m-2}{2}} \int_0^{\infty} v^{\frac{m+n-2}{2}} e^{-\frac{1}{2}\left(1+\frac{mF}{n}\right)v} dv \\ &= \frac{\left(\frac{m+n-2}{2}\right)!}{\left(\frac{m-2}{2}\right)! \left(\frac{n-2}{2}\right)!} \left(\frac{m}{n}\right)^{\frac{m}{2}} \frac{F^{\frac{m-2}{2}}}{\left(1+\frac{mF}{n}\right)^{\frac{m+n}{2}}} \quad F > 0 \quad (7) \end{aligned}$$

función en la que intervienen los parámetros m y n . Estos reciben también el nombre de *grados de libertad*; por ello (7) se denomina

función de densidad F con m y n grados de libertad; el número de grados de libertad de la variante u en el numerador de F es siempre el primero que se menciona.

Teorema 10-4.—Sea u una variante ji cuadrado con m grados de libertad y v otra variante ji cuadrado con n grados de libertad, y supongamos que u y v son independientes. La variable aleatoria

$$F = \frac{u/m}{v/n}$$

se distribuye según una distribución F con m y n grados de libertad. La función de densidad es la dada en (7).

En la tabla V se dan cinco puntos de la rama superior de la distribución acumulativa de F . Pueden hallarse tablas más completas en la referencia citada al pie de la tabla V, y en *Tablas estadísticas para investigaciones biológicas, agrícolas y médicas*, de Fisher y Yates (Aguilar, S. A. de Ediciones, Madrid, 1954). Los recíprocos de los números de la tabla V proporcionan cinco puntos de la rama inferior de la distribución acumulativa. En general, para calcular una integral de la forma

$$P(a < F < b) = \int_a^b h(F) dF$$

se transforma esta distribución en la beta. La transformación adecuada es

$$w = \frac{mF/n}{1 + (mF/n)} \quad (8)$$

la cual transforma (7) en una distribución beta con parámetros $\alpha = (m-2)/2$ y $\beta = (n-2)/2$.

10-6. Distribución «t» de Student.—Otra distribución de considerable importancia práctica es la de una variante con distribución normal dividida por la raíz cuadrada de otra variante independiente de la anterior, con distribución ji cuadrado. De modo más preciso: si x tiene una distribución normal de media μ y varianza σ^2 , y u , una distribución ji cuadrado con k grados de libertad, siendo x y u independientes, se trata de hallar la distribución de

$$t = \frac{(x - \mu)/\sigma}{\sqrt{u/k}} \quad (1)$$

en la que haciendo

$$y = \frac{x - \mu}{\sigma}$$

toma la forma $t = \frac{y}{\sqrt{u/k}}$. La función de densidad de y y u es

$$f(y, u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} \frac{1}{[(k-2)/2]! 2^{k/2}} u^{(k-2)/2} e^{-\frac{1}{2}u} \quad (2)$$

y hallamos la distribución de t por el mismo procedimiento utilizado en la sección anterior. Sustituimos y en función de $t (y = t\sqrt{u/k})$ en (2) e integramos la función resultante respecto de u . El resultado final es

$$h(t) = \frac{[(k-1)/2]!}{\sqrt{k\pi} [(k-2)/2]!} \frac{1}{[1 + (t^2/k)]^{(k+1)/2}} \quad -\infty < t < \infty \quad (3)$$

distribución con un solo parámetro k , que suele también designarse con el nombre de grados de libertad de la distribución. Como $[(x - \mu)/\sigma]^2$ tiene la distribución ji cuadrado con un grado de libertad, es evidente, por (1), que t^2 tiene la distribución F con uno y k grados de libertad. La forma acumulativa de la distribución se encuentra parcialmente tabulada en la tabla IV.

Teorema 10-5.—Sea y una variante normal con media 0 y varianza 1 y u una variante ji cuadrado con k grados de libertad, y supongamos que u e y son independientes. La variable aleatoria

$$t = \frac{y\sqrt{k}}{\sqrt{u}}$$

se distribuye según una t de Student con k grados de libertad. La función de densidad es la dada en (3).

10-7. Distribución de las medias muestrales en densidades binomiales y de Poisson.—En las secciones precedentes hemos estudiado dos métodos que permiten hallar distribuciones de funciones de variables aleatorias continuas descritas en la sección 10-1. Ahora vamos a detallar la técnica correspondiente a variantes discretas en dos casos de particular interés. Si $x_1, x_2, \dots, x_n$ es una muestra de tamaño n de la densidad binomial puntual

$$f(x) = p^x q^{1-x} \quad x=0, 1 \quad 0 \leq p \leq 1 \quad (1)$$

la densidad conjunta de las x es

$$f(x_1, x_2, \dots, x_n) = p^{\sum x_i} q^{n - \sum x_i} \quad x_i = 0, 1 \quad (2)$$

La media muestral

$$\bar{x} = \frac{1}{n} \sum x_i$$

es función de las variables aleatorias, y es evidente que los únicos valores posibles de $\bar{x}$ son $0, 1/n, 2/n, \dots, 1$. La probabilidad, $g(j/n)$, de que $\bar{x}$ tome el valor j/n , se obtiene sumando (2) para todos los conjuntos $(x_1, x_2, \dots, x_n)$ tales que $(1/n)\sum x_i = j/n$, o tales que $\sum x_i = j$. Para todos estos conjuntos, $f(x_1, x_2, \dots, x_n)$ tiene el mismo valor, $p^j q^{n-j}$; por tanto, la suma se calcula multiplicando este valor por el número de conjuntos $(x_1, x_2, \dots, x_n)$ que cumplan la condición indicada. El número de estos es igual al de permutaciones de j unos y $n-j$ ceros, o sea $\binom{n}{j}$; por tanto,

$$g\left(\frac{j}{n}\right) = \binom{n}{j} p^j q^{n-j} \quad \frac{j}{n} = 0, \frac{1}{n}, \frac{2}{n}, \dots, 1 \quad (3)$$

como ya vimos en la sección 7-7.

Teorema 10-6.—Sea $x_1, x_2, \dots, x_n$ una muestra aleatoria de la binomial puntual

$$f(x) = p^x q^{1-x} \quad x = 0, 1; \quad 0 \leq p \leq 1$$

La densidad de $\bar{x} = \frac{1}{n} \sum x_i$ es

$$g(\bar{x}) = \binom{n}{n\bar{x}} p^{n\bar{x}} q^{n-n\bar{x}} \quad \bar{x} = 0, \frac{1}{n}, \frac{2}{n}, \dots, \frac{n-1}{n}, 1$$

Análogamente, hallaríamos la distribución de la media de una muestra $x_1, x_2, \dots, x_n$, de una población de Poisson. La densidad conjunta de las observaciones viene dada por

$$f(x_1, x_2, \dots, x_n) = \frac{e^{-n\mu} \mu^{\sum x_i}}{\prod_i x_i!} \quad x_i = 0, 1, 2, \dots \quad (4)$$

si se designa por μ el parámetro de la distribución. La media muestral $\bar{x}$ puede tomar, evidentemente, cualquiera de los valores j/n ,

en donde $j=0, 1, 2, \dots$ Para un valor particular j/n , las x deberán cumplir la condición $\sum x_i = j$; por tanto,

$$\begin{aligned} g\left(\frac{j}{n}\right) &= \sum_{\sum x_i = j} \frac{e^{-n\mu} \mu^j}{\prod x_i!} \\ &= e^{-n\mu} \mu^j \sum_{\sum x_i = j} \frac{1}{\prod x_i!} \end{aligned} \quad (5)$$

Esta suma puede efectuarse utilizando el teorema polinomial, según el cual, si se hacen todas las $x_i=1$ en la ecuación (2-11-2), se tiene

$$\sum \frac{j!}{\prod x_i!} = n^j$$

Por tanto, la suma es igual a $n^j/j!$, y la distribución deseada viene dada por

$$g\left(\frac{j}{n}\right) = \frac{e^{-n\mu}(n\mu)^j}{j!} \quad \bar{x} = \frac{j}{n} = 0, \frac{1}{n}, \frac{2}{n}, \dots \quad (6)$$

Esta función puede expresarse explícitamente en función de $\bar{x}$:

$$g(\bar{x}) = \frac{e^{-n\mu}(n\mu)^{n\bar{x}}}{(n\bar{x})!} \quad (7)$$

Observemos que, por existir correspondencia biunívoca entre $\bar{x}=j/n$ y $j=\sum x_i$, la densidad de j es

$$h(j) = \frac{e^{-n\mu}(n\mu)^j}{j!} \quad j=0, 1, 2, \dots$$

Teorema 10-7.—Sea $x_1, x_2, \dots, x_n$ una muestra aleatoria de la distribución de Poisson con parámetro μ . La densidad de $\bar{x}$ es

$$g(\bar{x}) = \frac{e^{-n\mu}(n\mu)^{n\bar{x}}}{(n\bar{x})!} \quad \bar{x}=0, \frac{1}{n}, \frac{2}{n}, \dots$$

*** 10-8. Distribución, en muestras grandes, de estimadores máximo-verosímiles.**—Hemos estudiado varios problemas particu-

* Si el lector omitió el estudio de las secciones 9-2 y 9-3, también deberá omitir esta sección.

lares de la teoría del muestreo, no solo como ilustración de los métodos utilizados para hallar distribuciones en el muestreo, sino porque las distribuciones particulares que se obtienen son importantes en estadística aplicada. Estas se designan algunas veces con el nombre de «distribuciones de muestras pequeñas», aunque, naturalmente, también son válidas para muestras grandes; lo que se quiere indicar es, simplemente, su validez en el caso de muestras pequeñas. En esta sección consideraremos una distribución mucho más general, en el sentido de que es más o menos independiente de la forma de la distribución de la población, aunque solo es válida para muestras grandes.

Para una clase muy amplia de distribuciones, el estimador máximo-verosímil tiene una distribución aproximadamente normal respecto al valor verdadero del parámetro como media para muestras grandes. Esto representa un instrumento poderoso en la resolución de muchos problemas importantes de estadística aplicada, como veremos en los capítulos siguientes. El teorema siguiente se aplica tanto a distribuciones discretas como continuas, si la densidad satisface ciertas condiciones de regularidad.

Teorema 10-8.—*Los estimadores máximo-verosímiles $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_k$, de los parámetros de una distribución dada por $f(x; \theta_1, \theta_2, \dots, \theta_k)$ para muestras de tamaño n , tienen, para muestras grandes, una distribución que se aproxima a la normal multivariante de medias $\theta_1, \theta_2, \dots, \theta_k$, y matriz nR en la forma cuadrática, en donde*

$$r_{ij} = -E \left[\frac{\partial^2}{\partial \theta_i \partial \theta_j} \log f(x; \theta_1, \theta_2, \dots, \theta_k) \right]$$

Las varianzas y covarianzas de los estimadores son $(1/n)V$, con

$$V = R^{-1}$$

El teorema no depende en absoluto del hecho de que se hayan empleado distribuciones univariantes. La variante x puede sustituirse por el conjunto de variantes $(x, y, z, \dots)$, en cualquiera de las expresiones de esta sección.

Ejemplo 10-4.—Como aclaración del uso del teorema que acabamos de enunciar, hallaremos la distribución en muestras grandes de los estimadores de los dos parámetros de la distribución normal. Escribiremos esta en la forma

$$f(x; \theta_1, \theta_2) = \frac{1}{\sqrt{2\pi\theta_2}} e^{-(1/2\theta_2)(x-\theta_1)^2} \quad (1)$$

Para muestras de tamaño n hemos visto que los estimadores máximo-verosímiles son

$$\hat{\theta}_1 = \frac{1}{n} \sum x_i \quad (2)$$

$$\hat{\theta}_2 = \frac{1}{n} \sum (x_i - \hat{\theta}_1)^2 \quad (3)$$

De acuerdo con el teorema anterior, estos estimadores tendrán una distribución aproximadamente normal en muestras grandes, con medias θ_1 y θ_2 y matriz $nR = (nr_{ij})$ en la forma cuadrática, siendo

$$r_{ij} = -E \left(\frac{\partial^2 \log f}{\partial \theta_i \partial \theta_j} \right) \quad (4)$$

Puesto que

$$\log f = -\frac{1}{2} \log 2\pi - \frac{1}{2} \log \theta_2 - \frac{1}{2\theta_2} (x - \theta_1)^2 \quad (5)$$

las derivadas que se precisan son

$$\begin{aligned} \frac{\partial^2 \log f}{\partial \theta_1^2} &= -\frac{1}{\theta_2} \\ \frac{\partial^2 \log f}{\partial \theta_1 \partial \theta_2} &= -\frac{x - \theta_1}{\theta_2^2} \\ \frac{\partial^2 \log f}{\partial \theta_2^2} &= \frac{1}{2\theta_2^2} - \frac{(x - \theta_1)^2}{\theta_2^3} \end{aligned}$$

y como

$$E(x) = \theta_1 \quad E(x - \theta_1)^2 = \theta_2$$

se ve fácilmente que las r_{ij} son

$$R = \begin{pmatrix} \frac{1}{\theta_2} & 0 \\ 0 & \frac{1}{2\theta_2^2} \end{pmatrix} \quad (6)$$

La distribución de los estimadores para muestras grandes será, por tanto,

$$g(\hat{\theta}_1, \hat{\theta}_2) = \frac{1}{2\pi} \frac{n}{\sqrt{2\theta_2^3}} e^{-\frac{n}{2} \left[\frac{(\hat{\theta}_1 - \theta_1)^2}{\theta_2} + \frac{(\hat{\theta}_2 - \theta_2)^2}{2\theta_2^2} \right]} \quad (7)$$

con varianzas y covarianzas para muestras grandes dadas por

$$\left(\frac{1}{n} \sigma_{ij} \right) = \begin{pmatrix} \frac{\theta_2}{n} & 0 \\ 0 & \frac{2\theta_2^2}{n} \end{pmatrix} \quad (8)$$

Puesto que $\sigma_{12}=0$, se ve que los estimadores están distribuidos independientemente para muestras grandes; claro es que ya hemos visto en la sección 10-4 que, en realidad, aquellos son independientes cualquiera que sea el tamaño de la muestra. La distribución de $\hat{\theta}_1$ en muestras grandes es exactamente la normal, tal como viene dada en (7). Pero la distribución exacta de $\hat{\theta}_2$ es la distribución gamma, cualquiera que sea el tamaño de la muestra, lo que parece estar en contraposición con la distribución normal dada en (7). No obstante, se demuestra que la distribución exacta de $\hat{\theta}_2$ se aproxima a la forma normal

$$\frac{1}{\sqrt{2\pi}} \sqrt{\frac{n}{2}} \frac{1}{\theta_2} e^{-\frac{n}{2} \frac{(\hat{\theta}_2 - \theta_2)^2}{\theta_2^2}}$$

cuando n crece.

Ejemplo 10-5.—Como segundo ejemplo vamos a hallar la distribución en muestras grandes de los estimadores de los parámetros de una distribución polinomial.

Supongamos que los elementos de una población pueden clasificarse en $k+1$ categorías, $A_1, A_2, \dots, A_{k+1}$. Representaremos cada elemento por el conjunto de variables $(x_1, x_2, \dots, x_{k+1})$ donde, si el elemento pertenece a A_i , $x_i=1$ y todas las demás x son iguales a cero. Si la probabilidad de que un elemento elegido al azar pertenezca a A_i es p_i , la distribución conjunta de las x viene dada por

$$f(x_1, x_2, \dots, x_{k+1}) = p_1^{x_1} p_2^{x_2} \dots p_{k+1}^{x_{k+1}} \quad x_i = 0, 1; \sum x_i = 1 \quad (9)$$

en donde $\sum p_i = 1$. Sumando $f(x_1, x_2, \dots, x_{k+1})$ para todos los valores posibles de las x , esto es, $(1, 0, 0, \dots, 0)$, $(0, 1, 0, 0, \dots, 0)$, $(0, 0, 1, \dots, 0)$ y así sucesivamente, se obtiene

$$\sum f(x_1, \dots, x_{k+1}) = \sum_{i=1}^{k+1} p_i = 1$$

La distribución (9) es una distribución multivariante con k parámetros funcionalmente independientes; tomaremos como paráme-

tros $p_1, p_2, \dots, p_k$, y adoptaremos p_{k+1} como símbolo para $1 - p_1 - p_2 - \dots - p_k$.

Supongamos extraída una muestra de tamaño n , y sea n_i el número de sus elementos que pertenecen a A_i ; se verifica que $\sum n_i = n$, y la verosimilitud de la muestra es

$$L(p_1, \dots, p_k) = \prod_{i=1}^{k+1} p_i^{n_i}$$

y su logaritmo,

$$L^* = \log L(p_1, p_2, \dots, p_k) = \sum_{i=1}^{k+1} n_i \log p_i \quad (10)$$

Los estimadores se hallan igualando a cero las primeras derivadas de L^* y resolviendo respecto de los parámetros. Estas ecuaciones son

$$\begin{aligned} \frac{\partial L^*}{\partial p_1} &= \frac{n_1}{\hat{p}_1} - \frac{n_{k+1}}{\hat{p}_{k+1}} = 0 \\ \frac{\partial L^*}{\partial p_2} &= \frac{n_2}{\hat{p}_2} - \frac{n_{k+1}}{\hat{p}_{k+1}} = 0 \end{aligned} \quad (11)$$

etcétera, recordando que $\hat{p}_{k+1}$ representa $1 - \hat{p}_1 - \hat{p}_2 - \dots - \hat{p}_k$. Multiplicando la primera ecuación por $\hat{p}_1 \hat{p}_{k+1}$, la segunda por $\hat{p}_2 \hat{p}_{k+1}$, etc., y sumando los resultados, se obtiene $\hat{p}_{k+1} = n_{k+1}/n$, o sea

$$\hat{p}_i = \frac{n_i}{n} \quad i = 1, 2, \dots, k+1 \quad (12)$$

Queremos hallar la distribución aproximada de los estimadores en (12) para muestras grandes. Aplicando el teorema 10-8 sabemos que la distribución es normal y que sus medias son p_i . Basta hallar los coeficientes r_{ij} de la forma cuadrática. Por el teorema 10-8,

$$r_{ij} = -E \left(\frac{\partial^2}{\partial p_i \partial p_j} \log f \right) \quad (13)$$

Derivando $\log f$, resulta

$$\begin{aligned} -\frac{\partial^2}{\partial p_i \partial p_j} \log f &= -\frac{x_{k+1}}{p_{k+1}^2} & \text{si } i \neq j \\ &= -\frac{x_i}{p_i^2} - \frac{x_{k+1}}{p_{k+1}^2} & \text{si } i = j \end{aligned} \quad (14)$$

y tomando valores esperados,

$$E(x_i) = \sum_{x_i} x_i \prod_1^{k+1} p_i^{x_i} = p_i \quad i = 1, 2, \dots, k \quad (15)$$

$$E(x_{k+1}) = \sum_{x_i} x_{k+1} \prod_1^{k+1} p_i^{x_i} = p_{k+1}$$

Así, pues,

$$\begin{aligned} r_{ij} &= \frac{1}{p_{k+1}} & \text{si } i \neq j \\ &= \frac{1}{p_i} + \frac{1}{p_{k+1}} & \text{si } i = j \end{aligned} \quad (16)$$

y estas dos relaciones se condensan en una sola, utilizando el símbolo δ_{ij}

$$r_{ij} = \frac{\delta_{ij}}{p_i} + \frac{1}{p_{k+1}} \quad i, j = 1, 2, \dots, k \quad (17)$$

Puede demostrarse que el valor del determinante $|R|$ es $\frac{1}{\prod_1^{k+1} p_i}$, por lo que la distribución aproximada para muestras grandes de los estimadores es

$$\begin{aligned} g(\hat{p}_1, \hat{p}_2, \dots, \hat{p}_k) &= \\ &= \left(\frac{1}{2\pi} \right)^{\frac{k}{2}} \sqrt{\frac{1}{\prod_1^{k+1} p_i} n^{\frac{k}{2}} e^{-\frac{1}{2} \sum_{i=1}^k \sum_{j=1}^k n \left(\frac{\delta_{ij}}{p_i} + \frac{1}{p_{k+1}} \right) (\hat{p}_i - p_i)(\hat{p}_j - p_j)}} \end{aligned} \quad (18)$$

La inversa de R tiene por elementos

$$\begin{aligned} \sigma_{ii} &= p_i(1 - p_i) \\ \sigma_{ij} &= -p_i p_j \quad i \neq j \quad i, j = 1, 2, \dots, k \end{aligned} \quad (19)$$

como se comprueba calculando el producto $V \cdot R$. Las varianzas y covarianzas de los estimadores para muestras grandes se obtienen, por tanto, multiplicando (19) por $1/n$. En realidad, estas son las varianzas y covarianzas exactas, cualquiera que sea el tamaño de la muestra.

10-9. Distribución de estadísticos ordinales.—Sea $x_1, \dots, x_n$ una muestra aleatoria procedente de la densidad $f(x)$ y supongamos que deseamos hallar la distribución conjunta de $y_1, y_2, \dots, y_n$, en donde las y_i son las mismas x_i dispuestas en orden de magnitud creciente, $y_1 < y_2 < \dots < y_n$. Por sencillez nos limitaremos a considerar en esta sección el caso $n=3$, pero los métodos son completamente generales; supondremos que $f(x)$ es la densidad de una variable aleatoria continua x .

Sea x_1, x_2, x_3 una muestra de tamaño 3 de $f(x)$, $-\infty < x < \infty$. Transformemos las x en las y como sigue:

$$\begin{array}{ll}
 y_1 = x_1; y_2 = x_2; y_3 = x_3 & \text{si } x_1 < x_2 < x_3 \\
 y_1 = x_1; y_2 = x_3; y_3 = x_2 & \text{si } x_1 < x_3 < x_2 \\
 y_1 = x_2; y_2 = x_1; y_3 = x_3 & \text{si } x_2 < x_1 < x_3 \\
 y_1 = x_2; y_2 = x_3; y_3 = x_1 & \text{si } x_2 < x_3 < x_1 \\
 y_1 = x_3; y_2 = x_1; y_3 = x_2 & \text{si } x_3 < x_1 < x_2 \\
 y_1 = x_3; y_2 = x_2; y_3 = x_1 & \text{si } x_3 < x_2 < x_1
 \end{array} \quad (1)$$

Las seis regiones $x_1 < x_2 < x_3$; $x_1 < x_3 < x_2$; $\dots$; $x_3 < x_2 < x_1$ son disjuntas y la unión de estas seis regiones y de ciertos conjuntos que tienen probabilidad cero, tales como $x_1 = x_2 = x_3$, etc., es el espacio tridimensional $-\infty < x_1 < \infty$, $-\infty < x_2 < \infty$, $-\infty < x_3 < \infty$.

Vamos a hallar la densidad conjunta de y_1, y_2, y_3 para cada una de las seis regiones anteriores. La densidad conjunta de las x es

$$\begin{array}{ll}
 q(x_1, x_2, x_3) = f(x_1)f(x_2)f(x_3) & -\infty < x_1 < \infty \\
 & -\infty < x_2 < \infty \\
 & -\infty < x_3 < \infty
 \end{array} \quad (2)$$

Para la región $x_1 < x_2 < x_3$, la densidad de las y es

$$h(y_1, y_2, y_3) = f(y_1)f(y_2)f(y_3) \quad y_1 < y_2 < y_3 \quad (3)$$

ya que el jacobiano es igual a 1. Para la región $x_1 < x_3 < x_2$, la densidad de las y es

$$h(y_1, y_2, y_3) = f(y_1)f(y_3)f(y_2) \quad y_1 < y_2 < y_3 \quad (4)$$

Análogamente, para las otras regiones de (1) tenemos (el jacobiano es igual a 1 para todas las regiones):

$$h(y_1, y_2, y_3) = f(y_2)f(y_1)f(y_3) \quad y_1 < y_2 < y_3 \quad (5)$$

$$h(y_1, y_2, y_3) = f(y_2)f(y_3)f(y_1) \quad y_1 < y_2 < y_3 \quad (6)$$

$$h(y_1, y_2, y_3) = f(y_3)f(y_1)f(y_2) \quad y_1 < y_2 < y_3 \quad (7)$$

$$h(y_1, y_2, y_3) = f(y_3)f(y_2)f(y_1) \quad y_1 < y_2 < y_3 \quad (8)$$

Pero cada parte de la densidad en (3) a (8) es la misma, y está definida sobre la misma región $y_1 < y_2 < y_3$. Por tanto, teniendo en cuenta lo dicho anteriormente, la densidad conjunta es la suma de estas seis:

$$h(y_1, y_2, y_3) = 6f(y_1)f(y_2)f(y_3) \quad -\infty < y_1 < y_2 < y_3 < \infty$$

o

$$h(y_1, y_2, y_3) = 3! f(y_1)f(y_2)f(y_3) \quad -\infty < y_1 < y_2 < y_3 < \infty \quad (9)$$

Para generalizar lo anterior a una muestra de tamaño n , observemos que la única modificación que ha de hacerse es que existen $n!$ regiones de la clase $x_1 < x_2 < \dots < x_n$, obtenidas por permutación de las x_i . Se tiene así el siguiente teorema:

Teorema 10-9.—Sea $x_1, x_2, \dots, x_n$ una muestra aleatoria de la densidad $f(x)$, $-\infty < x < \infty$. La densidad conjunta de los estadísticos ordinales $y_1, y_2, \dots, y_n$ es

$$h(y_1, y_2, \dots, y_n) = n! f(y_1) \dots f(y_n) \quad -\infty < y_1 < y_2 < \dots < y_n < \infty$$

Cabe deducir muchas densidades interesantes y útiles a partir de esta distribución conjunta. Así, p. ej., puede hallarse la densidad de la variante mayor y_n , o de la menor y_1 , o bien de la variante que ocupa el lugar t contado desde la mayor y_{n-t} , etc. Ilustraremos el uso de esta distribución hallando la densidad del recorrido $y_n - y_1 = R$. Obtengamos primero la densidad conjunta de las variantes mayor y menor, y_n e y_1 ; es decir,

$$p(y_1, y_n) = \int_{y_1}^{y_n} \dots \int_{y_1}^{y_4} \int_{y_1}^{y_3} h(y_1, \dots, y_n) dy_2 dy_3 \dots dy_{n-1} \quad -\infty < y_1 < y_n < \infty$$

o

$$p(y_1, y_n) = n! f(y_1)f(y_n) \int_{y_1}^{y_n} \dots \int_{y_1}^{y_4} \int_{y_1}^{y_3} f(y_2)f(y_3) \dots f(y_{n-1}) dy_2 dy_3 \dots dy_{n-1}$$

Realizaremos esta integración término a término. En primer lugar,

$$\int_{y_1}^{y_3} f(y_2) dy_2 = F(y_3) - F(y_1)$$

donde $F(y)$ es

$$\int_{-\infty}^y f(t) dt$$

y, en consecuencia, $F'(y) = f(y)$ en los puntos donde exista derivada. A continuación,

$$\int_{y_1}^{y_4} [F(y_3) - F(y_1)] f(y_3) dy_3 = \frac{[F(y_4) - F(y_1)]^2}{2}$$

Luego,

$$\frac{1}{2} \int_{y_1}^{y_5} [F(y_4) - F(y_1)]^2 f(y_4) dy_4 = \frac{1}{3!} [F(y_5) - F(y_1)]^3$$

Continuando de este modo, resulta:

$$p(y_1, y_n) = n(n-1)[F(y_n) - F(y_1)]^{n-2} f(y_1) f(y_n) \quad -\infty < y_1 < y_n < \infty$$

Para hallar la distribución del recorrido $R = y_n - y_1$, hacemos la transformación

$$R = y_n - y_1$$

$$S = y_1$$

y la densidad de R , S es (el valor absoluto del jacobiano es la unidad)

$$k(R, S) = n(n-1)[F(R+S) - F(S)]^{n-2} f(S) f(R+S)$$

La densidad de R puede obtenerse integrando $k(R, S)$ respecto de S .

Ejemplo 10-6.—Supongamos $f(x) = 1$, $0 < x < 1$ y sea $x_1, \dots, x_6$ una muestra aleatoria de tamaño 6 procedente de esta distribución rectangular:

$$F(x) = \int_0^x dt = x$$

y, por tanto,

$$F(R+S) = R+S \quad F(S) = S$$

La densidad conjunta de R y S es

$$k(R, S) = 6 \cdot 5 (R+S-S)^4 = 30R^4 \quad \begin{aligned} 0 < S < 1-R \\ 0 < R < 1 \end{aligned}$$

y la densidad de R ,

$$g(R) = 30 \int_0^{1-R} R^4 dS = 30R^4(1-R) \quad 0 < R < 1$$

10-10. Recorrido «studentizado».—Sea $x_1, \dots, x_n$ una muestra aleatoria de tamaño n de una distribución normal de media 0 y varianza 1, y v^2 , una variable ji cuadrado con m grados de libertad, y supongamos que las x y v^2 son independientes. La variable aleatoria

$$q = \frac{y_n - y_1}{v/\sqrt{m}}$$

se conoce con el nombre de *recorrido studentizado* y es importante en ciertos problemas de estadística aplicada que veremos en los capítulos siguientes (las y son estadísticos ordinales de las x). Sea $N(x)$ la distribución acumulativa de la normal de media 0 y varianza 1, y sea $\phi(x)$ la función de densidad de tal distribución. Por la sección 10-9, la distribución del recorrido de las x es

$$g(R) = n(n-1) \int_{-\infty}^{\infty} [N(R+S) - N(S)]^{n-2} \phi(R+S) \phi(S) dS \quad (1)$$

donde

$$R = y_n - y_1$$

La distribución conjunta de R y v^2 , donde v^2 es una variante ji cuadrado con m grados de libertad, es

$$h(R, v^2) = g(R) \cdot p(v^2) \quad 0 < R < \infty, 0 < v^2 < \infty$$

Pasamos de las variables R, v^2 a las q, u mediante la transformación

$$\frac{R}{v/\sqrt{m}} = q \quad \frac{v^2}{m} = u \quad (2)$$

El jacobiano es $m\sqrt{u}$, y la densidad conjunta de q y u ,

$$k(q, u) = g(q\sqrt{u}) \cdot p(um) \cdot m\sqrt{u} \quad 0 < q < \infty, 0 < u < \infty \quad (3)$$

La densidad de q se obtiene integrando $k(q, u)$ respecto de u :

$$f(q) = \int_0^{\infty} g(q\sqrt{u}) \cdot p(um) \cdot m\sqrt{u} du \quad 0 < q < \infty \quad (4)$$

Para valores de n mayores que 2, esta función es bastante complicada, pero la integral

$$\int_{q_1}^{\infty} f(q) dq = \alpha$$

ha sido tabulada para $\alpha=0,01$; $0,05$ y $0,10$. Las tablas VI a VIII dan q_α para distintos valores de n y m . Todo lo anterior se resume en el siguiente teorema:

Teorema 10-10.—Sea $x_1, x_2, \dots, x_n$ una muestra aleatoria de tamaño n procedente de una distribución normal de media 0 y varianza 1, y v^2 , una variante ji cuadrado independiente con m grados de libertad. Entonces

$$q = \frac{y_n - y_1}{v/\sqrt{m}}$$

se distribuye como el recorrido studentizado con n y m grados de libertad, siendo $y_n - y_1$ el recorrido de las x_i . La densidad es la dada en (4).

PROBLEMAS

1. Aplíquese el método de la ecuación (10-1-4) al ejemplo 10-3.
2. Si la distribución de x viene dada por $f(x)=2x$, $0 < x < 1$, hállese la distribución de $u=(3x-1)^2$.
3. Si la distribución de x viene dada por $f(x)=1$, $0 < x < 1$, hállese la distribución de $\bar{x}$ para muestras x_1, x_2 de tamaño 2. Obsérvese que el recorrido de x_2 para un valor dado de $\bar{x}$ es $0 < x_2 < 2\bar{x}$, cuando $\bar{x} < 1/2$, y $2\bar{x}-1 < x_2 < 1$ para $\bar{x} > 1/2$.
4. Si la distribución de x es la normal con media μ y varianza σ^2 , demuéstrese por transformación de la variante que $u=[(x-\mu)/\sigma]^2$ admite una distribución ji cuadrado con un grado de libertad.
5. Obténgase la distribución de la media de una muestra de tamaño n de una población normal de media -2 y varianza 4, utilizando la función generatriz de momentos.
6. Si $\chi_1^2, \chi_2^2, \chi_3^2, \dots, \chi_k^2$ representan variables independientes ji cuadrado, con $n_1, n_2, \dots, n_k$ grados de libertad, respectivamente, demuéstrese utilizando la función generatriz de momentos que $u=\sum \chi_i^2$ tiene la distribución ji cuadrado con $n=\sum n_i$ grados de libertad.
7. Mediante un razonamiento análogo al utilizado para deducir la distribución ji cuadrado, y teniendo en cuenta que

$$|(1-2t)r_{ij}| = (1-2t)^k |r_{ij}|,$$

pruébese que la forma cuadrática de una distribución normal k variante tiene la distribución ji cuadrado con k grados de libertad.

8. Hállense la media y la varianza de una variante ji cuadrado con k grados de libertad.

9. Utilícese la integral de la distribución F , extendida a todo el recorrido, para hallar una identidad entre los parámetros m y n , y úsese esta identidad para obtener la media y la varianza de F .

10. Hállese el nivel de probabilidad del 0,95 para F con dos y cuatro grados de libertad, por integración directa de la función de densidad.

11. Demuéstrese que

$$w = \frac{mF/n}{1 + (mF/n)}$$

transforma la distribución F en la distribución beta.

12. Pruébese, transformando la variante de la distribución t , que $u = t^2$ tiene la distribución F .

13. Si $x_1, x_2, \dots, x_n$ es una muestra aleatoria de una población normal, demuéstrese que

$$u = \frac{\bar{x} - \mu}{\sqrt{\frac{\sum (x_i - \bar{x})^2}{n(n-1)}}}$$

tiene la distribución t con $n-1$ grados de libertad.

14. Si x_1 y x_2 es una muestra aleatoria de tamaño dos, de una población con $f(x) = e^{-x}$, $x > 0$, demuéstrese que $u = x_1 + x_2$ y $v = x_1/x_2$ tienen distribuciones independientes.

15. Si x, y, z tienen la función de densidad conjunta

$$f(x, y, z) = \frac{6}{(1 + x + y + z)^4} \quad x, y, z > 0$$

hállese la distribución de $u = x + y + z$.

16. Si x_1 y x_2 es una muestra aleatoria de tamaño dos de una población con distribución uniforme en el intervalo unitario, hállese la distribución de $u = x_1 x_2$.

17. Si x e y tienen la distribución normal bivalente, pruébese que

$$u = \frac{x - \mu_x}{\sigma_x} + \frac{y - \mu_y}{\sigma_y}$$

y

$$v = \frac{x - \mu_x}{\sigma_x} - \frac{y - \mu_y}{\sigma_y}$$

tienen distribuciones normales independientes, con medias cero y varianzas $2(1+\rho)$ y $2(1-\rho)$.

18. Si x e y tienen distribuciones normales independientes, con medias cero y varianzas iguales a la unidad, demuéstrese que las distribuciones $u = x^2 + y^2$ y $v = x/y$ son independientes. ¿Cómo se llaman las distribuciones de u y de v ?

19. Pruébese que la distribución de Student se aproxima a la forma normal cuando el número de grados de libertad tiende a infinito.

20. Si $x_1, x_2, \dots, x_n$ es una muestra aleatoria de una población normal, hállese la distribución conjunta de

$$u = \sum_{i=1}^k x_i \quad y \quad v = \sum_{i=r}^n x_i \quad 0 < r < k < n$$

21. Si x e y tienen distribuciones independientes ji cuadrado, con m y n grados de libertad, respectivamente, demuéstrese que las distribuciones de $u = x + y$ y $v = x/y$ son independientes.

22. Considérense muestras de tamaño n de una distribución normal bivalente. Pruébese que

$$\frac{(\hat{\mu}_x + \hat{\mu}_y - \mu_x - \mu_y) \sqrt{n-1}}{\sqrt{\hat{\sigma}_x^2 + \hat{\sigma}_y^2 - 2\hat{\rho}\hat{\sigma}_x\hat{\sigma}_y}}$$

tiene la distribución de Student con $n-1$ grados de libertad.

23. Si x e y son las componentes horizontal y vertical de las desviaciones de un impacto respecto del centro de un blanco, y x e y tienen una distribución normal bivalente con media cero, $\rho=0,1$, y desviaciones estándar de 10 pulg, hállese la ecuación de una elipse que contenga un impacto con probabilidad 0,95. (Usese el resultado del problema 7.)

24. Hállese la media y la varianza de $(1/n)\sum(x_i - \bar{x})^2$ para muestras de tamaño n de una población normal, y demuéstrese que, al crecer n , tienden a la media σ^2 y a la varianza $2\sigma^4/n$ correspondientes a muestras grandes.

25. Si $x_1, x_2, \dots, x_k$ tienen distribuciones independientes y normales con medias μ_i y varianzas σ_i^2 , pruébese que

$$u = \sum_{i=1}^k a_i x_i$$

en donde las a_i son constantes, admite una distribución normal con media $\sum a_i \mu_i$ y varianza $\sum a_i^2 \sigma_i^2$. Dedúzcase la distribución de la media muestral de una población normal, haciendo $a_i = 1/k$.

26. Obténgase un resultado análogo al del problema 25, cuando las x_i tienen la distribución normal multivalente.

27. Hállese la distribución en muestras grandes del estimador del parámetro β en la distribución gamma.

28. Obténgase la distribución en muestras grandes del estimador del parámetro de una distribución de Poisson.

29. Si $(x_{1a}, x_{2a}, \dots, x_{ka})$, $a=1, 2, \dots, n$, es una muestra de tamaño n de una población polinomial con densidad

$$\prod_{i=1}^k p_i^{x_i} \quad x_i = 0, 1; \quad \sum x_i = 1; \quad \sum p_i = 1$$

hállese la distribución de las variantes $n_i = \sum_a x_{ia}$, así como sus varianzas y covarianzas.

*30. Compruébese que (σ_{ij}) , definida en la ecuación (10-8-19), es la inversa de (r_{ij}) dada por la ecuación (10-8-17).

*31. Calcúlese el valor del determinante de (σ_{ij}) en el problema 30.

32. Si $x_1, x_2, \dots, x_n$ tienen distribuciones normales independientes con la misma media, pero diferentes varianzas, $\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2$, pruébese que las distribuciones de

$$u = \frac{\sum x_i / \sigma_i^2}{\sum 1 / \sigma_i^2} \quad \text{y} \quad v = \sum (x_i - u)^2 / \sigma_i^2$$

son independientes. Demuéstrese también que u es normal mientras que v tiene la distribución ji cuadrado con $n-1$ grados de libertad.

33. Representemos por s^2 el cuadrado medio $\sum (x_i - \bar{x})^2 / (n-1)$ para muestras de tamaño n . Para tres muestras de poblaciones normales (con varianzas σ_1^2, σ_2^2 y σ_3^2), de tamaños respectivos n_1, n_2 y n_3 , obténgase la función de densidad conjunta de

$$u = \frac{s_1^2}{s_3^2} \quad \text{y} \quad v = \frac{s_2^2}{s_3^2}$$

en donde s_1^2, s_2^2 y s_3^2 son los cuadrados medios muestrales.

34. Considérese una muestra de tamaño n_1 de una población normal (con varianza σ_1^2) tal que su cuadrado medio sea s_1^2 , y sea una segunda muestra de tamaño n_2 de una segunda población normal (con media μ_2 y varianza σ_2^2) tal que tenga por media $\bar{x}$ y por cuadrado medio s_2^2 . Hállese la densidad conjunta de

$$u = \frac{\sqrt{n_2}(\bar{x} - \mu_2)}{s_2} \quad \text{y} \quad v = \frac{s_1^2}{s_2^2}$$

35. Dedúzcase la distribución del recorrido para una muestra aleatoria de tamaño 2 de una densidad normal, con media 0 y varianza 1.

36. Sea x_1, x_2, x_3 una muestra aleatoria de tamaño 3 de la densidad

$$f(x) = e^{-x} \quad x > 0$$

a) Hállese la densidad del menor estadístico ordinal y_1 .

b) Determinése el valor esperado de y_1 .

37. Hállese el valor esperado de q para $n=2$ en el teorema 10-10.

38. ¿Cuál es la probabilidad de que la mayor de dos observaciones aleatorias de una distribución continua exceda a la mediana?

* Este problema depende de la sección 10-8.

39. Si x_1 y x_2 es una muestra de tamaño 2 de una población con densidad $f(x)=e^{-x}$, $x > 0$, y si designamos con y_1 al menor de estos valores y con y_2 al mayor, ¿cuál es la densidad conjunta de y_1 e y_1+y_2 ?

40. Generalícese el resultado del problema 38 a muestras de tamaño n , llamando y_1 a la menor de las n observaciones e y_n a la mayor.

41. ¿Cuál es la densidad marginal de la observación menor, para muestras de tamaño n y la densidad del problema 36?

42. Si se consideran muestras aleatorias de tamaño n , procedentes de una población con densidad $f(x)$, ¿cuál es el valor esperado del área situada bajo $f(x)$, a la izquierda de la menor observación muestral?

BIBLIOGRAFIA

1. ANDERSON, R. L., y T. A. BANCROFT: *Statistical Theory in Research*, McGraw-Hill Book Company, Inc., Nueva York, 1952.
2. BRUNK, H.: *An Introduction to Mathematical Statistics*, Ginn & Co., Boston, 1960.
3. GRAYBILL, F. A.: *An Introduction to Linear Statistical Models*, vol. I, McGraw-Hill Book Company, Inc., Nueva York, 1961.
4. HOEL, P. G.: *Introduction to Mathematical Statistics*, John Wiley & Sons, Inc., Nueva York, 1956.
5. HOGG, R., y A. CRAIG: *Introduction to Mathematical Statistics*, The MacMillan Company, Nueva York, 1959.
6. KENDALL, M. G.: *The Advanced Theory of Statistics*, vols. I, II, Charles Griffin & Co., Ltd., Londres, 1946.
7. PARZEN, E.: *Modern Probability Theory and Its Applications*, John Wiley & Sons, Inc., Nueva York, 1960.

CAPITULO 11

ESTIMACION POR INTERVALOS

11-1. Intervalos confidenciales.—La estimación puntual de un parámetro no resulta de mucho valor si no se posee alguna medida del posible error cometido en la estimación. Toda estimación $\hat{\theta}$ de un parámetro θ debería acompañarse de cierto intervalo que incluyera a θ , p. ej., de la forma $(\hat{\theta} - d, \hat{\theta} + d)$, junto con alguna medida de seguridad de que el parámetro verdadero θ fuera interior a dicho intervalo. A menudo las estimaciones se dan de esta manera. Así, la carga electrónica puede estimarse que vale $(4,770 \pm 0,005)10^{-10}$ unidades electrostáticas, dándose a entender con ello que es muy poco probable que el primer factor sea exterior al intervalo 4,765 a 4,775. Un contable que se ocupe de los costes de una editorial, y que quiera tener en cuenta todos los factores que intervienen en el coste de producción de cierto libro (costes efectivos de producción, proporción de sostenimiento, proporción de sueldos directos, etc.), podrá estimar el coste en $83 \pm 4,5$ centavos por volumen, lo que significa que es muy probable que el coste correcto esté comprendido entre 78,5 y 87,5 centavos por volumen. La Oficina de Estadística del Trabajo puede estimar el número de parados en un momento dado en $2,4 \pm 0,3$ millones, teniendo bastante seguridad en que el número verdadero está comprendido entre 2,1 y 2,7 millones.

A fin de precisar a estas ideas, consideremos un caso particular. Supongamos una muestra (1,2; 3,4; 0,6; 5,6) de cuatro observaciones, extraída de una población normal de media desconocida μ y desviación estándar conocida 3. La estimación máximo-verosímil de μ es la media de las observaciones muestrales,

$$\bar{x} = 2,7 \quad (1)$$

Queremos determinar los límites superior e inferior entre los cuales queda comprendido, con bastante seguridad, el valor verdadero del parámetro. En general, para muestras de tamaño cuatro, procedentes de la distribución dada, la cantidad

$$y = \frac{\bar{x} - \mu}{3/2} \quad (2)$$

tendrá una distribución normal con media cero y varianza unidad; $\bar{x}$ es la media muestral, y $^{3/2}$ es $\sigma/\sqrt{n}$. Por tanto, la cantidad y tiene por función de densidad

$$f(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} \quad (3)$$

que es independiente del valor verdadero del parámetro desconocido, y se podrá calcular la probabilidad de que y esté situado entre dos números elegidos arbitrariamente. Así, p. ej.,

$$P(-1,96 < y < 1,96) = \int_{-1,96}^{1,96} f(y) dy = 0,95 \quad (4)$$

En esta relación, la desigualdad $-1,96 < y$, o bien

$$-1,96 < \frac{\bar{x} - \mu}{^{3/2}} \quad (5)$$

equivale a la desigualdad

$$\mu < \bar{x} + ^{3/2}(1,96) = \bar{x} + 2,94$$

y la desigualdad

$$y < 1,96$$

es equivalente a

$$\mu > \bar{x} - 2,94$$

cabe por tanto, volver a escribir (4) en la forma

$$P(\bar{x} - 2,94 < \mu < \bar{x} + 2,94) = 0,95 \quad (6)$$

y sustituyendo $\bar{x}$ por 2,7,

$$P(-0,24 < \mu < 5,64) = 0,95 \quad (7)$$

Podemos decir que estos límites obtenidos, $-0,24$ y $5,64$, contendrán el valor del parámetro verdadero, con una seguridad del 95%.

Debe examinarse cuidadosamente el significado de (6) y (7). La probabilidad de que el *intervalo aleatorio* $\bar{x} - 2,94$ a $\bar{x} + 2,94$ cubra a la media verdadera μ es 0,95. Esto es, si se extraen repetidamente de la población muestras de tamaño 4, y si se calcula para cada muestra el intervalo aleatorio $\bar{x} - 2,94$ a $\bar{x} + 2,94$, es de esperar que el 95% de estos intervalos contengan la media verdadera μ . Tenemos, pues, una gran confianza en que el intervalo $-0,24$ a $5,64$ cu-

bra la media verdadera. La medida de nuestra confianza es 0,95, porque *antes* de extraer la muestra, la probabilidad de que el intervalo que intentamos construir cubra la media verdadera es 0,95.

El intervalo $-0,24$ a $5,64$ recibe el nombre de *intervalo confidencial* o, más concretamente, intervalo confidencial del 95%; la probabilidad, en este caso 0,95, se denomina coeficiente confidencial o coeficiente de confianza.

Es posible obtener intervalos con cualquier grado de confianza que se desee. Así, puesto que

$$P(-2,58 < y < 2,58) = 0,99 \quad (8)$$

se obtiene un intervalo confidencial del 99% para la media verdadera considerando las desigualdades como antes, y sustituyendo $\bar{x} = 2,7$, con lo que resulta

$$P(-1,17 < \mu < 6,57) = 0,99 \quad (9)$$

Debe observarse que hay muchos intervalos posibles con la misma probabilidad. Así, p. ej., ya que

$$P(-1,68 < y < 2,70) = 0,95 \quad (10)$$

tenemos otro intervalo confidencial del 95% para μ , dado por

$$P(-1,35 < \mu < 5,22) = 0,95 \quad (11)$$

Este intervalo es inferior al antes obtenido, ya que su longitud 6,57 es superior a la longitud 5,88 del intervalo dado en (7), por lo que procura una información menos precisa sobre la situación de μ . Dos números cualesquiera a y b , tales que las ordenadas que les corresponden incluyan el 95% del área limitada por $f(y)$, determinarán un intervalo confidencial del 95%. En general, se desea que el intervalo confidencial sea lo más pequeño posible; esto se logra haciendo que a y b estén tan próximos como sea posible, ya que la relación $P(a < y < b) = 0,95$ da lugar a un intervalo confidencial de longitud $(\sigma/\sqrt{n})(b-a)$. La distancia $(b-a)$ se hace mínima para un área dada cuando $f(a) = f(b)$, como se ve claramente en la figura 11-1. Si el punto b se desplaza un poco hacia la izquierda, el a deberá moverse una distancia menor hacia la izquierda, a fin de que el área siga siendo la misma; esta operación disminuye la longitud del intervalo y continúa disminuyéndola mientras $f(b) < f(a)$. Como en este ejemplo $f(y)$ es simétrica respecto a $y=0$, el valor mínimo de $b-a$, para un valor prefijado del área, corresponde a $b = -a$. Por tanto, (7) da el intervalo confidencial más corto del

95%, y (9) da el intervalo confidencial más corto del 99%, ambos para el parámetro μ .

En muchos problemas no es posible construir los intervalos confidenciales más cortos para un coeficiente de confianza dado. En estos casos, resultará deseable hallar un intervalo confidencial que

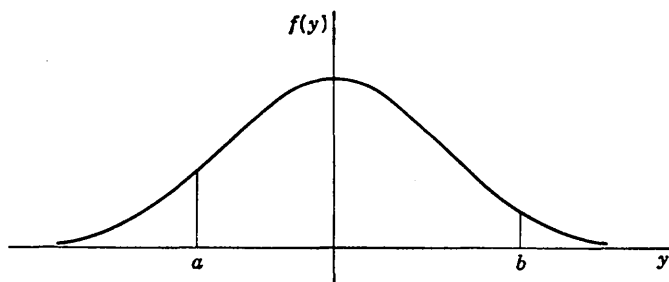


FIG. 11-1.

tenga la más corta longitud esperada, o que sea tal que haga mínima la probabilidad de que el intervalo confidencial cubra un valor μ^* , donde $\mu^* \neq \mu$. En este libro no examinaremos estos conceptos.

El método general que aquí exponemos es el siguiente: se halla, si es posible, una función de las observaciones muestrales y del parámetro a estimar (la función y anterior), cuya distribución sea independiente del parámetro y de otros parámetros cualesquiera. Entonces, cualquier afirmación probabilística de la forma $P(a < y < b) = \gamma$, en donde y es la función, dará lugar a una afirmación probabilística relativa al parámetro. Esta técnica es aplicable en muchos problemas importantes, pero hay también otros muchos en los que no puede aplicarse, por ser imposible hallar funciones de la forma deseada y cuya distribución no dependa de parámetros. Estos últimos problemas se abordan mediante una técnica más general que describiremos en la sección 11-5.

La idea de la estimación por intervalos puede generalizarse de modo que incluya la estimación simultánea de varios parámetros. Así, los dos parámetros de la distribución normal se estiman mediante una cierta región plana R , en el llamado espacio paramétrico, espacio de todas las combinaciones posibles de los valores de μ y σ^2 . Una región confidencial del 95% es una región que se puede construir a partir de la muestra, de tal forma que, si se extraen muestras repetidamente, construyendo una región para cada una de ellas, el 95% (por término medio) de estas regiones incluirán el punto paramétrico verdadero (μ_0, σ_0^2) . (Véase Fig. 11-2.)

Los intervalos y regiones confidenciales ilustran adecuadamente acerca de la incertidumbre de las inferencias. En (7) se hizo la inferencia de que el intervalo $-0,24$ a $5,64$ cubre el valor verdadero del parámetro, pero no se estableció de forma categórica. La medida $0,05$ de la incertidumbre de esta inferencia constituye parte esencial de la afirmación.

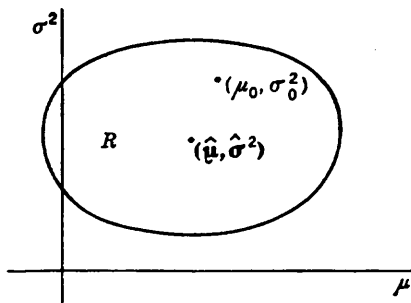


FIG. 11-2.

11-2. Intervalos confidenciales para la media de una distribución normal.—El método utilizado en la sección anterior no suele ser de posible utilización para estimar la media de una población normal, pues lo corriente es que se desconozca la varianza σ^2 . La función y toma la forma (para muestras de tamaño n)

$$y = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \quad (1)$$

y transformando las desigualdades:

$$P(-1,96 < y < 1,96) = 0,95 \quad (2)$$

se tiene

$$P\left(\bar{x} - 1,96 \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 1,96 \frac{\sigma}{\sqrt{n}}\right) = 0,95 \quad (3)$$

Para una muestra dada se conocen $\bar{x}$ y n , pero no σ , de modo que no será posible calcular límites para μ . Claro es que puede sustituirse σ por una estimación $\hat{\sigma}$; pero entonces la afirmación probabilística ya no sería exacta, y para muestras pequeñas podría ser muy errónea.

W. S. Gossett (que utilizó el seudónimo de *Student*) indicó el camino para resolver esta dificultad en una publicación clásica en que introdujo la distribución t . Se le considera como fundador de la teoría de la inferencia estadística exacta. La cantidad

$$t = \frac{\bar{x} - \mu}{\sqrt{\sum (x_i - \bar{x})^2 / n(n-1)}} \quad (4)$$

comprende solo el parámetro μ y tiene la distribución t con $n-1$ grados de libertad, sin incluir parámetros desconocidos. Por tanto, será posible hallar un número $t_{0,05}$ tal que

$$P(-t_{0,05} < t < t_{0,05}) = \int_{-t_{0,05}}^{t_{0,05}} f(t; n-1) dt = 0,90 \quad (5)$$

convirtiendo después las desigualdades para obtener

$$P \left[\bar{x} - t_{0,05} \sqrt{\frac{\sum(x_i - \bar{x})^2}{n(n-1)}} < \mu < \bar{x} + t_{0,05} \sqrt{\frac{\sum(x_i - \bar{x})^2}{n(n-1)}} \right] = 0,90 \quad (6)$$

donde los límites se calculan para cada muestra dada, obteniendo así un intervalo confidencial del 90%.

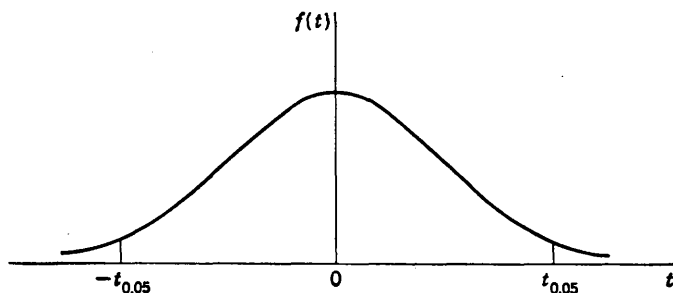


FIG. 11-3.

El número $t_{0,05}$ recibe el nombre de nivel del 5% de t , y sitúa a los puntos que separan un 5% del área limitada por $f(t)$ en cada rama de la curva. Cabe obtener otros intervalos confidenciales, empleando distintos niveles de t . Así, se puede hallar un intervalo confidencial del 98% usando el número $t_{0,01}$, que separa 0,01 del área en cada rama de la distribución (véase Fig. 11-3).

En este ejemplo la longitud del intervalo confidencial es

$$\begin{aligned} w &= \bar{x} + t_{0,05} \sqrt{\frac{\sum(x_i - \bar{x})^2}{n(n-1)}} - \bar{x} + t_{0,05} \sqrt{\frac{\sum(x_i - \bar{x})^2}{n(n-1)}} \\ &= 2t_{0,05} \sqrt{\frac{\sum(x_i - \bar{x})^2}{n(n-1)}} \end{aligned}$$

La longitud es una variable aleatoria, ya que es función de las variables aleatorias x_i . Es también función del tamaño n de la mues-

tra en que se basa el intervalo confidencial. Si este es muy amplio, quizá resulta poco útil aunque sea alta la probabilidad de que cubra al parámetro desconocido. Así, es preciso que el tamaño n de la muestra sea suficientemente grande para que siendo la probabilidad alta, la longitud resulte lo bastante pequeña para ser útil. La teoría sobre cuál sería el tamaño apropiado de la muestra cae fuera de los objetivos de este libro, pero el lector puede consultar los trabajos citados en la bibliografía del capítulo.

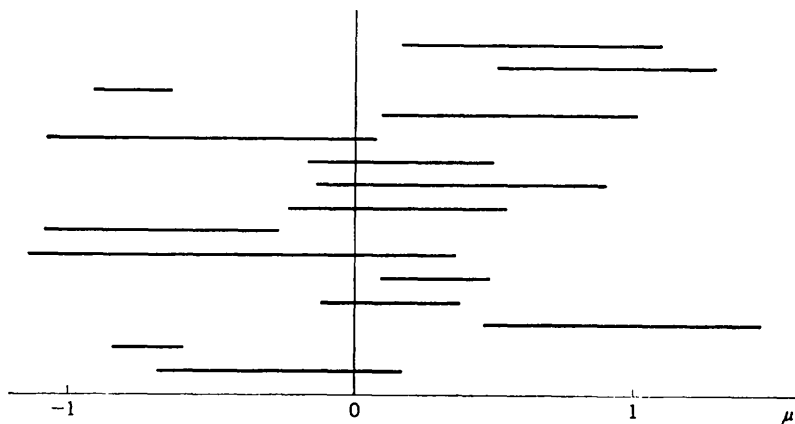


FIG. 11-4.

La figura 11-4 indica el resultado del cálculo de intervalos confidenciales del 50% para 15 muestras de tamaño cuatro, extraídas efectivamente de una población normal con media cero y varianza unidad. Los intervalos se representan por segmentos horizontales situados por encima del eje μ , y, como era de esperar, la mitad aproximadamente cubre la media verdadera cero. Análogamente, si se utilizaran intervalos confidenciales del 95%, sería de esperar que el 95% de ellos aproximadamente cubriría la media verdadera. Si se emplean sistemáticamente intervalos confidenciales del 95% para estimar parámetros, afirmando a la vez que el intervalo contiene el valor verdadero del parámetro, es de suponer que el 5% de estas afirmaciones resulten erróneas.

11-3. Intervalos confidenciales para la varianza de una distribución normal.—Para muestras de tamaño n de una población normal la cantidad

$$u = \frac{\sum (x_i - \bar{x})^2}{\sigma^2} \quad (1)$$

donde $\bar{x}$ es la media muestral, tiene la distribución ji cuadrado con $n-1$ grados de libertad. Por tanto, puede construirse un intervalo confidencial con coeficiente confidencial γ , hallando dos números a y b tales que

$$P(a < u < b) = \int_a^b f(\chi^2) d\chi^2 = \gamma \quad (2)$$

Transformando las desigualdades, obtenemos

$$P \left[\frac{\sum (x_i - \bar{x})^2}{b} < \sigma^2 < \frac{\sum (x_i - \bar{x})^2}{a} \right] = \gamma \quad (3)$$

que proporciona un intervalo confidencial para σ^2 .

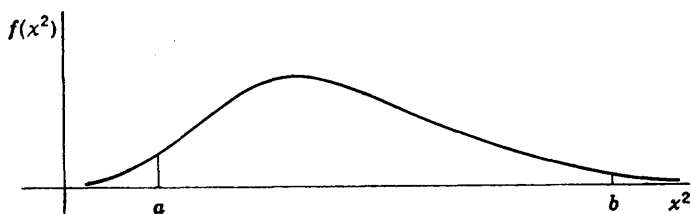


FIG. 11-5.

Puesto que la longitud de este es

$$\left(\frac{1}{a} - \frac{1}{b} \right) \sum (x_i - \bar{x})^2 \quad (4)$$

el intervalo confidencial más pequeño para una muestra dada se obtendría eligiendo a de modo que $[(1/a) - (1/b)]$ resultase mínimo para el valor elegido de γ . El cálculo necesario resulta muy laborioso. Las tablas ordinarias de la distribución ji cuadrado proporcionan números χ^2 tales que

$$P(u > \chi^2_\epsilon) = \int_{\chi^2_\epsilon}^{\infty} f(\chi^2) d\chi^2 = \epsilon \quad (5)$$

para valores elegidos de ϵ . Al construir, p. ej., un intervalo confidencial del 95 %, se suele elegir $a = \chi^2_{0,975}$ y $b = \chi^2_{0,025}$, esto es, se eligen a y b de modo que quede separado 0,025 del área en cada rama de la distribución. Esto da aproximadamente la longitud mínima del intervalo confidencial, a menos que el número de grados de libertad sea muy pequeño (véase Fig. 11-5).

11-4. Región confidencial para la media y la varianza de una distribución normal.—Al construir una región para la distribución conjunta de la media μ_0 y la varianza σ_0^2 de una distribución normal, cabría inclinarse a primera vista a utilizar las estimaciones individuales dadas por las distribuciones t y χ^2 . Así, p. ej., podría construirse una región 0,9025 ($=0,95^2$), como en la figura 11-6, haciendo uso de las dos relaciones:

$$P \left[\bar{x} - t_{0,025} \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n(n-1)}} < \mu_0 < \bar{x} + t_{0,025} \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n(n-1)}} \right] = 0,95 \quad (1)$$

$$P \left[\frac{\Sigma(x_i - \bar{x})^2}{\chi_{0,025}^2} < \sigma_0^2 < \frac{\Sigma(x_i - \bar{x})^2}{\chi_{0,975}^2} \right] = 0,95 \quad (2)$$

y suponiendo que la probabilidad de los dos sucesos fuera el producto de las probabilidades de cada uno. Esto no es correcto, puesto

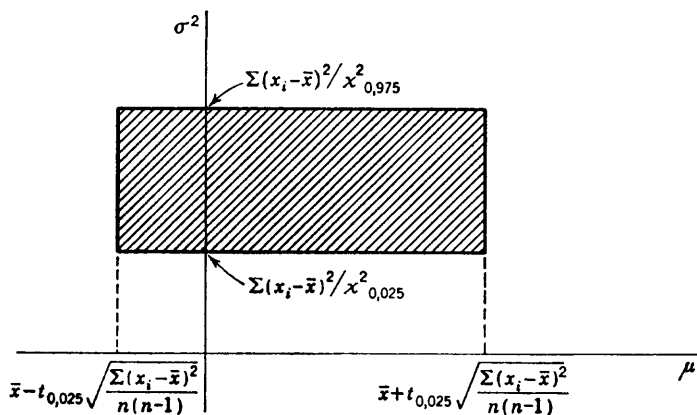


FIG. 11-6.

que las distribuciones de t y χ^2 no son independientes. La probabilidad conjunta de que ambos parámetros cubran los valores del parámetro verdadero no es igual al producto de las probabilidades correspondientes. Por tanto, la probabilidad de que la región rectangular de la figura 11-6 cubra al punto paramétrico verdadero (μ_0, σ_0^2) no es 0,9025.

Sin embargo es posible construir una región confidencial utilizando las distribuciones de $\bar{x}$ y $\Sigma(x_i - \bar{x})^2$, que son independientes.

Si, p. ej., se desea una región confidencial del 95%, pueden hallarse números a , a' y b' tales que

$$P\left(-a < \frac{\bar{x} - \mu_0}{\sigma_0/\sqrt{n}} < a\right) = \sqrt{0,95} \cong 0,975 \quad (3)$$

$$P\left[a' < \frac{\sum(x_i - \bar{x})^2}{\sigma_0^2} < b'\right] = \sqrt{0,95} \quad (4)$$

La probabilidad conjunta es

$$P\left[-a < \frac{\bar{x} - \mu_0}{\sigma_0/\sqrt{n}} < a, a' < \frac{\sum(x_i - \bar{x})^2}{\sigma_0^2} < b'\right] = 0,95 \quad (5)$$

debido a la independencia de las distribuciones. Las cuatro desigualdades de (5) determinan una región en el espacio paramétrico, fácil de determinar trazando las líneas que la limitan. Basta reemplazar los

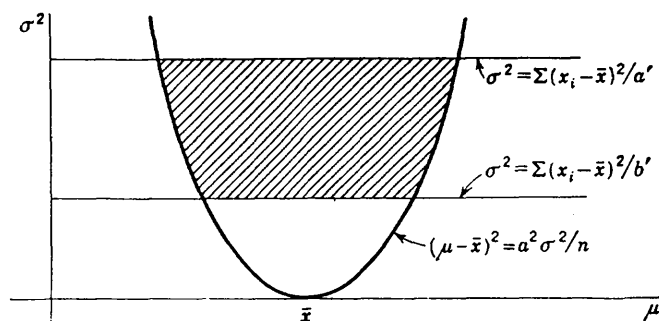


FIG. 11-7.

signos de desigualdad por otros de igualdad y representar cada una de las cuatro relaciones resultantes como funciones de μ y σ^2 en el espacio paramétrico. Resultará así una región como la que aparece rayada en la figura 11-7. Exactamente del mismo modo se obtendría una región confidencial para (μ_0, σ_0) ; las relaciones se representarían como funciones de σ en lugar de σ^2 , y la parábola de la figura 11-7 se transformaría en un par de rectas

$$\mu = \bar{x} \pm \frac{a\sigma}{\sqrt{n}}$$

que se cortarían en $\bar{x}$ sobre el eje de las μ .

La región que hemos construido no es la de área mínima, pero se construye fácilmente a partir de tablas y difiere poco de la región de área mínima, a menos que sea pequeño el tamaño de la muestra. La región mínima es, aproximadamente, de forma elíptica y difícil de construir.

11-5. Método general para la obtención de intervalos confidenciales.—El método utilizado en las secciones anteriores para la determinación de intervalos y regiones confidenciales obliga a encontrar funciones de la muestra y de los parámetros, distribuidas independientemente de estos. No obstante, es posible establecer intervalos confidenciales sin tener en cuenta la existencia previa de tales funciones.

Dada una población por $f(x; \theta)$ y un estimador $\hat{\theta}(x_1, x_2, \dots, x_n)$ para muestras de tamaño n (generalmente, se usará el estimador de máxima verosimilitud), determinaremos la distribución del estimador, que vendrá dada por $g(\hat{\theta}; \theta)$. Supongamos, para fijar ideas, que se desea un intervalo confidencial del 95%. Si se sustituye θ , en $g(\hat{\theta}; \theta)$, por el número arbitrario θ' , la distribución de $\hat{\theta}$ quedará completamente especificada, y será posible dar enunciados probabilísticos relativos a $\hat{\theta}$. En particular, será posible hallar dos números h_1 y h_2 tales que

$$P(\hat{\theta} < h_1) = \int_{-\infty}^{h_1} g(\hat{\theta}; \theta') d\hat{\theta} = 0,025 \quad (1)$$

$$P(\hat{\theta} > h_2) = \int_{h_2}^{\infty} g(\hat{\theta}; \theta') d\hat{\theta} = 0,025 \quad (2)$$

Claro es que los números h_1 y h_2 dependerán del número que sustituye a θ en $g(\hat{\theta}; \theta)$. En efecto, h_1 y h_2 son ciertas funciones de θ , esto es $h_1(\theta)$ y $h_2(\theta)$. Los valores de estas funciones para cualquier valor de θ vienen determinados por las dos ecuaciones anteriores. Evidentemente,

$$P[h_1(\theta) < \hat{\theta} < h_2(\theta)] = \int_{h_1(\theta)}^{h_2(\theta)} g(\hat{\theta}; \theta) d\hat{\theta} = 0,95 \quad (3)$$

Las funciones $h_1(\theta)$ y $h_2(\theta)$ pueden representarse en función de θ , como se ha hecho en la figura 11-8. Trazando una vertical por cualquier valor θ' de θ , esta cortará a ambas curvas en puntos que, proyectados sobre el eje de las $\hat{\theta}$, darán límites entre los cuales caerá $\hat{\theta}$, con probabilidad de 0,95.

Construidas las dos curvas $\theta = h_1(\hat{\theta})$ y $\theta = h_2(\hat{\theta})$, cabe obtener un intervalo confidencial para θ del siguiente modo: Se extrae una mues-

tra de tamaño n y se calcula el valor del estimador θ' . La horizontal trazada por el punto θ' del eje θ (Fig. 11-8) cortará a ambas curvas en puntos que pueden proyectarse sobre el eje θ y que llamaremos θ_1 y θ_2 , según se indica en la figura. Estos dos números definen el intervalo confidencial, pues se ve fácilmente que

$$P(\theta_2 < \theta < \theta_1) = 0,95 \quad (4)$$

Supongamos que estuviésemos extrayendo muestras de una población en que el valor de θ fuese θ' . La probabilidad de que la estimación $\hat{\theta}$ quede comprendida entre $h_1(\theta')$ y $h_2(\theta')$ es 0,95. Si la estimación cae entre estos dos límites, dicha horizontal cortará a la verti-

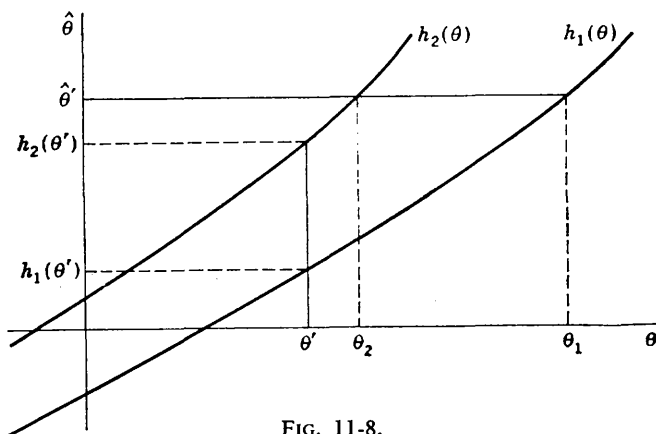


FIG. 11-8.

cal trazada por θ' en cierto punto situado entre las curvas, y el intervalo correspondiente (θ_2, θ_1) cubrirá a θ' . Si la estimación no cae entre $h_1(\theta')$ y $h_2(\theta')$, la horizontal no cortará a la vertical entre las curvas, y el intervalo correspondiente (θ_2, θ_1) no cubrirá a θ' . Se deduce, por tanto, que la probabilidad de que un intervalo (θ_2, θ_1) , construido por este método, cubra a θ' , es exactamente 0,95. Esta afirmación es cierta cualquiera que sea el valor de θ en la población. A veces, es posible determinar los límites θ_2 y θ_1 para una estimación dada, sin necesidad de hallar efectivamente las funciones $h_1(\theta)$ y $h_2(\theta)$. Con referencia a la figura 11-8, los límites para θ son los puntos θ_2 y θ_1 , tales que $h_1(\theta_1) = \theta'$ y $h_2(\theta_2) = \theta'$. Basándonos en la definición de h_1 y h_2 , diremos que θ_1 es el valor de θ para el cual

$$\int_{-\infty}^{\hat{\theta}'} g(\theta; \theta) d\theta = 0,025 \quad (5)$$

y θ_2 es el valor de θ para el cual

$$\int_{\hat{\theta}'}^{\infty} g(\theta; \theta) d\theta = 0,025 \quad (6)$$

Si es posible expresar los primeros miembros de estas dos ecuaciones explícitamente en función de θ , y si las ecuaciones pueden resolverse unívocamente respecto a θ , las raíces son los límites confidenciales del 95%, para θ .

Si $h_1(\theta)$ y $h_2(\theta)$ no son funciones monótonas de θ , el intervalo confidencial puede ser, en realidad, un conjunto de intervalos. Así p. ej., supongamos que las curvas de la figura 11-8 se inclinaran más hacia la derecha, de modo que la horizontal trazada por θ' volviera a cortarlas, p. ej., en los puntos θ_3 y θ_4 . El intervalo confidencial consistiría entonces en dos intervalos (θ_2, θ_1) y (θ_3, θ_4) . La afirmación sobre θ sería de la forma

$$P(\theta_2 < \theta < \theta_1, \text{ o } \theta_3 < \theta < \theta_4) = 0,95 \quad (7)$$

Sin embargo, en la mayoría de las situaciones que se plantean en la práctica habrá un intervalo único, o será posible elegir un intervalo único basándose en otros datos disponibles relativos al experimento que dio lugar a las observaciones muestrales.

El método aquí descrito para la obtención de intervalos confidenciales se extiende al caso de varios parámetros; pero la representación geométrica ya no es posible, ni siquiera para dos parámetros. Supongamos una distribución que dependa de dos parámetros θ_1 y θ_2 ; podemos hallar una región plana R en el plano θ_1, θ_2 tal que

$$P(\hat{\theta}_1, \hat{\theta}_2 \text{ en } R) = \int_R \int g(\theta_1, \theta_2; \theta_1, \theta_2) d\theta_1 d\theta_2 = 0,95 \quad (8)$$

Considerando todos los pares posibles de valores θ_1 y θ_2 limitaremos una región cuatridimensional en el espacio, $\theta_1, \theta_2, \hat{\theta}_1, \hat{\theta}_2$, que es análoga a la región bidimensional situada entre las curvas de la figura 11-8. Supongamos ahora que se extrae una muestra y se calculan las estimaciones θ'_1 y θ'_2 . La intersección de los dos hiperplanos $\hat{\theta}_1 = \theta'_1$ y $\hat{\theta}_2 = \theta'_2$ con la región cuatridimensional determinará una región bidimensional que, proyectada sobre el plano θ_1, θ_2 , será una región confidencial del 95% para θ_1, θ_2 .

Este razonamiento se generaliza para abarcar el caso de k parámetros. El método determinará una región confidencial para todos los parámetros de una distribución. Si se desea estimar algunos, pero no todos los parámetros de un conjunto de ellos, dicho método

no podrá usarse en general, pero en determinadas circunstancias, sí puede modificarse para adaptarse al problema en cuestión. Por ahora, no hay solución general del problema de construir regiones confidenciales para una parte del conjunto de k parámetros de una función de distribución, excepto en el caso de muestras grandes.

Ejemplo 11-1.—Como simple aclaración, consideraremos la estimación de α en

$$f(x; \alpha) = \frac{2}{\alpha^2} (\alpha - x) \quad 0 < x < \alpha \quad (9)$$

para muestras de tamaño uno. Si x es la observación, se encuentra como estimador máximo-verosímil $\hat{\alpha} = 2x$ resolviendo

$$\frac{\partial}{\partial \alpha} \left[\frac{2}{\alpha^2} (\alpha - x) \right] = 0$$

respecto a α . La distribución del estimador es

$$g(\hat{\alpha}; \alpha) = \frac{1}{2\alpha^2} (2\alpha - \hat{\alpha}) \quad 0 < \hat{\alpha} < 2\alpha \quad (10)$$

de forma que los intervalos confidenciales del 95% se obtienen determinando $h_1(\alpha)$ y $h_2(\alpha)$ de modo que

$$\int_0^{h_1(\alpha)} g(\hat{\alpha}; \alpha) d\hat{\alpha} = 0,025 \quad (11)$$

$$\int_{h_2(\alpha)}^{2\alpha} g(\hat{\alpha}; \alpha) d\hat{\alpha} = 0,025 \quad (12)$$

Las integraciones se efectúan fácilmente en este caso y dan, resolviendo las ecuaciones en h_1 y h_2 ,

$$h_1(\alpha) = 2(1 - \sqrt{0,975})\alpha \quad (13)$$

$$h_2(\alpha) = 2(1 - \sqrt{0,025})\alpha \quad (14)$$

La representación nos da rectas, como en la figura 11-9. Para una observación dada, como $x=2$, la estimación es $\hat{\alpha}'=4$ y el intervalo confidencial del 95% viene dado por

$$P \left(\frac{2}{1 - \sqrt{0,025}} < \alpha < \frac{2}{1 - \sqrt{0,975}} \right) = 0,95 \quad (15)$$

Realmente, como

$$u = \frac{2\alpha - \hat{\alpha}}{\alpha}$$

tiene una distribución independiente de α , en este problema no era necesario usar el método general. Bastaría haber hallado un inter-

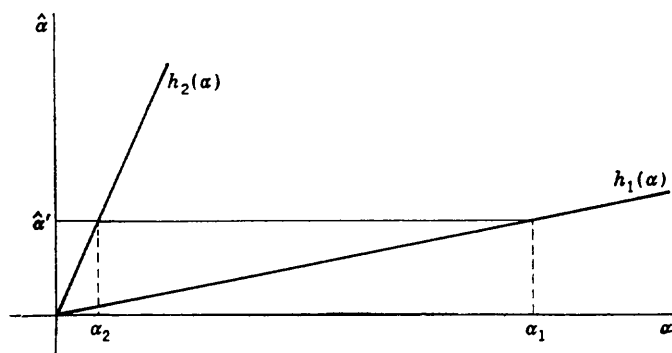


FIG. 11-9.

valo confidencial para α , obtener límites del 0,95 para u , y transformar después las desigualdades para deducir el correspondiente resultado relativo a α .

11-6. Intervalos confidenciales para el parámetro de una distribución binomial.—Aplicaremos el método general descrito en la sección precedente a un problema que exige su empleo. Si una muestra $x_1, x_2, \dots, x_n$ procede de una población binomial con

$$f(x; p) = p^x(1-p)^{1-x} \quad x=0, 1; \quad 0 \leq p \leq 1 \quad (1)$$

el estimador máximo-verosímil de p es

$$\hat{p} = \frac{y}{n} \quad (2)$$

en donde $y = \sum x_i$ puede tomar los valores 0, 1, 2, ..., n . La distribución de $\hat{p}$ viene dada por

$$g(\hat{p}; p) = \binom{n}{n\hat{p}} p^{n\hat{p}}(1-p)^{n(1-\hat{p})} \quad \hat{p} = 0, \frac{1}{n}, \frac{2}{n}, \dots, 1 \quad (3)$$

y no es posible hallar una función de $\hat{p}$ y p , cuya distribución sea independiente de p .

Volveremos a suponer, para fijar ideas, que el intervalo confidencial a construir es del 95%. El primer paso consiste en determinar las funciones $h_1(p)$ y $h_2(p)$. Así, para $p=0,4$, y de acuerdo con la sección anterior, buscaríamos un número $h_1(0,4)$, tal que

$$P[\hat{p} < h_1(0,4)] = \sum_{y=0}^{nh_1} \binom{n}{y} (0,4)^y (0,6)^{n-y} = 0,025 \quad (4)$$

No obstante, por tratarse de una distribución discreta, nh_1 deberá ser un entero, y será imposible lograr que la suma valga exactamente 0,025 para todo valor de p . Sin embargo, no nos preocuparemos por esto, ya que no necesitamos una curva $h_1(p)$ definida para todo valor de p . Los únicos puntos de interés son los que corresponden a valores posibles de $\hat{p}$. En efecto, es posible utilizar la técnica indicada por las ecuaciones (11-5-5) y (11-5-6), por disponerse inmediatamente de una expresión explícita para las probabilidades que figuran en el primer miembro de dichas ecuaciones. Suponiendo que tenemos una estimación

$$\hat{p}' = \frac{k}{n} \quad (5)$$

puede determinarse el límite superior confidencial p_1 , del 95%, hallando el valor de p para el cual

$$\sum_{y=0}^k \binom{n}{y} p^y (1-p)^{n-y} = 0,025 \quad (6)$$

siendo el límite inferior p_2 el valor de p para el cual

$$\sum_{y=k}^n \binom{n}{y} p^y (1-p)^{n-y} = 0,025 \quad (7)$$

Si es $k=0$, se toma cero como límite inferior, y si $k=n$, se toma uno como límite superior.

Para valores pequeños de n , las ecuaciones (6) y (7) pueden resolverse por tanteos, a fin de obtener las raíces p_1 y p_2 ; pero este cálculo se hace más prolijo a medida que aumenta n . Un método sencillo consiste en utilizar las tablas de Pearson para la función beta incompleta. La forma acumulativa de la distribución beta es

$$F(x; \alpha, \beta) = \frac{(\alpha + \beta + 1)!}{\alpha! \beta!} \int_0^x t^\alpha (1-t)^\beta dt \quad (8)$$

y por integración reiterada por partes se obtiene

$$F(x; \alpha, \beta) = - \sum_{i=0}^{\alpha} \binom{\alpha + \beta + 1}{i} x^i (1-x)^{\alpha + \beta + 1 - i} + 1 \quad (9)$$

Se deduce que las sumas binomiales parciales vienen dadas por la tabla de $F(x; \alpha, \beta)$. Podemos escribir la ecuación (6) del siguiente modo:

$$\sum_{y=0}^k \binom{n}{y} p^y (1-p)^{n-y} = 1 - F(p; k, n-k-1) = 0,025, \quad (10)$$

hallando inmediatamente en la tabla el valor de p que corresponde a $F=0,975$ para los valores dados de k y $n-k-1$. Análogamente, puesto que

$$\sum_k^n \binom{n}{y} p^y (1-p)^{n-y} = 1 - \sum_0^{k-1} \binom{n}{y} p^y (1-p)^{n-y}$$

se obtendrá el límite confidencial inferior escribiendo (7) en la forma

$$\sum_k^n \binom{n}{y} p^y (1-p)^{n-y} = F(p; k-1, n-k) = 0,025 \quad (11)$$

Para valores de n que excedan de los tabulados, puede emplearse la aproximación normal a la distribución binomial, y obtener intervalos confidenciales de p , tal como se indica en la sección siguiente, o bien utilizar las *Tables of the Binomial Probability Distribution* (National Bureau of Standards, Applied Mathematics Series 6, Washington D. C., 1950).

11-7. Intervalos confidenciales para muestras grandes.—Hemos visto en el capítulo 10 que, para muestras grandes, el estimador $\hat{\theta}$ máximo-verosímil para un parámetro θ de una distribución dada por $f(x; \theta)$ tiene, bajo condiciones bastante generales, una distribución aproximadamente normal respecto de θ . Cuando se satisfacen tales condiciones, se obtienen fácilmente intervalos confidenciales aproximados. La varianza del estimador en muestras grandes es

$$\sigma^2(\theta) = \frac{-1}{nE[\partial^2 \log f(x; \theta)/\partial \theta^2]} \quad (1)$$

en donde $\sigma^2(\theta)$ indica que es una función de θ , porque ordinariamente dependerá de este parámetro. Para muestras grandes, por tanto, puede determinarse un intervalo confidencial con probabilidad γ , convirtiendo las desigualdades en

$$P \left[-d_\gamma < \frac{\hat{\theta} - \theta}{\sigma(\theta)} < d_\gamma \right] \cong 2\gamma \quad (2)$$

en donde d_γ se ha elegido de modo que

$$\int_{-d_\gamma}^{d_\gamma} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt = 2\gamma$$

Ejemplo 11-2.—Como ejemplo, consideraremos la distribución binomial con parámetro p . La varianza $\hat{p}$ es

$$\sigma^2(p) = \frac{p(1-p)}{n} \quad (3)$$

Así se consigue un intervalo confidencial 2γ aproximado, convirtiendo las desigualdades de

$$P \left[-d_\gamma < \frac{\hat{p} - p}{\sqrt{p(1-p)/n}} < d_\gamma \right] \cong 2\gamma \quad (4)$$

para obtener

$$P \left[\frac{2n\hat{p} + d_\gamma^2 - d_\gamma \sqrt{4n\hat{p} + d_\gamma^2 - 4n\hat{p}^2}}{2(n + d_\gamma^2)} < p < \frac{2n\hat{p} + d_\gamma^2 + d_\gamma \sqrt{4n\hat{p} + d_\gamma^2 - 4n\hat{p}^2}}{2(n + d_\gamma^2)} \right] \cong 2\gamma \quad (5)$$

Estas expresiones de los límites quedan simplificadas si recordamos que, al deducir la distribución en muestras grandes, despreciábamos algunos términos que contenían el factor $1/\sqrt{n}$; esto es, la distribución normal asintótica es correcta solo dentro de términos de error de tamaño $k/\sqrt{n}$. Podemos, pues, despreciar los términos de este orden en los límites que figuran en (5), sin que ello afecte al grado de aproximación. Esto significa simplemente que es posible omitir todas las d_γ^2 que figuran en (5), por presentarse siempre unidas a un término con el factor n , por lo cual serán despreciables en

relación a n , cuando este sea grande, dentro del grado de aproximación admitido. Por tanto, (5) adopta la nueva forma

$$P \left[\hat{p} - d_\gamma \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < p < \hat{p} + d_\gamma \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right] \cong 2\gamma \quad (6)$$

En particular,

$$P \left[\hat{p} - 1,96 \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < p < \hat{p} + 1,96 \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right] \cong 0,95$$

da para p un intervalo confidencial aproximado del 95%, en el supuesto de tratarse de muestras grandes.

Observemos que (6) es, exactamente, la expresión que se hubiera obtenido al sustituir p por $\hat{p}$ en $\sigma^2(p)$. Esta sustitución supondría que

$$\frac{\hat{p} - p}{\sqrt{\hat{p}(1-\hat{p})/n}}$$

tiene una distribución aproximadamente normal, con media cero y varianza unidad. En efecto, lo que ocurre en general es que, en la distribución normal asintótica de un estimador máximo-verosímil $\hat{\theta}$, la varianza $\sigma^2(\hat{\theta})$ puede reemplazarse por un estimador $\sigma^2(\hat{\theta})$, sin que ello afecte apreciablemente al grado de aproximación. No demostraremos esta afirmación, aunque sí la utilizaremos, ya que simplifica mucho la conversión de las desigualdades que aparecen en una afirmación probabilística, a fin de obtener intervalos confidenciales.

Por tanto, para muestras grandes, se obtiene un intervalo confidencial aproximado, con coeficiente confidencial 2γ , a partir de

$$P[\hat{\theta} - d_\gamma \sigma(\hat{\theta}) < \theta < \hat{\theta} + d_\gamma \sigma(\hat{\theta})] \cong 2\gamma \quad (7)$$

en donde la distribución de $\hat{\theta}$ es asintóticamente normal, y la $\sigma(\hat{\theta})$ que aparece en esta expresión es la estimación máximo-verosímil de la desviación estándar de $\hat{\theta}$.

*11-8. Regiones confidenciales para muestras grandes.—

Cuando una distribución incluye varios parámetros ($\theta_1, \theta_2, \dots, \theta_k$), hemos visto en el capítulo 10 que, en condiciones bastantes generales, las estimaciones máximo-verosímiles en muestras grandes, ($\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_k$),

* Esta sección depende de la sección 10-8 y debe ser omitida si se omitió dicha sección 10-8.

tienen una distribución que es aproximadamente normal, con medias $(\theta_1, \theta_2, \dots, \theta_k)$ y coeficientes de la forma cuadrática dados por

$$r_{ij} = -nE \left[\frac{\partial^2 \log f(\mathbf{x}; \theta_1, \theta_2, \dots, \theta_k)}{\partial \theta_i \partial \theta_j} \right] \quad (1)$$

Según sabemos, los coeficientes serán, en general, funciones de las θ .

Hemos visto que la forma cuadrática de una distribución normal k -variante tiene la distribución ji cuadrado con k grados de libertad. Por consiguiente, la cantidad

$$u = \sum_{i=1}^k \sum_{j=1}^k r_{ij} (\hat{\theta}_i - \theta_i) (\hat{\theta}_j - \theta_j) \quad (2)$$

tiene una distribución que es aproximadamente la ji cuadrado con k grados de libertad, para muestras grandes. Aquí tampoco se perjudica el grado de aproximación sustituyendo las estimaciones de las θ_i por las $\hat{\theta}_i$ que figuran en r_{ij} ; la distribución de la cantidad

$$v = \Sigma \Sigma \hat{r}_{ij} (\hat{\theta}_i - \theta_i) (\hat{\theta}_j - \theta_j) \quad (3)$$

es también aproximadamente la ji cuadrado con k grados de libertad. La variante v permite establecer una región confidencial muy sencilla para las θ_i . Si $\chi^2_{1-\gamma}$ es el nivel $1-\gamma$ de la distribución ji cuadrado, se tiene

$$P(v < \chi^2_{1-\gamma}) = \gamma \quad (4)$$

que determina una región confidencial en el espacio paramétrico. Esta región queda definida por la ecuación

$$\Sigma \Sigma \hat{r}_{ij} (\hat{\theta}_i - \theta_i) (\hat{\theta}_j - \theta_j) = \chi^2_{1-\gamma} \quad (5)$$

que representa un elipsoide en el espacio $(\theta_1, \theta_2, \dots, \theta_k)$, con centro en $(\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_k)$.

Si lo que interesa es estimar solo una parte de un conjunto de k parámetros, p. ej., el conjunto $(\theta_1, \theta_2, \dots, \theta_r)$, siendo $r < k$, hallaremos primero la distribución marginal de los estimadores máximos-verosímiles para este conjunto. Si (a, b) son índices que toman los valores $1, 2, \dots, r$, los elementos $\bar{r}_{ij}$ de la matriz de la forma cuadrática de la distribución normal de $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_r$ en muestras grandes, vienen dados por

$$V_{11}^{-1} = (\bar{r}_{ij})$$

en donde la matriz V_{11} se obtiene prescindiendo de las últimas $k-r$ filas y columnas en V . Las $\bar{r}_{ij}$ serán, en general, funciones de todos los k parámetros originarios $\theta_1, \theta_2, \dots, \theta_k$. Si sustituimos las θ_i por las $\hat{\theta}_i$ en $\bar{r}_{ij}$, obtendremos los estimadores máximo-verosímiles $\hat{\bar{r}}_{ij}$ de las $\bar{r}_{ij}$. La forma cuadrática

$$w = \sum_a \sum_b \hat{\bar{r}}_{ab} (\hat{\theta}_a - \theta_a) (\hat{\theta}_b - \theta_b)$$

tiene una distribución que se aproxima a la ji cuadrado con r grados de libertad, y sirve para determinar una región confidencial elipsoidal en el espacio $\theta_1, \theta_2, \dots, \theta_r$ de estos parámetros.

Como ejemplo de la estimación de más de un parámetro, consideremos la estimación en muestras grandes de la media y de la varianza de una población normal. Hemos visto en la sección 10-8 que $\bar{x}$ y $\hat{\sigma}^2$ tienen distribuciones aproximadamente normales, con medias μ y σ^2 y matriz de la forma cuadrática

$$R = (r_{ij}) = \begin{pmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{n}{2\sigma^4} \end{pmatrix} \quad (6)$$

Si sustituimos σ^2 por $\hat{\sigma}^2$ en (6), la forma cuadrática es

$$v = \frac{n}{\hat{\sigma}^2} (\bar{x} - \mu)^2 + \frac{n}{2\hat{\sigma}^4} (\hat{\sigma}^2 - \sigma^2)^2 \quad (7)$$

cuya distribución, para muestras grandes, se aproxima a la ji cuadrado con dos grados de libertad. En particular, supongamos una muestra efectiva de 100 observaciones (3,4, 5,1, ..., 2,2) con

$$\begin{aligned} \bar{x} &= 1/100 \sum x_i = 4 \\ \hat{\sigma}^2 &= 1/100 \sum (x_i - \bar{x})^2 = 5 \end{aligned}$$

Ya que el nivel 0,05 de la ji cuadrado con dos grados de libertad es 5,99, una región confidencial del 95% para μ y σ^2 viene determinada por

$$P[20(4 - \mu)^2 + 2(5 - \sigma^2)^2 < 5,99] = 0,95 \quad (8)$$

Los valores de μ y σ^2 que satisfacen la desigualdad que figura en esta expresión son los puntos interiores a la elipse

$$20(4 - \mu)^2 + 2(5 - \sigma^2)^2 = 5,99$$

representada en la figura 11-10. Esta es la región confidencial del 95% para el punto paramétrico verdadero (μ_0, σ_0^2) . Antes de extraer la muestra, la probabilidad de que la región que se trataba de construir cubriese el punto paramétrico verdadero era, aproximadamente, 0,95.

Los intervalos y regiones confidenciales para muestras grandes que hemos obtenido en esta sección y en la precedente tienen una propiedad óptima, que indicaremos pero no demostraremos. En las

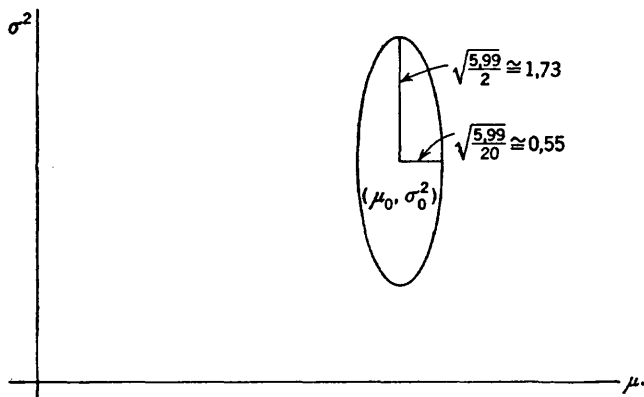


FIG. 11-10.

secciones anteriores de este capítulo nos ocupamos de hallar el intervalo más pequeño para una probabilidad dada. Así, el intervalo más pequeño del 95% para la media de una población normal, cuando se conoce σ , viene dado por

$$P\left(\bar{x} - \frac{1,96\sigma}{\sqrt{n}} < \mu < \bar{x} + \frac{1,96\sigma}{\sqrt{n}}\right) = 0,95$$

y la longitud del intervalo es $2 \times 1,96\sigma/\sqrt{n}$, en donde n es el tamaño de la muestra. Supongamos ahora que, en lugar de emplear $\bar{x} = (1/n)\sum x_i$ para construir el intervalo confidencial, utilizamos una sola de las observaciones; p. ej., la primera. El estimador es, simplemente,

$$\tilde{\mu} = x_1$$

y el intervalo confidencial viene dado por

$$P(\tilde{\mu} - 1,96\sigma < \mu < \tilde{\mu} + 1,96\sigma) = 0,95$$

cuya longitud es $2 \times 1,96\sigma$. Este intervalo es $\sqrt{n}$ veces mayor que el obtenido utilizando como estimador la media muestral.

Ahora bien: es evidente que la longitud de un intervalo confidencial correspondiente a un parámetro depende en gran medida de la función de las observaciones muestrales que se elija como estimador. La propiedad óptima de los intervalos y regiones para muestras grandes basados en estimadores de máxima verosimilitud es la siguiente:

Los intervalos y regiones confidenciales para muestras grandes, basados en estimadores máximo-verosímiles, son inferiores, por término medio, a los intervalos y regiones determinados por otros estimadores de los parámetros.

11-9. Intervalos confidenciales múltiples.—En las secciones anteriores hemos indicado que la interpretación frecuencial-probabilística de los intervalos confidenciales es la siguiente: En repetidos muestreos, $100(1-\alpha)$ por ciento de los intervalos confidenciales construidos contendrán el parámetro desconocido θ , donde $1-\alpha$ es el coeficiente confidencial. Para ilustrar esta interpretación con mayor precisión, supongamos que se extrae una muestra aleatoria de tamaño k de cada una de tres poblaciones normales de medias μ_1 , μ_2 y μ_3 , respectivamente, y varianza común σ^2 . Representemos esto por

Población 1: $n(x; \mu_1, \sigma^2)$	Población 2: $n(y; \mu_2, \sigma^2)$	Población 3: $n(z; \mu_3, \sigma^2)$	(1)
$x_1, x_2, \dots, x_k$	$y_1, y_2, \dots, y_k$	$z_1, z_2, \dots, z_k$	

Construiremos un intervalo confidencial del 95 % para $\mu_1 - \mu_2$, $\mu_2 - \mu_3$ y $\mu_1 - \mu_3$. Para hallar un intervalo confidencial para $\mu_1 - \mu_2$, tenemos en cuenta que $\bar{x} - \mu_1$ es normal, con media 0 y varianza σ^2/k ; $\bar{y} - \mu_2$ es normal con media 0 y varianza σ^2/k ; $\bar{y} - \mu_2$ y $\bar{x} - \mu_1$ son independientes, luego

$$w = (\bar{x} - \mu_1) - (\bar{y} - \mu_2) = (\bar{x} - \bar{y}) - (\mu_1 - \mu_2)$$

es normal con media 0 y varianza $2\sigma^2/k$ y, por tanto,

$$\frac{w}{\sqrt{2\sigma^2/k}}$$

es también normal, con media 0 y varianza 1. Si hacemos

$$s_1^2 = \frac{1}{k-1} \sum (x_i - \bar{x})^2$$

$$s_2^2 = \frac{1}{k-1} \sum (y_i - \bar{y})^2$$

$$s_3^2 = \frac{1}{k-1} \sum (z_i - \bar{z})^2$$

entonces

$$\frac{3(k-1)s^2}{\sigma^2} = \frac{(k-1)s_1^2 + (k-1)s_2^2 + (k-1)s_3^2}{\sigma^2} \quad (2)$$

se distribuye según una ji cuadrado con $3k-3$ grados de libertad, y s^2 es independiente de w . Por tanto,

$$t = \frac{w(\sqrt{k} \cdot \sigma)}{\sqrt{2s\sigma}} = \frac{w\sqrt{k}}{s\sqrt{2}}$$

se distribuye según una t de Student con $3(k-1)$ grados de libertad. Un intervalo confidencial del 95% para $\mu_1 - \mu_2$ es

$$P \left[(\bar{x} - \bar{y}) - t_{0,025} \sqrt{\frac{2s^2}{k}} < \mu_1 - \mu_2 < (\bar{x} - \bar{y}) + t_{0,025} \sqrt{\frac{2s^2}{k}} \right] = 0,95 \quad (3)$$

Por un proceso semejante se deduce que un intervalo confidencial del 95% para $\mu_1 - \mu_3$ es

$$P \left[(\bar{x} - \bar{z}) - t_{0,025} \sqrt{\frac{2s^2}{k}} < \mu_1 - \mu_3 < (\bar{x} - \bar{z}) + t_{0,025} \sqrt{\frac{2s^2}{k}} \right] = 0,95 \quad (4)$$

y, análogamente, un intervalo confidencial del 95% para $\mu_2 - \mu_3$ es

$$P \left[(\bar{y} - \bar{z}) - t_{0,025} \sqrt{\frac{2s^2}{k}} < \mu_2 - \mu_3 < (\bar{y} - \bar{z}) + t_{0,025} \sqrt{\frac{2s^2}{k}} \right] = 0,95 \quad (5)$$

Si se toman repetidos conjuntos de observaciones (1), y se calcula (3) para cada conjunto de $3k$ observaciones, entonces, para un número grande de repeticiones, el 95% de los intervalos confidenciales cubrirán a $\mu_1 - \mu_2$.

Si para cada conjunto de $3k$ observaciones se calcula el intervalo confidencial (4), para un número grande de repeticiones el

95% de estos intervalos cubrirán a $\mu_1 - \mu_3$. Análogamente, si para cada conjunto se calcula el intervalo confidencial (5), en un número grande de repeticiones, el 95% de los intervalos contendrán a $\mu_2 - \mu_3$. Deseamos calcular intervalos confidenciales para $\mu_1 - \mu_2$, $\mu_1 - \mu_3$ y $\mu_2 - \mu_3$, tales que la probabilidad de que los tres intervalos confidenciales resulten simultáneamente verdaderos sea, p. ej., el 95%. Si los tres intervalos dados por (3) a (5) fuesen independientes, en un número grande de repeticiones, para el $(0,95)^3$ de los conjuntos, (3) cubriría a $\mu_1 - \mu_2$, (4) cubriría a $\mu_1 - \mu_3$ y (5) cubriría a $\mu_2 - \mu_3$. Sin embargo, puesto que (3), (4) y (5) no son independientes, esta probabilidad no es $(0,95)^3$. Para resolver este problema definiremos el *coeficiente confidencial experimental*. Un conjunto de observaciones tales como (1) recibirá el nombre de *experimento*; puede haber t poblaciones en lugar de tres. En cada experimento, se calculan intervalos confidenciales para las $t(t-1)$ diferencias $\mu_i - \mu_j$. Si en el 95% de los experimentos la totalidad de los $t(t-1)$ intervalos confidenciales cubren a sus diferencias respectivas ($\mu_i - \mu_j$), diremos que el *coeficiente confidencial experimental* es 0,95.

Enunciaremos el siguiente teorema aunque no daremos su demostración.

Teorema 11-1.—Sea $v_1, v_2, \dots, v_n$ una muestra aleatoria de tamaño n de una población normal de media 0 y varianza σ^2 , y designemos por R el recorrido de estas variables aleatorias; es decir, $R = \max v_i - \min v_i$. Supongamos que vs^2/σ^2 es independiente de las v_i y está distribuida según una χ^2 cuadrado con ν grados de libertad. La variable aleatoria

$$q = \frac{R}{s}$$

se distribuye como el recorrido studentizado, con n y ν grados de libertad en el numerador y en el denominador, respectivamente.

La función frecuencial de q es bastante complicada y no se dará aquí, pero la cantidad q_α , definida por $P(q < q_\alpha) = 1 - \alpha$, puede obtenerse en las tablas VI a VIII para varios valores de n , ν y $\alpha = 0,01, 0,05$ y $0,10$.

Para ilustrar cómo puede emplearse este teorema, hallaremos un conjunto de intervalos confidenciales con un coeficiente confidencial experimental del 0,95. Consideremos las variables aleatorias (nos limitaremos al caso especial de tres)

$$\frac{3(k-1)s^2}{\sigma^2}, u_1, u_2, u_3$$

en donde s^2 está dada por (2), y u_1, u_2, u_3 son los estadísticos ordinarios de las tres variables aleatorias v_1, v_2, v_3 , con

$$v_1 = (\bar{x} - \mu_1)\sqrt{k} \quad v_2 = (\bar{y} - \mu_2)\sqrt{k} \quad v_3 = (\bar{z} - \mu_3)\sqrt{k}$$

Puesto que las v_i son variables normales independientes, de medias 0 y varianzas σ^2 , y dado que $3(k-1)s^2/\sigma^2$ es una variable ji cuadrado independiente, con $\nu=3(k-1)$ g. de l., utilizaremos el teorema 11-1 para demostrar que q se distribuye como el recorrido studentizado, con $n=3$ g. de l. en el numerador y $\nu=3(k-1)$ g. de l. en el denominador, siendo

$$q = \frac{R}{s} = \frac{u_3 - u_1}{s} = \frac{\text{máx } v_i - \text{mín } v_i}{s}$$

También

$$\begin{aligned} 1 - \alpha &= P(q < q_\alpha) = P\left(\frac{u_3 - u_1}{s} < q_\alpha\right) = \\ &= P\left(\frac{\text{máx } v_i - \text{mín } v_i}{s} < q_\alpha\right) = \\ &= P(\text{máx } v_i - \text{mín } v_i < s q_\alpha) \end{aligned} \quad (6)$$

Pero si $\text{máx } v_i - \text{mín } v_i < s q_\alpha$, se tienen las tres desigualdades siguientes:

$$|(\bar{x} - \mu_1) - (\bar{y} - \mu_2)| < \frac{s q_\alpha}{\sqrt{k}}$$

$$|(\bar{x} - \mu_1) - (\bar{z} - \mu_3)| < \frac{s q_\alpha}{\sqrt{k}}$$

y

$$|(\bar{y} - \mu_2) - (\bar{z} - \mu_3)| < \frac{s q_\alpha}{\sqrt{k}}$$

lo que implica

$$\begin{aligned} -\frac{s q_\alpha}{\sqrt{k}} &< (\bar{x} - \bar{y}) - (\mu_1 - \mu_2) < \frac{s q_\alpha}{\sqrt{k}} \\ -\frac{s q_\alpha}{\sqrt{k}} &< (\bar{x} - \bar{z}) - (\mu_1 - \mu_3) < \frac{s q_\alpha}{\sqrt{k}} \\ -\frac{s q_\alpha}{\sqrt{k}} &< (\bar{y} - \bar{z}) - (\mu_2 - \mu_3) < \frac{s q_\alpha}{\sqrt{k}} \end{aligned} \quad (7)$$

Si utilizamos (7) con (6), la probabilidad de que las seis desigualdades (8) sean verdaderas es $1 - \alpha$:

$$\begin{aligned}
 (\bar{x} - \bar{y}) - \frac{sq_a}{\sqrt{k}} &< \mu_1 - \mu_2 < (\bar{x} - \bar{y}) + \frac{sq_a}{\sqrt{k}} \\
 (\bar{y} - \bar{x}) - \frac{sq_a}{\sqrt{k}} &< \mu_2 - \mu_1 < (\bar{y} - \bar{x}) + \frac{sq_a}{\sqrt{k}} \\
 (\bar{x} - \bar{z}) - \frac{sq_a}{\sqrt{k}} &< \mu_1 - \mu_3 < (\bar{x} - \bar{z}) + \frac{sq_a}{\sqrt{k}} \\
 (\bar{z} - \bar{x}) - \frac{sq_a}{\sqrt{k}} &< \mu_3 - \mu_1 < (\bar{z} - \bar{x}) + \frac{sq_a}{\sqrt{k}} \\
 (\bar{y} - \bar{z}) - \frac{sq_a}{\sqrt{k}} &< \mu_2 - \mu_3 < (\bar{y} - \bar{z}) + \frac{sq_a}{\sqrt{k}} \\
 (\bar{z} - \bar{y}) - \frac{sq_a}{\sqrt{k}} &< \mu_3 - \mu_2 < (\bar{z} - \bar{y}) + \frac{sq_a}{\sqrt{k}}
 \end{aligned} \tag{8}$$

En el caso de haber más de tres poblaciones, serían válidas las mismas fórmulas, salvo que variarían los grados de libertad para q_a y que existirían $t(t-1)$ intervalos confidenciales.

Ejemplo 11-3.—Supongamos que un experimentador contrasta tres variedades de trigo, cultivando cada una de ellas en seis parcelas. Se registraron los siguientes datos, en bushels por acre,

$$\begin{aligned}
 [k=6, n=3, \nu=3(k-1)=15] \\
 \bar{x}=28,2 \quad \bar{y}=26,1 \quad \bar{z}=30,8 \quad s^2=3,84
 \end{aligned}$$

Para obtener intervalos confidenciales para $\mu_i - \mu_j$ con un coeficiente confidencial experimental del 0,95, calculamos

$$sq_{0,05}/\sqrt{k} = (3,67 \sqrt{0,6}) = 2,9$$

$q_{0,05}$ es el punto porcentual superior correspondiente a 0,05 para el recorrido studentizado en la tabla VII, para 3 y 15 g. de l. Los intervalos confidenciales con un coeficiente confidencial experimental del 0,95 son

$$\begin{aligned}
 -0,8 &< \mu_1 - \mu_2 < 5,0 \\
 -5,5 &< \mu_1 - \mu_3 < 0,3 \\
 -7,6 &< \mu_2 - \mu_3 < -1,8 \\
 -5,0 &< \mu_2 - \mu_1 < 0,8 \\
 -0,3 &< \mu_3 - \mu_1 < 5,5 \\
 1,8 &< \mu_3 - \mu_2 < 7,6
 \end{aligned}$$

Para intervalos confidenciales con un coeficiente confidencial experimentalmente fijado en condiciones distintas, remitimos al lector a los trabajos de Duncan, Scheffé y Tukey.

PROBLEMAS

1. Hállese un intervalo confidencial del 90% para la media de una distribución normal con $\sigma=3$, dada la muestra (3,3; -0,3; -0,6; -0,9). ¿Cuál sería el intervalo confidencial si σ fuese desconocido?

2. La resistencia a la rotura, expresada en libras, de cinco ejemplares de cuerda, cuyos diámetros sean 3/16 de pulgada, es de: 660, 460, 540, 580, 550. Estímese la resistencia media a la rotura mediante un intervalo confidencial del 95%, suponiendo distribuciones normales. Estímese el punto en que sería de esperar que se rompieran solamente el 5% de dichos ejemplares.

3. Con referencia al problema 2, estímese σ^2 mediante un intervalo confidencial del 90%; estímese también σ .

4. Con referencia al problema 2, constrúyase una región confidencial del 81%, para la estimación conjunta de μ y σ^2 ; de μ y σ .

5. Se han extraído cinco muestras de poblaciones que se suponen normales y con la misma varianza. Los valores de $(n-1)s^2 = \sum(x_i - \bar{x})^2$ y n , tamaño de la muestra, fueron

s^2 :	40	30	20	42	50
n :	6	4	3	7	8

Hállense los límites confidenciales del 98% para la varianza común a dichas poblaciones.

6. La observación mayor x' de una muestra de tamaño n , de una población rectangular $f(x)=1/\theta$ ($0 < x < \theta$), tiene por función de densidad

$$f(x') = \frac{n(x')^{n-1}}{\theta^n} \quad 0 < x' < \theta$$

Demuéstrese que $u=x'/\theta$ tiene una distribución independiente de θ . Utilizando u , hállese el intervalo confidencial más pequeño de θ para un coeficiente confidencial γ .

7. Calcúlese el intervalo confidencial del 95% para θ en el problema 6, dada la muestra (2,9; 1,8; 4,6; 1,9).

8. Para docimar dos nuevas variedades de maíz, en condiciones agrícolas normales, una sociedad de cereales eligió al azar ocho fincas en Iowa, plantando dichas variedades en parcelas experimentales de cada finca. La producción (en bushels por acre) en las ocho fincas fue

Variedad A:	86	87	56	93	84	93	75	79
Variedad B:	80	79	58	91	77	82	74	66

Suponiendo que la distribución conjunta de ambas producciones es normal, estime la diferencia entre las producciones medias mediante un intervalo confidencial del 90%. (Véase el problema 22 del Cap. 10).

9. Utilizando la función de densidad

$$f(x) = \frac{4x^3}{\theta^4} \quad 0 < x < \theta$$

para la mayor de las cuatro observaciones procedentes de una población rectangular, establézcase un sistema general de intervalos confidenciales del 95% para θ , hallando las funciones $h_1(\theta)$ y $h_2(\theta)$ y representándolas en el plano (θ, θ) . Hállese el intervalo correspondiente a la muestra dada en el problema 7. ¿Por qué difiere del intervalo hallado en dicho problema?

10. Con referencia al problema 9, represéntense las funciones $h_1(\theta)$ y $h_2(\theta)$ para muestras de tamaño ocho. Demuéstrese que, en general, las amplitudes de los intervalos decrecen al aumentar el tamaño de la muestra.

11. Se extrae la muestra (2,3; 1,2; 0,9; 3,2) de una población cuya distribución viene dada por $f(x) = \alpha e^{-\alpha x}$, $x > 0$. Hállese un intervalo confidencial del 90% para α .

12. Con referencia al problema 11, hállese intervalos confidenciales del 90% para la media y la varianza de la distribución. ¿Cuál es la probabilidad de que estos dos intervalos cubran, respectivamente, a la media y a la varianza verdaderas?

13. Se lanza una moneda tres veces y se obtienen una cara y dos cruces. Hállese un intervalo confidencial del 90% para la probabilidad de obtener cara.

14. Se lanza una moneda 400 veces, y se obtienen 175 caras y 225 cruces. Hállese un intervalo confidencial del 90% para la probabilidad de obtener cara. Hállese un intervalo confidencial del 99%. ¿Está bien construida la moneda?

15. Se pregunta a 2000 votantes cuál será su actitud respecto a una determinada propuesta política. 1200 favorecen la propuesta; 600 se oponen y 200 están indecisos. Suponiendo que la muestra fuese aleatoria y procedente de una población binomial, constrúyase una región confidencial del 95% para p_1 y p_2 , proporciones de individuos a favor y en contra de la propuesta. (Utilícense los resultados de la Sec. 10-8.)

16. Constrúyase una región confidencial del 95% como la de la figura 11-7 para el ejemplo utilizado en la sección 11-8 y compárese con la región de la figura 11-10.

17. Intégrese por partes [integrando $(1-t)^s$ y derivando t^r] para demostrar

$$\int_0^x t^r (1-t)^s dt = -\frac{1}{s+1} x^r (1-x)^{s+1} + \frac{r}{s+1} \int_0^x t^{r-1} (1-t)^{s+1} dt$$

18. Aplíquese repetidamente el resultado anterior para obtener la forma acumulativa de la distribución beta, $F(x; \alpha, \beta)$.

19. Demuéstrese que

$$F(x; \alpha, \beta) = \sum_{i=\alpha+1}^{\alpha+\beta+1} \binom{\alpha+\beta+1}{i} x^i (1-x)^{\alpha+\beta+1-i}$$

utilizando el resultado del problema 18. Esta es la forma que se hubiera obtenido si la integración se hubiese realizado por partes derivando $(1-t)^{\alpha}$ e integrando t^r .

20. Dada una muestra de tamaño 100 de una población normal, con $\hat{\mu}=3$, $\hat{\sigma}^2=0,25$, ¿cuál es la estimación máximo-verosímil del número α para el cual

$$\int_{\alpha}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-(1/2\sigma^2)(x-\mu)^2} dx = 0,05 ?$$

21. Hállese la distribución de $\hat{\mu}$ y $\hat{\sigma}$, para muestras grandes procedentes de una población normal. Como se sabe que $\hat{\mu}$ y $\hat{\sigma}$ tienen distribuciones independientes y normales con medias μ y σ , solo es necesario hallar su varianza.

22. Con referencia al problema anterior, hállese la distribución en muestras grandes de $\hat{\mu} + k\hat{\sigma}$, siendo k una constante dada. Utilícese esto para obtener un intervalo confidencial del 95 % para α , en el problema 20.

23. Encuéntrese un método para la estimación de la razón de varianzas de dos poblaciones normales por un intervalo confidencial.

24. Encuéntrese un método para estimar el parámetro de la distribución de Poisson por un intervalo confidencial.

25. Desarróllese con detalle la obtención de la ecuación (11-2-6).

26. ¿Cuál es la probabilidad de que la longitud de un intervalo confidencial t sea inferior a σ para muestras de tamaño 20?

27. Compárese la longitud media de un intervalo confidencial del 95 % para la media de una población normal, basada en la distribución t , con la longitud que tendría este intervalo si se conociera la varianza.

28. Demuéstrese que la longitud y la varianza de la longitud del intervalo confidencial t se aproximan a cero al aumentar el tamaño de la muestra.

29. ¿Qué tamaño de muestra habrá que extraer de una población normal para obtener una probabilidad del 0,95 de que un intervalo confidencial del 90 % (basado en t) para la media, tenga longitud inferior a $\sigma/5$?

30. Pruébese que la longitud del intervalo confidencial para σ (de una población normal) tiende a cero al aumentar el tamaño de la muestra.

31. Considérese una población normal *truncada*, cuya función de densidad sea

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma\alpha} e^{-\frac{1}{2}(x-\mu)^2/\sigma^2} \quad x < a$$

$$= 0 \quad x \geq a$$

donde

$$\alpha = \int_{-\infty}^a \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(x-\mu)^2/\sigma^2} dx$$

Demuéstrese que las esperanzas de

$$\frac{\partial}{\partial \mu} \log f(x) \quad \text{y} \quad \frac{\partial}{\partial \sigma} \log f(x)$$

son iguales a cero.

32. Con referencia al problema 31, sean $\hat{\mu}$ y $\hat{\sigma}$ los estimadores máximo-verosímiles de μ y σ . Demuéstrese que la matriz de coeficientes de la forma cuadrática de la distribución de $\hat{\mu}$ y $\hat{\sigma}$ en muestras grandes es

$$r_{ij} = \begin{pmatrix} \frac{n(1-tb-b^2)}{\sigma^2} & \frac{-nb(1+tb+t^2)}{\sigma^2} \\ \frac{-nb(1+tb+t^2)}{\sigma^2} & \frac{n(2-tb-t^2b^2-t^3b)}{\sigma^2} \end{pmatrix}$$

donde $b = \sigma f(a)$ y $t = (a - \mu)/\sigma$.

33. Supongamos que los datos siguientes son normales, con medias respectivas μ_1 , μ_2 , μ_3 y varianza σ^2 .

	1	60,3	61,2	60,5	60,2	59,7
Población	2	62,2	62,8	61,8	60,9	59,3
	3	60,2	61,8	62,3	61,7	61,8

Calcúlese un intervalo confidencial, con un coeficiente confidencial experimental del 95 %, para las diferencias $\mu_i - \mu_j$.

BIBLIOGRAFIA

1. ANDERSON, R. L., y T. A. BANCROFT: *Statistical Theory in Research*, McGraw-Hill Book Company, Inc., Nueva York, 1952.
2. CRAMÉR, H.: *Métodos matemáticos de estadística*. Aguilar, 4.^a ed., Madrid, 1967.

3. DUNCAN, D. B.: «Multiple range and multiple F -tests», *Biometrics*, volumen 11 (1955), págs. 1-42.
4. KENDALL, M. G.: *The Advanced Theory of Statistics*, vols. I, II, Charles Griffin & Co., Ltd., Londres, 1946.
5. SCHEFFÉ, H.: «A method for judging all contrasts in the analysis of variance», *Biometrika*, vol. 40 (1953), págs. 87-104.
6. SNEDECOR, G. W.: «Statistical Methods», 5.^a ed., Iowa State College Press, Ames, Iowa, 1956.
7. TUKEY, J. W.: *The Problem of Multiple Comparisons*, Princeton University, 1953.

CAPITULO 12

DOCIMASIA DE HIPOTESIS

12-1. Introducción.—La inferencia estadística comprende dos partes principales, a saber: la estimación de parámetros y la docimasia de hipótesis. En este capítulo estudiaremos la segunda de ellas, con el objeto de desarrollar métodos generales para la docimasia de hipótesis y su aplicación a algunos problemas corrientes. Estos métodos también se utilizarán en capítulos posteriores.

En la investigación experimental se pretende a veces simplemente estimar un parámetro; p. ej., puede que interese estimar la producción de un nuevo híbrido de maíz. Muchas veces, el objetivo final es la utilización de dicha estimación. Así ocurre cuando se quiere comparar la producción del nuevo híbrido con la correspondiente a una variedad conocida, a fin de recomendar la sustitución de esta por aquel, en caso de que parezca superior. Esto sucede corrientemente en la investigación; puede ocurrir que interese determinar si un método nuevo para cerrar lámparas aumenta la vida de estas; si un nuevo germicida resulta más efectivo en el tratamiento de cierta infección; si un método de conservación de alimentos es preferible a otros, en lo que se refiere a la conservación de vitaminas, etc.

Utilizando como ejemplo el caso de las lámparas, supongamos que la vida media de las fabricadas por medio de un proceso conocido es de 1400 h. Se desea docimar un nuevo procedimiento para la fabricación de lámparas. En este caso, el modelo estadístico es el siguiente: se trata de dos poblaciones de lámparas, la constituida por las fabricadas utilizando el proceso conocido y la constituida por las correspondientes al proceso que se propone. Sabemos (en virtud de numerosas investigaciones anteriormente efectuadas) que la media de la primera población es aproximadamente 1400. Se desea averiguar si la media de la segunda población es superior o inferior a 1400. Tradicionalmente, para resolver este problema, se establece la hipótesis de que una media es mayor que la otra. Basándose en una muestra de las poblaciones, se aceptará o rechazará la hipótesis. (Naturalmente, se confía en que el nuevo proceso es mejor y que la hipótesis será rechazada.)

Para docimar la hipótesis se fabrica cierto número de lámparas mediante el nuevo procedimiento, midiendo después su duración.

Supongamos que la media de esta muestra de observaciones es de 1550 h. Esto parece indicar que el nuevo proceso es mejor; pero supongamos que la estimación de la desviación estándar de la media es $\hat{\sigma}/\sqrt{n}$, igual a 125 (siendo n el tamaño de la muestra). Por tanto, el intervalo confidencial del 95% para la media de la segunda población (suponiendo la población normal) es aproximadamente de 1300 h a 1800 h. La media muestral 1550 podría proceder fácilmente de una población cuya media fuese 1400. No tenemos, pues, motivos suficientes para rechazar la hipótesis. Por otra parte, si $\hat{\sigma}/\sqrt{n}$ fuese igual a 25, podríamos rechazar la hipótesis con gran confianza y afirmar la superioridad del nuevo proceso de fabricación.

Se ve, pues, que la docimasia de hipótesis está relacionada íntimamente con el problema de la estimación. No obstante, resulta instructivo desarrollar la teoría de la docimasia independientemente de la de la estimación, al menos en principio.

La docimasia de hipótesis puede integrarse en la estructura del problema general de decisión de la siguiente forma: Existen dos acciones finales posibles, a_1 y a_2 . La acción apropiada a tomar depende del valor del parámetro desconocido θ , llamado algunas veces *estado de la naturaleza*, que es un elemento del espacio paramétrico Ω . El conjunto Ω puede descomponerse en dos conjuntos, ω_1 y ω_2 , tales que se elige la acción a_1 si θ pertenece a ω_1 , y la acción a_2 si θ pertenece a ω_2 . La pérdida asociada a la acción a y al estado de la naturaleza θ viene dada por $l(a; \theta)$, donde $l(a; \theta) \geq 0$ y

$$\begin{aligned} l(a_1; \theta) &= 0 & \text{si } \theta \text{ está en } \omega_1 \\ l(a_2; \theta) &= 0 & \text{si } \theta \text{ está en } \omega_2 \end{aligned} \quad (1)$$

Sea $s = (x_1, x_2, \dots, x_n)$ una muestra aleatoria procedente de $f(x; \theta)$, y S , el espacio muestral n -dimensional. Una estrategia (función de decisión) es una función d que asigna a cada posible muestra una acción de A , donde

$$A = \{a : a = a_1 \text{ o } a_2\}.$$

La acción que se toma es

$$a = d(x_1, x_2, \dots, x_n)$$

En este problema en el que existen solo dos acciones, cada estrategia d (función de decisión) puede representarse por una partición del espacio muestral n -dimensional en dos conjuntos disjuntos, S_1 y S_2 , siendo

$$S_2 = \bar{S}_1 = S - S_1,$$

tales que se toma la acción a_1 si el punto muestral s cae en S_1 , y la a_2 si s cae en S_2 . El riesgo (pérdida esperada) correspondiente a la estrategia d está dado por

$$\begin{aligned}
 R(d; \theta) &= \int \int \dots \int_S l[d(x_1, x_2, \dots, x_n); \theta] f(x_1; \theta) \dots f(x_n; \theta) dx_1 \dots dx_n \\
 &= \int \int \dots \int_{S_1} l[d(x_1, x_2, \dots, x_n); \theta] f(x_1; \theta) \dots f(x_n; \theta) \times \\
 &\quad dx_1 \dots dx_n + \int \int \dots \int_{S_2} l[d(x_1, x_2, \dots, x_n); \theta] f(x_1; \theta) \\
 &\quad \dots f(x_n; \theta) dx_1 \dots dx_n \\
 &= \int \int \dots \int_{S_1} l(a_1; \theta) f(x_1; \theta) \dots f(x_n; \theta) dx_1 \dots dx_n \\
 &\quad + \int \int \dots \int_{S_2} l(a_2; \theta) f(x_1; \theta) \dots f(x_n; \theta) dx_1 \dots dx_n \\
 &= l(a_1; \theta) \int \int \dots \int_{S_1} f(x_1; \theta) \dots f(x_n; \theta) dx_1 \dots dx_n \\
 &\quad + l(a_2; \theta) \int \int \dots \int_{S_2} f(x_1; \theta) \dots f(x_n; \theta) dx_1 \dots dx_n \\
 &= l(a_1; \theta) P(s \in S_1 | \theta) + l(a_2; \theta) P(s \in S_2 | \theta) \tag{2}
 \end{aligned}$$

donde $P(s \in S_1 | \theta)$ denota la probabilidad de que el punto muestral s caiga en S_1 cuando el valor del parámetro (estado de la naturaleza) es θ , y análogamente para $P(s \in S_2 | \theta)$.

Puesto que se toma la acción a_1 si s cae en S_1 y la a_2 si cae en S_2 , las probabilidades en la ecuación anterior son las correspondientes a adoptar las acciones a_1 y a_2 , respectivamente, cuando θ es el estado de la naturaleza. Se denominan *probabilidades de acción*.

Definición 12-1.—Sea S un espacio muestral n -dimensional, y S_1 y S_2 , una partición del espacio muestral, tal que si un punto muestral

$$s = (x_1, \dots, x_n)$$

cae en S_1 , se toma la acción a_1 , y si s cae en S_2 se adopta la acción a_2 . Las siguientes probabilidades se denominan *probabilidades de acción*:

$$P(s \in S_1 | \theta) \quad P(s \in S_2 | \theta)$$

donde $P(s \in S_i | \theta)$ es la probabilidad de que s caiga en S_i (probabilidad de que se tome la acción a_i) cuando el verdadero estado de la naturaleza es θ .

Si en la ecuación (2) calculamos el riesgo cuando θ pertenece a ω_1 , el cual designaremos por $R(d; \theta \in \omega_1)$, se obtiene:

$$R(d; \theta \in \omega_1) = l(a_1; \theta \in \omega_1)P(s \in S_1 | \theta \in \omega_1) + l(a_2; \theta \in \omega_1)P(s \in S_2 | \theta \in \omega_1) \quad (3)$$

Utilizando la ecuación (1), resulta

$$R(d; \theta \in \omega_1) = l(a_2; \theta \in \omega_1)P(s \in S_2 | \theta \in \omega_1) \quad (4)$$

Por un procedimiento análogo, calcularemos el riesgo cuando θ está en ω_2 , obteniendo

$$R(d; \theta \in \omega_2) = l(a_1; \theta \in \omega_2)P(s \in S_1 | \theta \in \omega_2) \quad (5)$$

Es decir, puesto que una de las dos pérdidas $l(a_1; \theta)$ y $l(a_2; \theta)$ es igual a 0, escribiremos el riesgo en la ecuación (2) en la forma

$$R(d; \theta) = l(\theta)\mathcal{E}(d; \theta) \quad (6)$$

donde

$$\begin{aligned} l(\theta) &= l(a_1; \theta) & \text{si } \theta \text{ está en } \omega_2 \\ &= l(a_2; \theta) & \text{si } \theta \text{ está en } \omega_1 \end{aligned} \quad (7)$$

siendo $l(\theta)$ la pérdida asociada con la acción incorrecta cuando el estado de la naturaleza es θ , y $\mathcal{E}(d; \theta)$, la probabilidad de error definida a continuación.

Definición 12-2. *Probabilidades de error.*—La probabilidad de error, designada por $\mathcal{E}(d; \theta)$ en la ecuación (6), es la probabilidad de adoptar la acción incorrecta. Es decir, es la probabilidad de tomar la acción a_1 si θ está en ω_2 , o bien de tomar la acción a_2 si θ está en ω_1 .

Si $\theta \in \omega_1$, esta probabilidad se expresará así:

$$\mathcal{E}_1(d; \theta) = P[(x_1, x_2, \dots, x_n) \in S_2 | \theta \in \omega_1] = P(s \in S_2 | \theta \in \omega_1)$$

que es la correspondiente a tomar la acción a_2 cuando θ está en ω_1 ; y si $\theta \in \omega_2$, la probabilidad de error puede escribirse:

$$\mathcal{E}_2(d; \theta) = P[(x_1, x_2, \dots, x_n) \in S_1 | \theta \in \omega_2] = P(s \in S_1 | \theta \in \omega_2)$$

que es la probabilidad de adoptar la acción a_1 cuando θ está en ω_2 . Definiremos ahora la *docimasia de hipótesis*.

Definición 12-3.—Los conjuntos ω_1 y ω_2 en la formulación anterior del problema de decisión pueden asociarse a la hipótesis o afirmación H_1 : « θ está en ω_1 » y a la hipótesis alternativa H_2 : « θ

está en ω_2 , respectivamente. La acción a_1 consiste en aceptar la hipótesis (aceptar H_1) y la acción a_2 en rechazar la hipótesis (rechazar H_1). La función de decisión d que, aplicada a los datos, conduce a la aceptación o rechazo de la hipótesis, se denomina *dócima de la hipótesis*.

El objetivo es encontrar la dócima (la función de decisión d) que hace mínimo el riesgo para cada valor de θ en Ω . Sin embargo, esto *no es generalmente posible*, sino que una función de decisión puede dar un riesgo mínimo para ciertos valores de θ , mientras que otra función de decisión puede hacer mínimo el riesgo para otros valores de θ , etc. Por tanto, puesto que θ es desconocido, hay que contar con la posibilidad de que no exista un método definido para determinar qué función da riesgo mínimo en un problema particular.

Otra dificultad inherente a la utilización de las ecuaciones (4) y (5) se debe a que en gran parte de los problemas de aplicación, donde un experimentador desea utilizar dócimas de hipótesis, la función de pérdida es totalmente desconocida, o bien no se conoce con la suficiente acuracidad para garantizar su empleo. Si la función de pérdida no es conocida, parece que un procedimiento razonable consistirá en utilizar una función de decisión que, en cierto sentido, minimice las probabilidades de error. El procedimiento tradicional es elegir una probabilidad α , usualmente en el entorno de 0,01, 0,05, 0,10, 0,20, y hallar la clase de funciones de decisión (o sea, determinar los conjuntos S_2) tales que se satisfaga

$$P(s \in S_2 | \theta \in \omega_1) \leq \alpha \quad (8)$$

Entonces, de la clase de dócimas que satisfacen a (8) se considera como «mejor» dócima aquella para la cual

$$P(s \in S_1 | \theta \in \omega_2) \quad (9)$$

es mínimo. En esta formulación, la cantidad $P(s \in S_2 | \theta \in \omega_1)$ de (8) se llama probabilidad de rechazar una hipótesis verdadera (rechazar la hipótesis H_1 cuando de hecho es cierta), y a veces se la denomina probabilidad de un error de tipo I, y (8) se escribe en la forma $P(I) \leq \alpha$. La cantidad $P(s \in S_1 | \theta \in \omega_2)$ de (9) se llama probabilidad de aceptar una hipótesis falsa (aceptar H_1 cuando no es cierta), pero algunas veces se denomina también probabilidad de un error de tipo II, y se escribe $P(II)$. Obsérvese que

$$\varepsilon_1(d; \theta) = P(I) \quad \text{y} \quad \varepsilon_2(d; \theta) = P(II)$$

La región S_2 recibe el nombre de *región de rechazo* o de *región crítica*, y S_1 , *región de aceptación*. Si la afirmación de (8) es verda-

dera, se dice que la extensión de la dócima es α . En lugar de la cantidad $P(s \in S_1 | \theta \in \omega_2)$ de (9) es a menudo más conveniente utilizar $P(s \in S_2 | \theta \in \omega_2)$, donde, evidentemente,

$$1 - P(s \in S_1 | \theta \in \omega_2) = P(s \in S_2 | \theta \in \omega_2) \quad (10)$$

que es la probabilidad de rechazar la hipótesis H_1 cuando de hecho es falsa. La cantidad $P(s \in S_2 | \theta)$ se denomina potencia de la dócima, designándose por $\beta(\theta)$, y es función de θ . Obsérvese que $\beta(\theta) = P(I)$ cuando $\theta \in \omega_1$. También $\beta(\theta) = 1 - P(II)$ si $\theta \in \omega_2$.

A primera vista puede parecer que *esta* formulación del problema de la docimasia de hipótesis no tiene en cuenta la función de pérdida. En realidad, no prescinde de ella *completamente*, puesto que llegar a un valor razonable para α requiere que el experimentador sopesa las consecuencias de cometer errores de los tipos I y II. La anterior formulación del problema ha recibido una atención preferente por parte de los estadísticos matemáticos y se emplea extensamente por los experimentadores.

Daremos a continuación un ejemplo para aclarar las ideas anteriores; luego estudiaremos con detalle el caso más simple en que ω_1 y ω_2 solo contienen un único punto.

Ejemplo 12-1.—Sea la variable aleatoria x distribuida normalmente con media μ y varianza 1, donde μ solo puede tomar los valores -1 ó 0 . Supongamos que se desea docimar la hipótesis $H_1: \mu = -1$ con la hipótesis alternativa $H_2: \mu = 0$. El espacio paramétrico Ω es $\Omega = \{\mu: \mu = -1, 0\}$; el espacio de acción contiene dos puntos, a_1 y a_2 , siendo a_1 la afirmación «se acepta $H_1: \mu = -1$ », y a_2 , la afirmación «se acepta $H_2: \mu = 0$ », o bien, «se rechaza H_1 ». También

$$\omega_1 = \{\mu: \mu = -1\}; \quad \omega_2 = \{\mu: \mu = 0\}$$

Supongamos que para este problema elegimos una función de pérdida definida por

$$l(a_1; \mu \in \omega_1) = 0$$

$$l(a_1; \mu \in \omega_2) = 1$$

$$l(a_2; \mu \in \omega_2) = 0$$

$$l(a_2; \mu \in \omega_1) = 4$$

e imaginemos que se utiliza una muestra de tamaño 1. Entonces, el espacio muestral S es unidimensional y está definido por

$$S = \{x: -\infty < x < \infty\}$$

Supongamos que se realiza una partición de S en S_1 y S_2 tales que

$$S_1 = \{x: -\infty < x < -1\}$$

$$S_2 = \{x: -1 \leq x < \infty\}$$

es decir, se toma la acción a_1 (aceptar $H_1: \mu = -1$) si $x < -1$ y se toma la acción a_2 (aceptar $H_2: \mu = 0$) si $x \geq -1$. La función de decisión d está definida por:

El valor de d es a_1 si $x < -1$

El valor de d es a_2 si $x \geq -1$

En virtud de (2), el riesgo para esta función de decisión es

$$R(d; \mu \in \omega_1) = 4 \cdot P(x \geq -1 | \mu = -1) = 4(0,50) = 2$$

$$R(d; \mu \in \omega_2) = 1 \cdot P(x < -1 | \mu = 0) = 1(0,16) = 0,16$$

En la tabla siguiente se dan las probabilidades de acción para este problema:

ESTADO DE LA NATURALEZA

		ω_1	ω_2
		$\mu = -1$	$\mu = 0$
Acción	a_1	0,50	0,16
	a_2	0,50	0,84

A partir de esta tabla se obtienen las probabilidades de error, que son:

$$\mathcal{E}(d; \mu) \begin{cases} P(\text{II}) = P(\text{acción } a_1 | \mu \in \omega_2) = P(x < -1 | \mu = 0) = 0,16 \\ P(\text{I}) = P(\text{acción } a_2 | \mu \in \omega_1) = P(x \geq -1 | \mu = -1) = 0,50 \end{cases} \quad (11)$$

Se observan dos cosas: Las probabilidades de acción pueden obtenerse a partir de las probabilidades de error, y viceversa, y, en general, estas probabilidades son función del parámetro si existe más de un punto en ω_1 u ω_2 .

Si la función de pérdida fuese desconocida, formularíamos el problema examinando los errores de tipo I y tipo II. Estos pueden obtenerse a partir de $\mathcal{E}(d; \mu)$. Cuando hablamos de la hipótesis nos referimos a H_1 , que en este ejemplo es $\mu = -1$; la hipótesis

alternativa es $H_2: \mu = 0$. Así, la cantidad $P(I)$, que es la probabilidad de rechazar la hipótesis cuando es verdadera, es

$$\begin{aligned} P(\text{rechazar } H_1 | H_1 \text{ es verdadera}) &= \\ &= P(\text{tomar la acción } a_2 | H_1 \text{ es verdadera}) = \\ &= P(x \in S_2 | H_1) = P(I) \end{aligned}$$

También $P(II)$, que es la probabilidad de aceptar la hipótesis cuando es falsa, es

$$\begin{aligned} P(\text{aceptar } H_1 | H_1 \text{ es falsa}) &= P(\text{aceptar } H_1 | H_2 \text{ es verdadera}) = \\ &= P(\text{tomar la acción } a_1 | H_2 \text{ es verdadera}) = \\ &= P(x \in S_1 | H_2) = P(II) \end{aligned}$$

Utilizando la misma región de rechazo que antes, por (11) obtenemos $P(II) = 0,16$. En este ejemplo, la extensión de la dócima es $\alpha = 0,50$. Si se quiere que α sea otro valor, tendremos que cambiar la región crítica S_2 . Observemos que cuando se emplean las expresiones «probabilidad de rechazar la hipótesis cuando es verdadera» y «probabilidad de aceptar la hipótesis cuando es falsa», la hipótesis a que nos referimos es H_1 y el experimentador puede tomar como $H_1: \mu = -1$ o $\mu = 0$; la hipótesis alternativa será la restante. Una vez que el experimentador ha decidido sobre H_1 , esta es la hipótesis que se utilizará en todo el análisis.

12-2. Dócima de una hipótesis simple contra una alternativa simple.—Una hipótesis $H: \theta \in \omega$ se llama simple si ω contiene un punto único. Así, si ω_1 consta del punto θ_1 y si ω_2 es el punto θ_2 , el problema se denomina docimar una hipótesis simple contra una alternativa simple.

Aquí, la función de riesgo para una estrategia d toma dos valores:

$$R(d; \theta_1) = l(\theta_1)P(I) \quad \text{y} \quad R(d; \theta_2) = l(\theta_2)P(II);$$

por tanto, para cada función de decisión d , el riesgo $R(d; \theta)$ puede representarse por un punto en un gráfico cuyas coordenadas sean $R(d; \theta_1)$ y $R(d; \theta_2)$. Análogamente, $\mathcal{E}(d; \theta)$ podrá representarse en un gráfico cuyas coordenadas son las probabilidades de error $P(I)$ y $P(II)$. Este último gráfico no implica la función de pérdida y es útil en aquellas aplicaciones donde esta función no se conoce perfectamente y $P(I)$ y $P(II)$ pueden utilizarse como se explicó en la sección anterior.

Ilustremos lo anterior con un problema simplificado.

Ejemplo 12-2.—Se desea apostar sobre el resultado de lanzar una moneda cuya probabilidad p de que salga cara se sabe que es $p_1=0,6$ ó $p_2=0,3$. El jugador decide docimar la hipótesis $H_1: p=p_1=0,6$ con la hipótesis alternativa $H_2: p=p_2=0,3$. Las dos acciones son: a_1 , aceptar H_1 (rechazar H_2) y a_2 , aceptar H_2 (rechazar H_1). La función de pérdida que consideraremos está dada por

$$l(a_1; p_1)=l(a_2; p_2)=0$$

$$l(a_2; p_1)=l(p_1)=1$$

y

$$l(a_1; p_2)=l(p_2)=2$$

El apostante decide hacer una observación mediante el lanzamiento de la moneda, siendo $x=1$ si sale cara y $x=0$ si sale cruz. Puesto que el espacio muestral contiene solo dos puntos, 0 y 1, existen únicamente cuatro posibles funciones de decisión, d_1, d_2, d_3, d_4 :

$$d_1: d_1(1)=a_1 \quad d_1(0)=a_1$$

$$d_2: d_2(1)=a_1 \quad d_2(0)=a_2$$

$$d_3: d_3(1)=a_2 \quad d_3(0)=a_1$$

$$d_4: d_4(1)=a_2 \quad d_4(0)=a_2$$

donde, p. ej., $d_4(0)=a_2$ significa: la función de decisión d_4 indica que se adoptará la acción a_2 cuando sea $x=0$, es decir, cuando sale cruz.

Los riesgos y probabilidades de error correspondientes están dados en la tabla siguiente, y los gráficos, en las figuras 12-1 y 12-2.

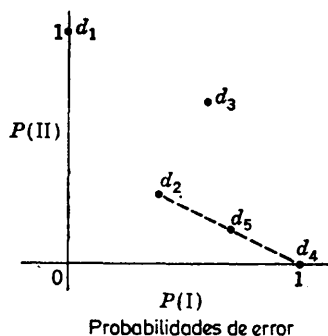


FIG. 12-1.

Probabilidades de error			Riesgo		
	$P(I)$	$P(II)$		$R(d; p_1)$	$R(d; p_2)$
d_1	0	1	d_1	0	2
d_2	0,4	0,3	d_2	0,4	0,6
d_3	0,6	0,7	d_3	0,6	1,4
d_4	1	0	d_4	1	0

Para aclarar cómo se calculan las probabilidades de error, obtengamos $\varepsilon_1(d_2; p_1)$. Se tiene, $\varepsilon_1(d_2; p_1)=P(\text{tomar la acción equivocada})$

cuando se utilizan d_2 y p_1) = $P(\text{tomar la acción } a_2 \text{ cuando se utilizan } d_2 \text{ y } p_1) = P(x=0|p_1) = 0,4$. Análogamente se calculan las restantes probabilidades de error. Para ilustrar el método de cálculo de los riesgos, obtenemos $R(d_3; p_2)$. Se tiene,

$$R(d_3; p_2) = \varepsilon_2(d_3; p_2)l(p_2) = (0,7)(2) = 1,4$$

Un cálculo más detallado de $R(d_3; p_2)$ conduce a

$$\begin{aligned} R(d_3; p_2) &= l(a_1; p_2)P(\text{acción } a_1|p_2) + l(a_2; p_2)P(\text{acción } a_2|p_2) \\ &= 2P(x=0|p_2) + 0 \cdot P(x=1|p_2) \\ &= (2)(0,7) = 1,4 \end{aligned}$$

como antes. El lector comprobará los restantes riesgos dados en la tabla. Observando estos riesgos para las cuatro funciones de

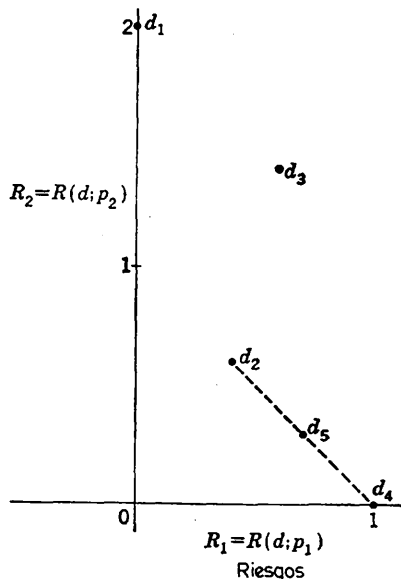


FIG. 12-2.

decisión, vemos que, aun en este ejemplo tan sencillo, en general, no es obvio que una estrategia sea preferible a otra. La estrategia d_2 parece que constituiría una buena elección, pero vemos que si el estado de la naturaleza verdadero es p_1 , d_1 es mejor (riesgo menor) que d_2 ; si el estado de la naturaleza verdadero es p_2 , d_4 es entonces mejor que d_2 . Puesto que el estado de la naturaleza es desconocido, no es posible seleccionar una estrategia que tenga riesgo mínimo para cada estado de la naturaleza. Sin embargo, existe un hecho evidente: la estrategia d_2 tiene menor riesgo que d_3 , con independencia de cuál sea el estado de la naturaleza; por tanto, descartaremos d_3 como posible estrategia y diremos que la estrategia d_2 domina (es mejor que) a la d_3 . Si representamos gráficamente los riesgos para las diferentes estrategias, como en la figura 12-2, donde hemos dibujado el punto (0,2) para d_1 , el punto (0,4, 0,6) para d_2 , etc., vemos que si una estrategia d_i domina a otra d_j , d_i se hallará debajo y a la izquierda de d_j . Si está debajo, tiene riesgo menor para el estado de la naturaleza p_2 ; si a la iz-

quierda, el riesgo es menor para el estado de la naturaleza p_1 . Luego a partir de la figura 12-2, vemos que d_2 domina a d_3 , pero que d_1 , d_2 y d_4 no dominan una a otra.

Supongamos que el apostante no se decide entre d_2 y d_4 y recurre a lanzar una moneda «perfecta» para elegir entre ellas. Entonces, si $H_1: p = p_1$ es correcta, tendrá un riesgo de 0,4 con probabilidad $\frac{1}{2}$ y un riesgo de 1 con probabilidad $\frac{1}{2}$. Esto equivale a un riesgo esperado de 0,7; análogamente, si $H_2: p = p_2$ es correcta, su riesgo será 0,3. Así, al decidir el jugador aleatorizar su elección de la función de decisión, se le ofrece una nueva estrategia d_5 , llamada *estrategia aleatorizada*, y representada por los riesgos 0,7 y 0,3. Evidentemente, es el punto medio del segmento cuyos extremos son los

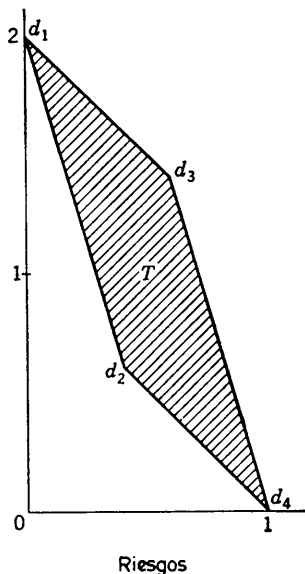


FIG. 12-4.

para mostrar que d_3 es una dócima no deseable. Naturalmente, las únicas dócimas razonables son las que corresponden a puntos

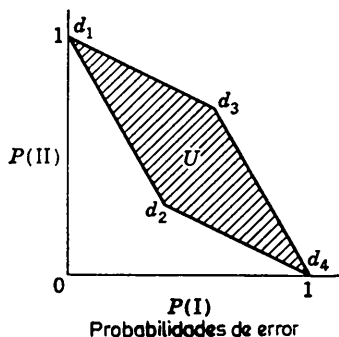


FIG. 12-3.

puntos del gráfico que corresponden d_2 y d_4 . Seleccionando entre d_2 y d_4 , con otras probabilidades distintas de $\frac{1}{2}$, $\frac{1}{2}$, cabe encontrar una estrategia aleatorizada correspondiente a un punto cualquiera del segmento rectilíneo que une estos dos puntos de riesgo.

Se dice que un conjunto de puntos es *convexo* si, dados dos puntos cualesquiera del mismo, el segmento que los une pertenece al conjunto. El conjunto T de todos los puntos de riesgo $[R(d; p_1), R(d; p_2)]$ obtenido al considerar todas las estrategias, incluso las aleatorizadas, es convexo. Lo mismo puede decirse del conjunto U de puntos de error $[P(I), P(II)]$. T y U son los conjuntos convexos más pequeños que contienen a los puntos correspondientes a las estrategias *no aleatorizadas* o *puras* originales. Estos conjuntos se han representado en las figuras 12-3 y 12-4.

Una ojeada a la figura 12-4 basta

de los segmentos rectilíneos que unen d_1 con d_2 , y d_2 con d_4 . Todas las restantes estrategias pueden ser también aprovechables. Introduciremos ahora la noción de *estrategia admisible* (función de decisión admisible).

Definición 12-4.—Una estrategia (función de decisión o dócima) d es admisible si no existe otra estrategia d^* tal que

$$R(d^*; \theta) \leq R(d; \theta) \quad \text{para todo } \theta \text{ de } \Omega$$

y

$$R(d^*; \theta) < R(d; \theta) \quad \text{para algún } \theta \text{ de } \Omega$$

Como se indicó anteriormente, no hay, en general, una función de decisión que dé riesgo mínimo para todos los valores de θ en Ω ; por tanto, se comprende que lo más razonable consiste en hallar la clase de las funciones de decisión admisibles y seleccionar una de ellas.

Para ayudar a encontrar la clase de estrategias admisibles, probaremos que *toda estrategia admisible es una estrategia de Bayes*, y que *toda estrategia de Bayes es una dócima de la razón de verosimilitud*. Por tanto, *toda estrategia admisible es una dócima de la razón de verosimilitud*. En consecuencia, si es posible hallar la clase de dócimas de la razón de verosimilitud, esta incluirá todas las estrategias admisibles; la obtención de la clase de dócimas de la razón de verosimilitud es, frecuentemente, bastante fácil. Dedicaremos el resto de esta sección al desarrollo de estas ideas. Recordemos que nos limitamos a considerar una hipótesis *simple* y una alternativa *simple*.

Definiremos ahora una estrategia de Bayes (véase Cap. 8).

Definición 12-5.—*Estrategia de Bayes.* Una estrategia d es una estrategia de Bayes correspondiente a probabilidades «a priori» h_1 y $h_2 = 1 - h_1$ ($h_i \geq 0$) si hace mínimo $B(d)$, donde

$$B(d) = E[R(d; \theta)] = h_1 R(d; \theta_1) + h_2 R(d; \theta_2)$$

Esbozaremos la demostración del siguiente teorema.

Teorema 12-1.—*Para docimar una hipótesis simple contra una alternativa simple, toda estrategia admisible es una estrategia de Bayes.*

Demostración.—En primer lugar, refiriéndonos a la figura 12-5, observamos que la estrategia de Bayes correspondiente a h_1 y h_2 puede representarse geoméricamente dibujando la recta

$$h_1 R_1 + h_2 R_2 = c$$

y desplazándola mediante la variación de c , paralelamente a sí misma, hasta que toque a T . El punto (o puntos) donde toca a T co-

responde a la estrategia de Bayes. Como h_1 varía desde 0 hasta 1, la pendiente de la recta lo hace, desde 0 hasta $-\infty$. Una propiedad de los conjuntos convexos es que, dado cualquier punto del contorno, existe una recta que pasa por ese punto en la que se *apoya* el conjunto. Luego para toda estrategia admisible, es decir, para

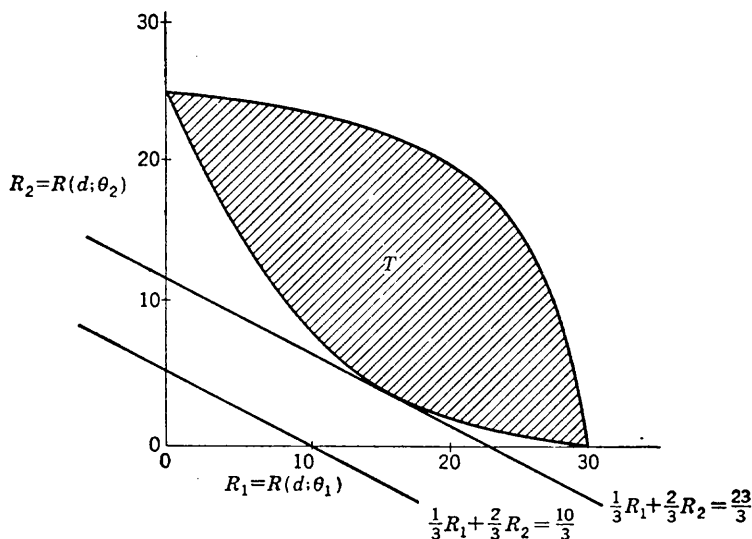


FIG. 12-5.

cualquier punto de contorno inferior de T , existe una recta de apoyo que pasa por dicho punto. Por tanto, puede trazarse esta recta con pendiente no positiva, y expresarse en la forma

$$h_1 R_1 + h_2 R_2 = c$$

donde h_1 y h_2 son probabilidades posibles *a priori* (o sea, $0 \leq h_i \leq 1$). Por consiguiente, la estrategia admisible es una estrategia de Bayes.

Aunque el teorema 12-1 se aplica a problemas más generales, el caso especial de docimar una hipótesis simple contra una alternativa simple nos lleva a un resultado particularmente interesante; es decir, toda estrategia de Bayes es una *dócima de la razón de verosimilitud*. Definiremos qué se entiende por dócima de la razón de verosimilitud.

Definición 12-6. *Dócima de la razón de verosimilitud.*—Una dócima basada en una muestra aleatoria $x_1, x_2, \dots, x_n$ de la densidad $f(x; \theta)$ para docimar $H_1: \theta = \theta_1$ contra $H_2: \theta = \theta_2$ es una dócima

de la razón de verosimilitud, si existe un número k tal que la dócima permite

$$\begin{array}{lll} \text{aceptar } H_1 & (\text{acción } a_1) & \text{si } \lambda > k \\ \text{rechazar } H_1 & (\text{acción } a_2) & \text{si } \lambda < k \end{array}$$

y

$$\text{una de las dos acciones} \quad \text{si } \lambda = k$$

donde λ es la razón de verosimilitud dada por

$$\lambda = t(x_1, \dots, x_n) = \frac{f(x_1; \theta_1)f(x_2; \theta_1) \dots f(x_n; \theta_1)}{f(x_1; \theta_2)f(x_2; \theta_2) \dots f(x_n; \theta_2)} \quad (1)$$

Discutiremos brevemente esta definición. Puesto que θ_1 y θ_2 son números conocidos, la desigualdad $\lambda > k$, para k fijo, define un conjunto S_1 . Otra manera de formular esto es: para un valor de k fijo, existe un conjunto de x que satisface la desigualdad $\lambda > k$. Este conjunto de x forma la región de aceptación S_1 , y el conjunto de x definido por $\lambda < k$ es la región crítica (región de rechazo) S_2 para este valor particular de k . El lector recordará que utilizamos indistintamente los términos función de decisión, estrategia y dócima; todos ellos se refieren a una función para pasar de las observaciones al espacio de acción. La dócima de la razón de verosimilitud, si existe, define las regiones S_1 y S_2 , y la dócima es: adoptar la acción a_1 si s está en S_1 , y adoptar la acción a_2 si s está en S_2 , donde s es el punto n -dimensional $(x_1, \dots, x_n)$.

Ejemplo 12-3.—Sea x una muestra de tamaño 1 de una distribución normal con media μ y varianza 1. Docimaremos la hipótesis $H_1: \mu = -1$ con la hipótesis alternativa $H_2: \mu = 0$. La razón de verosimilitud es, por (1),

$$\lambda = \frac{f(x; -1)}{f(x; 0)} = \frac{e^{-\frac{1}{2}(x+1)^2}}{e^{-\frac{1}{2}x^2}} = e^{-\frac{1}{2}(x^2+2x+1-x^2)} = e^{-\frac{1}{2}(2x+1)}$$

A título ilustrativo, seleccionaremos $k = e^{\frac{1}{2}}$; entonces $\lambda > k$ es

$$e^{-\frac{1}{2}(2x+1)} > e^{\frac{1}{2}}$$

o sea

$$e^{-x} \cdot e^{-\frac{1}{2}} > e^{\frac{1}{2}}$$

o bien

$$e^{-x} > e$$

Existe un conjunto, que llamamos S_1 , sobre el eje x que satisface a $e^{-x} > e$. Dicho conjunto es

$$S_1 = \{x: -\infty < x < -1\}$$

ya que tomando en $e^{-x} > e$ logaritmos en base e , se tiene $-x > 1$; este es el conjunto donde se toma la acción a_1 (aceptar H_1).

Teorema 12-2.—*Para docimar la hipótesis simple $H_1: \theta = \theta_1$ contra la alternativa simple $H_2: \theta = \theta_2$, toda estrategia de Bayes es una dócima de la razón de verosimilitud.*

Analicemos e ilustremos este teorema antes de dar una demostración formal del mismo. Cabe interpretar que la razón de verosimilitud λ es una medida de cómo la evidencia confirma H_1 . Así, es razonable aceptar H_1 cuando λ es suficientemente grande. Obsérvese que el ser «suficientemente grande» puede depender de factores tales como las pérdidas debidas al error y el grado de confianza previa, si la hay, en la hipótesis.

En el ejemplo 12-2, donde solo hay una observación,

$$\lambda = \frac{p_1}{p_2} = 2 \quad \text{si } x = 1$$

y

$$\lambda = \frac{1-p_1}{1-p_2} = \frac{4}{7} \quad \text{si } x = 0$$

Ejemplo 12-4.—Como otro ejemplo de la dócima de la razón de verosimilitud, elijamos una muestra aleatoria de tamaño n para docimar la hipótesis de que la media es 0 contra la alternativa de que la media μ es 2, en una densidad normal con $\sigma^2 = 1$. Esto puede describirse como una dócima de

$$H_1: \mu = \mu_1 = 2$$

contra la alternativa

$$H_2: \mu = \mu_2 = 0$$

con

$$f(x_i; 2) = \frac{1}{\sqrt{2\pi}} e^{-(x_i-2)^2/2} \quad i = 1, 2, \dots, n$$

y

$$f(x_i; 0) = \frac{1}{\sqrt{2\pi}} e^{-x_i^2/2} \quad i = 1, 2, \dots, n$$

y

$$\lambda = \frac{\prod_{i=1}^n f(x_i; \mu_1)}{\prod_{i=1}^n f(x_i; \mu_2)} = \exp(2n\bar{x} - 2n)$$

La d cima de la raz n de verosimilitud consiste en aceptar H_1 si

$$\lambda = \exp(2n\bar{x} - 2n) > k$$

Tomando logaritmos y simplificando, se tiene

$$2n\bar{x} > \log k + 2n$$

o bien

$$\bar{x} > \frac{1}{2n} \log k + 1$$

Si hacemos $(1/2n) \log k + 1 = c$, aceptamos entonces H_1 si $\bar{x} > c$, y rechazamos H_1 si $\bar{x} < c$.

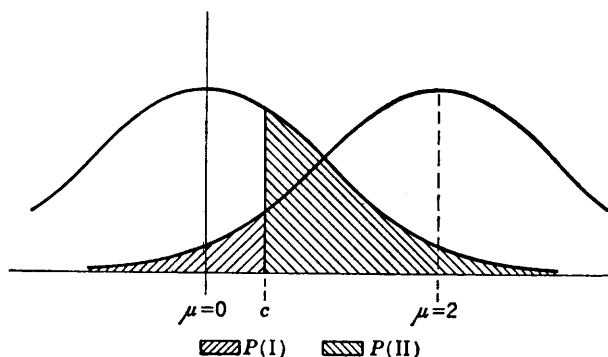


FIG. 12-6.

Puesto que la d cima de la raz n de verosimilitud puede expresarse en funci n de $\bar{x}$, las probabilidades de error, $P(I)$ y $P(II)$, vendr n representadas por las  reas rayadas de la figura 12-6. Las curvas de esta figura son las densidades de $\bar{x}$ bajo las hip tesis H_1 y H_2 . Obs rvese que cuando c crece, $P(II)$ disminuye, pero $P(I)$ aumenta. Para cada c se obtiene un punto sobre el contorno inferior del conjunto U de puntos de error para el problema (comp rese con la Fig. 12-3). Este contorno inferior se llama a veces *curva de error*.

Hallar un punto de esta curva para un valor dado de $P(I)$, p. ej., 0,05, cuando $n=4$, es particularmente sencillo. Observemos que para $\mu = \mu_1 = 2$,

$$P(I) = P(\bar{x} < c | \mu = 2) = 0,05$$

si

$$c = \mu_1 - \frac{1,645}{\sqrt{4}} = 2 - 0,8225 = 1,1775$$

Luego

$$P(II) = P(\bar{x} > 1,1775 | \mu = 0) = 0,009$$

Otro problema interesante es el de encontrar un tamaño n de la muestra tal que $P(I)$ y $P(II)$ tomen valores predeterminados; p. ej., 0,02 y 0,01, respectivamente. Entonces para que $P(I) = 0,02$ se necesita que

$$c = \mu_1 - \frac{2,054}{\sqrt{n}} = 2 - \frac{2,054}{\sqrt{n}}$$

Para $P(II) = 0,01$, se necesita que

$$c = \mu_2 + \frac{2,323}{\sqrt{n}} = \frac{2,323}{\sqrt{n}}$$

De aquí,

$$\begin{aligned} 2 - \frac{2,054}{\sqrt{n}} &= \frac{2,323}{\sqrt{n}} \\ 2 &= \frac{4,377}{\sqrt{n}} \end{aligned}$$

y

$$n = \left(\frac{4,377}{2} \right)^2 = 4,8$$

Además

$$c = \frac{2,323}{\sqrt{n}} = 1,06$$

Luego bastaría una muestra de tamaño $n=5$ para tener probabilidades respectivas de error ligeramente menores que 0,02 y 0,01. Aquí la dócima consistiría en aceptar $H_1: \mu = 2$ si $\bar{x} > c$, donde c es, aproximadamente, 1,06.

Queda por demostrar que las estrategias de Bayes son dócimas de la razón de verosimilitud. Para hacerlo, observemos que

$$\begin{aligned} P(I) &= \int \int \dots \int_{S_1} f(x_1; \theta_1) \dots f(x_n; \theta_1) dx_1 \dots dx_n \\ &= 1 - \int \int \dots \int_{S_1} f(x_1; \theta_1) \dots f(x_n; \theta_1) dx_1 \dots dx_n \end{aligned}$$

$$P(II) = \int \int \dots \int_{S_1} f(x_1; \theta_2) \dots f(x_n; \theta_2) dx_1 \dots dx_n$$

$$B(d) = h_1 R(d; \theta_1) + h_2 R(d; \theta_2)$$

$$= h_1 l(\theta_1) P(I) + h_2 l(\theta_2) P(II)$$

$$\begin{aligned} &= h_1 l(\theta_1) + \int \int \dots \int_{S_1} \left[-h_1 l(\theta_1) \prod_{i=1}^n f(x_i; \theta_1) + h_2 l(\theta_2) \prod_{i=1}^n \right. \\ &\quad \left. f(x_i; \theta_2) \right] dx_1 \dots dx_n \end{aligned}$$

La minimización de $B(d)$ consiste en seleccionar el conjunto S_1 que hace mínima la integral anterior. Evidentemente, esto se consigue introduciendo en S_1 todos los puntos $(x_1, x_2, \dots, x_n)$ para los que el integrando es negativo, y excluyendo de tal conjunto todos aquellos para los cuales es positivo. Por tanto, la estrategia de Bayes implica la acción a_1 si

$$0 > -h_1 l(\theta_1) \prod_{i=1}^n f(x_i; \theta_1) + h_2 l(\theta_2) \prod_{i=1}^n f(x_i; \theta_2)$$

o si

$$\lambda = \frac{\prod_{i=1}^n f(x_i; \theta_1)}{\prod_{i=1}^n f(x_i; \theta_2)} > \frac{h_2 l(\theta_2)}{h_1 l(\theta_1)} = k$$

Análogamente, debemos tomar la acción a_2 si $\lambda < k$, y podrá adoptarse cualquier acción, o bien hacer una elección aleatorizada entre ellas, si $\lambda = k$.

Con esto damos por terminado el análisis del caso de docimar una hipótesis simple contra una alternativa simple. Este caso no es de mucha utilidad en estadística aplicada, pero nos ha servido

como introducción a la teoría de la docimasia de hipótesis. Hemos destacado el hecho de que una de las mayores dificultades en la docimasia de hipótesis consiste en seleccionar una dócima «mejor»; es decir, una dócima que dé riesgo mínimo para todos (uno y otro) los valores del parámetro. Para eludir en parte esta dificultad, definimos una dócima admisible y probamos que la dócima de la razón de verosimilitud es un método para la obtención de la clase que contiene todas las dócimas admisibles. Una vez obtenida esta clase, un experimentador podrá seleccionar una dócima de ella utilizando criterios adicionales.

Otra dificultad en la docimasia de hipótesis la representa el hecho de que, en gran parte de los problemas, el experimentador no conoce la función de pérdida con un grado elevado de acuracidad. Esta dificultad queda parcialmente eludida si se conoce de forma aproximada la importancia relativa de los errores de los tipos I y II. Este es el método tradicional utilizado en el pasado por los experimentadores y lo expondremos más adelante en este capítulo para otros problemas.

12-3. Hipótesis compuestas.—En la práctica, la mayor parte de los problemas de docimasia implican *hipótesis compuestas*. Estas hipótesis son de la forma $H_1: \theta \in \omega_1$, con la alternativa $H_2: \theta \in \omega_2$, en donde ω_1 , y/o ω_2 contienen más de un elemento. Así, p. ej., volvamos a examinar el ejemplo de las lámparas de la sección 12-1 desde el punto de vista de la teoría de la decisión. Se consideran dos acciones: la primera consiste en decidir que las nuevas lámparas representan un perfeccionamiento sustancial en relación con las antiguas, y adoptar el nuevo proceso de fabricación (a_1); la otra, a_2 , equivale a persistir en el proceso primitivo. Para precisar las ideas, supongamos que la vida de las lámparas fabricadas con el nuevo proceso tiene una distribución exponencial, con media desconocida $\mu > 0$; es decir,

$$\begin{aligned} f(x; \mu) &= \frac{1}{\mu} e^{-x/\mu} && \text{para } x > 0 \\ &= 0 && \text{para } x \leq 0 \end{aligned}$$

El fabricante, comparando el beneficio esperado con el costo de modificar sus instalaciones, se decide por un valor de μ , por encima del cual resultaría provechoso introducir el nuevo proceso, mientras que no lo sería por debajo de él. Imaginemos que este valor es $\mu = 1450$. El fabricante debe, naturalmente, calcular las pérdidas $l(a_1; \mu)$ y $l(a_2; \mu)$ y representarlas, tal como se ha hecho en la figura 12-7.

Interesa docimar la hipótesis $H_1: \mu > 1450$ contra la alternativa $H_2: \mu \leq 1450$. Para una dócima dada d , se calcula la probabilidad de aceptar H_1 para cada valor de μ . En la figura 12-8 se ha trazado un gráfico que da esta probabilidad (aproximadamente) para cuatro dócimas basadas en muestras de tamaño 400:

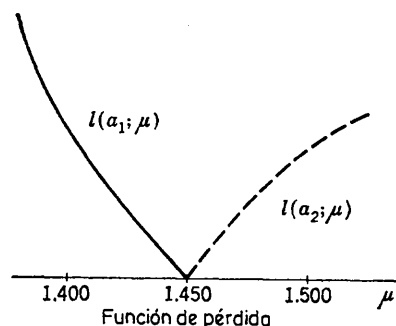


FIG. 12-7.

d_1 :	aceptar H_1	si $\bar{x} > 1400$
d_2 :	aceptar H_1	si $\bar{x} > 1425$
d_3 :	aceptar H_1	si $\bar{x} > 1450$
d_4 :	aceptar H_1	si $\bar{x} > 1475$

Para calcular estas probabilidades, aplicamos el teorema central del límite, teniendo en cuenta que la distribución exponencial posee media μ y desviación típica μ . Luego $\bar{x}$ se distribuye aproximadamente según la normal de media μ y desviación típica $\mu/20$.

Estas curvas se contrastan con la probabilidad *ideal* de aceptar H_1 , es decir, la función que vale 0 para $\mu \leq 1450$ y 1 para $\mu > 1450$. La función ideal no puede obtenerse con una muestra finita, y las

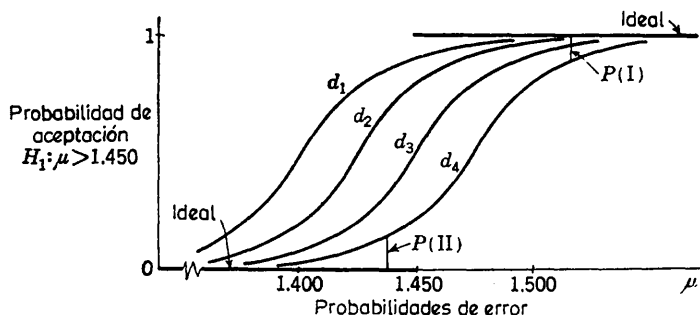


FIG. 12-8.

desviaciones respecto a ella representan las probabilidades de error. Obsérvese que mediante un cambio de estrategia se puede reducir la probabilidad de error a un lado de 1450, si bien a costa de incrementarla al otro lado de dicho valor. En la docimasia de hipótesis compuestas la situación resulta mucho más compleja que cuando las hipótesis son simples. En el caso compuesto, las dócimas admisibles son difíciles o imposibles de obtener. En este caso,

nos contentaremos, en general, con un análisis de las probabilidades de error $P(I)$ y $P(II)$, e intentaremos hallar dúcimas que de cierta manera las controlen.

En el resto del capítulo nos limitaremos a considerar la siguiente formulación del problema de la docimasia de hipótesis: Para docimar la hipótesis $H_1: \theta \in \omega_1$ contra la hipótesis alternativa $H_2: \theta \in \omega_2$, buscaremos una dúcima tal que para un α seleccionado ($0 < \alpha < 1$)

$$P(I) \leq \alpha \quad \text{para todo } \theta \in \omega_1$$

y

$$1 - P(II) = \beta(\theta) \text{ es máximo para todo } \theta \in \omega_2 \quad (1)$$

En otras palabras, trataremos de hallar dúcimas que tengan potencia máxima para una extensión fijada α . La primera solución a este problema fue dada por Neyman y Pearson en el caso de hipótesis simple contra alternativa simple. Se conoce con el nombre de lema fundamental de Neyman-Pearson, y damos a continuación su enunciado, aunque omitimos la demostración.

Teorema 12-3.—*La región crítica R_k de extensión α que hace máxima la potencia de la dúcima de la hipótesis $H_1: \theta = \theta_1$, contra la alternativa $H_2: \theta = \theta_2$, donde $x_1, \dots, x_n$ es una muestra aleatoria de tamaño n de $f(x; \theta)$, se obtiene hallando la región R_k (si existe) que satisface a*

$$\lambda = \frac{f(x_1; \theta_1)f(x_2; \theta_1) \dots f(x_n; \theta_1)}{f(x_1; \theta_2)f(x_2; \theta_2) \dots f(x_n; \theta_2)} < k \quad (2)$$

para un número fijo k , y tal que

$$\iint_{R_k} \dots \int f(x_1; \theta_1)f(x_2; \theta_1) \dots f(x_n; \theta_1) dx_1 dx_2 \dots dx_n = \alpha \quad (3)$$

Esto, evidentemente, constituye una aplicación de la razón de verosimilitud.

A primera vista no parece claro cómo (3) implica k , pero la región en que se verifica (2) cambia al variar k , y cuando esto ocurre puede haber una región (un valor de k) que satisface a (3). Esto se pondrá de relieve con detalle cuando consideremos un ejemplo. Es importante insistir en que este teorema proporciona una región crítica más potente (de extensión α) para docimar *solo* que θ es un punto *único* contra la alternativa de que θ es también un *único* punto. El teorema *no* da necesariamente un método para hallar una región crítica más potente de extensión α cuando ω_1 u ω_2 contienen más de un punto. Veremos más adelante que *algunas veces* puede

utilizarse en tales situaciones, que, evidentemente, son los casos más útiles. Es decir, un experimentador puede desear docimar que la diferencia de rendimientos medios de dos variedades de trigo es cero contra la alternativa de que es positiva. O un fabricante deseeará quizá docimar la hipótesis $H_1: \mu \leq 0$ contra la alternativa $H_2: \mu > 0$, donde μ es la diferencia de eficacia media de dos medicamentos. En estos casos, ω_1 u ω_2 (o ambos, ω_1 y ω_2) contienen más de un punto. Existen cuatro casos distintos: 1) ω_1 contiene un punto y ω_2 contiene un punto; 2) ω_1 contiene un punto y ω_2 contiene más de un punto; 3) ω_1 contiene más de un punto y ω_2 contiene solo uno; 4) ω_1 y ω_2 contienen ambos más de un punto. En general, el lema de Neyman-Pearson se aplica únicamente al caso 1, pero veremos que algunas veces es también útil en otros casos.

Ejemplo 12-5.—Como ejemplo sencillo de la utilización del lema de Neyman-Pearson, docimemos la hipótesis de que en una densidad normal con $\sigma^2=1$, la media es 0 contra la hipótesis alternativa de que la media es 2 (véase Ej. 12-4). Esto puede escribirse

$$H_1: \mu = 0 \quad H_2: \mu = 2$$

$$f(x_i; 0) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x_i^2} \quad i = 1, 2, \dots, n$$

y

$$f(x_i; 2) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x_i-2)^2} \quad i = 1, 2, \dots, n$$

y también

$$\lambda = \frac{\prod_{i=1}^n f(x_i; \mu_1)}{\prod_{i=1}^n f(x_i; \mu_2)} = \exp \left[+\frac{1}{2} \sum_{i=1}^n (x_i^2 - 4x_i + 4) \right]$$

Por (2), necesitamos hallar la región del espacio n -dimensional tal que

$$\exp \left[+\frac{1}{2} \sum_{i=1}^n (x_i^2 - 4x_i + 4) \right] < k$$

Tomando logaritmos y simplificando resulta

$$-2n\bar{x} < \log k - 2n$$

o bien

$$\bar{x} > -\frac{1}{2n} \log k + 1$$

Si hacemos $-(1/2n) \log k + 1 = c$, la región de rechazo R_k es tal que $\bar{x} > c$. Evidentemente, c es desconocido, puesto que lo es k . Sin embargo, conocemos la *forma* de la región de rechazo. La distribución de $\bar{x}$, cuando H_1 es verdadera, es normal con media 0; por tanto, en virtud de (3), tenemos para, p. ej., $\alpha = 0,05$,

$$\frac{\sqrt{n}}{\sqrt{2\pi}} \int_c^\infty \exp\left(-\frac{n\bar{x}^2}{2}\right) d\bar{x} = 0,05 \quad (4)$$

Si hacemos $y = \sqrt{n} \bar{x}$, (4) se transforma en

$$\frac{1}{\sqrt{2\pi}} \int_{c\sqrt{n}}^\infty e^{-y^2/2} dy = 0,05$$

y, por la tabla II, $\sqrt{n}c = 1,645$ y $c = 1,645/\sqrt{n}$. En este problema no es preciso conocer el valor de k , puesto que la región crítica depende de c . Hemos demostrado que la región crítica de extensión 0,05 con máxima potencia, de la dócima $H_1: \mu = 0$ contra la alternativa $H_2: \mu = 2$, es $\bar{x} > 1,645/\sqrt{n}$. En consecuencia, deberá rechazarse H_1 si $\bar{x} > 1,645/\sqrt{n}$.

Aunque el lema fundamental de Neyman-Pearson se aplica específicamente a problemas que implican una hipótesis simple contra una alternativa simple, demostraremos que, *algunas veces*, también puede utilizarse ventajosamente en hipótesis compuestas.

12-4. Docimasia de $\theta \leq \theta_1$ contra $\theta > \theta_1$ para densidades con un parámetro único θ .—En estadística aplicada existen muchas densidades que contienen un parámetro único desconocido, tales como la binomial, la de Poisson, la normal de media conocida, la normal de varianza conocida, la exponencial, etc. Muchas veces un experimentador desea docimar la hipótesis $H_1: \theta \leq \theta_1$ con la hipótesis alternativa $H_2: \theta > \theta_1$, siendo θ_1 conocido y donde la densidad es $f(x; \theta)$.

Ejemplo 12-6.—Como ejemplo consideremos una densidad normal con varianza 1 y media desconocida μ . Supongamos que deseamos docimar la hipótesis $H_1: \mu = \mu_1$ contra la hipótesis alternativa $H_2: \mu > \mu_1$, donde μ_1 es conocido. Seleccionemos un número μ_2 , $\mu_2 > \mu_1$, y formemos una nueva hipótesis $H_1^*: \mu = \mu_1$ contra la alternativa $H_2^*: \mu = \mu_2$, donde μ_2 es un valor tal que $\mu_2 > \mu_1$. La razón de verosimilitud para una muestra aleatoria de tamaño n se reduce a

$$\lambda = \exp -^{1/2}[2n\bar{x}(\mu_2 - \mu_1) + n(\mu_1^2 - \mu_2^2)]$$

Por el teorema 12-3, la región crítica es aquel conjunto de las x tal que $\lambda < k$. Después de algunas simplificaciones, resulta:

$$-n\bar{x}(\mu_2 - \mu_1) < \log k + \frac{n}{2}(\mu_1^2 - \mu_2^2)$$

Puesto que $\mu_2 - \mu_1 > 0$, podemos dividir por $-n(\mu_2 - \mu_1)$ y obtener

$$\bar{x} > -\frac{\log k}{n(\mu_2 - \mu_1)} + \frac{1}{2}(\mu_1 + \mu_2)$$

Si hacemos

$$-\frac{\log k}{n(\mu_2 - \mu_1)} + \frac{1}{2}(\mu_1 + \mu_2) = c$$

la región crítica es, pues, de la forma $\bar{x} > c$, *independientemente* del valor de μ_2 (en tanto sea $\mu_2 > \mu_1$). El valor particular que toma c depende del asignado a α , y en modo alguno depende del valor que toma μ_2 si $\mu_2 > \mu_1$, y, si se rechaza H_1 cuando $\bar{x} > c$, esta será una dócima de $H_1: \mu = \mu_1$ contra la alternativa $H_2: \mu > \mu_1$. Calculamos c por medio de

$$P(I) = 0,05 = -\frac{\sqrt{n}}{\sqrt{2\pi}} \int_c^\infty \exp \left[-\frac{n}{2}(\bar{x} - \mu_1)^2 \right] d\bar{x}$$

lo que da $c = 1,645/\sqrt{n} + \mu_1$. Esta se denomina dócima *uniformemente más potente* (dócima UMP) de extensión α de la hipótesis H_1 contra la H_2 . Definamos una dócima UMP.

Definición 12-7.—Una dócima de la hipótesis $H_1: \theta \in \omega_1$, contra la alternativa $H_2: \theta \in \omega_2$, se dice que es una dócima UMP de extensión α si su región crítica R es tal que

$$\begin{aligned} P(I) &\leq \alpha && \text{para todo } \theta \text{ de } \omega_1 \\ \beta(\theta) &= 1 - P(II) \text{ es máximo} && \text{para cada } \theta \text{ de } \omega_2 \end{aligned} \quad (1)$$

En la formulación de la docimasia de hipótesis dada en (12-3-1), una dócima UMP es la «mejor» dócima.

A continuación daremos un teorema bastante útil para determinar una dócima UMP de $H_1: \theta \leq \theta_1$ contra la hipótesis alternativa $H_2: \theta > \theta_1$.

Teorema 12-4.—Sea $x = (x_1, \dots, x_n)$ una muestra aleatoria de una densidad con un único parámetro θ en un intervalo Ω , y sea

$f(x; \theta)$ la densidad conjunta de las variables aleatorias. Supongamos que $f(x; \theta)$ puede escribirse así:

$$f(x; \theta) = s(\theta)U(x)e^{v(x)t(\theta)} \quad (2)$$

donde $t(\theta)$ es una función estrictamente creciente de θ en Ω . Si existe una constante c tal que $P[v(x) > c | \theta_1] = \alpha$ para un α dado y comprendido entre 0 y 1, R es entonces una región crítica UMP de extensión α para docimar $H_1: \theta \leq \theta_1$ contra $H_2: \theta > \theta_1$, donde $R = \{x: v(x) > c\}$. Si $t(\theta)$ es una función estrictamente decreciente de θ en Ω y si existe una constante c tal que $P[v(x) < c | \theta_1] = \alpha$ para un α dado entre 0 y 1, R es una región crítica UMP de extensión α para docimar la H_1 anterior contra la H_2 , siendo $R = \{x: v(x) < c\}$.

Ejemplo 12-7.—Docimemos $H_1: \theta \leq 1$ contra $H_2: \theta > 1$ mediante una muestra x de tamaño uno de la densidad $\theta e^{-\theta x}$, $\theta > 0$. La densidad de x puede escribirse en la forma (2) haciendo $s(\theta) = \theta$, $U(x) = 1$, $v(x) = x$, $t(\theta) = -\theta$; luego $t(\theta)$ es una función estrictamente decreciente de θ . Deseamos hallar una dócima de extensión $\alpha = 0,05$ y, por tanto, hacemos

$$P[v(x) < c | \theta = 1] = P[v(x) < c | 1] = 0,05$$

y obtenemos

$$0,05 = P(x < c | 1) = \int_0^c e^{-x} dx = -e^{-c} + 1$$

o bien $e^{-c} = 0,95$, lo que da $c = -\log_e 0,95$. Luego una dócima UMP de H_1 contra H_2 es: Rechazar H_1 si $x < -\log_e 0,95$.

Este teorema también puede utilizarse para obtener una dócima UMP de la hipótesis $H_1: \theta = \theta_1$ contra la alternativa $H_2: \theta > \theta_1$. Para docimar la hipótesis $H_1: \theta \geq \theta_1$ contra la alternativa $H_2: \theta < \theta_1$, se empleará el teorema para hallar una dócima UMP invirtiendo en el mismo el sentido de todas las desigualdades. También cabe utilizarlo para docimar $H_1: \theta = \theta_1$ contra la alternativa $H_2: \theta < \theta_1$ si se invierten las desigualdades que figuran en él. Muchas de las densidades con un parámetro único θ , que tienen importancia en estadística aplicada, pueden escribirse en la forma (2), por lo que el teorema resulta aplicable.

12-5. Docimasia de la hipótesis $H_1: \theta_1 \leq \theta \leq \theta_2$ contra la hipótesis alternativa $H_2: \theta > \theta_2, \theta < \theta_1$.—Una dócima que ha sido muy aplicada en diversos campos de la ciencia es $H_1: \theta = \theta_0$ contra $H_2: \theta \neq \theta_0$. Así, p. ej., sea θ la diferencia media de rendimientos entre dos variedades de trigo. Algunas veces es deseable docimar la hipótesis $H_1: \theta = 0$ contra la $H_2: \theta \neq 0$; es decir, docimar si las

dos variedades son diferentes en sus rendimientos medios. Sin embargo, en esta situación y en otras muchas en que θ puede variar de forma continua en un intervalo, es inconcebible que θ sea exactamente igual a 0 (o sea, que las variedades sean idénticas en sus rendimientos medios). Con todo, la dócima establecería: ¿son los dos recidimientos idénticos (hasta una diez mil millonésima de bushel, etc.)? Se ve que en muchos casos resulta más realista para un experimentador seleccionar un intervalo en torno a θ_0 (p. ej., $\theta_1 \leq \theta \leq \theta_2$) y docimar $H_1: \theta_1 \leq \theta \leq \theta_2$ contra la alternativa $H_2: \theta < \theta_1; \theta > \theta_2$. Así, p. ej., en la ilustración anterior puede ser factible hacer $\theta_1 = -1/2$ y $\theta_2 = 1/2$ y docimar si la diferencia de rendimientos medios de las dos variedades se halla incluida entre $-1/2$ bushel y $+1/2$ bushel, contra la alternativa de que no se encuentre en este intervalo. Quizá sea difícil o imposible de obtener una dócima UMP para la hipótesis anterior, pero si $f(x; \theta)$ es una densidad con un único parámetro, el estimador máximo-verosímil $\hat{\theta}$ podrá utilizarse algunas veces para construir una dócima, y comparar la potencia de esta con la función de potencia ideal para una dócima de extensión α . Para algunas densidades cabe utilizar una dócima de la forma siguiente:

$$\begin{aligned} &\text{rechazar } H_1 \text{ si } \hat{\theta} \text{ no está en el intervalo} & c_1 \leq \hat{\theta} \leq c_2 \\ &\text{aceptar } H_1 \text{ si } \hat{\theta} \text{ está en el intervalo} & c_1 \leq \hat{\theta} \leq c_2 \end{aligned} \quad (1)$$

en donde c_1 y c_2 se eligen de tal manera que la dócima tenga extensión α . A menudo c_1 y c_2 podrán elegirse de forma que

$$\int_{c_1}^{c_2} g(\theta; \theta_1) d\theta = \int_{c_1}^{c_2} g(\theta; \theta_2) d\theta = 1 - \alpha$$

en donde $g(\hat{\theta}; \theta)$ es la densidad de $\hat{\theta}$ cuando el parámetro es θ . La potencia de la dócima es

$$\beta(\theta) = 1 - \int_{c_1}^{c_2} g(\hat{\theta}; \theta) d\hat{\theta} \quad \text{para } \theta \text{ en } \Omega = \{\theta: b_0 < \theta < b_1\} \quad (2)$$

Esta potencia puede compararse con la función de potencia ideal, dada por

$$\begin{aligned} \beta(\theta) &= 1 & b_0 < \theta < \theta_1 \\ \beta(\theta) &= \alpha & \theta_1 \leq \theta \leq \theta_2 \\ \beta(\theta) &= 1 & b_1 > \theta > \theta_2 \end{aligned}$$

Si la potencia (2) de la dócima (1) no se desvía de la *ideal* en más de lo que le es posible tolerar al experimentador, la dócima será útil aun cuando no sea UMP.

Hay muchos problemas de docimasia de hipótesis para los cua-

les no existe una d cima UMP; el que hemos tratado en esta secci n es uno de ellos. En estos casos resulta a veces posible restringir la clase de d cimas y hallar en la clase restringida una que sea UMP. Una de tales clases es la de d cimas insesgadas.

Definici n 12-8. *D cimas insesgadas.*—Una d cima de la hip tesis $H_1: \theta \in \omega_1$ contra la hip tesis alternativa $H_2: \theta \in \omega_2$ se dice que es una d cima insesgada de extensi n α si

$$P(I) \leq \alpha \quad \text{para todo } \theta \text{ de } \omega_1$$

y

$$\beta(\theta) \geq \alpha \quad \text{para todo } \theta \text{ de } \omega_2$$

En consecuencia, la probabilidad en una d cima insesgada de rechazar la hip tesis H_1 cuando es falsa es al menos tan grande como la de rechazar la hip tesis H_1 cuando es verdadera. En muchos aspectos esta restricci n parece razonable para establecer una d cima. Si en esta clase restringida existe una d cima UMP, se tendr  entonces una d cima insesgada UMP (d cima IUMP).

12-6. D cima de la raz n de verosimilitud generalizada.—

En numerosos casos no se aplicar  la teor a precedente, y no existe una d cima UMP ni IUMP. En tales circunstancias quiz  sea posible restringir la clase de d cimas a n m s que a las d cimas insesgadas y buscar d cimas UMP en la clase restringida. Otro tratamiento del problema consiste en utilizar un m todo general para construir una d cima que tenga propiedades deseables. Tal d cima es la de la raz n de verosimilitud generalizada, que definiremos m s adelante.

En el resto de este cap tulo consideraremos una familia general de distribuciones $f(x; \theta_1, \theta_2, \dots, \theta_k)$; la hip tesis a docimar, denominada hip tesis nula y designada por H_0 , establece que los par metros pertenecen a cierto subespacio ω del espacio param trico Ω .

La hip tesis alternativa se designar  por H_a , pero no se formular  expl citamente cuando quede suficientemente claro por el contexto cu l es esta hip tesis. Sea $x_1, x_2, \dots, x_n$ una muestra de tama o n de una poblaci n con densidad $f(x; \theta_1, \theta_2, \dots, \theta_k)$, donde Ω , espacio param trico, es la totalidad de los puntos que puede tomar $(\theta_1, \theta_2, \dots, \theta_k)$. Bas ndose en esta muestra, supongamos que deseamos docimar la hip tesis $H_0: (\theta_1, \theta_2, \dots, \theta_k)$ es un punto de ω . La hip tesis alternativa ser  $H_a: (\theta_1, \theta_2, \dots, \theta_k)$ es un punto de $\bar{\omega}$, donde $\bar{\omega}$ es $\Omega - \omega$.

La verosimilitud de la muestra es

$$L = \prod_{i=1}^n f(x_i; \theta_1, \theta_2, \dots, \theta_k) \quad (1)$$

La verosimilitud, como función de los parámetros, tendrá de ordinario un máximo al variar los parámetros sobre la totalidad del espacio paramétrico Ω ; designemos este valor máximo por $L(\theta_1, \theta_2, \dots, \theta_k)$ o, más brevemente, por $L(\hat{\Omega})$. En el subespacio ω , L tendrá también de ordinario un valor máximo, que designaremos por $L(\hat{\omega})$.

Definición 12-9.—*La razón de verosimilitud generalizada es el cociente*

$$\lambda = \frac{L(\hat{\omega})}{L(\hat{\Omega})}$$

en donde $L(\hat{\omega})$ es el máximo de la función de verosimilitud, respecto a los parámetros, en la región ω , y $L(\hat{\Omega})$, el máximo de la función de verosimilitud, también respecto a los parámetros, en la región Ω .

La denominaremos simplemente razón de verosimilitud cuando no exista peligro de confusión con la razón de verosimilitud que figura en la definición 12-6. Esta cantidad es necesariamente un número no negativo; L es no negativo por ser una razón de funciones de densidad y $L(\hat{\omega})$ será menor o, a lo sumo, igual a $L(\hat{\Omega})$, ya que hay menos libertad para maximizar L en ω que en Ω . La razón λ es función solo de las observaciones muestrales; no depende de ningún parámetro, y λ varía entre 0 y 1.

Ejemplo 12-8.—Sea $f(x; \mu)$ una densidad normal con varianza 1 y media desconocida μ , y $x_1, x_2, \dots, x_n$, una muestra aleatoria que consta de n observaciones. Docimaremos la hipótesis nula

$$H_0: \mu = 3 \quad \text{contra la alternativa} \quad H_a: \mu \neq 3$$

El punto $\mu = 3$ es ω , mientras que todo el eje μ es Ω . La verosimilitud es

$$L = \left(\frac{1}{\sqrt{2\pi}} \right)^n e^{-\frac{1}{2} \sum (x_i - \mu)^2}$$

que puede escribirse

$$L = \left(\frac{1}{\sqrt{2\pi}} \right)^n \exp \left[-\frac{1}{2} \sum (x_i - \bar{x})^2 - \frac{n}{2} (\bar{x} - \mu)^2 \right]$$

Por supuesto, el valor máximo de esta cantidad en Ω se obtiene haciendo $\mu = \bar{x}$, lo que da

$$L(\Omega) = \left(\frac{1}{\sqrt{2\pi}} \right)^n \exp \left[-\frac{1}{2} \sum (x_i - \bar{x})^2 \right]$$

Puesto que en este ejemplo ω es un punto (la hipótesis nula es simple), no existe oportunidad para que varíe μ , y el valor mayor de L en ω es, simplemente, su único valor:

$$\begin{aligned} L(\hat{\omega}) &= \left(\frac{1}{\sqrt{2\pi}} \right)^{n/2} \exp \left[-\frac{1}{2} \sum (x_i - 3)^2 \right] \\ &= \left(\frac{1}{\sqrt{2\pi}} \right)^n \exp \left[-\frac{1}{2} \sum (x_i - \bar{x})^2 - \frac{n}{2} (\bar{x} - 3)^2 \right] \end{aligned}$$

La razón de verosimilitud es ahora

$$\lambda = \exp \left[-\frac{n}{2} (\bar{x} - 3)^2 \right]$$

Si $\bar{x}$ se aproxima mucho a 3, la muestra resulta razonablemente compatible con H_0 , y λ no diferirá apenas de 1. Si $\bar{x}$ es mucho mayor o menor que 3, la muestra no será compatible con H_0 y λ valdrá aproximadamente cero.

Evidentemente, la región crítica adecuada para la docimasia de H_0 es un intervalo,

$$0 < \lambda < A$$

en donde A es cierto número (menor que 1) elegido de modo que proporcione el control deseado del error de tipo I.

Si tomamos 0,05 como probabilidad de un error de tipo I, tenemos que determinar A de tal modo que

$$\int_0^A g(\lambda|H_0) d\lambda = 0,05$$

en donde $g(\lambda|H_0)$ es la densidad de λ cuando H_0 es verdadera, es decir, para este problema, si es $\mu=3$. En nuestro caso $\bar{x}$ es normal con media 3 y varianza $1/n$, y $(\bar{x}-3)\sqrt{n}$ se distribuye como una ji cuadrado con un grado de libertad. En consecuencia, $-2 \log \lambda$ es también una ji cuadrado con un grado de libertad. Sea $y = -2 \log \lambda$; entonces

$$0,05 = \int_0^A g(\lambda|H_0) d\lambda = \int_{-2 \log A}^{\infty} h(y) dy = \int_{3,84}^{\infty} h(y) dy$$

puesto que $h(y)$ es una densidad ji cuadrado con un g. de l. La re-

gión crítica es, por tanto, $(\bar{x} - 3)^2 n > 3,84$, o la región definida por las desigualdades

$$\bar{x} > 3 + \frac{1,96}{\sqrt{n}}$$

$$\bar{x} < 3 - \frac{1,96}{\sqrt{n}}$$

Este ejemplo ilustra la situación general. Si el estimador de máxima verosimilitud coincide con o se halla próximo a ω , la muestra se considerará consistente con H_0 , y la razón λ diferirá poco de 1. Si el estimador está a cierta distancia de ω , la muestra no será acorde con H_0 y λ será pequeño. Esto queda resumido como sigue:

Teorema 12-5.—*La región crítica para la docimasia de una hipótesis simple H_0 , mediante una dócima de la razón de verosimilitud, es $0 < \lambda < A$, en donde A se determina por*

$$\int_0^A g(\lambda|H_0) d\lambda = \alpha$$

siendo α la probabilidad de error de tipo I.

Si la distribución de λ no está tabulada y sí lo está la de una función de λ , puede utilizarse esta.

Teorema 12-6.—*Sea λ una razón de verosimilitud para docimar una hipótesis simple H_0 , e $y = u(\lambda)$ una función monótona creciente (o decreciente) de λ ; la dócima basada en y es equivalente a la de la razón de verosimilitud. La región crítica para la dócima basada en y es*

$$u(0) < y < u(A) [u(A) < y < u(0)].$$

Demostración.—La región crítica es $0 < \lambda < A$, en donde A está determinada por

$$\alpha = \int_0^A g(\lambda|H_0) d\lambda$$

Si $y = u(\lambda)$ es una función monótona creciente de λ , entonces haremos

$$\alpha = \int_0^A g(\lambda|H_0) d\lambda = \int_{u(0)}^{u(A)} h(y|H_0) dy$$

siendo $h(y|H_0)$ la densidad de y cuando H_0 es verdadera. Luego la región crítica es

$$u(0) < y < u(A).$$

El sentido de las desigualdades se invierte si $u(\lambda)$ es una función monótona decreciente de λ . Este teorema se ha aplicado en el ejemplo 12-8, en donde

$$y = -2 \log \lambda = (\bar{x} - 3)^2 n$$

es una función monótona decreciente de λ ; por tanto, puede utilizarse $(\bar{x} - 3)^2 n$ como función de docimasia.

Para precisar la región crítica para λ es necesario conocer la distribución de λ cuando H_0 es verdadera. Si H_0 es una hipótesis simple [p. ej., ω es un punto $(\theta_1^0, \theta_2^0, \dots, \theta_k^0)$], habrá entonces una sola distribución determinada para λ ; pero si H_0 es compuesta, la distribución para λ puede no ser única. Es perfectamente posible que la distribución de λ sea diferente para distintos puntos paramétricos en ω . Si es este el caso, se elige A de manera que la integral del teorema 12-5 sea menor o igual que α para todos los valores de los parámetros en ω .

Existe una solución satisfactoria del problema de la docimasia de hipótesis cuando se trabaja, como es frecuente, con muestras grandes. La solución se basa en un teorema que no podemos probar dado el carácter avanzado de su demostración.

Teorema 12-7.—Sea $x_1, x_2, \dots, x_n$ una muestra aleatoria de tamaño n de una densidad $f(x; \theta_1, \theta_2, \dots, \theta_k)$ que satisface condiciones generales de regularidad, e imaginemos que Ω es k -dimensional. Supongamos que se desea docimar la hipótesis

$$H_0: \theta_1 = \theta_1^0; \theta_2 = \theta_2^0; \dots; \theta_t = \theta_t^0 \quad t < k$$

en donde $\theta_1^0, \theta_2^0, \dots, \theta_t^0$ representan números conocidos. Cuando H_0 es verdadera, $-2 \log \lambda$ se distribuye aproximadamente, si n es grande, como una ji cuadrado con t grados de libertad.

Puesto que $-2 \log \lambda$ crece al decrecer λ y tiende a infinito cuando λ tiende a cero, la región crítica para $-2 \log \lambda$ es la rama de la derecha de la distribución ji cuadrado. Por tanto, si se trata de una muestra grande y deseamos docimar una hipótesis nula, p. ej., con probabilidad 0,05 de error de tipo I, solo será necesario calcular $-2 \log \lambda$ y compararlo con el nivel 0,05 de la ji cuadrado; si $-2 \log \lambda$ excede de este nivel, H_0 se rechazará; en caso contrario, se aceptará.

12-7. Dósimas relativas a la media de una población normal.—Las ideas anteriores pueden ilustrarse mediante un problema práctico muy corriente: Se trata de docimar si la media de una población normal tiene un valor predeterminado. Supondremos una muestra de n observaciones, $x_1, x_2, \dots, x_n$, procedente de una

población normal de media μ y varianza σ^2 . Queremos docimar la hipótesis nula:

$$H_0: \mu = \mu_0 \quad 0 < \sigma^2 < \infty \quad H_a: \mu \neq \mu_0 \quad 0 < \sigma^2 < \infty$$

en donde μ_0 es un número dado. El espacio paramétrico Ω es el semiplano

$$\Omega = \{\mu, \sigma^2: -\infty < \mu < \infty; 0 < \sigma^2 < \infty\}$$

El subespacio ω , caracterizado por la hipótesis nula, es la recta vertical $\mu = \mu_0$; es decir,

$$\omega = \{\mu, \sigma^2: \mu = \mu_0; 0 < \sigma^2 < \infty\}$$

siendo μ_0 cierto número dado.

Docimaremos H_0 mediante la razón de verosimilitud. La verosimilitud es

$$L = \left(\frac{1}{\sqrt{2\pi\sigma}} \right)^n e^{-\frac{1}{2}\sum (x_i - \mu)/\sigma^2} \quad (1)$$

Hemos visto que los valores de μ y σ^2 que hacen máxima a L en Ω son

$$\begin{aligned} \hat{\mu} &= \frac{1}{n} \sum x_i = \bar{x} \\ \hat{\sigma}^2 &= \frac{1}{n} \sum (x_i - \bar{x})^2 \end{aligned} \quad (2)$$

Sustituyendo estos valores en L , resulta

$$L(\hat{\Omega}) = \left[\frac{1}{(2\pi/n)\sum (x_i - \bar{x})^2} \right]^{n/2} e^{-(n/2)} \quad (3)$$

Para hacer máximo L en ω , ponemos $\mu = \mu_0$, y queda como único parámetro σ^2 ; el valor de σ^2 que hace ahora máximo a L se encuentra fácilmente, y es

$$\hat{\sigma}^2 = \frac{1}{n} \sum (x_i - \mu_0)^2$$

que da

$$L(\hat{\omega}) = \left[\frac{1}{(2\pi/n)\sum (x_i - \mu_0)^2} \right]^{n/2} e^{-(n/2)} \quad (4)$$

La razón de [12-12] a [12-11] es la razón de verosimilitud:

$$\lambda = \left[\frac{\sum (x_i - \bar{x})^2}{\sum (x_i - \mu_0)^2} \right]^{n/2} \quad [12-13]$$

La siguiente etapa consiste en obtener la distribución de λ en la hipótesis H_0 y utilizar esta distribución para determinar un número A tal que la región crítica $0 < \lambda < A$ dé la probabilidad deseada, p. ej., 0,05, de rechazar H_0 cuando ésta es verdadera.

La distribución de λ es fácil de obtener en este caso. La suma de cuadrados que aparece en el denominador de [12-13] puede escribirse en la forma

$$\sum (x_i - \mu_0)^2 = \sum (x_i - \bar{x})^2 + n(\bar{x} - \mu_0)^2$$

de modo que λ puede expresarse así:

$$\lambda = \left\{ \frac{1}{1 + [n(\bar{x} - \mu_0)^2 / \sum (x_i - \bar{x})^2]} \right\}^{n/2} \quad [12-14]$$

Podemos recordar (Sec. 11-2) que la fracción del denominador es precisamente

$$\frac{t^2}{n-1}$$

en donde t tiene la distribución t con $n-1$ grados de libertad, cuando H_0 es cierta. Para obtener la distribución de λ basta transformar la distribución t mediante la sustitución

$$\lambda = \left\{ \frac{1}{1 + [t^2 / (n-1)]} \right\}^{n/2} \quad [12-15]$$

En realidad no es necesario obtener la distribución de λ , ya que se trata de una función monótona de t^2 y la dócima puede efectuarse igualmente tomando como criterio t^2 que tomando λ . Puesto que $t^2 = 0$ cuando $\lambda = 1$, y t^2 tiende a infinito cuando λ tiende a cero, una región crítica de la forma $0 < \lambda < A$ será equivalente a una región crítica $t^2 > B$, en donde B puede determinarse a partir de A , mediante la ecuación [12-15]. Los valores críticos de t son, por tanto, los valores extremos positivos o negativos, y la región crítica 0,05 para t está constituida por el par de intervalos.

$$t < -t_{0.05} \quad \text{y} \quad t > t_{0.05}$$

siendo $t_{0.05}$ el número para el cual

$$\int_{t_{0.05}}^{\infty} f(t; n-1) dt = 0,025 \quad [12-16]$$

en donde $f(t; n-1)$ representa la distribución t con $n-1$ grados de libertad. La d6cima de H_0 puede, por tanto, efectuarse del siguiente modo: se calcula la cantidad

$$\frac{\sqrt{n(n-1)}(\bar{x}-\mu_0)}{\sqrt{\sum(x_i-\bar{x})^2}} \quad [12-17]$$

Si 6sta est1 situada entre $-t_{0.05}$ y $t_{0.05}$, se acepta H_0 ; en otro caso, se rechaza H_0 .

Merece la pena observar la conexi3n que existe entre esta d6cima y la estimaci3n de intervalo de confianza de la media. Suponiendo que la media de la poblaci3n muestreada sea μ , un intervalo confidencial del 95% para μ ser1 precisamente el conjunto de valores de μ para los cuales

$$-t_{0.05} < \frac{\sqrt{n(n-1)}(\bar{x}-\mu)}{\sqrt{\sum(x_i-\bar{x})^2}} < t_{0.05} \quad [12-18]$$

Por tanto, la d6cima de H_0 equivale a la siguiente: Se construye un intervalo confidencial para la media de la poblaci3n. Si μ_0 est1 situado en ese intervalo, se acepta H_0 ; si μ_0 no est1 en dicho intervalo, se rechaza H_0 .

Podemos observar tambi3n que el teorema del final de la secci3n precedente da la distribuci3n correcta de λ para muestras grandes. Dado que

$$-2 \log \lambda = n \log \left(1 + \frac{t^2}{n-1} \right) \quad [12-19]$$

y puesto que el desarrollo en serie de $\log(1+x)$ es

$$\log(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \dots \quad \text{para } -1 < x < 1 \quad [12-20]$$

tenemos

$$-2 \log \lambda = \frac{n}{n-1} t^2 - \frac{n}{(n-1)^2} \frac{t^4}{2} + \frac{n}{(n-1)^3} \frac{t^6}{3} - \dots$$

para cualquier valor fijo de t por grande que sea, suponiendo que n se toma suficientemente grande para que $t^2/(n-1)$ sea menor que la uni-

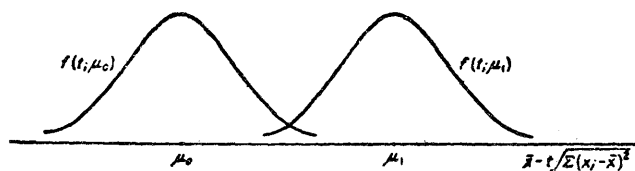


FIG. 62.

dad. El primer término de esta serie se reduce esencialmente a t^2 , y los demás tienden a cero al crecer n . Así, pues, para valores grandes de n , $-2 \log \lambda$ es, aproximadamente, t^2 . Además, t tiene una distribución que se aproxima a la normal para muestras grandes (Sec. 11-7) con media cero y varianza unidad, si la media verdadera es μ_0 ; por tanto, la distribución de t^2 se aproxima a la ji cuadrado con un grado de libertad. Esto está de acuerdo con el teorema, puesto que Ω es un plano y tiene $k=2$ dimensiones, mientras que ω es una línea y tiene $r=1$ dimensión.

Dócimas de una sola rama para la media.—La dócima que acabamos de construir recibe el nombre de dócima de dos ramas para la media, indicando así el hecho de que la región crítica consta de ambos extremos de la distribución t . Esta dócima no es uniformemente prepotente, y en realidad no existe una dócima de tal clase para la hipótesis nula dada. Si consideramos una sola de las alternativas a μ_0 , p. ej., $\mu = \mu_1$, en donde $\mu_1 > \mu_0$, las dos distribuciones t se representan en la figura 62. La mejor región crítica para t , dada una probabilidad de error del tipo I de 0,05, es evidentemente el intervalo $t > t_{0.10}$, que deja a un lado el 5% del área limitada por $f(t; \mu_0)$ sobre la rama de la derecha. Esta será la mejor región crítica para cualquier valor de μ superior a μ_0 . La potencia $P_1(\mu)$ de esta dócima se representa en la figura 63, junto con la potencia $P_2(\mu)$ de la dócima de dos ramas. La dócima de una rama es, sin duda, mejor que la de dos ramas para alternativas $\mu > \mu_0$, y es una dócima uniformemente prepotente para tales alternativas; pero para las $\mu < \mu_0$, la dócima de una rama no resulta conveniente; la potencia (probabilidad de rechazar μ_0 cuando μ es la media verdadera) tiende a cero cuando μ se aleja de μ_0 hacia la izquierda.

Existen muchas situaciones prácticas en que debe emplearse la dócima de una rama. Refiriéndonos de nuevo al ejemplo de las lámparas

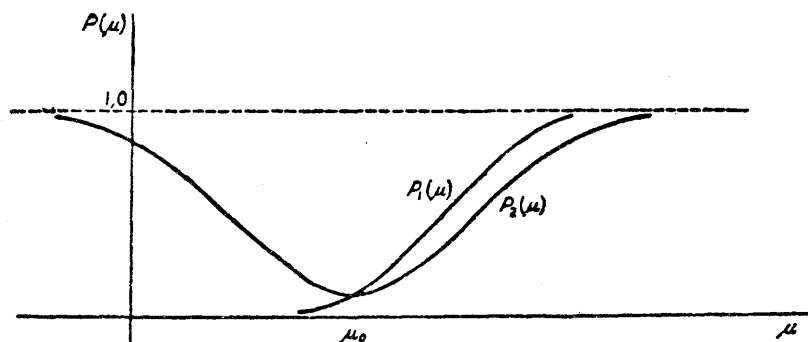


FIG. 63.

en que el proceso conocido de fabricación producía lámparas cuya vida media era de unas 1400 horas, cualquier proceso nuevo sólo interesaría si diese lugar a lámparas con una vida media mayor. Habría que docimar la hipótesis nula $\mu = 1400$, utilizando la dócima de una rama. Claro es que no habría perjuicio en aceptar $\mu = 1400$ cuando en realidad fuese μ inferior a 1400, ya que en cualquiera de estos dos casos dejaría de aplicarse el proceso propuesto. En otros problemas, la dócima de una rama sería la apropiada; p. ej., podría pensarse en un nuevo proceso

que redujese el coste de producción media por unidad; se docimaría la hipótesis nula según la cual el coste medio θ en el nuevo proceso fuese igual al coste medio θ_0 en el proceso conocido, contra la alternativa $\theta < \theta_0$; pero si se comparasen dos procesos propuestos a fin de elegir el mejor de ellos para ulteriores investigaciones y trabajos, la dócima apropiada sería la de dos ramas.

12-8. Diferencias entre medias de dos poblaciones normales.

En muchas situaciones es necesario comparar dos medias, siendo ambas desconocidas; en la sección anterior suponíamos conocida una de ellas. Si, p. ej., deseáramos comparar dos procesos propuestos en la fabricación de lámparas habría que basar la comparación en estimaciones de las medias de ambos procesos. Al confrontar la producción que se obtiene con una estirpe nueva de trigo híbrido con la de otra estirpe cono-

nueva de maíz híbrido con la de otra estirpe conocida, habrá que usar así mismo estimaciones de las dos producciones medias, por ser imposible conocer la producción media de la estirpe conocida para las condiciones climáticas en que habrá de desenvolverse la nueva estirpe. Es necesario comparar las dos estirpes plantándolas en la misma época y sobre el mismo tipo de terreno, obteniendo así estimaciones de las producciones medias de ambas en condiciones análogas. Claro es que todo ello da un carácter especial a la comparación; para que esta fuese completa, se requerirían dósimas extendidas a lo largo de cierto número de años y sobre una determinada variedad de tipos de terreno.

El problema general es el siguiente: Tenemos dos poblaciones normales, una de variante x_1 , con media μ_1 y varianza σ_1^2 y otra de variante x_2 , con media μ_2 y varianza σ_2^2 . Sobre la base de dos muestras procedentes cada una de dichas poblaciones hay que docimar la hipótesis nula:

$$H_0: \mu_1 = \mu_2 \quad \sigma_1^2 > 0, \sigma_2^2 > 0$$

$$H_a: \mu_1 \neq \mu_2 \quad \sigma_1^2 > 0, \sigma_2^2 > 0$$

Aquí el espacio paramétrico Ω es de cuatro dimensiones. La distribución conjunta de x_1 y x_2 queda especificada al asignar valores a las cuatro cantidades $(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2)$. El subespacio ω es tridimensional porque solo hay que especificar valores para las tres cantidades $(\mu_2, \sigma_1^2, \sigma_2^2)$, de forma que la distribución conjunta quede determinada por completo en la hipótesis de ser $\mu_1 = \mu_2 = \mu$.

Supondremos que hay m observaciones $(x_{11}, x_{12}, \dots, x_{1m})$ en la muestra de la primera población y n observaciones $(x_{21}, x_{22}, \dots, x_{2n})$ en la de la segunda. La verosimilitud es

$$L = \left(\frac{1}{2\pi\sigma_1^2} \right)^{\frac{m}{2}} e^{-\frac{1}{2} \sum_1^m \left(\frac{x_{1i} - \mu_1}{\sigma_1} \right)^2} \left(\frac{1}{2\pi\sigma_2^2} \right)^{\frac{n}{2}} e^{-\frac{1}{2} \sum_1^n \left(\frac{x_{2j} - \mu_2}{\sigma_2} \right)^2}$$

y su valor máximo en Ω se ve fácilmente que es

$$L(\hat{\Omega}) = \left[\frac{m}{2\pi \sum_1^m (x_{1i} - \bar{x}_1)^2} \right]^{m/2} \left[\frac{n}{2\pi \sum_1^n (x_{2j} - \bar{x}_2)^2} \right]^{n/2} e^{-(m/2)} e^{-(n/2)} \quad (2)$$

Si hacemos μ_1 y μ_2 iguales a μ , y tratamos de maximizar L respecto de μ , σ_1^2 y σ_2^2 , encontraremos que la estimación de μ viene dada como raíz de una ecuación cúbica y es una función muy com-

plicada de las observaciones. La razón de verosimilitud λ que resulta será, por tanto, una función complicada y calcular su distribución constituye una tarea realmente penosa que implica la razón de ambas varianzas. Esto hace que sea imposible determinar una región crítica $0 < \lambda < A$ para una probabilidad dada de un error de tipo I, ya que suele desconocerse la razón de las varianzas poblacionales. Para orillar esta dificultad cabe emplear varios artificios; pero no diremos más acerca de este problema. Claro es que el criterio siguiente puede utilizarse para muestras grandes. La raíz de la ecuación cúbica se calcula por métodos numéricos en un ejemplo dado, y se obtiene después el valor de λ . La distribución de $-2 \log \lambda$ se aproximará a la ji cuadrado con un grado de libertad.

Si es posible suponer que las dos poblaciones tienen la misma varianza, el problema se simplifica relativamente. El espacio paramétrico Ω es el tridimensional de coordenadas (μ_1, μ_2, σ^2) , mientras que ω , para la hipótesis nula $\mu_1 = \mu_2$, es bidimensional de coordenadas (σ^2, μ) . En Ω hallamos

$$\hat{\mu}_1 = \bar{x}_1 \quad \hat{\mu}_2 = \bar{x}_2$$

$$\hat{\sigma}^2 = \frac{1}{m+n} \left[\sum_1^m (x_{1i} - \bar{x}_1)^2 + \sum_1^n (x_{2j} - \bar{x}_2)^2 \right]$$

de modo que

$$L(\hat{\Omega}) = \left\{ \frac{m+n}{2\pi[\Sigma(x_{1i} - \bar{x}_1)^2 + \Sigma(x_{2j} - \bar{x}_2)^2]} \right\}^{(m+n)/2} e^{-[(m+n)/2]} \quad (3)$$

En ω

$$\hat{\mu} = \frac{1}{m+n} \left(\sum_1^m x_{1i} + \sum_1^n x_{2j} \right) = \frac{m\bar{x}_1 + n\bar{x}_2}{m+n}$$

$$\begin{aligned} \hat{\sigma}^2 &= \frac{1}{m+n} \left[\sum (x_{1i} - \hat{\mu})^2 + \sum (x_{2j} - \hat{\mu})^2 \right] \\ &= \frac{1}{m+n} \left[\sum (x_{1i} - \bar{x}_1)^2 + \sum (x_{2j} - \bar{x}_2)^2 + \frac{mn}{m+n} (\bar{x}_1 - \bar{x}_2)^2 \right] \end{aligned}$$

lo que da

$$\begin{aligned} L(\hat{\omega}) &= \\ &= \left\{ \frac{m+n}{2\pi \left[\sum (x_{1i} - \bar{x}_1)^2 + \sum (x_{2j} - \bar{x}_2)^2 + \frac{mn}{m+n} (\bar{x}_1 - \bar{x}_2)^2 \right]} \right\}^{(m+n)/2} e^{-[(m+n)/2]} \quad (4) \end{aligned}$$

y, finalmente,

$$\lambda = \left[\frac{1}{1 + \frac{mn}{m+n} \frac{(\bar{x}_1 - \bar{x}_2)^2}{\Sigma(x_{1i} - \bar{x}_1)^2 + \Sigma(x_{2j} - \bar{x}_2)^2}} \right]^{(m+n)/2} \quad (5)$$

La última expresión se parece mucho a su homóloga de la sección anterior; esta dócima puede efectuarse también en función de una cantidad que tiene la distribución t . Sabemos que $\bar{x}_1$ y $\bar{x}_2$ tienen distribuciones normales independientes, con medias μ_1 y μ_2 y varianzas σ^2/m y σ^2/n . Con referencia al problema 25 del capítulo 10 se ve fácilmente que la distribución de $u = \bar{x}_1 - \bar{x}_2$ es normal, con media $\mu_1 - \mu_2$ y varianza $\sigma^2[(1/m) + (1/n)]$. En la hipótesis nula, la media de u será igual a cero. Las cantidades $\Sigma(x_{1i} - \bar{x}_1)^2/\sigma^2$ y $\Sigma(x_{2j} - \bar{x}_2)^2/\sigma^2$ tienen distribuciones independientes ji cuadrado con $m-1$ y $n-1$ grados de libertad, respectivamente. Por tanto, su suma v tiene una distribución ji cuadrado con $m+n-2$ grados de libertad. Puesto que en la hipótesis nula

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sigma \sqrt{(1/m) + (1/n)}}$$

se distribuye normalmente, con media cero y varianza unidad, la cantidad

$$\begin{aligned} t &= \frac{z}{\sqrt{v/(m+n-2)}} \\ &= \frac{\sqrt{mn/(m+n)} (\bar{x}_1 - \bar{x}_2)}{\sqrt{[\Sigma(x_{1i} - \bar{x}_1)^2 + \Sigma(x_{2j} - \bar{x}_2)^2]/(m+n-2)}} \end{aligned} \quad (6)$$

tiene una distribución t con $m+n-2$ grados de libertad. La razón de verosimilitud es

$$\lambda = \left\{ \frac{1}{1 + [t^2/(m+n-2)]} \right\}^{(m+n)/2} \quad (7)$$

y su distribución está determinada por la distribución t . Claro es que esta dócima se efectuaría basándose en t más bien que en λ . Una región crítica del 5% para t es $t^2 > (t_{0.025})^2$. Si deseamos doci-
mar $H_0: \mu_1 = \mu_2$ con la alternativa $H_a: \mu_1 < \mu_2$ y con $\alpha = 0.05$ (p. ej., si la primera población se refiere al rendimiento de una variedad de maíz de uso corriente, mientras la segunda se refiere al rendimiento de otra variedad propuesta para reemplazar aquella), la región crítica será $t < -t_{0.05}$.

Observemos que es posible determinar un intervalo confidencial para la diferencia $\mu_1 - \mu_2$ de las medias poblacionales, mediante el uso de la distribución t . La cantidad

$$y = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sigma \sqrt{(1/m) + (1/n)}}$$

se distribuye normalmente, con media cero y varianza unidad, de modo que

$$t = \frac{y}{\sqrt{v/(m+n-2)}}$$

tiene una distribución t con $m+n-2$ grados de libertad. Como en t no interviene σ^2 , sino solo el parámetro $\theta = \mu_1 - \mu_2$, puede obtenerse un intervalo confidencial para θ . Los límites superior e inferior para un intervalo del 90%, p. ej., se obtendrían resolviendo respecto a θ las ecuaciones

$$t = \pm t_{0,05}$$

12.9. Dótimas de la varianza de una distribución normal.

Para docimar la hipótesis nula de que la varianza de una población normal toma un valor determinado σ_0^2 , basándose en una muestra de tamaño n (es decir, $H_0: \sigma^2 = \sigma_0^2$ con la alternativa $H_a: \sigma^2 \neq \sigma_0^2$), empezamos por hacer máxima

$$L = \left(\frac{1}{2\pi\sigma^2} \right)^{n/2} e^{-\frac{1}{2}\Sigma[(x_i - \mu)/\sigma]^2} \quad (1)$$

en Ω , que es un espacio bidimensional, y en ω , que es la recta $\sigma^2 = \sigma_0^2$. El cociente de estos máximos se obtiene fácilmente y es

$$\lambda = \left(\frac{u}{n} \right)^{n/2} e^{-\frac{1}{2}u(n-n)} \quad (2)$$

en donde

$$u = \sum_{i=1}^n \frac{(x_i - \bar{x})^2}{\sigma_0^2} \quad (3)$$

Como se sabe que u tiene una distribución ji cuadrado con $n-1$ grados de libertad, se obtiene la distribución de λ transformando la distribución ji cuadrado mediante (2). Sin embargo, puede efectuarse esta dócima utilizando como criterio u . Al representar la ecuación (2) (Fig. 12-9) se ve que una región crítica para λ , de la forma $0 < \lambda < A$, corresponde al par de intervalos $0 < u < a$ y

$b < u < \infty$ para u , siendo a y b tales que las ordenadas de (2) sean iguales.

También aquí, si $H_a: \sigma^2 < \sigma_0^2$, o bien $H_a: \sigma^2 > \sigma_0^2$, se emplearían las dócimas de una sola rama, $u < \chi_{0.95}^2$ o $u > \chi_{0.05}^2$.

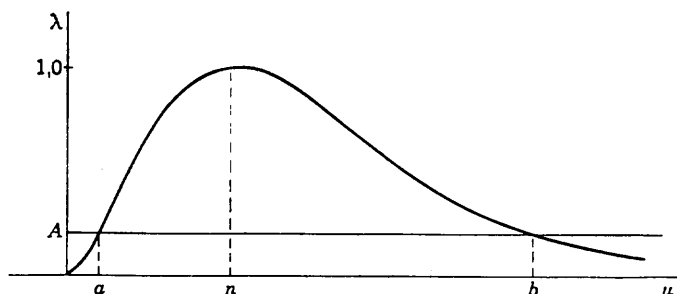


FIG. 12-9.

Igualdad de dos varianzas.—Dadas muestras procedentes de dos poblaciones normales, de medias y varianzas (μ_1, σ_1^2) y (μ_2, σ_2^2) , respectivamente, podemos docimar

$$H_0: \sigma_1^2 = \sigma_2^2 \quad H_a: \sigma_1^2 \neq \sigma_2^2 \quad (4)$$

La razón de verosimilitud es igual a

$$\lambda = \frac{\left[\frac{m+n}{2\pi[\Sigma(x_{1i} - \bar{x}_1)^2 + \Sigma(x_{2j} - \bar{x}_2)^2]} \right]^{(m+n)/2}}{\left[\frac{m}{2\pi\Sigma(x_{1i} - \bar{x}_1)^2} \right]^{m/2} \left[\frac{n}{2\pi\Sigma(x_{2j} - \bar{x}_2)^2} \right]^{n/2}} \quad (5)$$

en donde la notación es la misma que en la sección anterior. Este criterio puede escribirse así:

$$\lambda = \frac{(m+n)^{(m+n)/2}}{m^{m/2} n^{n/2}} \frac{\left(\frac{m-1}{n-1} F \right)^{m/2}}{\left(1 + \frac{m-1}{n-1} F \right)^{(m+n)/2}} \quad (6)$$

siendo F la razón de varianzas,

$$F = \frac{(n-1)\Sigma(x_{1i} - \bar{x}_1)^2}{(m-1)\Sigma(x_{2j} - \bar{x}_2)^2} \quad (7)$$

la cual tiene la distribución F con $m-1$ y $n-1$ grados de libertad cuando H_0 es cierta. Representando λ como función de F , es evi-

dente que la región crítica $0 < \lambda < A$ corresponde a una dócima de dos ramas relativas a F . Es costumbre hacer que ambas ramas tengan áreas iguales (aunqu ello no proporciona la dócima mejor de todas) porque las tabulaciones de F facilitan la determinación de esta región. De nuevo, si $H_a: \sigma_1^2 < \sigma_2^2$ o $H_a: \sigma_2^2 < \sigma_1^2$, resultan apropiadas dócimas de una sola rama.

12-10. Dócima de la bondad del ajuste.—Si una población tiene la distribución polinomial

$$f(x_i; p_i) = \prod_{i=1}^k p_i^{x_i} \quad x_i = 0, 1; \quad \sum x_i = 1; \quad \sum p_i = 1 \quad (1)$$

como sucede en el muestreo con reemplazamiento para una población de individuos que puedan clasificarse en k clases, se plantea corrientemente el problema de docimar si las probabilidades tienen valores numéricos determinados. Así, el resultado del lanzamiento de un dado suele clasificarse en seis clases, y puede interesarnos docimar, en una muestra de observaciones, si el dado está bien construido, o sea, si

$$p_i = 1/6 \quad i = 1, 2, \dots, 6$$

Supongamos que se hacen n observaciones de una población cuya distribución sea (1), y que el número de observaciones correspondientes a la clase i -ésima es n_i ($\sum n_i = n$). La verosimilitud de la muestra es

$$L = \prod_{i=1}^k p_i^{n_i} \quad (2)$$

y vamos a docimar la hipótesis nula:

$$H_0: p_i = p_{0i}$$

donde los p_{0i} son números dados. El espacio paramétrico Ω tiene $k-1$ dimensiones (dadas $k-1$ de las p_i , la restante se determina por la relación $\sum p_i = 1$), siendo ω un punto. Se halla fácilmente que L es máxima en Ω cuando

$$\hat{p}_i = \frac{n_i}{n} \quad (3)$$

Por tanto,

$$L(\hat{\Omega}) = \frac{1}{n^n} \prod_{i=1}^k n_i^{n_i} \quad (4)$$

El valor máximo de L en ω coincide con su valor único,

$$L(\hat{\omega}) = \prod_1^k p_{0i}^{n_i} \quad (5)$$

La razón de verosimilitud es

$$\lambda = n^n \prod_1^k \left(\frac{p_{0i}}{n_i} \right)^{n_i} \quad (6)$$

y la región crítica es $0 < \lambda < A$, elegido A de modo que dé la probabilidad deseada para un error de tipo I. Para valores pequeños de n , se tabula directamente la distribución de λ con objeto de determinar A . Para valores grandes de n , podemos basarnos en el hecho de que $-2 \log \lambda$ tiene aproximadamente la distribución ji cuadrado con $k-1$ grados de libertad. La aproximación de la ji cuadrado resulta sorprendentemente buena, aun en el caso de que n sea pequeño, si se supone $k > 2$.

Otra dócima corrientemente utilizada para H_0 fue propuesta por Karl Pearson antes que se desarrollase la teoría general de la docimasia de hipótesis. El criterio de esta dócima es

$$u = \sum \frac{(n_i - np_{0i})^2}{np_{0i}} \quad (7)$$

la cual, para muestras grandes, tiene aproximadamente la distribución ji cuadrado con $k-1$ grados de libertad, si H_0 es verdadera. El razonamiento para utilizar (7) como criterio es, en pocas palabras, el siguiente: en muestras grandes la distribución aproximada de las $\hat{p}_i = n_i/n$ ($i=1, 2, \dots, k-1$) es normal, y su expresión es

$$f(\hat{p}_1, \hat{p}_2, \dots, \hat{p}_{k-1}) = \left(\frac{n}{2\pi} \right)^{\frac{k-1}{2}} \sqrt{\frac{1}{\prod_1^k p_i}} e^{-\frac{1}{2} \sum \sum n \left(\frac{\delta_{ij}}{p_i} + \frac{1}{p_k} \right) (\hat{p}_i - p_i)(\hat{p}_j - p_j)} \quad (8)$$

según se deduce de la ecuación (10-8-18) al sustituir k por $k-1$. Hemos visto en el capítulo 10 que la forma cuadrática de una distribución normal multivariante con $k-1$ variantes admite una distribución ji cuadrado con $k-1$ grados de libertad; por tanto,

$$v = \sum_1^{k-1} \sum_1^{k-1} n \left(\frac{\delta_{ij}}{p_i} + \frac{1}{p_k} \right) (\hat{p}_i - p_i)(\hat{p}_j - p_j) \quad (9)$$

tiene aproximadamente esta distribución para muestras grandes. Sumando esta expresión respecto a j y recordando que

$$p_k = 1 - \sum_{i=1}^{k-1} p_i$$

hallamos

$$v = \sum_{i=1}^k \frac{n(\hat{p}_i - p_i)^2}{p_i} \quad (10)$$

o bien

$$v = \sum_{i=1}^k \frac{(n_i - np_i)^2}{np_i} \quad (11)$$

que coincide con (7) si los valores verdaderos de las p_i son p_{0i} . Pero supongamos que las p_i verdaderas son p_{1i} , de las cuales algunas al menos son diferentes de las p_{0i} ; entonces

$$v = \sum \frac{(n_i - np_{1i})^2}{np_{1i}} \quad (12)$$

tiene aproximadamente la distribución ji cuadrado, con valor esperado $k-1$. Se ve fácilmente que el valor esperado de la cantidad

$$u = \sum \frac{(n_i - np_{0i})^2}{np_{0i}} \quad (13)$$

es

$$E(u) = \sum \frac{1}{np_{0i}} [np_{1i}(1 - p_{1i}) + n^2(p_{1i} - p_{0i})^2] \quad (14)$$

el cual es con seguridad superior a $k-1$ para valores de n suficientemente grandes, e incluso superior a $k-1$ cualquiera que sea n , ya que si se hace mínima $E(u)$ respecto a las p_{0i} , se ve que el mínimo ocurre para $p_{0i} = p_{1i}$, y es, por tanto, $k-1$. Resulta ahora evidente la razón del uso de u como criterio docimásico. Si las p_i verdaderas son p_{0i} , u tendrá aproximadamente la distribución ji cuadrado, mientras que si las p_i verdaderas no son p_{0i} , la distribución de u tendrá un valor medio superior, que tiende a infinito al crecer n . Por tanto, en la docimasia de H_0 es razonable la utilización de u como criterio, y la rama derecha de la distribución como región crítica.

Hemos estudiado el criterio χ^2 de Pearson debido a interés histórico y por ser todavía de uso corriente en la docimasia de H_0 .

En realidad, equivale a la dócima de máxima verosimilitud en muestras grandes. Para comprobar que así sucede, tal vez lo más sencillo sea escribir λ en la siguiente forma:

$$\lambda = K \frac{n!}{\prod n_i!} \prod p_{0_i}^{n_i}$$

en donde

$$K = \frac{n^n}{\prod n_i^n} \frac{\prod n_j!}{n!}$$

Si se cambian las variantes de (8), poniendo n_i en lugar de $\hat{p}_i$, la función no varía, excepto en lo que se refiere al factor $n^{(k-1)/2}$, ya que $n d\hat{p}_i = dn_i$. Se deduce de la sección 10-8 que $\lambda n^{k-1}/K$ tiende a (8). Utilizando la fórmula de Stirling (Cap. 2) para las factoriales de K , se demuestra que K/n^{k-1} reduce el coeficiente de la exponencial de (8) a términos de orden $1/\sqrt{n}$. Por tanto, $-2 \log \lambda$ es asintóticamente equivalente a u .

12-11. Dócimas de independencia en tablas de contingencia.

Una tabla de contingencia es una clasificación múltiple. Así, en un análisis de la opinión pública, los individuos entrevistados pueden clasificarse en relación con su actitud respecto a determinada cuestión política y de acuerdo también con su sexo, obteniendo una tabla de la forma siguiente:

	A favor	En contra	Indecisos
Hombres	1154	475	243
Mujeres	1083	442	362

Esta es una tabla de contingencia 2×3 . Los individuos se han clasificado según dos criterios, uno con dos categorías y otro con tres. Las seis clasificaciones distintas reciben el nombre de *casillas*. Se hubiera obtenido una tabla de contingencia triple si los individuos se hubiesen clasificado además según un tercer criterio, p. ej., el grupo de renta anual. Si existieran cinco grupos de renta (p. ej., menos de \$1000, \$1000 a \$3000, ...), la tabla de contingencia se denominaría tabla $2 \times 3 \times 5$ y habría 30 casillas donde colocar a una persona cualquiera. A menudo, resulta conveniente considerar las casillas como cubos en un bloque de dimensiones 2, 3 y 5 unidades de anchura, longitud y altura, respectivamente. Si, además, los individuos se clasificasen según ocho regiones geográficas, se ob-

tendría una tabla de contingencia cuádruple $2 \times 3 \times 5 \times 8$, con 240 casillas en un bloque cuatridimensional de aristas respectivas 2, 3, 5 y 8 unidades. La tabla de contingencia proporciona una técnica para la investigación de relaciones cuya existencia se sospecha. Supongamos que se presume que la reacción de hombres y mujeres a determinada propuesta política será distinta; en este caso, se construye una tabla como la anterior, docimando la hipótesis nula, según la cual las actitudes resultarían independientes del sexo; otro ejemplo sería el de un genético que sospechase la herencia de la propensión a determinada enfermedad; clasificaría una muestra de individuos de acuerdo con los siguientes criterios: 1) según que tengan o no la enfermedad; 2) según que la tengan o no sus padres; 3) según que sus madres sufran o no dicha enfermedad. En la tabla de contingencia $2 \times 2 \times 2$ docimaría la hipótesis nula, según la cual la clasificación 1) es independiente de 2) y 3); o bien, supongamos un investigador médico que sospechase que determinado medio favorece una enfermedad dada, y clasificase los individuos: 1) según que tengan o no tengan esta enfermedad, y 2) según que estén o no sometidos a dichas condiciones. Un ingeniero industrial emplearía una tabla de contingencia para descubrir si dos tipos de defectos de determinado producto se deben a una misma causa o a causas diferentes. Es evidente que esta técnica constituye un instrumento muy útil en cualquier campo de investigación.

Tablas de contingencia doble.—Supondremos n individuos o conceptos clasificados según dos criterios A y B , de modo que hay r clasificaciones $A_1, A_2, \dots, A_r$ en A y s clasificaciones $B_1, B_2, \dots, B_s$ en B , y que el número de individuos pertenecientes a A_i y B_j es n_{ij} . Tenemos así una tabla de contingencia $r \times s$ con frecuencias abso-

	B_1	B_2	B_3	$\dots$	B_s
A_1	n_{11}	n_{12}	n_{13}	$\dots$	n_{1s}
A_2	n_{21}	n_{22}	n_{23}	$\dots$	n_{2s}
A_3	n_{31}	n_{32}	n_{33}	$\dots$	n_{3s}
$\vdots$					
A_r	n_{r1}	n_{r2}	n_{r3}	$\dots$	n_{rs}

(1)

lutas n_{ij} en las casillas, siendo $\sum n_{ij} = n$. Utilizaremos además la notación $n_{i.}$ para representar los totales de fila y $n_{.j}$ para designar los totales de columna:

$$n_{i.} = \sum_j n_{ij} \quad n_{.j} = \sum_i n_{ij}$$

Es obvio que

$$\sum_i n_{i.} = \sum_j n_{.j} = n$$

Construiremos ahora un modelo probabilístico para el problema que deseamos estudiar. Consideremos los n individuos como una muestra de tamaño n de una población polinomial, con probabilidades p_{ij} ($i=1, 2, \dots, r$; $j=1, 2, \dots, s$). La distribución probabilística para una sola observación (Sec. 10-8) es

$$f(x_{11}, x_{12}, \dots, x_{rs}) = \prod_{ij} p_{ij}^{x_{ij}} \quad x_{ij} = 0, 1; \quad \sum_{i,j} x_{ij} = 1 \quad (2)$$

Queremos docimar la hipótesis nula de que las clasificaciones A y B son independientes, esto es que la probabilidad de que un individuo caiga en B_j no está afectada por la clase A a la cual pertenece dicho individuo. Utilizando el simbolismo del capítulo 2, escribiremos

$$P(B_j|A_i) = P(B_j) \quad P(A_i|B_j) = P(A_i)$$

o bien

$$P(A_i, B_j) = P(A_i)P(B_j)$$

Si designamos las probabilidades marginales $P(A_i)$ por p_i ($i=1, 2, \dots, r$) y las probabilidades marginales $P(B_j)$ por q_j , la hipótesis nula es

$$H_0: p_{ij} = p_i q_j \quad \sum p_i = 1, \sum q_j = 1 \quad (3)$$

Cuando la hipótesis nula no es cierta, se dice que hay *interacción* entre ambos criterios de clasificación.

El espacio paramétrico completo Ω para la distribución (1) tiene $rs - 1$ dimensiones (dadas todas las p_{ij} menos una, la restante viene determinada por $\sum_{ij} p_{ij} = 1$), mientras que en la hipótesis H_0 tenemos

un espacio paramétrico ω de $r - 1 + s - 1$ dimensiones. La verosimilitud correspondiente a una muestra de tamaño n es

$$L = \prod_{ij} p_{ij}^{n_{ij}} \quad (4)$$

y el máximo en Ω se presenta para

$$\hat{p}_{ij} = \frac{n_{ij}}{n} \quad (5)$$

En ω ,

$$L = \prod_{ij} (p_i q_j)^{n_{ij}} = \left(\prod_i p_i^{n_{i.}} \right) \left(\prod_j q_j^{n_{.j}} \right) \quad (6)$$

y su máximo corresponde a

$$\hat{p}_i = \frac{n_{i.}}{n} \quad \hat{q}_j = \frac{n_{.j}}{n} \quad (7)$$

La razón de verosimilitud será, por tanto,

$$\lambda = \frac{\left(\prod_i n_{i.}^{n_{i.}} \right) \left(\prod_j n_{.j}^{n_{.j}} \right)}{n^n \prod_{i,j} n_{ij}^{n_{ij}}} \quad (8)$$

La distribución de λ en la hipótesis nula no es única, porque la hipótesis es compuesta y la distribución exacta de λ comprende los parámetros desconocidos p_i y q_j . Sin embargo, para muestras grandes, tenemos una dócima, porque en tal caso $-2 \log \lambda$ tiene una distribución próxima a la ji cuadrado con

$$rs - 1 - (r + s - 2) = (r - 1)(s - 1)$$

grados de libertad; basándose en esta distribución queda determinada una región crítica única para λ .

Si se trata de hallar una dócima que pueda utilizarse cuando la muestra no es grande, cabe preguntarse por qué el criterio de una dócima da lugar a una distribución única para muestras grandes, siendo así que la distribución depende de parámetros desconocidos, que toman valores cualesquiera dentro de ciertos campos de variabilidad. La respuesta es que estos parámetros no son realmente desconocidos; pueden estimarse, y, al aumentar el tamaño de la muestra, sus estimaciones tienden a los valores verdaderos. En el límite, cuando n tiende a infinito, se conocen exactamente los valores de los parámetros, y entonces la distribución de λ llega a ser única. Es única porque se elige como valor verdadero del parámetro el correspondiente a un punto determinado de ω , de

modo que se da a las n_{ij} una distribución única, y la distribución de λ viene determinada por aquella.

Parece razonable emplear un procedimiento análogo en la obtención de una dócima para muestras pequeñas, y definir así una distribución para λ utilizando las estimaciones de los parámetros desconocidos. En el problema que ahora consideramos, dado que las estimaciones de las p_i y q_j vienen dadas por (7), bastaría sustituir estos valores en la función de distribución de las n_{ij} y emplear esta distribución para obtener la de λ . Pero con ello no resolveríamos por completo la cuestión; la región crítica dependería de los totales marginales $n_{i.}$ y $n_{.j}$; por tanto, la probabilidad de un error de tipo I variará de muestra a muestra para cualquier región crítica determinada, $0 < \lambda < A$.

Esta dificultad puede resolverse por un procedimiento que merece un estudio detallado, no solo por su propio interés, sino por el que este problema presenta en estadística aplicada. Para abreviar, designemos por $f(n_{ij})$ la función de densidad o de cuantía conjunta de todas las n_{ij} ; por $g(n_{i.}, n_{.j})$ la marginal de todas las $n_{i.}$ y $n_{.j}$, y por

$$f(n_{ij}|n_{i.}, n_{.j}) = \frac{f(n_{ij})}{g(n_{i.}, n_{.j})}$$

la condicional de las n_{ij} , dados los totales marginales. En la hipótesis nula, esta distribución condicional es independiente de los parámetros desconocidos (como inmediatamente veremos); los estimadores $n_{i.}/n$ y $n_{.j}/n$ constituyen un conjunto suficiente para p_i y q_j . Esto nos permite la construcción de la dócima.

La distribución conjunta de las n_{ij} se reduce a la polinomial

$$f(n_{11}, n_{12}, \dots, n_{rs}) = \frac{n!}{\prod_{ij} n_{ij}!} \prod_{ij} p_{ij}^{n_{ij}} \quad (9)$$

en Ω , y en ω (nos interesa la distribución de λ en la hipótesis H_0), esta da lugar a

$$f(n_{11}, n_{12}, \dots, n_{rs}) = \frac{n!}{\prod_{ij} n_{ij}!} \left(\prod_i p_i^{n_{i.}} \right) \left(\prod_j q_j^{n_{.j}} \right) \quad (10)$$

Para obtener la distribución condicional deseada, empezaremos por

hallar la distribución de las $n_{i.}$ y $n_{.j}$, lo que se consigue sumando (10) para todos los conjuntos de n_{ij} tales que

$$\sum_i n_{ij} = n_{.j} \quad \sum_j n_{ij} = n_{i.} \quad (11)$$

Para totales marginales determinados, en la suma solo interviene el factor $1/\Pi n_{ij}!$ de (10), de manera que tenemos que sumar dicho factor para todas las n_{ij} sujetas a la condición (11). La suma deseada se obtiene comparando los coeficientes de $\Pi x_i^{n_{i.}}$ en la expresión

$$(x_1 + \dots + x_r)^{n_{.1}} (x_1 + \dots + x_r)^{n_{.2}} \dots (x_1 + \dots + x_r)^{n_{.r}} = (x_1 + \dots + x_r)^n \quad (12)$$

En el segundo miembro, el coeficiente de $\Pi x_i^{n_{i.}}$ es

$$\frac{n!}{\prod_i n_{i.}!} \quad (13)$$

En el primero hay términos con coeficientes de la forma

$$\frac{n_{.1}!}{\prod_i n_{i1}!} \frac{n_{.2}!}{\prod_i n_{i2}!} \dots \frac{n_{.r}!}{\prod_i n_{ir}!} = \frac{\prod_j n_{.j}!}{\prod_{ij} n_{ij}!} \quad (14)$$

donde n_{ij} es el exponente de x_i en la polinomial j -ésima. En esta expresión, las n_{ij} satisfacen las condiciones (11). La primera condición se satisface en virtud del teorema polinomial, mientras que se cumple la segunda porque exigimos que la potencia de x_i en estos términos sea $n_{i.}$. La suma de todos estos coeficientes (14) debe ser igual a (13); por tanto, escribiremos

$$\sum \frac{1}{\Pi n_{ij}!} = \frac{n!}{\prod_i n_{i.}! \prod_j n_{.j}!} \quad (15)$$

Esta es precisamente la suma que necesitamos, ya que evidentemente existe un coeficiente y solo uno de la forma de (14) en el primer miembro de (12) para toda tabla de contingencia (1) con totales marginales dados. La distribución de las $n_{i.}$ y $n_{.j}$ es, por tanto,

$$g(n_{i.}, n_{.j}) = \frac{(n!)^2}{(\Pi n_{i.}!)(\Pi n_{.j}!)} \left(\prod_i p_i^{n_{i.}} \right) \left(\prod_j q_j^{n_{.j}} \right) \quad (16)$$

lo que incidentalmente muestra que las $n_{i.}$ tienen distribuciones independientes de las $n_{.j}$, resultado inesperado, ya que $n_{1.}$ y $n_{.1}$, p. ej., tienen en común la variante n_{11} .

La distribución condicional de las n_{ij} , dados los totales marginales, se obtiene dividiendo (10) por (16), lo que da

$$f(n_{11}, n_{12}, \dots, n_{rs} | n_{1.}, n_{2.}, \dots, n_{.s}) = \frac{(\prod n_{i.}!)(\prod n_{.j}!)}{n! \prod n_{ij}!} \quad (17)$$

que, afortunadamente, no comprende parámetros desconocidos e indica que los estimadores son suficientes.

Para ver cómo puede construirse una dócima, consideremos la situación general en que un criterio λ para una dócima determinada tiene una distribución $u(\lambda; \theta)$ que comprende un parámetro desconocido θ . Si θ posee un estimador suficiente $\hat{\theta}$, la distribución conjunta de λ y $\hat{\theta}$ vendrá dada por

$$v(\lambda, \hat{\theta}; \theta) = v_1(\lambda | \hat{\theta}) v_2(\hat{\theta}; \theta) \quad (18)$$

y en la distribución condicional de λ , dado $\hat{\theta}$, no intervendrá θ . Si usamos la distribución condicional, podremos hallar un número $A(\hat{\theta})$ para todo $\hat{\theta}$ tal que, p. ej.,

$$\int_0^{A(\hat{\theta})} v_1(\lambda | \hat{\theta}) d\lambda = 0,05$$

En el plano $\lambda\hat{\theta}$, la curva $\lambda = A(\hat{\theta})$ determina, juntamente con la recta $\lambda = 0$, una región R . (Véase Fig. 12-10.) La probabilidad de que una muestra dé lugar a un par de valores $(\lambda, \hat{\theta})$ que correspondan a un punto de R , es exactamente 0,05, porque

$$\begin{aligned} P[(\lambda, \hat{\theta}) \text{ en } R] &= \int_{-\infty}^{\infty} \int_0^{A(\hat{\theta})} v(\lambda, \hat{\theta}; \theta) d\lambda d\hat{\theta} \\ &= \int_{-\infty}^{\infty} \left[\int_0^{A(\hat{\theta})} v_1(\lambda | \hat{\theta}) d\lambda \right] v_2(\hat{\theta}; \theta) d\hat{\theta} \quad (19) \\ &= \int_{-\infty}^{\infty} 0,05 v_2(\hat{\theta}; \theta) d\hat{\theta} \\ &= 0,05 \end{aligned}$$

Por tanto, docimaremos la hipótesis usando $\hat{\theta}$ en unión de λ . La región crítica es una región del plano, en vez del intervalo.

$0 < \lambda < A$, y una región tal que cualquiera que sea el valor desconocido de θ , el error del tipo I tiene una probabilidad predeterminada. En cualquier situación dada, la dócima constituye en realidad una dócima condicional; observamos $\hat{\theta}$ y docimamos λ mediante un intervalo $0 < \lambda < A(\hat{\theta})$, utilizando la distribución condicional de λ , dado $\hat{\theta}$. Obsérvese que esto es lícito solo si θ tiene un estadístico suficiente.

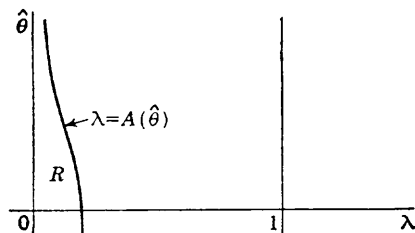


FIG. 12-10.

Evidentemente, la técnica anterior es aplicable cuando θ es un conjunto de parámetros en vez de un parámetro único, y tiene un conjunto de estadísticos suficientes. En particular, esta técnica puede emplearse en la docimasia del criterio (8) para la hipótesis nula

de una tabla de contingencia doble. Nos limitamos a hacer uso de la distribución condicional (17), determinando un intervalo $0 < \lambda < A(n_{i.}; n_{.j})$ que tenga la probabilidad deseada para un error de tipo I para los totales marginales observados. En las aplicaciones de esta dócima nos encontramos con que es necesario realizar un cálculo extremadamente penoso para determinar la distribución de λ , a menos que r , s y los totales marginales sean muy pequeños. No obstante, se demuestra que la aproximación correspondiente a muestras grandes puede utilizarse sin error apreciable, excepto en el caso en que tanto r como s sean iguales a 2. Para este caso se han establecido otras simplificaciones aproximadas (véase, p. ej., Fisher y Yates, *Tables for Statisticians and Biometricians*, Oliver & Boyd, Ltd., Edimburgo y Londres, 1938); pero no seguiremos ocupándonos de esta cuestión.

Si la distribución (17) se sustituye por su aproximación normal multivariante, puede demostrarse que el criterio

$$u = \sum_{i,j} \frac{[n_{ij} - (n_{i.}n_{.j}/n)]^2}{n_{i.}n_{.j}/n} \quad (20)$$

tiene aproximadamente la distribución ji cuadrado con $(r-1)(s-1)$ grados de libertad, y constituye un criterio razonable para la docimasia de H_0 de (3). Este criterio fue propuesto por primera vez (por Karl Pearson) para la docimasia de esta hipótesis, que difiere de $-2 \log \lambda$ en términos de orden $1/\sqrt{n}$. Por consiguiente, ambos criterios son en esencia equivalentes, a menos que n sea pequeño.

El razonamiento para probar que u es un criterio razonable es totalmente análogo al empleado en la sección anterior para justificar (12-10-7).

Tablas de contingencia triple.—Si los elementos de una población se pueden clasificar según tres criterios A , B , C , con clasificaciones

$$A_i (i=1, 2, \dots, s_1), B_j (j=1, 2, \dots, s_2) \text{ y } C_k (k=1, 2, \dots, s_3)$$

una muestra de n individuos podrá clasificarse a su vez en una tabla de contingencia triple $s_1 \times s_2 \times s_3$. Representaremos por p_{ijk} las probabilidades asociadas con las casillas individuales; por n_{ijk} , los números de elementos muestrales en dichas casillas, y, como antes, indicaremos los totales marginales sustituyendo por un punto el índice sumado; esto es,

$$n_{i..k} = \sum_{j=1}^{s_2} n_{ijk} \quad n_{..k} = \sum_{i=1}^{s_1} \sum_{j=1}^{s_2} n_{ijk} \quad (21)$$

En relación con esta tabla hay cuatro hipótesis que pueden docimarse. Cabe docimar si los tres criterios son mutuamente independientes, en cuyo caso la hipótesis nula es

$$p_{ijk} = p_i q_j r_k \quad (22)$$

o bien si cualquiera de dichos criterios es independiente de los otros dos; así, para docimar si la clasificación B es independiente de A y C , estableceremos como hipótesis nula

$$p_{ijk} = p_{ik} q_j \quad (23)$$

El procedimiento a seguir para la docimasia de estas hipótesis es totalmente análogo al empleado para las tablas dobles. La verosimilitud de la muestra es

$$L = \prod_{ijk} p_{ijk}^{n_{ijk}} \quad \sum_{ijk} p_{ijk} = 1 \quad \sum_{ijk} n_{ijk} = n \quad (24)$$

En Ω , el máximo de L se presenta para

$$\hat{p}_{ijk} = \frac{n_{ijk}}{n} \quad (25)$$

de modo que

$$L(\hat{\Omega}) = \frac{1}{n^n} \prod_{ijk} n_{ijk}^{n_{ijk}} \quad (26)$$

Para docimar (23), p. ej., sustituiremos (23) en (24), hallando después el máximo de L respecto a las $p_{ik}q_j$; resulta así:

$$\hat{p}_{ik} = \frac{n_{i,k}}{n} \quad \hat{q}_j = \frac{n_{.j}}{n} \quad (27)$$

y

$$L(\hat{\omega}) = \frac{1}{n^{2n}} \left(\prod_{ik} n_{i,k}^{n_{i,k}} \right) \left(\prod_j n_{.j}^{n_{.j}} \right) \quad (28)$$

La razón de verosimilitud λ viene dada por el cociente de (28) y (26), y, en muestras grandes, $-2 \log \lambda$ tiene la distribución ji cuadrado con

$$s_1 s_2 s_3 - 1 - [(s_1 s_3 - 1) + s_2 - 1] = (s_1 s_3 - 1)(s_2 - 1)$$

grados de libertad. También la distribución en muestras grandes es completamente adecuada para muchos propósitos.

PROBLEMAS

1. La variable aleatoria x se distribuye normalmente con media μ y varianza 1. Se desea docimar $H_1: \mu=6$ con la hipótesis alternativa $H_2: \mu=7$. Formúlese la cuestión como un problema de decisión de dos acciones y defínase el espacio de acción.

2. En el problema 1, defínase el espacio paramétrico.

3. Se toma una muestra aleatoria de tamaño 3 de la densidad del problema 1. Defínase una función de decisión para este problema.

4. En el problema 3, ¿puede ser $|\bar{x}|$ una función de pérdida? ¿Lo podrá ser $(\bar{x} - \mu)$? Razónense las respuestas.

5. En el problema 1, sea d una función de decisión definida como sigue ($\bar{x}$ está basada en una muestra aleatoria de tamaño 4):

Se toma la acción a_1 , es decir, se acepta H_1 , si $\bar{x} < 7$.

Se toma la acción a_2 , es decir, se acepta H_2 , si $\bar{x} \geq 7$.

Hállense $P(I)$ y $P(II)$.

6. Dense las probabilidades de acción para el problema 5.

7. Defínense las probabilidades de error para el problema 5.

8. Una función de pérdida para el problema 1 es (la acción a_1 consiste en aceptar H_1 , y la acción a_2 , en aceptar H_2):

$$\begin{aligned} l(a_1; \mu=6) &= 0 & l(a_2; \mu=6) &= 2 \\ l(a_1; \mu=7) &= 1 & l(a_2; \mu=7) &= 0 \end{aligned}$$

Hállese y represéntese el riesgo para las siguientes funciones de decisión (x es una observación única):

Función de decisión	Tomar la acción a_1 si	Tomar la acción a_2 si
d_1	$x < 6,0$	$x \geq 6,0$
d_2	$x < 6,5$	$x \geq 6,5$
d_3	$x < 7,0$	$x \geq 7,0$
d_4	$x < 7,5$	$x \geq 7,5$
d_5	$x > 6,5$	$x \leq 6,5$

9. En el problema 8, ¿están dominadas algunas de las cinco estrategias d_i por alguna otra? Si es así, ¿cuáles de ellas?

10. Hállese y represéntese el riesgo en el problema 8 para la siguiente función de decisión,

$$\begin{aligned} d_c: & \text{tomar la acción } a_1 & \text{si } x < c \\ & \text{tomar la acción } a_2 & \text{si } x \geq c \end{aligned}$$

como función de c para $-\infty < c < \infty$.

11. Un fabricante de televisores tiene establecido cierto sistema de inspección para cada aparato. Si este pasa la inspección se clasifica como modelo de primera clase y se vende a un precio más alto que los de la segunda clase, o sea, aquellos aparatos que son rechazados en la inspección. Sea A el espacio de acción

$$A = \{a: a = a_1 \text{ o } a_2\},$$

en donde definimos:

$$\begin{aligned} a_1: & \text{vender como modelo de 1.ª clase;} \\ a_2: & \text{vender como modelo de 2.ª clase.} \end{aligned}$$

La producción anual de televisores se distribuye en dos clases distintas, que clasificaremos como θ_1 : grado superior; θ_2 : grado inferior. Después de tener en cuenta los costes de fabricación y de otro tipo, se considera adecuada para este problema la siguiente función de pérdida:

$$\begin{aligned} l(a_1; \theta_1) &= 0 & l(a_2; \theta_1) &= 3 \\ l(a_1; \theta_2) &= 5 & l(a_2; \theta_2) &= 0 \end{aligned}$$

El sistema de inspección se halla a cargo de un ingeniero que examina

cada televisor y lo clasifica como excelente, bueno o mediano. Las probabilidades de cada clase son:

Estado de la naturaleza	Probabilidad de clasificar un televisor como		
	1 (excelente)	2 (bueno)	3 mediano
θ_1	$1/2$	$1/3$	$1/6$
θ_2	$1/9$	$1/2$	$7/18$

Enumérense las ocho estrategias puras disponibles para este problema.

12. En el problema 11, calcúlese el riesgo de cada estrategia pura para cada estado de la naturaleza. Representéense estos riesgos.

13. En el problema 11, sean las estrategias 1 y 2, respectivamente,

	Clasificación		
	1 (excelente)	2 (bueno)	3 (mediano)
d_1	a_1	a_2	a_2
d_2	a_1	a_1	a_2

Si elegimos d_1 con probabilidad $1/2$ y d_2 con igual probabilidad, hállese el riesgo de la estrategia aleatorizada d_g que resulta.

14. En el problema 13, calcúlese el riesgo para la estrategia aleatorizada d_p , en donde se elige d_1 con probabilidad p y d_2 con probabilidad $1-p$, siendo $0 \leq p \leq 1$. Representéense gráficamente este riesgo como función de p .

15. Supongamos que x se distribuye normalmente con media μ y varianza 1, e imaginemos que se desea docimar la hipótesis H_1 contra la alternativa H_2 , con

$$H_1: \mu < 0 \quad H_2: \mu \geq 0$$

La función de pérdida para este problema es

$$\begin{aligned} l(a_1; \mu) &= 0 & \text{si } \mu < 0 \\ l(a_1; \mu) &= \mu^2 & \text{si } \mu \geq 0 \\ l(a_2; \mu) &= \mu^2 & \text{si } \mu < 0 \\ l(a_2; \mu) &= 0 & \text{si } \mu \geq 0 \end{aligned}$$

en donde la acción a_i consiste en aceptar la hipótesis H_i . Hállese el riesgo para cada una de las dos funciones de decisión, d_1 y d_2 , siguientes:

d_1 : tomar la acción a_2 si $x \geq 2$; tomar la acción a_1 si $x < 2$

d_2 : tomar la acción a_2 si $x \geq 3$; tomar la acción a_1 si $x < 3$.

16. En el problema 15, representese el riesgo para ambas funciones de decisión como función de μ .

17. Dada la muestra $(-0,2; -0,9; -0,6; 0,1)$ de una población normal con varianza 1, docímese si la media poblacional es menor que cero con nivel de significación del 0,05 (esto es, con probabilidad 0,05 de un error de tipo I). Es decir, docímese $H_0: \mu \leq 0$ al nivel 0,05 con relación a $H_a: \mu > 0$.

18. Dada la muestra $(-4,4; 4,0; 2,0; -4,8)$ de una población normal con varianza 4 y la muestra $(6,0; 1,0; 3,2; -0,4)$ de otra población también normal con varianza 5, docímese al nivel 0,05 que las medias difieren en una cantidad no superior a una unidad. (Empléese el método de la sección 12-5.) Representése la función de potencia para esta dócima, así como la función de potencia ideal.

19. Un metalúrgico ha hecho cuatro determinaciones del punto de fusión del manganeso, obteniendo los valores de 1269, 1271, 1263, 1265 °C. Docímese la hipótesis de que la media μ de esta población difiere en menos de 5 unidades del valor publicado 1260 °C, al nivel 0,05. (Supóngase la normalidad y $\sigma^2=5$).

20. Representése la función de potencia para una dócima de la hipótesis nula $H_0: -1 < \mu < 1$ en una población normal de varianza conocida, utilizando tamaños muestrales 1, 4, 16, 64. (Utilícese la desviación estándar σ como unidad de medida sobre el eje μ , y una probabilidad de 0,05 para un error de tipo I; Sec. 12-5.) Representése la función de potencia ideal.

21. Sea $x_1, x_2, \dots, x_n$ una muestra aleatoria de tamaño n de una densidad normal de varianza conocida. ¿Cuál es la mejor región crítica R para docimar la hipótesis nula de que la media es 6 contra la alternativa de que esta es 4?

22. Utilícese el teorema 12-4 para docimar $H_0: \sigma^2 < 10$ contra $H_a: \sigma^2 \geq 10$, en una muestra de tamaño n de una población normal con media cero.

23. Al docimar entre dos valores μ_0 y μ_1 para la media de una población normal, demuéstrese que las probabilidades para ambos tipos de error se hacen tan pequeñas como queramos tomando una muestra suficientemente grande.

24. Un fabricante de cigarrillos envió a dos laboratorios sendas muestras de tabaco, que se suponen idénticas. Cada laboratorio hizo cinco determinaciones del contenido de nicotina en miligramos, obteniendo: a) 24, 27, 26, 21, 24, y b) 27, 28, 23, 31, 26. ¿Medían los dos laboratorios una misma cosa? (Se supone normalidad y varianza común.)

25. El metalúrgico del problema 19, tras evaluar la magnitud de los diversos errores que pueden presentarse en su técnica experimental, decidió que sus medidas deberían tener una desviación estándar de dos grados o menos. ¿Son compatibles los datos con esta hipótesis al nivel 0,05? (Es decir, docímese $H_0: \sigma \leq 2$.)

26. Docímese la hipótesis de que las dos muestras del problema 18 proceden de poblaciones con igual varianza al nivel 0,05.

27. La función de potencia para docimar que las medias de dos poblaciones normales son iguales depende de los valores de estas dos medias, μ_1 y μ_2 , y es, por consiguiente, una superficie. Pero el valor numérico de la función depende solo de la diferencia $\theta = \mu_1 - \mu_2$, de modo que puede representarse adecuadamente por una curva, que llamamos $\beta(\theta)$. Represéntese $\beta(\theta)$ cuando se extraen muestras de tamaño 4 de una población con varianza igual a 2, y muestras de tamaño 2 de otra población con varianza 3, para dócimas al nivel 0,01.

28. Dadas las muestras (1,8; 2,9; 1,4; 1,1) (5,0; 8,6; 9,2) de poblaciones normales, docímese si las varianzas son iguales al nivel 0,05.

29. Dada una muestra de tamaño 100 con $\bar{x}=2,7$ y $\Sigma(x_i - \bar{x})^2 = 225$, docímese la hipótesis nula:

$$H_0: \mu=3 \quad \text{y} \quad \sigma^2=2,5$$

al nivel 0,01, suponiendo que la población es normal.

30. Docímese la hipótesis $\mu=\sigma^2$, al nivel 0,01, haciendo uso de la muestra del problema anterior.

31. Docímese al nivel 0,01, haciendo uso de la muestra del problema 29, si el punto α del 95% para la distribución poblacional es 3, con relación a las alternativas $\alpha < 3$. El punto del 95% es el número α tal que

$$\int_{-\infty}^{\alpha} f(x) dx = 0,95,$$

siendo $f(x)$ la función de densidad de la población; es obvio que en este ejemplo, donde la distribución se supone normal, $\alpha = \mu + 1,645 \sigma$.

32. Se extrae una muestra de tamaño n de cada una de k poblaciones normales con la misma varianza. Dedúzcase el criterio de la razón de verosimilitud para docimar la hipótesis de que todas las medias son iguales a cero. Demuéstrese que tal criterio es una función de una razón que tiene la distribución F .

33. Dedúzcase el criterio de la razón de verosimilitud para docimar si la correlación de una distribución normal bivalente es igual a cero.

34. Si $x_1, x_2, \dots, x_n$, son observaciones de poblaciones normales con varianzas conocidas $\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2$, ¿cómo se docimaría que todas sus medias son iguales?

35. Un periódico de cierta ciudad observó que habían mejorado mucho las condiciones de tráfico, porque en el último año el número de accidentes automovilísticos fatales había sido solo 9, mientras que el promedio por año para todos los anteriores era 15. ¿Es posible que las condiciones fuesen aproximadamente las mismas que antes? Supóngase que el número de accidentes en un año dado tiene la distribución de Poisson.

36. Se docimaron seis ejemplares, de un pie de longitud, de un alambre aislado a alto voltaje para encontrar los lugares de aislamiento insuficiente. El número de tales lugares fue 2, 0, 1, 1, 3, 2. El fabricante afirma que la calidad es tal que hay menos de 120 de dichos defectos por cada 100 pies. ¿Cumple la hornada de la cual proceden los ejemplares docimados con lo afirmado por el fabricante para un nivel de significación del 0,05? (Úsese la distribución de Poisson.)

37. Un psiquiatra empleado en cierta clínica observa en una reunión de los médicos de esta que aproximadamente el 40% de todos los dolores crónicos de cabeza pertenecen al tipo psicossomático. Sus colegas, escépticos, mezclan varias píldoras de harina y agua, y las dan a todos los pacientes inscritos en la clínica, diciendo que constituyen una nueva medicina para el dolor de cabeza, y solicitan su opinión. Recogidas todas las opiniones, pueden clasificarse con bastante precisión del siguiente modo: 1) mejor que la aspirina, 8; 2) poco más o menos como la aspirina, 3; 3) más lentas que la aspirina, 1; 4) no producen ningún efecto, 29. Aunque los doctores quedaron sorprendidos por tales resultados, acusaron, no obstante, al psiquiatra de exagerado. ¿Estaba justificada su acusación?

38. Se lanza 300 veces un dado, obteniendo los resultados siguientes:

punto obtenido	1	2	3	4	5	6
frecuencia	43	49	56	45	66	41

¿Están de acuerdo los datos obtenidos, para un nivel de 0,05, con la hipótesis de que el dado está bien construido?

39. De 64 descendientes de cierto cruce entre cobayas, 34 resultaron rojos, 10 negros y 20 blancos. Según el modelo proporcionado por la genética, estos números deberían estar en la proporción 9:3:4. ¿Están de acuerdo los datos con dicho modelo, para un nivel del 0,05?

40. El resultado medio de un destacado jugador de pelota base descendió de 0,313 un año a 0,280 el año siguiente. Jugó 374 veces el primer año y 268 el segundo. ¿Puede mantenerse la hipótesis, para un nivel de 0,05, de que sus aptitudes de jugador eran las mismas durante dichos dos años?

41. Hállense la media y la varianza de n_{ij} en la distribución condicional (12-11-17).

42. Demuéstrese que el valor esperado de u definido por (12-11-20) es $n(r-1)(s-1)/(n-1)$ en la distribución condicional (12-11-17).

43. Utilizando los datos del problema 40, supóngase que se tiene una muestra de tamaño 374 de una población binomial, y de 268, de otra. Obténgase el criterio λ para docimar si la probabilidad de *batear* es la misma para ambas poblaciones. ¿Cuál es el resultado de la comparación de esta dócima con la dócima ordinaria para una tabla de contingencia 2×2 ?

44. La progenie de cierta pareja se clasificó en tres grupos, de acuerdo con determinado atributo físico, obteniéndose los números 10, 53, 46. Según un modelo genético, las frecuencias deberían estar en la proporción

$p^2:2p(1-p):(1-p)^2$. Son compatibles los datos con dicho modelo, para un nivel de 0,05?

45. 1000 individuos se clasificaron como sigue, de acuerdo con su sexo y según que sufrieran o no de daltonismo:

	Masculino	Femenino
Normales	442	514
Con daltonismo	38	6

Según el modelo genético, estos números deberían tener frecuencias relativas dadas por

$$\frac{p}{2} \quad \frac{p^2}{2} + pq$$

$$\frac{q}{2} \quad \frac{q^2}{2}$$

en donde $q=1-p$ es la proporción de genes defectuosos en la población. ¿Son compatibles los datos con el modelo?

46. Considerando la tabla del problema 45 como una tabla de contingencia 2×2 , docímese la hipótesis de independendencia entre el daltonismo y el sexo.

47. Gilby clasificó 1725 escolares según su inteligencia y de acuerdo con el supuesto nivel económico familiar. La clasificación dio el siguiente resultado:

	Torpes	Inteligentes	Muy capaces
Muy bien vestidos	81	322	233
Bien vestidos	141	457	153
Mal vestidos	127	163	48

Docímese la independendencia para un nivel de 0,01.

48. Una vacuna, que se supone eficaz para prevenir los resfriados, se docimó en 500 individuos, comparando los resultados durante un año con los correspondientes a 500 individuos no tratados. Se obtuvieron los siguientes resultados:

	Ningún resfriado	Un resfriado	Más de un resfriado
Vacunados	252	145	103
No vacunados	224	136	140

Docímese, para un nivel del 0,05, si pueden considerarse iguales las dos poblaciones trinomiales.

49. Obténgase el criterio general λ para docimar la independencia en una tabla $r \times s$ cuando se ha fijado de antemano, como en el problema 48, una serie de totales marginales (totales de fila, p. ej.). Cada fila se considera como una muestra procedente de una población polinomial s -ple, con probabilidades p_{ij} tales que

$$\sum_i p_{ij} = 1,$$

para todos los valores de i . La hipótesis de independencia, en este caso, es

$$p_{1j} = p_{2j} = p_{3j} = \dots = p_{rj}$$

para todos los valores de j . ¿Cuántos grados de libertad tiene $-2 \log \lambda$?

50. Según un modelo facilitado por la genética, la proporción de individuos que tienen los cuatro tipos de grupos sanguíneos debe ser la siguiente:

$$O: q^2$$

$$A: p^2 + 2pq$$

$$B: r^2 + 2qr$$

$$AB: 2pr$$

en donde $p + q + r = 1$. Dada la muestra: O, 374; A, 436; B, 132; AB, 58. ¿Cómo podría docimarse si el modelo es correcto?

51. Dadas las frecuencias de clase

$$n_{ijk} (i=1, 2, \dots, r; j=1, 2, \dots, s; k=1, 2, \dots, t)$$

en una clasificación triple, docímese si los tres criterios de clasificación son independientes. ¿Cuántos grados de libertad tiene $-2 \log \lambda$?

52. Galton investigó 78 familias, clasificando los hijos según que tuvieran o no ojos claros, tuvieran o no un progenitor de ojos claros y tuvieran o no uno de los abuelos de ojos claros. Resultó así la siguiente tabla $2 \times 2 \times 2$:

		Abuelos			
		Claros		No claros	
		Padres			
		Claros	No claros	Claros	No claros
Hijos	Claros	1928	552	596	508
	No claros	303	395	225	501

Docímese la independencia completa para un nivel del 0,01. Docímese si la clasificación de los hijos es independiente de las otras dos clasificaciones al nivel 0,01.

53. Obténgase el criterio λ para docimar si la clasificación i es independiente de la clasificación jk en una tabla de contingencia triple, cuando los totales marginales $n_{i..}$ se han fijado de antemano. Las probabilidades satisfacen las relaciones

$$\sum_{jk} p_{ijk} = 1$$

para todo valor de i , y la hipótesis nula es

$$p_{ijk} = p_{2jk} = \dots = p_{rjk} \quad \text{o, simplemente,} \quad p_{ijk} = p_{jk}$$

¿Cuántos grados de libertad tiene $-2 \log \lambda$?

54. Obténgase la dócima de independencia completa en la situación descrita en el problema 53. La hipótesis nula es $p_{ijk} = p_i q_k$. ¿Cuántos grados de libertad tiene $-2 \log \lambda$? ¿Cómo resulta esta dócima, comparada con la correspondiente al caso en que los $n_{i..}$ no se han fijado de antemano?

55. Calcúlese la distribución exacta de λ para una tabla de contingencia 2×2 con totales marginales $n_{1.} = 4$; $n_{2.} = 7$; $n_{.1} = 6$; $n_{.2} = 5$. ¿Cuál es la probabilidad exacta de que $-2 \log \lambda$ exceda a 3,84, nivel al 0,05 de la ji cuadrado para un grado de libertad?

BIBLIOGRAFIA

1. BLACKWELL, D., y M. A. GIRSCHICK: *Theory of Games and Statistical Decisions*, John Wiley & Sons, Inc., Nueva York, 1954.
2. CHERNOFF, H., y L. E. MOSES: *Elementary Decision Theory*, John Wiley & Sons, Inc., Nueva York, 1959.
3. FISHER, R. A.: «On the mathematical foundations of theoretical statistics», *Philosophical Transactions of the Royal Society, London*, series A, vol. 222 (1922), págs. 309-368.
4. FRASER, D. A. S.: *Nonparametric Methods in Statistics*, John Wiley & Sons, Inc., Nueva York, 1957.
5. KENDALL, M. G.: «The Advanced Theory of Statistics», vol. 2, Charles Griffin & Co., Ltd., Londres, 1946.
6. LEHMANN, E. L.: «On families of admissible tests», *Annals of Mathematical Statistics*, vol. 18 (1947), págs. 97-104.
7. LEHMANN, E. L.: «Some principles of the theory of hypothesis testing», *Annals of Mathematical Statistics*, vol. 21 (1950), págs. 1-26.
8. LEHMANN, E. L.: «A general concept of unbiasedness», *Annals of Mathematical Statistics*, vol. 22 (1951), págs. 587-597.
9. LEHMANN, E. L.: «Ordered families of distributions», *Annals of Mathematical Statistics*, vol. 26 (1955), págs. 399-419.
10. LEHMANN, E. L.: *Testing Statistical Hypotheses*, John Wiley & Sons, Inc., Nueva York, 1959.

11. NEYMAN, J., y E. S. PEARSON: «On the use and interpretation of certain test criteria for purposes of statistical inference», *Biometrika*, volumen 20A (1928), págs. 175-240, 263-294.
12. NEYMAN, J., y E. S. PEARSON: «On the problem of the most efficient tests of statistical hypotheses», *Philosophical Transactions of the Royal Society of London*, series A, vol. 231 (1933), págs. 289-337.
13. STEIN, C. M.: «A property of some tests of composite hypotheses», *Annals of Mathematical Statistics*, vol. 22 (1951), págs. 475-476.
14. STUDENT (W. S. GOSSET): «On the probable error of a mean», *Biometrika*, vol. 6 (1908), págs. 1-25.
15. WALD, Abraham: «Contributions to the theory of statistical estimation and testing hypotheses», *Annals of Mathematical Statistics*, volumen 10 (1939), págs. 299-326.
16. WALD, Abraham: *Statistical Decision Functions*, John Wiley & Sons, Inc., Nueva York, 1950.
17. WEISS, L.: *Statistical Decision Theory*, McGraw-Hill Book Company, Inc., Nueva York., 1961.

CAPITULO 13

REGRESION E HIPOTESIS LINEALES

13-1. Introducción.—Estudiaremos en este capítulo modelos lineales, de los que los dos ejemplos siguientes constituyen casos especiales.

Ejemplo 13-1.—La distancia s que ha recorrido una partícula durante el tiempo t está dada por la fórmula $s = \alpha + \beta t$, en donde β es la velocidad media, y α , la posición (inicial) en el instante $t = 0$. Si α y β son desconocidas, puede observarse s para dos valores distintos de t , y resolver el sistema en α y β de las dos ecuaciones resultantes. Así, p. ej., supongamos que s vale 4 cuando es $t = 1$ y 11 para $t = 4$. Esto da $2 = \alpha + \beta$ y $11 = \alpha + 4\beta$, cuya solución es $\alpha = -1$ y $\beta = 3$; por tanto, $s = -1 + 3t$. Imaginemos que, por alguna razón, no es posible observar con exactitud la distancia, sino que existe un error de medida de naturaleza aleatoria. Por tanto, s no podrá observarse, pero sí podemos observar y , siendo $y = s + e$, en donde e es un error aleatorio cuya media es 0. Sustituyendo s , tenemos

$$y = \alpha + \beta t + e \quad (1)$$

en donde y es una variable aleatoria observable, t una variable matemática (no aleatoria) también observable y e es una variable aleatoria que no puede ser observada; α y β son parámetros desconocidos. No es posible determinar α y β por observación de dos conjuntos de valores de y y t , como antes hicimos para s y t , ya que no existe una relación funcional entre y y t . El objetivo en este modelo es hallar α y β y calcular, por tanto, $s = \alpha + \beta t$ para varios valores de t . Puesto que s está sujeto a errores y no puede observarse, no conocemos α y β ; pero, observando varios conjuntos de valores de y y t , podremos obtener estimaciones de α , β y s por medio de métodos estadísticos. Este tipo de proceso constituye un modelo de *relación funcional* con un error de medida.

Ejemplo 13-2.—Como otro ejemplo, fijémonos en la relación existente entre la estatura h y el peso w de individuos pertenecientes a determinada ciudad. Evidentemente, no hay una relación funcional entre las variables w y h , pero sí existe entre ellas cierto tipo de conexión. Considerémoslas como variables aleatorias y postulamos que (w, h) tiene una distribución normal bivalente. En

tal supuesto, el valor esperado de h para un valor dado de w está dado por

$$E(h|w) = \alpha + \beta w \quad (2)$$

en donde α y β son funciones de los parámetros en una densidad normal bivalente. Aunque no existe una relación funcional entre h y w , si su distribución conjunta es normal, existirá una relación funcional lineal entre los pesos y el valor medio de las estaturas. Así, diremos que h y w admiten una distribución conjunta normal, y escribiremos

$$E(h|w) = \alpha + \beta w$$

o bien,

$$h_w = \alpha + \beta w + e$$

Este es un modelo de regresión y aunque proviene de un problema algo diferente que el de la relación funcional del ejemplo 13-1, ambos son casos especiales de un modelo estadístico lineal que examinaremos en este capítulo.

13-2. Modelos lineales simples.—Los dos ejemplos de la sección anterior originaron modelos que se incluyen en una clase general de modelos lineales y, aunque las situaciones que representan son distintas, desde un punto de vista estadístico son análogas. En esta sección nos ocuparemos de una clase de modelos que reciben la denominación de *modelos lineales simples*.

Definición 13-1.—Sean $y_1, y_2, \dots, y_n$ variables aleatorias observables incorrelacionadas y tales que $y_i = \alpha + \beta x_i + e_i$, en donde: α y β son parámetros desconocidos; x_i , variables matemáticas (no aleatorias), y e_i , variables aleatorias no observables e incorrelacionadas, con media 0 y varianza σ^2 (σ^2 no es función de α, β o x_i). Estos supuestos definen un modelo lineal simple.

Analizaremos dos casos en relación con este modelo lineal simple:

Caso A: Las variables aleatorias se distribuyen normalmente.

Caso B: Las variables aleatorias no se distribuyen normalmente.

En ambos casos nos interesará la estimación puntual, y en el caso A examinaremos además la estimación puntual y por intervalos, así como la docimasia de hipótesis.

Caso A: Estimación puntual.—Esta situación se representa en la figura 13-1; para un valor dado de x , la variable aleatoria y se distribuye normalmente con media $\alpha + \beta x$ y varianza σ^2 . Sea (y_i, x_i) , $i=1, 2, \dots, n$, una muestra de las y , junto con los valores correspon-

dientes de x . Puede ocurrir que algunas de las x_i sean iguales, como sucedería si se obtuviera más de un valor y de cualquier distribución dada. Es conveniente representar las x por símbolos diferentes, aunque algunas de ellas sean iguales. No obstante, es necesario que haya, al menos, dos valores diferentes de x . Claro es que no será

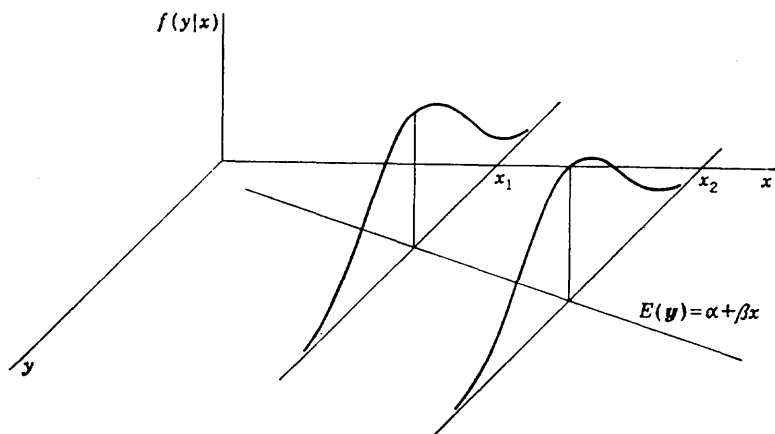


FIG. 13-1.

posible efectuar la estimación de α y β a partir de una muestra procedente de un solo miembro de la familia de distribuciones. Se empleará el método de máxima verosimilitud para la estimación de los parámetros. La verosimilitud es

$$L = f(\mathbf{e}_1, \dots, \mathbf{e}_n) = \left(\frac{1}{\sqrt{2\pi}\sigma} \right)^n \exp \left(-\frac{1}{2\sigma^2} \sum \mathbf{e}_i^2 \right) \quad (1)$$

$$= \left(\frac{1}{\sqrt{2\pi}\sigma} \right)^n \exp \left\{ \left(-\frac{1}{2\sigma^2} \right) \sum [y_i - (\alpha + \beta x_i)]^2 \right\} \quad (2)$$

y su logaritmo

$$\log L = -\frac{n}{2} \log 2\pi - \frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_i [y_i - (\alpha + \beta x_i)]^2 \quad (3)$$

Derivando esta expresión respecto a α , β y σ^2 e igualando a cero las derivadas, resultan las relaciones

$$n\hat{\sigma}^2 = \sum (y_i - \hat{\alpha} - \hat{\beta}x_i)^2 \quad (4)$$

$$\sum (y_i - \hat{\alpha} - \hat{\beta}x_i) = 0 \quad (5)$$

$$\sum x_i (y_i - \hat{\alpha} - \hat{\beta}x_i) = 0, \quad (6)$$

sistema que debe resolverse para obtener los parámetros desconocidos. Las dos últimas ecuaciones reciben el nombre de *ecuaciones normales* y determinan los coeficientes $\hat{\alpha}$ y $\hat{\beta}$. Las ecuaciones en $\hat{\alpha}$ y $\hat{\beta}$ son lineales, y su resolución no presenta ninguna dificultad. Hagamos

$$\bar{x} = \frac{1}{n} \sum x_i \quad \bar{y} = \frac{1}{n} \sum y_i \quad (7)$$

Las soluciones de (5), (6) y (7) pueden escribirse:

$$\hat{\beta} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} \quad (8)$$

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} \quad (9)$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum (y_i - \hat{\alpha} - \hat{\beta}x_i)^2 \quad (10)$$

Se obtienen así los estimadores puntuales deseados de los parámetros desconocidos. Observemos que la resolución no podría efectuarse si todas las x_i fuesen iguales, ya que se anularía el denominador de (8).

Distribución de los estimadores.—Puesto que $\hat{\alpha}$ y $\hat{\beta}$ son funciones lineales de las y_i (cuya distribución es normal), se deduce que $\hat{\alpha}$ y $\hat{\beta}$ tendrán una distribución normal bivalente. Cabría determinar esta distribución obteniendo las medias, varianzas y covarianza de $\hat{\alpha}$ y $\hat{\beta}$; sin embargo, la hallaremos de otra manera. El objetivo principal consiste en probar que la distribución de $\hat{\alpha}$ y $\hat{\beta}$ es independiente de la de $\hat{\sigma}^2$, y al hacer esto se deducirá incidentalmente la distribución que buscamos.

Calcularemos la función generatriz de los momentos trivariantes, a saber:

$$m(s_1, s_2, s_3) = E \left[\exp \left(s_1 \frac{\hat{\alpha} - \alpha}{\sigma} + s_2 \frac{\hat{\beta} - \beta}{\sigma} + s_3 \frac{n\hat{\sigma}^2}{\sigma^2} \right) \right] \quad (11)$$

$$= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \left(\frac{1}{2\pi\sigma^2} \right)^{n/2} \exp \left[s_1 \frac{\hat{\alpha} - \alpha}{\sigma} + s_2 \frac{\hat{\beta} - \beta}{\sigma} + s_3 \frac{n\hat{\sigma}^2}{\sigma^2} - \frac{1}{2\sigma^2} \sum (y_i - \alpha - \beta x_i)^2 \right] \prod dy_i \quad (12)$$

para las tres variantes $(\hat{\alpha} - \alpha)/\sigma$, $(\hat{\beta} - \beta)/\sigma$ y $n\hat{\sigma}^2/\sigma^2$. El primer paso para el cálculo de la integral es transformar las variantes y_i en

$$z_i = \frac{1}{\sigma} (y_i - \alpha - \beta x_i) \quad (13)$$

lo que suprime el factor $1/\sigma^n$ del integrando y transforma el exponente en el integrando de (12) en

$$\sum_{i=1}^n c_i z_i - (1/2) \sum_{i,j=1}^n r_{ij} z_i z_j \quad (14)$$

siendo

$$c_i = \frac{s_1[(\sum x_i^2/n) - \bar{x}x_i] + s_2(x_i - \bar{x})}{\sum (x_i - \bar{x})^2} = a_i s_1 + b_i s_2 \quad (15)$$

y

$$r_{ij} = \delta_{ij}(1 - 2s_3) + 2s_3[na_i a_j + n\bar{x}(a_i b_j + a_j b_i) + b_i b_j \sum x_i^2] \quad (16)$$

en donde las a_i y b_i quedan definidas por (15), y δ_{ij} es uno o cero según que i sea o no igual a j . Ahora tenemos que calcular una integral de la forma

$$\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \left(\frac{1}{2\pi} \right)^{n/2} \exp \left[\sum c_i z_i - (1/2) \sum \sum r_{ij} z_i z_j \right] \prod dz_i \quad (17)$$

la cual, salvo el factor $\sqrt{|(r_{ij})|}$, es precisamente la integral de la ecuación (9-3-20) con las μ_i iguales a 0. El valor de la integral se ha dado en el teorema 9.13 y se deduce que

$$m(s_1, s_2, s_3) = \frac{\exp [(1/2) \sum \sum \sigma_{ij} c_i c_j]}{\sqrt{|(r_{ij})|}} \quad (18)$$

La reducción algebraica de (18) puede efectuarse como sigue. Puesto que

$$a_i + \bar{x}b_i = \frac{1}{n}$$

la ecuación (16) se escribirá

$$r_{ij} = \delta_{ij}(1 - 2s_3) + 2s_3 \left[b_i b_j \sum (x_i - \bar{x})^2 + \frac{1}{n} \right] \quad (19)$$

o bien

$$r_{ij} = \delta_{ij}(1 - 2s_3) + 2s_3 \left(d_i d_j + \frac{1}{n} \right) \quad (20)$$

en donde

$$d_i = \frac{x_i - \bar{x}}{\sqrt{\sum (x_i - \bar{x})^2}} \quad (21)$$

de modo que $\Sigma(d_i) = 0$ y $\Sigma(d_i^2) = 1$. No es difícil comprobar que

$$|(r_{ij})| = (1 - 2s_3)^{n-2} \quad (22)$$

y que los elementos de la inversa de la matriz (r_{ij}) son

$$\sigma_{ij} = \frac{\delta_{ij}}{1 - 2s_3} - \frac{2s_3}{1 - 2s_3} \left(d_i d_j + \frac{1}{n} \right) \quad (23)$$

Estas dos últimas relaciones permiten escribir (18) del modo siguiente:

$$m(s_1, s_2, s_3) = \frac{\exp \left[(1/2) \Sigma (x_i - \bar{x})^2 \right] [s_1^2 (1/n) \Sigma x_i^2 - 2\bar{x} s_1 s_2 + s_2^2]}{(1 - 2s_3)^{(n-2)/2}} \quad (24)$$

La forma de la función generatriz de momentos (24) nos permite deducir varias conclusiones importantes. Recordando que s_1 está asociada con $(\hat{\alpha} - \alpha)/\sigma$, s_2 con $(\hat{\beta} - \beta)/\sigma$ y s_3 con $n\hat{\sigma}^2/\sigma^2$, observamos:

1. Que el par de variantes $\hat{\alpha}$ y $\hat{\beta}$ están distribuidas independientemente de $\hat{\sigma}^2$, ya que $m(s_1, s_2, s_3)$ se descompone en el producto de una sola función de s_3 y una sola función de s_1 y s_2 (véase Sec. 10-4). Hacemos

$$m(s_1, s_2, s_3) = m_1(s_1, s_2) m_2(s_3) \quad (25)$$

2. Que la forma funcional de $m_1(s_1, s_2)$ es la de la función generatriz de momentos de una distribución normal bivalente (teorema 9-1); por tanto, $\hat{\alpha}$ y $\hat{\beta}$ tienen conjuntamente una distribución normal de medias respectivas α y β , y varianzas y covarianzas

$$\sigma_{11} = \sigma_{\hat{\alpha}}^2 = \frac{\sigma^2 \Sigma x_i^2}{n \Sigma (x_i - \bar{x})^2} \quad (26)$$

$$\sigma_{22} = \sigma_{\hat{\beta}}^2 = \frac{\sigma^2}{\Sigma (x_i - \bar{x})^2} \quad (27)$$

$$\sigma_{12} = \text{cov}(\hat{\alpha}, \hat{\beta}) = -\frac{\sigma^2 \bar{x}}{\Sigma (x_i - \bar{x})^2} \quad (28)$$

La inversa de la matriz de estas varianzas y covarianzas es

$$\begin{pmatrix} \frac{n}{\sigma^2} & \frac{n\bar{x}}{\sigma^2} \\ \frac{n\bar{x}}{\sigma^2} & \frac{\Sigma x_i^2}{\sigma^2} \end{pmatrix}; \quad (29)$$

estos son los coeficientes de la forma cuadrática en la distribución de $(\hat{\alpha} - \alpha)$ y $(\hat{\beta} - \beta)$.

3. Que $\hat{\alpha}$ y $\hat{\beta}$ están distribuidas independientemente si las x_i se eligen de modo que sea $\bar{x} = 0$.

4. Que la forma cuadrática de la distribución conjunta de $\hat{\alpha}$ y $\hat{\beta}$

$$Q = \frac{1}{\sigma^2} [n(\hat{\alpha} - \alpha)^2 + 2n\bar{x}(\hat{\alpha} - \alpha)(\hat{\beta} - \beta) + \sum x_i^2 (\hat{\beta} - \beta)^2] \quad (30)$$

se distribuye según una ji cuadrado con dos grados de libertad.

5. Que $m_2(s_3)$ es la función generatriz de momentos de una distribución ji cuadrado con $n-2$ grados de libertad, y que, por consiguiente, $n\hat{\sigma}^2/\sigma^2$ admite la misma distribución (Sec. 10-3).

6. Descomponiendo el exponente de (2) en producto y utilizando el criterio de Neyman para la suficiencia, se deduce que $\hat{\alpha}$, $\hat{\beta}$, $\hat{\sigma}^2$ son estadísticos suficientes de α , β , σ^2 . También se demuestra que son completos.

Teorema 13-1.—*Para el caso A del modelo lineal de la definición 13-1 los estimadores $\hat{\alpha}$, $\hat{\beta}$, $[n/(n-2)]\hat{\sigma}^2$, definidos en (8) a (10), son:*

1) *insesgados de varianza mínima; 2) consistentes; 3) asintóticamente eficientes; 4) suficientes y completos; 5) $n\hat{\sigma}^2/\sigma^2$ se distribuye según una ji cuadrado con $n-2$ g. de l.; 6) $\hat{\alpha}$, $\hat{\beta}$ tienen la distribución normal bivalente, con matriz de covarianzas cuyos elementos están dados por (26) a (28); 7) $\hat{\alpha}$, $\hat{\beta}$ son independientes de $\hat{\sigma}^2$.*

Regiones confidenciales y docimasia de hipótesis.—En estos modelos el interés principal suele centrarse en los coeficientes α y β . Por supuesto, la estimación de σ^2 o la docimasia de hipótesis sobre σ^2 no ofrece ninguna dificultad, ya que la distribución ji cuadrado citada proporciona directamente los intervalos confidenciales y las dócimas.

Para obtener un intervalo confidencial para α , basta observar que la distribución marginal de $\hat{\alpha}$ es normal, con media α y varianza dada por (26). En consecuencia,

$$u = \frac{\hat{\alpha} - \alpha}{\sigma} \sqrt{\frac{n \sum (x_i - \bar{x})^2}{\sum x_i^2}}$$

tiene una distribución normal con media cero y varianza 1. Puesto

que las distribuciones de u y $n\hat{\sigma}^2/\sigma^2$ son independientes, de la sección 10-6 se deduce que

$$t = \frac{\sigma u}{\sqrt{n\hat{\sigma}}/\sqrt{n-2}} = (\hat{\alpha} - \alpha) \sqrt{\frac{n(n-2)\sum(x_i - \bar{x})^2}{\sum x_i^2 \sum (y_i - \hat{\alpha} - \hat{\beta}x_i)^2}} \quad (31)$$

tiene la distribución t con $n-2$ grados de libertad. Como α es la única incógnita en esta expresión, pueden transformarse las desigualdades que figuran en

$$P(-t_{\epsilon/2} < t < t_{\epsilon/2}) = 1 - \epsilon$$

obteniendo un intervalo confidencial con probabilidad $1 - \epsilon$ para α . La cantidad t proporciona también un criterio para la docimasia de hipótesis relativas a α , del mismo modo que sucede para la media de una distribución normal. Así, para docimar si la recta de regresión $E(y) = \alpha + \beta x$ pasa por el origen del plano yx , bastará hacer $\alpha = 0$ en (31) y observar si $|t| < t_{\epsilon/2}$, en donde ϵ es el nivel de significación elegido. También es posible efectuar dócimas de una sola rama.

De manera completamente análoga se construyen intervalos confidenciales para β y dócimas relativas a β . Se ve en seguida que

$$t = (\hat{\beta} - \beta) \sqrt{\frac{(n-2)\sum(x_i - \bar{x})^2}{\sum(y_i - \hat{\alpha} - \hat{\beta}x_i)^2}} \quad (32)$$

tiene también la distribución t con $n-2$ grados de libertad e incluye solo el parámetro desconocido β . Para docimar, p. ej., si las medias de la familia de distribuciones normales considerada son independientes del parámetro observable, pondríamos $\beta = 0$ en (32) y observaríamos si $|t| < t_{\epsilon/2}$, en donde ϵ es el nivel de significación elegido.

Para la estimación simultánea de α y β cabe utilizar el hecho de que

$$F = \frac{Q}{n\hat{\sigma}^2/\sigma^2} \cdot \frac{n-2}{2} \quad (33)$$

donde Q está definido por (30), tiene la distribución F con 2 y $n-2$ grados de libertad (Sec. 10-5) y depende solo de los parámetros desconocidos α y β . Se ve fácilmente que la desigualdad

$$P(F < F_\epsilon) = 1 - \epsilon$$

define para α y β una región confidencial elíptica en el plano α, β .

Para docimar si α y β toman ciertos valores predeterminados α_0 y β_0 , haríamos $\alpha = \alpha_0$ y $\beta = \beta_0$ en (33) y observaríamos si el valor resultante de F excede o no a F_c .

Todas estas dócimas relativas a α y β podrían también haberse obtenido por el método de la razón de verosimilitud.

Merece observarse que la acuracidad de la observación de α y β depende de la elección de las x_i . Así, pues, la varianza de $\hat{\alpha}$ será la menor posible cuando se elijan las x_i de modo que sea $\bar{x} = 0$. Ya que

$$\Sigma(x_i - \bar{x})^2 = \Sigma x_i^2 - n\bar{x}^2$$

el menor valor posible $\sigma_{\hat{\alpha}}^2$ [Ec. (26)] será σ^2/n y se presenta en $\bar{x} = 0$. Evidentemente, el intervalo confidencial para α será, por término medio, más corto para un n dado cuando sea $\bar{x} = 0$. Evidentemente, la varianza de $\hat{\beta}$ [Ec. (27)] puede hacerse más pequeña eligiendo valores muy separados para las x_i . En efecto, si x_1 es el menor valor práctico de x y x_2 es el mayor, la estimación de β será mejor cuando la mitad de los valores de y se elijan en cada uno de estos dos valores de x , si n es par. Sin embargo, en la práctica, se plantean a menudo dudas sobre la linealidad del modelo, y es deseable docimar esta linealidad. En tal caso es necesario poseer observaciones para más de dos valores de x .

13.3. Predicción.—Supongamos que $E(y) = y_x = \alpha + \beta x$ ha sido estimada mediante $\hat{y}_x = \hat{\alpha} + \hat{\beta}x$, sobre la base de una muestra de n observaciones. Queremos ahora predecir el valor de y para un valor dado x_0 de x . Así, si y es la estatura de un hijo adulto, y x , la del padre, una muestra de observaciones proporcionará estimaciones $\hat{\alpha}$ y $\hat{\beta}$ para una función lineal. Un futuro padre, de estatura x_0 , desea predecir la estatura de su hijo. La estatura predicha es, desde luego, $\hat{y}_0 = \hat{\alpha} + \hat{\beta}x_0$. O para considerar un problema diferente, sea y la demanda de cierta mercancía y x su precio al por mayor dos meses antes, o bien el precio al por mayor de algún ingrediente de dicha mercancía en tal fecha. Se desea predecir la demanda con dos meses de anticipación. Basándose en datos del pasado se relacionará un conjunto de pares de observaciones (x_i, y_i) , donde y_i es la demanda en un momento dado y x_i el precio al por mayor dos meses antes, y se estimarán los coeficientes α y β en un modelo lineal. Si x_0 es el precio al por mayor actual, la demanda prevista dos meses más tarde será $y_0 = \hat{\alpha} + \hat{\beta}x_0$.

El valor de la predicción depende de la magnitud de su posible error, y tendremos este en cuenta obteniendo un *intervalo de predicción*, lo que es análogo a un intervalo confidencial. La variante y_0 es una variable aleatoria con distribución normal, de media $\alpha + \beta x_0$ y varianza σ^2 . El valor predicho $\hat{y}_0 = \hat{\alpha} + \hat{\beta}x_0$ contiene dos

fuentes de error: en primer lugar, $\hat{\alpha} + \hat{\beta}x_0$ es una estimación de la media de y_0 , y, naturalmente, el valor real de y_0 puede desviarse de su media; en segundo lugar, la media estimada está sometida a los errores del muestreo aleatorio inherentes a $\hat{\alpha}$ y $\hat{\beta}$. Si se conociesen exactamente α , β y σ , un intervalo de predicción del 95% para y_0 sería

$$\alpha + \beta x_0 - 1,96\sigma \quad \text{a} \quad \alpha + \beta x_0 + 1,96\sigma$$

puesto que, para una distribución normal, la probabilidad de que y_0 no difiera en más de $1,96\sigma$ de su media es 0,95. Puesto que todos estos parámetros, excepto x_0 , son desconocidos, intentaremos construir un intervalo basado en sus estimaciones.

La variante

$$u = y_0 - \hat{\alpha} - \hat{\beta}x_0 \quad (1)$$

tiene necesariamente una distribución normal, ya que es función lineal de las variantes normales y_0 , $\hat{\alpha}$, $\hat{\beta}$. Por tanto, la distribución de u queda determinada cuando se dan su media y su varianza. Puesto que

$$E(y_0) = \alpha + \beta x_0 \quad E(\hat{\alpha}) = \alpha \quad \text{y} \quad E(\hat{\beta}) = \beta$$

resulta que

$$E(u) = 0$$

La varianza de u es, por consiguiente,

$$\begin{aligned} \sigma_u^2 &= E(u^2) \\ &= E(y_0 - \hat{\alpha} - \hat{\beta}x_0)^2 \\ &= \sigma_{y_0}^2 + \sigma_{\hat{\alpha}}^2 + x_0^2 \sigma_{\hat{\beta}}^2 + 2x_0 E[(\hat{\alpha} - \alpha)(\hat{\beta} - \beta)] \end{aligned} \quad (2)$$

si se recuerda que y_0 es independiente de $\hat{\alpha}$ y $\hat{\beta}$; $\sigma_{y_0}^2$ coincide con σ^2 , varianza de la distribución normal, y los otros términos de (2) vienen dados por (13-2-26) a (13-2-28), de forma que

$$\begin{aligned} \sigma_u^2 &= \sigma^2 \left[1 + \frac{(1/n)\Sigma x_i^2 + x_0^2 - 2x_0\bar{x}}{\Sigma(x_i - \bar{x})^2} \right] \\ &= \sigma^2 \left[1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\Sigma(x_i - \bar{x})^2} \right] \\ &= \sigma^2 \left[\frac{n+1}{n} + \frac{(x_0 - \bar{x})^2}{\Sigma(x_i - \bar{x})^2} \right] \end{aligned} \quad (3)$$

Un intervalo de predicción del 95% para u es precisamente $1,96\sigma_u$ a $1,96\sigma_u$; pero este incluye aún un parámetro descono-

cido σ , que aparece en σ_u . Eliminaremos σ utilizando la distribución t . La variante u/σ_u tiene una distribución normal con media cero y varianza 1, y su distribución es independiente de $n\hat{\sigma}^2/\sigma^2$. Por tanto,

$$t = \frac{u/\sigma_u}{\sqrt{n\hat{\sigma}^2/(n-2)\sigma^2}} \quad (4)$$

tiene la distribución t con $n-2$ grados de libertad y no incluye parámetros desconocidos. Pueden transformarse las desigualdades de

$$P(-t_{\epsilon/2} < t < t_{\epsilon/2}) = 1 - \epsilon$$

para determinar un intervalo de predicción de $100(1-\epsilon)$ por ciento para y_0 . El intervalo viene dado por

$$P(\hat{\alpha} + \hat{\beta}x_0 - A < y_0 < \hat{\alpha} + \hat{\beta}x_0 + A) = 1 - \epsilon \quad (5)$$

en donde

$$A = t_{\epsilon/2}\hat{\sigma} \sqrt{\frac{n}{n-2} \left[\frac{n+1}{n} + \frac{(x_0 - \bar{x})^2}{\sum(x_i - \bar{x})^2} \right]} \quad (6)$$

Deben subrayarse varias propiedades del intervalo de predicción:

1) Por término medio, la longitud del intervalo es superior a $2t_{\epsilon/2}\sigma$, con independencia del tamaño de la muestra empleada para estimar α y β . Esto resulta totalmente razonable, debido a que estamos prediciendo una observación única y_0 que tiene una distribución normal con desviación estándar σ .

2) La longitud media del intervalo de predicción aumenta cuando x_0 se aleja de $\bar{x}$. Si es posible, los valores seleccionados de x_i , con el fin de obtener observaciones para la estimación de parámetros, han de elegirse de forma que tengan un valor medio próximo a x_0 .

3) La relación (5) se verifica solo para una predicción única basada en las estimaciones $\hat{\alpha}$, $\hat{\beta}$, $\hat{\sigma}$. La relación tiene sentido únicamente si α , β y σ vuelven a estimarse cada vez que se haga una predicción sobre y_0 . La afirmación probabilística toma en cuenta la variación en el muestreo, tanto en las estimaciones como en y_0 , y si se utilizan repetidas veces las estimaciones originales (sin permitirles variar), la afirmación carece de precisión.

Es fácil generalizar la técnica anterior para tener en cuenta la predicción de media de una muestra de tamaño m observada para $x=x_0$. Sea $y'_1, y'_2, \dots, y'_m$ una muestra de m observaciones en x_0 , con media $\bar{y}'$. La media de

$$v = \bar{y}' - \hat{\alpha} - \hat{\beta}x_0$$

es cero, y su varianza σ^2 es la misma que en (3), salvo que $(n+1)/n$ se sustituye por $(1/m) + (1/n)$. La variante

$$t = \frac{v/\sigma_v}{\sqrt{n\hat{\sigma}^2/(n-2)\sigma^2}}$$

tiene la distribución t con $n-2$ grados de libertad y no contiene parámetros desconocidos; por tanto, puede emplearse para construir un intervalo de predicción para $\bar{y}'$.

13-4. Discriminación.—La cuestión de la discriminación es un problema de estimación y en cierto modo el inverso del de predicción. En esta se trata de predecir y conociendo x_0 , sobre la base de estimaciones de α , β , σ . En la discriminación se desea estimar x_0 cuando se ha observado y . El tipo general de los problemas que surgen en los ensayos biológicos posee este carácter. Así, p. ej., para medir la concentración de cierta vitamina se observa el incremento de peso de un pollo de una semana, al enriquecer su alimento durante varios días mediante dosis diarias de vitaminas. Un fabricante de dicha vitamina podría determinar la fuerza de una nueva hornada del siguiente modo: representemos por y lo que se gana en peso y por x la concentración. Utilizando material de concentración conocida alimentaríamos varios pollos con concentraciones distintas x_i ($i=1, 2, \dots, n$) y observaría las ganancias de peso y_i . Al mismo tiempo otros pollos recibirían la vitamina desde el principio con concentración desconocida x_0 y se observarían sus aumentos de peso, y'_j ($j=1, 2, \dots, m$). Sobre la base de estos datos se desea estimar el parámetro x_0 .

La cuestión general de la clasificación es un problema de discriminación; p. ej., los antropólogos efectúan medidas y sobre cráneos de edad conocida x y estiman después la edad x_0 de un cráneo de edad desconocida mediante medidas y' . Los taxónomos usan la misma técnica para discriminar entre variedades de plantas de apariencia semejante.

Utilizando la notación anterior y el modelo de la sección 13-2, la verosimilitud de las observaciones $y_1, y_2, \dots, y_n$ e $y'_1, y'_2, \dots, y'_m$ es

$$L = \left(\frac{1}{\sqrt{2\pi}\sigma} \right)^{m+n} \exp \left[-\frac{1}{2\sigma^2} \sum_1^n (y_i - \alpha - \beta x_i)^2 - \frac{1}{2\sigma^2} \sum_1^m (y'_j - \alpha - \beta x_0)^2 \right] \quad (1)$$

y derivando sucesivamente el logaritmo de esta expresión respecto a σ^2 , α , β y x_0 , se pueden determinar sin dificultad las estimaciones máximo-verosímiles de estos parámetros; estas son:

$$\hat{\beta} = \frac{\sum (y_i - \bar{y})(x_i - \bar{x})}{\sum (x_i - \bar{x})^2} \quad (2)$$

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} \quad (3)$$

$$\hat{\sigma}^2 = \frac{1}{m+n} \left[\sum (y_i - \hat{\alpha} - \hat{\beta}x_i)^2 + \sum (y'_j - \bar{y}')^2 \right] \quad (4)$$

$$\hat{x}_0 = \frac{\bar{y} - \hat{\alpha}}{\hat{\beta}} \quad (5)$$

donde

$$\bar{x} = \frac{1}{n} \sum x_i \quad \bar{y} = \frac{1}{n} \sum y_i \quad \bar{y}' = \frac{1}{m} \sum y'_j$$

Las ecuaciones (2) y (3) son las mismas que las (13-2-8) y (13-2-9); la ecuación (5) proporciona la estimación puntual deseada de x_0 .

También es fácil construir un intervalo confidencial para x_0 . La cantidad

$$v = \bar{y}' - \hat{\alpha} - \hat{\beta}x_0 \quad (6)$$

se distribuye normalmente, por ser función lineal de variantes normales; su media es cero, y su varianza,

$$\sigma_v^2 = \sigma^2 \left[\frac{1}{m} + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum (x_i - \bar{x})^2} \right] \quad (7)$$

según se halló en la sección 13-3. Las dos sumas de (4) tienen distribuciones ji cuadrado cuando se dividen por σ^2 , la primera con $n-2$ y la segunda con $m-1$ grados de libertad. Estas dos distribuciones son independientes por ser funciones de muestras independientes; por tanto, su suma se distribuye según una ji cuadrado con $m+n-3$ grados de libertad. Además, ambas distribuciones son evidentemente independientes de v . Se deduce, pues, que

$$t = \frac{v/\sigma_v}{\sqrt{(m+n)\hat{\sigma}^2/(m+n-3)\sigma^2}} \quad (8)$$

tiene la distribución t y proporciona un intervalo confidencial para x_0 , por ser el único parámetro desconocido que aparece en (8).

Hemos considerado un problema de discriminación muy simplificado, pero que surge con frecuencia en la práctica. El problema

general está relacionado con el caso de que cada observación conste de varias componentes ($y_1, y_2, \dots, y_k$) con una distribución normal multivariante, de medias

$$\alpha_1 + \beta_1 x, \alpha_2 + \beta_2 x, \dots, \alpha_k + \beta_k x$$

Dadas estimaciones de las α y las β , basadas en una muestra de observaciones ($y_{1i}, y_{2i}, \dots, y_{ki}$), se desea estimar x_0 para una sola observación ($y_{10}, y_{20}, \dots, y_{k0}$). Tendremos que abandonar este problema, debido a que su resolución resulta muy laboriosa mediante métodos elementales.

13.5. Estimación puntual. Caso B.—Sea el modelo

$$y = \alpha + \beta x + e$$

en donde y es un variable aleatoria observable con $E(y) = \alpha + \beta x$; x es una variable matemática (no aleatoria) conocida y α y β , parámetros desconocidos. Se eligen n valores de x y se selecciona para cada x_i un valor y_i de una densidad de media $\alpha + \beta x_i$ y cuya varianza es σ^2 .

Haremos las siguientes hipótesis:

Las variables aleatorias $y_1, y_2, \dots, y_n$ son variantes incorrelacionadas, con medias $E(y_i) = \alpha + \beta x_i$ y varianza σ^2 , independiente de x_i, α y β . Esto implica que las e_i son variables aleatorias incorrelacionadas, con $E(e_i) = 0$ y $\text{var}(e_i) = \sigma^2$, lo que puede escribirse

$$\begin{aligned} E(y_i) &= \alpha + \beta x_i \\ \text{cov}(y_i, y_j) &= 0 \quad \text{si } j \neq i \\ \text{var}(y_i) &= \sigma^2 \end{aligned} \tag{1}$$

Puesto que la densidad de las variables aleatorias y_i (como la de e_i) no está determinada, no es posible obtener estimadores de máxima verosimilitud de α y β . En tales situaciones cabe recurrir al método de estimación de *mínimos cuadrados*.

Definición 13-2.—Sean n puntos (y_i, x_i), $i=1, 2, \dots, n$, que satisfacen el modelo lineal simple $y_i = \alpha + \beta x_i + e_i$. Los estimadores mínimo-cuadráticos de α y β son los valores de α y β que hacen

mínima la suma $\sum_{i=1}^n e_i^2$ de cuadrados de los errores.

Sustituyendo e_i , se obtiene

$$L(\alpha, \beta) = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \alpha - \beta x_i)^2$$

Evidentemente, los valores de α y β que hacen mínima esta expresión son los mismos que hacen máxima a L en la ecuación (13-2-2). Por tanto, los estimadores mínimo-cuadráticos son:

$$\begin{aligned}\hat{\alpha} &= \bar{y} - \hat{\beta}\bar{x} \\ \hat{\beta} &= \frac{\sum(y_i - \bar{y})(x_i - \bar{x})}{\sum(x_i - \bar{x})^2}\end{aligned}\quad (2)$$

El método de mínimos cuadrados no proporciona estimador de σ^2 , pero un estimador para σ^2 , basado en los estimadores mínimos cuadrados de α y β , es

$$\hat{\sigma}^2 = \frac{1}{n-2} \sum (y_i - \hat{\alpha} - \hat{\beta}x_i)^2 \quad (3)$$

Si se supone que las y_i se distribuyen normalmente, el estimador de máxima verosimilitud de σ^2 es $[(n-2)/n]\hat{\sigma}^2$, si bien este estimador es sesgado. Cuando las y_i son variables normales, las propiedades de los estimadores de máxima verosimilitud (corregidos de sesgo) de α , β y σ^2 se han dado en el teorema 13-1.

La primera de dichas propiedades es que los estimadores son insesgados de varianza mínima. Es decir, en la clase de todos los estimadores insesgados de α , β , σ^2 , los estimadores $\hat{\alpha}$, $\hat{\beta}$, $\hat{\sigma}^2$ de (3) tienen varianza mínima. Los estimadores mínimo-cuadráticos (2) y (3) no gozan de tal propiedad; pero, a continuación, se define una propiedad equivalente.

Definición 13-3.—Consideremos la clase de los estimadores de α y β que son funciones lineales de las variables aleatorias y_i . En tal clase, solo tendremos en cuenta la subclase de los estimadores insesgados. Si existen, en esta clase restringida, estimadores de α y β con varianza menor que cualesquiera otros estimadores de α y β , los llamaremos «estimadores insesgados lineales óptimos» (óptimo se refiere a varianza mínima).

Demostraremos un importante teorema para el modelo lineal simple, en el caso de ser válidas las hipótesis (1). Esta proposición se suele conocer con el nombre de teorema de Gauss-Markoff.

Teorema 13-2.—Sea el modelo $y_i = \alpha + \beta x_i + e_i$ tal que las x_i son variables matemáticas conocidas, y las y_i , variables aleatorias observables. Supongamos también que las variables aleatorias no observables e_i están incorrelacionadas y tienen media 0 y varianza σ^2 . Los estimadores mínimo-cuadráticos (2) de α y β son estimadores insesgados lineales óptimos.

Demostración.—Desarrollaremos la demostración solo para α , dado que la correspondiente a β es análoga. Puesto que hemos res-

tringido la clase de estimadores a los lineales, se tiene que $\hat{\alpha} = \sum a_p y_p$. Hay que determinar las constantes a_p tales que:

a) $E(\hat{\alpha}) = \alpha$, es decir, insesgado.

b) Varianza ($\hat{\alpha}$) es un mínimo para todos los estimadores que satisfagan a a).

Por a),

$$\alpha = E(\hat{\alpha}) = \sum a_p (E y_p) = \sum a_p (\alpha + \beta x_p)$$

Esto proporciona las dos ecuaciones que deben cumplirse:

$$\begin{aligned} \sum a_p &= 1 \\ \sum a_p x_p &= 0 \end{aligned} \quad (4)$$

Ahora bien:

$$\begin{aligned} \text{var}(\hat{\alpha}) &= E(\hat{\alpha} - \alpha)^2 = E(\sum a_p y_p - \alpha)^2 = E[\sum a_p (\alpha + \beta x_p + e_p) - \alpha]^2 \\ &= E(\alpha \sum a_p + \beta \sum a_p x_p + \sum a_p e_p - \alpha)^2 \end{aligned}$$

Por la restricción (4), esta expresión se transforma en

$$\text{var}(\hat{\alpha}) = E \left(\sum a_p e_p \right)^2 = E \left(\sum_p a_p^2 e_p^2 + \sum_{\substack{p, q \\ p \neq q}} a_p a_q e_p e_q \right)$$

La cantidad $E(e_p e_q)$ es 0 si $p \neq q$, ya que, por hipótesis, las e_i están incorrelacionadas y sus medias son iguales a cero. Por tanto,

$$\text{var}(\hat{\alpha}) = \sigma^2 \sum a_p^2$$

Puesto que σ^2 es constante, para hacer mínima $\text{var}(\hat{\alpha})$ se precisa que lo sea también $\sum a_p^2$. Por tanto, hay que hallar las constantes a_p que hacen mínima a $\sum a_p^2$, sujetas a las restricciones (4). Empleando la teoría de los multiplicadores de Lagrange, tendremos que minimizar la expresión

$$L = \sum a_p^2 - \lambda_1 (\sum a_p - 1) - \lambda_2 (\sum a_p x_p)$$

Tomando derivadas, resulta

$$\begin{aligned} \frac{\partial L}{\partial a_t} &= 2a_t - \lambda_1 - \lambda_2 x_t = 0 & t = 1, 2, \dots, n \\ \frac{\partial L}{\partial \lambda_1} &= - \sum a_p + 1 = 0 \\ \frac{\partial L}{\partial \lambda_2} &= - \sum a_p x_p = 0 \end{aligned} \quad (5)$$

Si sumamos las n primeras ecuaciones, obtenemos (teniendo en cuenta que $\sum a_i = 1$)

$$2 = n\lambda_1 + \lambda_2 \sum x_i \quad (6)$$

Si multiplicamos la ecuación p -ésima de (5) por x_p y sumamos, se deduce que

$$2\sum x_p a_p = \lambda_1 \sum x_p + \lambda_2 \sum x_p^2$$

o bien, puesto que $\sum a_p x_p = 0$,

$$\lambda_1 = -\lambda_2 \frac{\sum x_i^2}{\sum x_i} \quad (7)$$

Si sustituimos este valor en (6), resulta

$$\lambda_2 = \frac{-2\sum x_i/n}{\sum x_i^2 - n\bar{x}^2} = \frac{-2\bar{x}}{\sum (x_i - \bar{x})^2}$$

y

$$\lambda_1 = \frac{2\sum x_i^2/n}{\sum (x_i - \bar{x})^2}$$

Sustituyendo λ_1 y λ_2 en la ecuación t -ésima de (5) y despejando a_t , se tiene

$$a_t = \frac{\sum x_i^2/n - \bar{x}x_t}{\sum (x_i - \bar{x})^2}$$

El estimador insesgado lineal óptimo de α es, por tanto,

$$\hat{\alpha} = \sum a_t y_t = \frac{\bar{y}\sum x_i^2 - \bar{x}\sum y_i x_i}{\sum (x_i - \bar{x})^2} = \bar{y} - \hat{\beta}\bar{x}$$

que es el dado por el método de mínimos cuadrados, con lo que queda terminada la demostración. Una demostración análoga es válida también para β .

*** 13-6. El modelo lineal general.**—En esta sección estudiaremos lo que a veces se denomina «regresión múltiple». Deducire-

* En esta sección se utilizan matrices con bastante frecuencia. El lector no familiarizado con la teoría de matrices deberá omitirla.

mos las distribuciones de estadísticos adecuados para la estimación y docimasia de ciertos parámetros de los modelos lineales generales.

Consideremos la función de densidad de una variable aleatoria y , $f(y; x_1, x_2, \dots, x_p; \beta_1, \beta_2, \dots, \beta_p)$, que depende de p cantidades conocidas $x_1, x_2, \dots, x_p$ y de p parámetros desconocidos β_i . Esta función de densidad se representará por $f(y; x; \beta)$, $f(y; \beta)$ o $f(y)$. En todo lo que sigue supondremos que

$$E(y) = \sum_{i=1}^p \beta_i x_i$$

y que la varianza de y es igual a σ^2 , que no depende de las x_i ni de las β_i . El punto

$$P'_j = (y_j, x_{j1}, x_{j2}, \dots, x_{jp})$$

del espacio de $(p+1)$ dimensiones representará una observación de esta distribución; es decir, cuando se hace una observación y , se supondrán dados los valores x_i correspondientes. En toda la sección, las x_i se considerarán constantes conocidas, no variables aleatorias.

Si en $f(y; x; \beta)$ hacemos la transformación $e = y - \sum \beta_i x_i$, e será una variable aleatoria tal que $E(e) = 0$ y $E(e)^2 = \sigma^2$. Esto puede escribirse

$$y = \sum_{i=1}^p \beta_i x_i + e \quad (1)$$

y constituye una generalización del modelo lineal simple de las secciones anteriores.

Para estimar las β_i , se tomará una muestra aleatoria de tamaño n procedente de la distribución $f(y; x; \beta)$. La muestra se designará por $P'_1, P'_2, \dots, P'_n$, y la relación entre el sistema de observaciones podrá escribirse:

$$y_j = \sum_{i=1}^p \beta_i x_{ji} + e_j \quad j = 1, 2, \dots, n \quad (2)$$

o bien, en forma de vector,

$$Y = X\beta + e \quad (3)$$

en donde

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \quad X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix}$$

$$\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix} \quad e = \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix} \quad (4)$$

Si examinamos $Y = X\beta + e$ con más detalle, vemos que debemos seleccionar primero (bien sea al azar o por diseño) un conjunto de $x, x_{11}, x_{12}, \dots, x_{1p}$, y elegir al azar una observación y_1 de la distribución

$$f(y; x_1 = x_{11}, x_2 = x_{12}, \dots, x_p = x_{1p}).$$

A continuación se selecciona otro conjunto de $x, x_{21}, x_{22}, \dots, x_{2p}$ y se toma al azar una observación y_2 de la distribución

$$f(y; x_1 = x_{21}, x_2 = x_{22}, \dots, x_p = x_{2p}).$$

Se repite este proceso hasta que se hayan extraído n valores de y . Todavía no se ha hecho ninguna estipulación sobre las x_{ij} , y no es necesario que sean todas distintas. Sin embargo, en casos especiales, pondremos algunas restricciones a la matriz de las x_{ij} . Por tanto, siempre que aparezca el modelo $Y = X\beta + e$ se supondrá que se construyó por el proceso de muestreo antes definido. Se deducirán estimadores de las β_i y de σ sobre la base de la matriz observada X y del vector aleatorio observado Y .

En la siguiente definición se formula lo anterior:

Definición 13-4.—El modelo $Y = X\beta + e$ [en donde las cantidades dadas en (4) son tales que Y es un vector observado aleatorio; e , un vector aleatorio; X , una matriz $n \times p$ de cantidades fijas conocidas, y β , un vector $p \times 1$ de parámetros desconocidos] se denomina *modelo lineal general*.

Si es p la característica de X , el modelo se dice que es de característica completa. Examinaremos a continuación dos casos relacionados con la distribución del vector e .

Caso A: $\mathbf{e}$ se distribuye según la normal p -variante con media $\mathbf{0}$ y covarianza $\sigma^2 I$.

Caso B: $\mathbf{e}$ es un vector aleatorio tal que $E(\mathbf{e}) = \mathbf{0}$ y

$$\text{cov}(\mathbf{e}) = E(\mathbf{e}\mathbf{e}') = \sigma^2 I$$

en donde σ^2 es desconocida [$\text{cov}(\mathbf{e})$ designa la matriz de covarianzas de $\mathbf{e}$].

El caso A equivale a decir que cada $\mathbf{e}_i$ se distribuye normalmente con media cero y varianza igual a σ^2 , y que las $\mathbf{e}_i$ son conjuntamente independientes. El caso B es equivalente a afirmar que el valor esperado de cada $\mathbf{e}_i$ es 0, que las $\mathbf{e}_i$ están incorrelacionadas y que cada $\mathbf{e}_i$ tiene varianza común desconocida σ^2 . El A suele denominarse caso de la teoría normal.

Estimación puntual.—Consideraremos separadamente para uno y otro caso las estimaciones puntuales de los parámetros en el modelo lineal general.

Caso A: Estimaciones de β y σ^2 bajo la teoría normal.—Puesto que estamos estudiando el caso en que el vector de errores $\mathbf{e}$ se distribuye normalmente, utilizaremos el método de máxima verosimilitud para estimar $\beta_1, \beta_2, \dots, \beta_p$ y σ^2 . La ecuación de verosimilitud es

$$f(\mathbf{e}; \beta, \sigma^2) = \frac{1}{(2\pi)^{n/2} \sigma^n} e^{-\mathbf{e}'\mathbf{e}/2\sigma^2} = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-[(\mathbf{Y}-\mathbf{X}\beta)'(\mathbf{Y}-\mathbf{X}\beta)/2\sigma^2]} \quad (5)$$

Tomando logaritmos, se tiene

$$\log f(\mathbf{e}; \beta, \sigma^2) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} (\mathbf{Y} - \mathbf{X}\beta)'(\mathbf{Y} - \mathbf{X}\beta)$$

Las estimaciones máximo-verosímiles de β y σ^2 son las soluciones de las ecuaciones

$$\begin{aligned} \frac{\partial \log f(\mathbf{e}; \beta, \sigma^2)}{\partial \beta_1} &= 0 \\ \frac{\partial \log f(\mathbf{e}; \beta, \sigma^2)}{\partial \beta_2} &= 0 \\ &\dots\dots\dots \\ \frac{\partial \log f(\mathbf{e}; \beta, \sigma^2)}{\partial \beta_p} &= 0 \\ \frac{\partial \log f(\mathbf{e}; \beta, \sigma^2)}{\partial \sigma^2} &= 0 \end{aligned} \quad (6)$$

Para hallar las derivadas de

$$(\mathbf{Y} - \mathbf{X}\beta)'(\mathbf{Y} - \mathbf{X}\beta)$$

respecto a cada β_i , se multiplica y se obtiene

$$(\mathbf{Y} - \mathbf{X}\beta)'(\mathbf{Y} - \mathbf{X}\beta) = \mathbf{Y}'\mathbf{Y} - 2\mathbf{Y}'\mathbf{X}\beta + \beta'\mathbf{X}'\mathbf{X}\beta$$

ya que $\beta'\mathbf{X}'\mathbf{Y}$ es un escalar y, por tanto, igual a $\mathbf{Y}'\mathbf{X}\beta$. Escribiremos esto en la forma

$$\begin{aligned} (\mathbf{Y} - \mathbf{X}\beta)'(\mathbf{Y} - \mathbf{X}\beta) &= \mathbf{Y}'\mathbf{Y} + 2\mathbf{A}\beta + \beta'\mathbf{S}\beta \\ &= \mathbf{Y}'\mathbf{Y} + 2 \sum_{i=1}^p a_i \beta_i + \sum_{i=1}^p \sum_{\substack{j=1 \\ i \neq j}}^p \beta_i \beta_j s_{ij} + \sum_{i=1}^p \beta_i^2 s_{ii} \end{aligned}$$

en donde a_i es el elemento i -ésimo de $\mathbf{A}$ (igual a $-\mathbf{Y}'\mathbf{X}$), y s_{ij} , el ij -ésimo elemento de $\mathbf{X}'\mathbf{X}$. Si tomamos derivadas respecto a β_t , se obtiene

$$\begin{aligned} \frac{\partial[(\mathbf{Y} - \mathbf{X}\beta)'(\mathbf{Y} - \mathbf{X}\beta)]}{\partial \beta_t} &= 2a_t + \sum_{\substack{i=1 \\ i \neq t}}^p \beta_i s_{it} + \sum_{\substack{j=1 \\ j \neq t}}^p \beta_j s_{tj} + 2\beta_t s_{tt} \\ &= 2a_t + 2 \sum_{i=1}^p \beta_i s_{it} \end{aligned}$$

Si se hace lo mismo para cada t ($t=1, 2, \dots, p$) y se iguala a cero el resultado, obtenemos:

$$\frac{\partial[(\mathbf{Y} - \mathbf{X}\beta)'(\mathbf{Y} - \mathbf{X}\beta)]}{\partial \beta_1} = 2a_1 + 2 \sum_i \beta_i s_{i1} = 0$$

$$\frac{\partial[(\mathbf{Y} - \mathbf{X}\beta)'(\mathbf{Y} - \mathbf{X}\beta)]}{\partial \beta_2} = 2a_2 + 2 \sum_i \beta_i s_{i2} = 0$$

$$\dots\dots\dots \frac{\partial[(\mathbf{Y} - \mathbf{X}\beta)'(\mathbf{Y} - \mathbf{X}\beta)]}{\partial \beta_p} = 2a_p + 2 \sum_i \beta_i s_{ip} = 0$$

que da

$$2\mathbf{A}' + 2\mathbf{S}\beta = 0$$

o bien

$$-2X'Y + 2X'X\beta = 0$$

Si designamos por $\hat{\beta}$ y $\hat{\sigma}^2$ las soluciones de las ecuaciones resultantes, se tiene

$$X'X\hat{\beta} = X'Y$$

y

$$\hat{\sigma}^2 = \frac{1}{n} (Y - X\hat{\beta})'(Y - X\hat{\beta})$$

La relación matricial $X'X\hat{\beta} = X'Y$ se llama ecuación normal y desempeña un papel de la máxima importancia en nuestra teoría. Puesto que X tiene característica p , también es p la característica de $X'X$ y, por tanto, existe su inversa. En consecuencia,

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_p \end{pmatrix} = S^{-1}X'Y$$

en donde $S = X'X$. Puesto que $\hat{\beta}$ y $\hat{\sigma}^2$ son estimadores de máxima verosimilitud, son consistentes y asintóticamente eficientes, pero habrá que comprobar si son insesgados o suficientes. Para ver si $\hat{\beta}$ es insesgado, procederemos como sigue:

$$E(\hat{\beta}) = E(S^{-1}X'Y) = S^{-1}X'E(Y) = S^{-1}X'E(X\beta + e) = S^{-1}X'X\beta = \beta$$

Luego $\hat{\beta}$ es una estimación insesgada de β . Omitimos la demostración, pero puede probarse que

$$\tilde{\sigma}^2 = \frac{n}{n-p} \hat{\sigma}^2 = \frac{(Y - X\hat{\beta})'(Y - X\hat{\beta})}{n-p}$$

es un estimador insesgado de σ^2 .

Examinemos a continuación la propiedad de suficiencia. La frecuencia conjunta de las e_i puede escribirse:

$$f(e) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp \left[-\frac{1}{2\sigma^2} (Y - X\beta)'(Y - X\beta) \right]$$

La identidad en β

$$(Y - X\beta)'(Y - X\beta) = (Y - X\hat{\beta})'(Y - X\hat{\beta}) + (\beta - \hat{\beta})'S(\beta - \hat{\beta})$$

puede establecerse fácilmente. Utilizando esta identidad, la densidad conjunta es

$$\begin{aligned} f(e) &= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp \left[-\frac{(Y - X\beta)'(Y - X\beta)}{2\sigma^2} \right] \\ &= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp \left[-\frac{(n-p)\hat{\sigma}^2 + (\beta - \hat{\beta})'S(\beta - \hat{\beta})}{2\sigma^2} \right] \end{aligned}$$

y empleando la definición de suficiencia del capítulo 8, resulta evidente que $\hat{\sigma}^2$, $\hat{\beta}_1$, $\hat{\beta}_2$, ..., $\hat{\beta}_p$ forman un conjunto de estimadores que son conjuntamente suficientes para σ^2 , β_1 , β_2 , ..., β_p . No daremos la demostración, pero puede probarse que los estimadores $\hat{\sigma}^2$, $\hat{\beta}_1$, $\hat{\beta}_2$, ..., $\hat{\beta}_p$ son completos.

Puesto que $\hat{\beta}$ es igual al producto de una matriz constante $S^{-1}X'$ y de un vector Y que se distribuye normalmente, por medio del problema 9-21 se demuestra que $\hat{\beta}$ tiene la distribución normal p -variante. Hemos visto ya que la media de $\hat{\beta}$ es β . La matriz de covarianzas de $\hat{\beta}$ es

$$\text{cov}(\hat{\beta}) = E(\hat{\beta} - \beta)(\hat{\beta} - \beta)' = E(S^{-1}X'Y - \beta)(S^{-1}X'Y - \beta)'$$

Si sustituimos Y por $X\beta + e$, resulta

$$\begin{aligned} \text{cov}(\hat{\beta}) &= E[S^{-1}X'(X\beta + e) - \beta][S^{-1}X'(X\beta + e) - \beta]' = E(S^{-1}X'e)(S^{-1}X'e)' \\ &= E(S^{-1}X'ee'XS^{-1}) = S^{-1}X'E(ee')XS^{-1} = \sigma^2 S^{-1}X'XS^{-1} = \sigma^2 S^{-1} \end{aligned}$$

Luego $\hat{\beta}$ se distribuye según la normal p -variante, con media β y matriz de covarianzas $\sigma^2 S^{-1}$. Aunque omitimos la demostración, puede probarse que

$$(n-p) \frac{\hat{\sigma}^2}{\sigma^2} = \frac{(Y - X\hat{\beta})'(Y - X\hat{\beta})}{\sigma^2}$$

se distribuye según una ji cuadrado con $n-p$ grados de libertad.

Lo anterior queda resumido en el teorema siguiente:

Teorema 13-3.—Si $Y = X\beta + e$ es un modelo lineal general de característica completa, y si e se distribuye según la normal p -variante con media 0 y matriz de covarianzas $\sigma^2 I$,

$$\hat{\beta} = S^{-1}X'Y \quad \hat{\sigma}^2 = \frac{Y'(I - XS^{-1}X')Y}{n-p}$$

gozan de las siguientes propiedades:

1. Consistentes.
2. Asintóticamente eficientes.
3. Insesgados de varianza mínima.

4. Suficientes, completos.
5. $\hat{\beta}$ se distribuye según la normal p -variante con media β y matriz de covarianzas $\sigma^2 S^{-1}$.
6. $(n-p)\hat{\sigma}^2/\sigma^2$ se distribuye según la ji cuadrado con $n-p$ grados de libertad.
7. $\hat{\beta}$ y $\hat{\sigma}^2$ son independientes.

Estimación de β y σ^2 , caso B: El teorema de Gauss-Markoff.— Como en la sección 13-5, utilizaremos el método de mínimos cuadrados para estimar β cuando no se conoce la función de densidad de $\mathbf{Y}$. Es decir, hallaremos el valor $\hat{\beta}$ de β tal que sea mínima la suma de cuadrados

$$\sum_{i=1}^n e_i^2$$

Esto da

$$\sum_{i=1}^n e_i^2 = \mathbf{e}'\mathbf{e} = (\mathbf{Y} - \mathbf{X}\beta)'(\mathbf{Y} - \mathbf{X}\beta)$$

El valor de β que hace mínimo $\mathbf{e}'\mathbf{e}$ es la solución de

$$\frac{\partial(\mathbf{e}'\mathbf{e})}{\partial\beta_i} = 0 \quad t = 1, 2, \dots, p$$

La resolución de estas ecuaciones da (véase caso A)

$$2\mathbf{X}'\mathbf{Y} - 2\mathbf{X}'\mathbf{X}\hat{\beta} = \mathbf{0}.$$

La estimación de β por el método de mínimos cuadrados es, por tanto,

$$\hat{\beta} = S^{-1}\mathbf{X}'\mathbf{Y}$$

que, naturalmente, es la misma que la estimación máximo-verosímil de la teoría normal. Hacer mínima la suma de cuadrados, $\mathbf{e}'\mathbf{e}$, no proporciona una estimación de σ^2 . Sin embargo, la estimación de σ^2 , basada en la estimación de β por mínimos cuadrados e insesgada, viene dada por

$$\hat{\sigma}^2 = \frac{(\mathbf{Y} - \mathbf{X}\hat{\beta})'(\mathbf{Y} - \mathbf{X}\hat{\beta})}{n-p} = \frac{\mathbf{Y}'(\mathbf{I} - \mathbf{X}S^{-1}\mathbf{X}')\mathbf{Y}}{n-p}$$

La consideración que sigue se refiere a las *propiedades* de los estimadores obtenidos por mínimos cuadrados. Puesto que no está

determinada la distribución del vector aleatorio $\mathbf{e}$, en general, no será posible examinar la «bondad» del estimador $\hat{\beta}$ respecto a todas las funciones, sino que deberemos limitarnos a un subconjunto de estas; p. ej., ya que $\hat{\beta}$ es función lineal de las y_i , examinaremos $\hat{\beta}$ para ver qué propiedades posee respecto a todas las funciones *lineales* de las y_i . Cuando el vector $\mathbf{e}$ se distribuye normalmente, demostramos que $\hat{\beta}_i$ tiene varianza menor que cualquier otro estimador insesgado de β_i , independientemente de la función de las y_i que se utilice como estimador de β .

Para el estimador de mínimos cuadrados no es posible hacer una afirmación tan categórica, pero sí podemos comparar la «bondad» de $\hat{\beta}_i$ con otros estimadores que sean funciones *lineales* de las observaciones y_i . No obstante, si nos interesase como estimador de β_i una función general $h(y_1, y_2, \dots, y_n)$, en condiciones bastante generales podríamos entonces desarrollar la función $h(y_1, y_2, \dots, y_n)$ en serie de Taylor y tomar como primera aproximación el término lineal del desarrollo. Aunque hay varias razones que aconsejan restringirnos a las funciones lineales de las observaciones y_i como estimadores de β_i , existen otras muchas por las que *no* deberíamos proceder así. Un teorema que afirma la calidad de las estimaciones del método de mínimos cuadrados en el modelo $\mathbf{Y} = \mathbf{X}\beta + \mathbf{e}$ es el de Gauss-Markoff.

Teorema 13-4.—*Si el modelo lineal general $\mathbf{Y} = \mathbf{X}\beta + \mathbf{e}$ es tal que se dan las siguientes condiciones acerca del vector aleatorio $\mathbf{e}$:*

- a) $E(\mathbf{e}) = 0$
- b) $E(\mathbf{e}\mathbf{e}') = \sigma^2 \mathbf{I}$

la estimación insesgada lineal (funciones lineales de las y_i) óptima (varianza mínima) de β es la dada por el método de mínimos cuadrados; es decir, $\hat{\beta} = \mathbf{S}^{-1}\mathbf{X}'\mathbf{Y}$ es la estimación insesgada lineal óptima de β .

Omitimos la demostración de este teorema, que, en líneas generales, coincide con la del teorema de Gauss-Markoff para el modelo lineal simple.

Hasta ahora la discusión se ha centrado en la estimación de las β_i . Sin embargo, interesa también con frecuencia estimar ciertas funciones de las β_i . Por la propiedad de invarianza de los estimadores de máxima verosimilitud, este procedimiento es válido para obtener el estimador máximo-verosímil de cualquier función de las β_i con inversa única. La situación no es tan simple en el caso de estimadores dados por el método de mínimos cuadrados. Hay, sin embargo, un caso importante en el que es válida la propiedad invariante para los estimadores de mínimos cuadrados; es el caso de fun-

ciones lineales de las β_i . En el teorema siguiente se precisa esta cuestión.

Teorema 13-5.—*En el modelo lineal general dado en el teorema 13-4, la estimación insesgada lineal óptima de cualquier combinación lineal de las β_i es la misma combinación lineal de las estimaciones insesgadas lineales óptimas de las β_i . Es decir, la estimación insesgada lineal óptima de $t'\beta$ (siendo t un vector constante $p \times 1$) es $t'\hat{\beta}$, en donde $\hat{\beta}$ es la estimación insesgada lineal óptima de β ; o sea, $t'\hat{\beta} = t'S^{-1}X'Y$.*

La demostración es análoga a la del teorema 13-4 y se propone como ejercicio al lector.

Ejemplo 13-3.—Un modelo lineal simple está dado por

$$y_i = \beta_1 + \beta_2 x_i + e_i \quad i = 1, 2, \dots, n$$

en donde β_1 y β_2 son constantes escalares desconocidas, y las x_i , constantes escalares dadas. Por tanto, mediante la ecuación (4), encontramos que ($p=2$)

$$X = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \quad \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$$

Es fácil ver que

$$S = X'X = \begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix}$$

$$S^{-1} = \frac{1}{n\sum(x_i - \bar{x})^2} \begin{pmatrix} \sum x_i^2 & -\sum x_i \\ \sum x_i & n \end{pmatrix}$$

y

$$X'Y = \begin{pmatrix} \sum y_i \\ \sum x_i y_i \end{pmatrix}$$

Luego

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = S^{-1}X'Y = \frac{1}{n\sum(x_i - \bar{x})^2} \begin{pmatrix} (\sum x_i^2)(\sum y_i) - (\sum x_i)(\sum x_i y_i) \\ -(\sum x_i)(\sum y_i) + n\sum y_i x_i \end{pmatrix}$$

o bien

$$\hat{\beta}_2 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2} \quad \hat{\beta}_1 = \bar{y} - \hat{\beta}_2 \bar{x}$$

β_2 es la pendiente, y β_1 , el valor de $E(y)$ para $x=0$.

Puesto que $\text{cov}(\hat{\beta}) = S^{-1}\sigma^2$, vemos que

$$\text{cov}(\hat{\beta}_1, \hat{\beta}_2) = -\frac{\sigma^2 \sum x_i}{n \sum (x_i - \bar{x})^2}$$

$$\text{var } \hat{\beta}_1 = \frac{\sigma^2 \sum x_i^2}{n \sum (x_i - \bar{x})^2}$$

y

$$\text{var } \hat{\beta}_2 = \frac{\sigma^2}{\sum (x_i - \bar{x})^2}$$

Estos resultados son exactamente los dados en la sección 13-2; se han repetido aquí para mostrar que también pueden emplearse matrices en aquel caso.

Estimación de β_i por intervalos.—Dedicaremos ahora nuestra atención al problema de la estimación por intervalos de β_i , aunque solo para el caso A.

Para hallar un intervalo confidencial $1-\alpha$ de β_i , utilizamos el hecho de que la distribución de $\hat{\beta}_i$ es $n(\beta_i, c_{ii}\sigma^2)$, en donde c_{ii} es el elemento ij -ésimo de $C=S^{-1}$. Luego,

$$y = \frac{\hat{\beta}_i - \beta_i}{\sigma \sqrt{c_{ii}}}$$

se distribuye según la $n(y; 0, 1)$ y es independiente de $(n-p)\hat{\sigma}^2/\sigma^2$, la cual tiene una distribución ji-cuadrado con $n-p$ grados de libertad. Se deduce que

$$u = \frac{\hat{\beta}_i - \beta_i}{\sigma \sqrt{c_{ii}}} \sqrt{\frac{\sigma^2}{\hat{\sigma}^2}} = \frac{\hat{\beta}_i - \beta_i}{\sqrt{\hat{\sigma}^2 c_{ii}}}$$

se distribuye según la t de Student con $n-p$ grados de libertad. Así,

$$\int_{-t_{\alpha/2}}^{t_{\alpha/2}} t(u) du = P \left(-t_{\alpha/2} < \frac{\hat{\beta}_i - \beta_i}{\sqrt{\hat{\sigma}^2 c_{ii}}} < t_{\alpha/2} \right) = 1 - \alpha$$

donde $t(u)$ es la función de densidad de la t de Student. Después de algunas transformaciones se obtiene

$$P(\hat{\beta}_i - t_{\alpha/2} \sqrt{c_{ii} \hat{\sigma}^2} < \beta_i < \hat{\beta}_i + t_{\alpha/2} \sqrt{c_{ii} \hat{\sigma}^2}) = 1 - \alpha; \quad (7)$$

la cantidad entre paréntesis es un intervalo confidencial $1-\alpha$ de β_i , y la longitud de este es $2t_{\alpha/2} \sqrt{c_{ii} \hat{\sigma}^2}$.

Décima de la hipótesis $\beta_1 = \beta_2 = \dots = \beta_r = 0$ ($r \leq p$) en el modelo $Y = X\beta + e$ de la definición 13-4.—Esta es una décima muy útil e importante; p. ej., para el caso $p=3$, el modelo es

$$y = x_1\beta_1 + x_2\beta_2 + x_3\beta_3 + \epsilon$$

Supongamos ahora que un experimentador desea saber si puede reemplazar dicho modelo por este otro:

$$y = x_1\beta_1 + x_2\beta_2 + \epsilon$$

En otras palabras, quiere docimar la hipótesis $\beta_3 = 0$ en el modelo sin establecer ninguna condición acerca del valor de β_1 o del de β_2 . Enunciaremos un teorema que permitirá docimar cualquier conjunto de las β_i igual a cero en el modelo lineal general $Y = X\beta + e$.

En el modelo

$$Y = X\beta + e$$

descomponemos la matriz X y el vector β en la forma

$$X = (X_1, X_2) \quad \beta = \begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix}$$

en donde las dimensiones de X_1 y de γ_1 son $n \times r$ y $r \times 1$, respectivamente. El modelo puede así escribirse

$$Y = X\beta + e = (X_1, X_2) \begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix} + e$$

o bien

$$Y = X_1\gamma_1 + X_2\gamma_2 + e$$

Deseamos una décima de la hipótesis $\gamma_1 = 0$ sin fijar condición alguna acerca de γ_2 . Puesto que

$$\gamma_1 = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_r \end{pmatrix}$$

esta décima equivale a docimar la hipótesis $\beta_1 = \beta_2 = \dots = \beta_r = 0$ ($r \leq p$) en el modelo lineal general:

$$y_j = \sum_{i=1}^p x_{ji}\beta_i + e_j \quad j = 1, 2, \dots, n$$

En el teorema que sigue se formula la d cima de la hip tesis $H_0: \gamma_1 = \mathbf{0}$ y su construcci n:

Teorema 13-6.—*Particionemos el modelo $Y = X\beta + e$, dado en la definici n 13-1, de modo que*

$$Y = X_1\gamma_1 + X_2\gamma_2 + e$$

en donde γ_1 es de dimensi n $r \times 1$, y supongamos que e se distribuye seg n la normal p -variante con media $\mathbf{0}$ y matriz de covarianzas $\sigma^2 I$. Para docimar entonces la hip tesis $H_0: \gamma_1 = \mathbf{0}$ por la d cima de la raz n de verosimilitud se procede del siguiente modo:

1. Se obtienen las ecuaciones normales $X'X\hat{\beta} = X'Y$, y a partir de ellas se calcula $\hat{\beta}'X'Y$.

2. A partir del modelo $Y = X_2\gamma_2 + e$ (llamado modelo restringido por la hip tesis H_0) se obtienen las ecuaciones normales $X_2'X_2\tilde{\gamma}_2 = X_2'Y$ y por medio de ellas se calcula $\tilde{\gamma}_2'X_2'Y$. Obs rvase que las ecuaciones normales reducidas se deducen de las ecuaciones normales originales, $X'X\hat{\beta} = X'Y$, haciendo $\hat{\beta}_1 = \hat{\beta}_2 = \dots = \hat{\beta}_r = 0$ y utilizando  nicamente las  ltimas $p - r$ ecuaciones.

3. Se calcula

$$u = \frac{\hat{\beta}'X'Y - \tilde{\gamma}_2'X_2'Y}{Y'Y - \hat{\beta}'X'Y} \frac{n-p}{r} \quad (8)$$

y se rechaza H_0 si $u > F_{\alpha}(r, n-p)$, en donde $F_{\alpha}(r, n-p)$ es el punto porcentual 100α superior de la distribuci n F con r y $n-p$ grados de libertad. Esto da una probabilidad α de error del tipo 1.

La demostraci n de este teorema exige utilizar recursos que sobrepasan el alcance de este libro; en la referencia 3 de la bibliograf a del cap tulo se da una demostraci n.

Las cantidades que intervienen en este teorema se suelen disponer en una tabla an loga a la 13-1.

TABLA 13-1. TABLA DE AN LISIS DE LA VARIANZA

Fuente	Grados de libertad	Suma de cuadrados	Cuadrado medio	F
Total	n	$Y'Y$		
Reducci�n debida a β	p	$\hat{\beta}'X'Y$		
Reducci�n debida a γ_2	$p-r$	$\tilde{\gamma}_2'X_2'Y$		
Reducci�n debida a γ_1 (ajustada para γ_2)	r	$\hat{\beta}'X'Y - \tilde{\gamma}_2'X_2'Y = rs_1^2$	s_1^2	$\frac{s_1^2}{s_2^2}$
Error	$n-p$	$Y'Y - \hat{\beta}'X'Y = (n-p)s_2^2$	s_2^2	

Ilustraremos este teorema con algunos ejemplos.

Ejemplo 13-4.—Para aclarar el teorema anterior, emplearemos un ejemplo de un modelo $y = \beta_1 + \beta_2 x_2 + \beta_3 x_3 + e$. Las matrices apropiadas para este modelo son

$$X'Y = \begin{pmatrix} \sum y_i \\ \sum x_{2i} y_i \\ \sum x_{3i} y_i \end{pmatrix} \quad X = \begin{pmatrix} 1 & x_{21} & x_{31} \\ 1 & x_{22} & x_{32} \\ 1 & x_{23} & x_{33} \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ 1 & x_{2n} & x_{3n} \end{pmatrix}$$

$$X'X = \begin{pmatrix} n & \sum_i x_{2i} & \sum_i x_{3i} \\ \sum_i x_{2i} & \sum_i x_{2i}^2 & \sum_i x_{2i} x_{3i} \\ \sum_i x_{3i} & \sum_i x_{2i} x_{3i} & \sum_i x_{3i}^2 \end{pmatrix}$$

Obsérvese que $x_{1i} = 1$ para todo i . Por sencillez, utilizaremos datos artificiales. Supongamos que $Y'Y = \sum y_i^2 = 6540$, y que

$$S = X'X = \begin{pmatrix} 16 & 8 & 4 \\ 8 & 6 & 2 \\ 4 & 2 & 6 \end{pmatrix} \quad X'Y = \begin{pmatrix} 80 \\ 120 \\ 40 \end{pmatrix} \quad p=3, n=16 \quad (9)$$

Las ecuaciones normales $X'X\hat{\beta} = X'Y$ son

$$\begin{aligned} 16\hat{\beta}_1 + 8\hat{\beta}_2 + 4\hat{\beta}_3 &= 80 \\ 8\hat{\beta}_1 + 6\hat{\beta}_2 + 2\hat{\beta}_3 &= 120 \\ 4\hat{\beta}_1 + 2\hat{\beta}_2 + 6\hat{\beta}_3 &= 40 \end{aligned} \quad (10)$$

Resulta que

$$C = S^{-1} = \begin{pmatrix} 0,20 & -0,25 & -0,05 \\ -0,25 & 0,50 & 0 \\ -0,05 & 0 & 0,20 \end{pmatrix}$$

y que

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{pmatrix} = S^{-1}X'Y = \begin{pmatrix} 0,20 & -0,25 & -0,05 \\ -0,25 & 0,50 & 0 \\ -0,05 & 0 & 0,20 \end{pmatrix} \begin{pmatrix} 80 \\ 120 \\ 40 \end{pmatrix} = \begin{pmatrix} -16 \\ 40 \\ 4 \end{pmatrix}$$

Luego, las estimaciones son $\hat{\beta}_1 = -16$, $\hat{\beta}_2 = 40$, $\hat{\beta}_3 = 4$ y $\hat{\beta}'X'Y = 3680$, y

$$\hat{\sigma}^2 = \frac{(Y - X\hat{\beta})(Y - X\hat{\beta})}{n - p} = \frac{Y'Y - \hat{\beta}'X'Y}{n - p} = \frac{6540 - 3680}{13} = 220$$

Para docimar la hipótesis $\beta_2 = \beta_3 = 0$, igualamos β_2 y β_3 a cero en el modelo, con lo que las ecuaciones normales reducidas son $X'_2X_2\tilde{y}_2 = X'_2Y$, que, por el teorema 13-6, se deducen a partir de (10) haciendo $\hat{\beta}_2 = \hat{\beta}_3 = 0$ y eliminando las dos últimas ecuaciones (las correspondientes a $\hat{\beta}_2$ y $\hat{\beta}_3$). Así se obtienen las ecuaciones normales reducidas (en este caso particular, una sola)

$$16\tilde{\beta}_1 = 80$$

que da $\tilde{\beta}_1 = 5$. Repetimos: Las ecuaciones normales reducidas pueden hallarse siempre eliminando las filas y columnas de $X'X$ correspondientes a los elementos nulos de la β que se docima, y eliminando las filas de $X'Y$ que corresponden a dichos elementos. Así, en este ejemplo, se trata de docimar que los dos últimos elementos de β son iguales a 0 y, por tanto, suprimimos las dos últimas filas y columnas de $X'X$, quedando 16; tachamos los dos últimos elementos de $X'Y$, y nos queda 80. La ecuación normal reducida correspondiente a $X'_2X_2\tilde{y}_2 = X'_2Y$ es así $n\tilde{\beta}_1 = \Sigma y_i$. A continuación calculamos

$$\tilde{y}_2X'_2Y = \tilde{\beta}_1(\Sigma y_i) = (5)(80) = 400$$

Por consiguiente,

$$\begin{aligned}\hat{\beta}'X'Y &= 3680 \\ \tilde{y}_2X'_2Y &= \tilde{\beta}_1(\Sigma y_i) = 400 \\ \hat{\beta}X'Y - \tilde{y}_2X'_2Y &= 3280\end{aligned}$$

Suma de cuadrados del error = $Y'Y - \hat{\beta}X'Y = 2860$.

La cantidad

$$u = \frac{\frac{3280}{2}}{\frac{2860}{13}} = \frac{1640}{220} = 7,5$$

es mayor que $F_{0,01}$; por tanto, se rechaza la hipótesis al nivel del 1%. Las cantidades pueden resumirse en una tabla de análisis de la varianza (véase tabla 13-2).

TABLA 13-2. TABLA DE ANÁLISIS DE LA VARIANZA PARA DOCIMAR $\beta_2=\beta_3=0$

Fuente	Grados de libertad	Suma de cuadrados	Cuadrado medio	F
Total	16	6540		
Reducción debida a $(\beta_1, \beta_2, \beta_3)$...	3	3680		
Reducción debida a (β_1)	1	400		
Reducción debida a (β_2, β_3) ajustada para β_1	2	3280	1640	7,5
Error	13	2860	220	

Ejemplo 13-5.—Para aclarar aún más la aplicación del teorema 13-6, vamos a docimar la hipótesis $\beta_3=0$ en el ejemplo 13-4. Naturalmente, las ecuaciones normales $\hat{\beta}'X'Y$ son las mismas que en el ejemplo anterior. Luego

$$\hat{\beta}'X'Y=3680,$$

y la suma de cuadrados del error es

$$Y'Y - \hat{\beta}'X'Y=2860.$$

Para hallar las ecuaciones normales reducidas, suprimimos la fila y la columna correspondientes a $\hat{\beta}_3$ (última fila y última columna) de $X'X$ y de $X'Y$. Esto da

$$X_2'X_2\tilde{\gamma}_2=X_2'Y$$

como

$$\begin{pmatrix} 16 & 8 \\ 8 & 6 \end{pmatrix} \begin{pmatrix} \tilde{\beta}_1 \\ \tilde{\beta}_2 \end{pmatrix} = \begin{pmatrix} 80 \\ 120 \end{pmatrix}$$

Luego

$$\tilde{\gamma}_2 = \begin{pmatrix} \tilde{\beta}_1 \\ \tilde{\beta}_2 \end{pmatrix} = \begin{pmatrix} 0,19 & -0,25 \\ -0,25 & 0,50 \end{pmatrix} \begin{pmatrix} 80 \\ 120 \end{pmatrix} = \begin{pmatrix} -14,8 \\ 40,0 \end{pmatrix}$$

y $r=1$, $n=16$, $p=3$; también

$$\tilde{\gamma}_2'X_2'Y = (-14,8)(80) + (120)(40) = 3616$$

El valor de F es menor que 1.

PROBLEMAS

1. Supongamos que los datos siguientes satisfacen el modelo

$$y_i = \beta_1 + \beta_2 x_i + e_i$$

y	-6,1	-0,5	7,2	6,9	-0,2	-2,1	-3,9	3,8
x	-2,0	0,6	1,4	1,3	0,0	-1,6	-1,7	0,7

Hállense: a) $X'X$; b) $X'Y$; c) $\hat{\beta}' = (\hat{\beta}_1, \hat{\beta}_2)$.

2. En el problema anterior, determínese un intervalo confidencial del 95% para β_2 .

3. En el modelo de la definición 13-1, hállese $\text{var}(\hat{\alpha} + \hat{\beta}x_0)$ para un valor fijo x_0 .

4. Para el caso A sobre los errores, en el modelo de la definición 13-1, hállese la distribución de $(\hat{\alpha} + \hat{\beta}x_0)$ para un valor fijo x_0 .

5. Demuéstrese en el problema anterior que

$$[(\hat{\alpha} + \hat{\beta}x_0) - (\alpha + \beta x_0)] \sqrt{a\hat{\sigma}^2}$$

se distribuye según la t de Student con $n-2$ grados de libertad, donde $a\sigma^2$ es $\text{var}(\hat{\alpha} + \hat{\beta}x_0)$ en dicho problema.

6. Utilícense los resultados del problema 5 para definir un intervalo confidencial del 0,95 para $(\alpha + \beta)$, con los datos del problema 1.

7. En el problema 1, documense la hipótesis $H_0: \beta_2 = 0$ con probabilidad 0,05 de error de tipo I.

*8. Utilizando el teorema 13-3, pruébese que

$$u = \frac{(\hat{\beta} - \beta)' S(\hat{\beta} - \beta)}{\sigma^2}$$

se distribuye según la χ^2 cuadrado con p grados de libertad.

*9. Por medio del teorema 13-3 y del resultado del problema anterior, demuéstrese que

$$v = \frac{(\hat{\beta} - \beta)' S(\hat{\beta} - \beta)}{p\hat{\sigma}^2}$$

admite una distribución F con p y $n-p$ grados de libertad.

*10. Utilizando el problema 9 y los datos del problema 1, hállese una región confidencial del 0,95 para β_1 y β_2 . Representétese gráficamente esta región.

* Los problemas señalados con asterisco dependen del contenido de la sección 13.6.

*11. Con los datos:

y	12,1	11,9	10,2	8,0	7,7	5,3	7,9	7,8	5,5	2 6
x_1	0	1	2	3	4	5	6	7	8	9
x_2	7	4	44	6	4	2	1	1	1	0

ajústese el modelo

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + e,$$

y hállese un intervalo confidencial del 0,95 para σ^2 .

*12. Hállese un intervalo confidencial del 95% para el β_1 del problema precedente.

*13. Docímese la hipótesis nula de que β_2 del problema 11 sea 0.

*14. Docímese la hipótesis nula de que $\beta_1 + 10\beta_2 = 0$ en el problema 11.

*15. Utilizando solo las dos primeras filas de datos del problema 11, ajústese el modelo

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + e$$

y docímese la hipótesis nula de que sea $\beta_2 = 0$.

*16. El ajuste de polinomios, tal como el cuadrático del problema 15, se simplifica considerablemente, si los valores están igualmente espaciados, utilizando *polinomios ortogonales*. Sea $x=0, 1, \dots, n$. Los tres primeros polinomios ortogonales son

$$P_1 = x - \frac{n}{2}$$

$$P_2 = \left(x - \frac{n}{2}\right)^2 - \frac{n(n+2)}{12}$$

$$P_3 = \left(x - \frac{n}{2}\right)^3 - \frac{3n(n+2)-4}{20} \left(x - \frac{n}{2}\right)$$

Demuéstrese que

$$\sum_x P_1 P_2 = \sum_x P_1 P_3 = \sum_x P_2 P_3 = 0$$

*17. Resuélvase de nuevo el problema 15, ajustando ahora el modelo

$$y = \beta_0 + \beta_1 P_1 + \beta_2 P_2 + e,$$

en donde P_1 y P_2 se definen en el problema anterior.

18. Si y_1 e y_2 tienen una distribución normal bivalente, ¿cuáles son los coeficientes (en función de σ_{11} , σ_{22} y ρ) de la función de regresión para la distribución condicional de y_1 ? ¿Y para la distribución condicional de y_2 ? Si las dos rectas de regresión se estiman a partir de la misma muestra, ¿serán, en general, diferentes?

19. Si y_1 , y_2 , y_3 tienen una distribución normal trivalente, ¿cuáles son los coeficientes de la función de regresión para la distribución condicional de y_1 , dados y_2 e y_3 , en función de las varianzas y correlaciones?

20. La variable aleatoria bidimensional (y, x) tiene una densidad normal bivalente con correlación ρ igual a 0. Sea (y_1, x_1) , ..., (y_n, x_n) una muestra aleatoria de tamaño n de esta densidad y supóngase que

$$\hat{\rho} = \frac{\sum (y_i - \bar{y})(x_i - \bar{x})}{\sqrt{\sum (y_i - \bar{y})^2 \sum (x_i - \bar{x})^2}}$$

Demuéstrese que $\hat{\rho}$ tiene la densidad

$$f(\hat{\rho}) = \frac{[(n-3)/2]! (1-\hat{\rho}^2)^{(n-4)/2}}{\sqrt{\pi} [(n-4)/2]!} \quad -1 < \hat{\rho} < 1$$

21. Con referencia al problema 20, transfórmese $\hat{\rho}$ en la nueva variante

$$t = \hat{\rho} \sqrt{\frac{n-2}{1-\hat{\rho}^2}}$$

y pruébese que tiene la distribución de *Student* con $n-2$ grados de libertad, de modo que pueden utilizarse las tablas t para docimar la hipótesis nula $\rho=0$.

22. Supóngase que los datos del problema 1 proceden de una población normal bivalente, y docímese la hipótesis nula $\rho=0$.

23. Si ρ no es cero, la distribución de $\hat{\rho}$ no es una función sencilla, pero se ha tabulado para n , tamaño de la muestra, inferior a 25. Fisher ha demostrado que

$$z = \frac{1}{2} \log \frac{1+\hat{\rho}}{1-\hat{\rho}}$$

tiene una distribución aproximadamente normal, con media

$$\epsilon = \frac{1}{2} \log \frac{1+\rho}{1-\rho}$$

y varianza $1/(n-3)$; utilizando este resultado, hállese aproximadamente un intervalo confidencial del 95% para la ρ del problema 22.

24. Examinense detenidamente los detalles de la demostración del teorema 13-2 para el parámetro β .

25. En el problema 2, hállese la longitud esperada del intervalo confidencial.

*26. Para el modelo y los datos del problema 11, deseamos docimar la hipótesis $H_0: \beta_1 = \beta_2 = 0$. Si se utiliza el teorema 13-6: a) identifíquense γ_1 y X_1 ; b) identifíquense γ_2 y X_2 .

*27. En el problema 26, hállese las ecuaciones normales $X'X\hat{\beta} = X'Y$ y $\hat{\beta}'X'Y$.

*28. Obténganse en el problema 26 las ecuaciones normales reducidas $X'_2X_2\tilde{\gamma}_2 = X'_2Y$ y $\tilde{\gamma}'_2X'_2Y$.

*29. En el teorema 13-6, compruébese que

$$\hat{\beta}'X'Y - \tilde{\gamma}'_2X'_2Y$$

es igual a

$$Y'[XS^{-1}X' - X_2(X'_2X)^{-1}X'_2]Y.$$

30. Una variante y se distribuye alrededor de $\alpha + \beta x$ según la densidad

$$f(y|x) = 1 \quad \alpha + \beta x - \frac{1}{2} < y < \alpha + \beta x + \frac{1}{2}.$$

Hállese la estimación máximo-verosímil de α y β , dada la muestra de cuatro puntos (y, x) : (0,3; 1), (-0,7; 2), (-1,7; 3), (-1,8; 4). Compárese con la recta de mínimos cuadrados.

31. Si x e y tienen la distribución trinomial

$$f(x, y) = \frac{m!}{x!y!(m-x-y)!} p^x q^y (1-p-q)^{m-x-y}$$

hállese la función de regresión de y sobre x .

BIBLIOGRAFIA

1. ANDERSON, R. L., y T. A. BANCROFT: *Statistical Theory in Research*, McGraw-Hill Book Company, Inc., Nueva York, 1952.
2. FISHER, R. A.: «The goodness of fit and regression formulae, and the distribution of regression coefficients», *Journal of the Royal Statistical Society*, vol. 85, pt. IV (1922); reproducido en *Contributions to Mathematical Statistics*, John Wiley & Sons, Inc., Nueva York, 1950.
3. GRAYBILL, F. A.: *An Introduction to Linear Statistical Models*, vol. I. McGraw-Hill Book Company, Inc., Nueva York, 1961.
4. KEMPTHORNE, O.: *Design and Analysis of Experiments*, John Wiley & Sons, Inc., Nueva York, 1952.
5. KOŁODZIEJCZYK, S.: «On an important class of statistical hypotheses», *Biometrika*, vol. 27 (1935).

6. KULLBACK, S.: *Information Theory and Statistics*, John Wiley & Sons, Inc., Nueva York, 1959.
7. MANN, H. B.: *Analysis and Design of Experiments*, Dover Publications, Nueva York, 1949.
8. RAO, C. R.: *Advanced Statistical Methods in Biometric Research*, John Wiley & Sons, Inc., Nueva York, 1952.
9. WILKS, S. S.: *Mathematical Statistics*, Princeton University Press, Princeton, N.J., 1943.

CAPITULO 14

MODELOS DE DISEÑO EXPERIMENTAL *

14-1. Introducción.—En el capítulo anterior hemos expuesto las técnicas que un experimentador utilizaría si estuviese interesado en encontrar una fórmula a partir de la cual pudiera predecir el valor de un factor en estudio, por medio del valor de otro factor relacionado con él. Así, p. ej., el experimentador puede desear predecir la dureza y de un metal, conociendo la temperatura T y el tiempo t de cierta operación química. El modelo sería de la forma

$$y = f(T, t) + e$$

En este capítulo consideraremos una situación algo diferente, en la que el interés no se centra en predecir el valor de un factor utilizando valores de factores relacionados con él, sino en comparar los *efectos* de dos o más factores.

Ejemplo 14-1.—Supongamos que el director de una fábrica desea comprar máquinas para realizar cierta operación en un proceso de producción. Hay tres compañías que fabrican tales máquinas y toma a prueba una de cada compañía, a fin de determinar cuál de las tres es la más apropiada a sus fines. Supongamos también que cada máquina es manejada por un operario. El director hace que manejen las máquinas varios de sus hombres durante unos días para descubrir con cuál de las tres se consigue mayor producción por día. En este sencillo experimento la medida es el número de productos por día, y el factor único es el tipo de máquina.

Imaginemos que se utilizan seis hombres en el experimento, asignando al azar dos para cada máquina, y que cada hombre trabaja un día en la máquina particular que le ha sido asignada. Habrá entonces dos observaciones para cada una de las tres máquinas, siendo cada observación la cantidad producida por la máquina en un día. Los datos podrían ser los que aparecen en la tabla de la página siguiente. Lo interesante es saber si las máquinas son o no diferentes en cuanto al número de elementos que son capaces de producir.

* En este capítulo se hace extenso uso del álgebra de matrices, por lo que se aconseja al lector que repase sus conocimientos a este respecto.

Número de máquina		
1	2	3
64	41	65
39	48	57

El modelo para este experimento puede escribirse así

$$\begin{aligned}
 y_{11} &= \mu + \tau_1 + e_{11} \\
 y_{12} &= \mu + \tau_1 + e_{12} \\
 y_{21} &= \mu + \tau_2 + e_{21} \\
 y_{22} &= \mu + \tau_2 + e_{22} \\
 y_{31} &= \mu + \tau_3 + e_{31} \\
 y_{32} &= \mu + \tau_3 + e_{32}
 \end{aligned} \tag{1}$$

en donde y_{ij} es la j -ésima observación (según cierto método de identificación) de la i -ésima máquina; μ es una media general, τ_i , el efecto de la i -ésima máquina, y e_{ij} , errores aleatorios no observables introducidos por variaciones incontrolables, tales como diferencias entre los operadores, temperatura, etc. Un modelo realista deberá cumplir la restricción $\tau_1 + \tau_2 + \tau_3 = 0$, y la media general μ será $E[1/6 \sum y_{ij}]$. Sin embargo, para probar los teoremas generales, no impondremos esta condición. Esto se analizará con detalle más adelante. El modelo (1) se expresa de forma más compacta así:

$$y_{ij} = \mu + \tau_i + e_{ij} \quad i = 1, 2, 3; \quad j = 1, 2 \tag{2}$$

Con notación matricial, el sistema de ecuaciones (1) se escribirá

$$\begin{pmatrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{22} \\ y_{31} \\ y_{32} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \tau_3 \end{pmatrix} + \begin{pmatrix} e_{11} \\ e_{12} \\ e_{21} \\ e_{22} \\ e_{31} \\ e_{32} \end{pmatrix} \tag{3}$$

o bien

$$Y = X\beta + e$$

en donde Y es un vector 6×1 de observaciones; β , un vector 4×1 de parámetros desconocidos; e , un vector 6×1 de errores no observables, y X , una matriz cuyos elementos son ceros y unos. Si el lector se entretiene en realizar las operaciones indicadas en el sistema (3), comprobará que es equivalente al (!). Obsérvese que este modelo, $Y = X\beta + e$, es el mismo que el del capítulo anterior, salvo que los elementos de X son ceros y unos, y que la matriz X de (3) es de orden 6×4 , pero su característica es 3. Por tanto, $X'X$ tiene también característica 3 y, en consecuencia, carece de inversa. Esto constituye cierta dificultad que, en general, no se plantea en la teoría de modelos lineales generales expuesta en el capítulo 13.

14-2. Modelo de diseño experimental.—Aunque utilizaremos (14.1.3) como ejemplo, el modelo general de diseño experimental se da en la siguiente definición.

Definición 14-1.—Sea Y un vector aleatorio observable $n \times 1$ tal que

$$Y = X\beta + e \quad (1)$$

en donde X es una matriz $n \times p$ conocida, de característica r , con $r < p \leq n$, y X contiene solo ceros y unos; β es un vector de parámetros desconocidos, y e otro vector de variables aleatorias no observables. Por definición, este es un modelo de diseño experimental.

Obsérvese que, puesto que la característica de X es r , también será r la de $X'X$ y, por tanto, esta última no tiene inversa. En general, un modelo de diseño experimental se escribirá en la forma

$$y_{ij\dots m} = \mu_{ij\dots m} + e_{ij\dots m} \quad (1a)$$

en donde $\mu_{ij\dots m}$ es una combinación lineal de los parámetros. Así, en el ejemplo 14.1, $\mu_{ij} = \mu + \tau_i$.

Haremos las siguientes hipótesis acerca de las e_{ij} .

Caso A: e_{ij} son variables normales incorrelacionadas, con media 0 y varianza σ^2 .

Caso B: e_{ij} son variables aleatorias incorrelacionadas (no necesariamente normales) con media 0 y varianza σ^2 .

Estimación puntual.—Tanto si utilizamos el caso A y máxima verosimilitud, como si empleamos mínimos cuadrados en el caso B, se obtiene el mismo sistema de ecuaciones al igualar a cero las derivadas respecto a las β_i :

$$X'X\hat{\beta} = X'Y \quad (2)$$

También obtenemos

$$\tilde{\sigma}^2 = \frac{1}{n} (Y - X\hat{\beta})(Y - X\hat{\beta}) \quad (3)$$

Al deducir las ecuaciones normales (2) y la ecuación (3) para $\tilde{\sigma}^2$, no hemos tropezado con el hecho de que $X'X$ carece de inversa; sin embargo, cuando $X'X$ tiene inversa, las ecuaciones normales poseen solución única,

$$\hat{\beta} = (X'X)^{-1}X'Y$$

que son los estimadores puntuales de los elementos de β . La situación es algo diferente cuando $X'X$ no tiene inversa, y hay que distinguir dos casos: 1.º, no existe ningún vector $\hat{\beta}$ que satisfaga la ecuación (2), y 2.º, existe un número infinito de vectores que satisfacen a (2). Puede demostrarse que se verifica precisamente el último caso; es decir, existe un número infinito de vectores $\hat{\beta}$ que satisfacen las ecuaciones normales (2). Esto se prueba utilizando un teorema del álgebra de matrices, en virtud del cual si una matriz cuadrada $X'X$ no tiene inversa y su característica es igual a la de la matriz $(X'X|X'Y)$, existe entonces un número infinito de vectores $\hat{\beta}$ que satisfacen a (2). La matriz $(X'X|X'Y)$ es la propia matriz $X'X$ con $X'Y$ como columna adicional.

Esta no es una situación muy satisfactoria; dos experimentadores que utilizasen el mismo modelo e idénticas observaciones, llegarían a las mismas ecuaciones normales, pero cada uno de ellos obtendría una estimación diferente de β_i .

Nos interesarán, si existen, estimaciones insesgadas de las β_i (elementos de β) y, por tanto, será necesario ver qué soluciones $\hat{\beta}$ de las ecuaciones normales (2) son estimadores insesgados. Cualquier solución de (2) debe ser una función lineal de las y ; por tanto, escribiremos

$$\hat{\beta} = AY \quad (4)$$

en donde A es una matriz $p \times n$ de constantes, las cuales pueden depender de los elementos de X . Por ser insesgado, debemos tener $E(\hat{\beta}) = \beta$, o bien

$$\beta = E(\hat{\beta}) = E(AY) = E[A(X\beta + e)] = AX\beta + AE(e) = AX\beta$$

Luego $AX\beta = \beta$ se cumple para todos los valores de β_i , lo que implica que $AX = I$. Pero I es la matriz unitaria $p \times p$, cuya característica es p , y la característica del producto de dos matrices no puede ser mayor que la de cualquiera de ambas. Por la definición

14-1, la característica de X es $r < p$; por tanto, la de AX será menor que p . Luego no existe una matriz A tal que $E(A\mathbf{Y}) = \beta$ y, por tanto, no hay un estimador insesgado de β . Así, en el ejemplo 14-1 no existen estimadores insesgados de μ , τ_1 , τ_2 y τ_3 . En la mayoría de los casos de un modelo de diseño experimental se está interesado en estimar ciertas combinaciones lineales de los parámetros y, en consecuencia, examinaremos estas.

Definición 14-2.—Sea λ un vector $p \times 1$ de constantes conocidas; se dirá que la combinación lineal de β , dada por

$$\lambda'\beta = \sum_{i=1}^p \lambda_i \beta_i,$$

es una función estimable si existe una combinación lineal de las y cuyo valor esperado es $\lambda'\beta$.

En otras palabras, $\lambda'\beta$ es estimable si existe un vector a de dimensión $n \times 1$ tal que $E(a'\mathbf{Y}) = \lambda'\beta$.

Particionemos la matriz X en la forma

$$X = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix}$$

en donde X_i es un vector $1 \times p$ que forma la i -ésima fila de X . Nos interesa saber si $X_i\beta$ es estimable para todo i . Por la definición 14-2, el conjunto $X_1\beta$, $X_2\beta$, ..., $X_n\beta$ constituye un conjunto de n funciones estimables si existen n vectores A_1 , A_2 , ..., A_n , tales que $E(A_i\mathbf{Y}) = X_i\beta$; en otras palabras, si existe una matriz A , de dimensión $n \times p$, tal que $E(A\mathbf{Y}) = X\beta$, ya que si A_i es la i -ésima fila de A y $X_i\beta$ es estimable, se tendrá:

$$E(A\mathbf{Y}) = E \begin{pmatrix} A_1 \\ A_2 \\ \vdots \\ A_n \end{pmatrix} \mathbf{Y} = \begin{pmatrix} E(A_1\mathbf{Y}) \\ E(A_2\mathbf{Y}) \\ \vdots \\ E(A_n\mathbf{Y}) \end{pmatrix} = \begin{pmatrix} X_1\beta \\ X_2\beta \\ \vdots \\ X_n\beta \end{pmatrix} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix} \beta = X\beta$$

Sea $A = I$; entonces

$$E(I\mathbf{Y}) = E(\mathbf{Y}) = E(X\beta + \mathbf{e}) = E(X\beta) + E(\mathbf{e}) = X\beta.$$

Teorema 14-1.— $X\beta$ representa un conjunto de n funciones estimables; es decir, cada elemento de $E(Y)$ es estimable.

Representando (1) por (1a), el teorema 14-1 nos dice que $E(y_{ij} \dots_m)$ es estimable. En el ejemplo 14-1, $E(y_{ij})$ es estimable para todo i y j ; por tanto, $\mu + \tau_1$, $\mu + \tau_2$ y $\mu + \tau_3$ son estimables.

Teorema 14-2.— $(X'X)\beta$ representa un conjunto de p funciones estimables.

Demostración.—Particionemos la matriz X' , de dimensión $p \times n$, en p filas, y sea X_i^* el vector que figura como i -ésima fila; entonces

$$X' = \begin{pmatrix} X_1^* \\ X_2^* \\ \vdots \\ X_p^* \end{pmatrix}$$

Ahora bien,

$$E(X'Y) = X'E(Y) = X'(X\beta) = X'X\beta.$$

En consecuencia,

$$E(X'Y) = E \begin{pmatrix} X_1^* \\ X_2^* \\ \vdots \\ X_p^* \end{pmatrix} Y = \begin{pmatrix} EX_1^*Y \\ EX_2^*Y \\ \vdots \\ EX_p^*Y \end{pmatrix} = \begin{pmatrix} X_1^*X\beta \\ X_2^*X\beta \\ \vdots \\ X_p^*X\beta \end{pmatrix} = X'X\beta$$

Luego $X_i^*X\beta$ es estimable, puesto que $E(X_i^*Y) = X_i^*X\beta$. Pero el i -ésimo elemento de $X'X\beta$ es $X_i^*X\beta$; por tanto, $X'X\beta$ forma un conjunto de p funciones estimables.

Vamos a hallar a continuación el conjunto de «todas» las funciones estimables.

Definición 14-3.—Sean $\theta_1, \theta_2, \dots, \theta_t$ vectores $p \times 1$ tales que $\theta_1'\beta, \theta_2'\beta, \dots, \theta_t'\beta$ son estimables y que la característica de la matriz $p \times t$ ($t \leq p$)

$$(\theta_1, \theta_2, \dots, \theta_t)$$

es igual a t ; se dice entonces que $\theta_1'\beta, \theta_2'\beta, \dots, \theta_t'\beta$ son funciones estimables linealmente independientes.

Puesto que $X'X\beta$ es un conjunto de p funciones estimables y dado que $X'X$ posee característica r , en el modelo lineal general (1) hay al menos r funciones estimables linealmente independientes,

siendo r la característica de X . Supongamos que existen $k > r$ funciones estimables linealmente independientes dadas por $\theta'_1\beta$, $\theta'_2\beta$, ..., $\theta'_k\beta$. Este conjunto puede escribirse así

$$L'\beta = \begin{pmatrix} \theta'_1 \\ \theta'_2 \\ \vdots \\ \vdots \\ \theta'_k \end{pmatrix} \beta$$

en donde L' es una matriz $k \times p$ de característica $k > r$. Por las definiciones 14-2 y 14-3, existe una matriz A , de dimensión $k \times n$, tal que $E(AY) = L'\beta$. Sin embargo, $E(AY) = AX\beta$ y, por tanto, $AX = L'$; pero AX tiene característica menor o igual que la característica r de X , lo que contradice el supuesto de que es $k > r$. Queda así demostrado el siguiente teorema:

Teorema 14-3.—*En el modelo lineal general (1) existen exactamente r funciones estimables linealmente independientes, si r es la característica de X .*

Puesto que $X'X\beta$ es un conjunto de p funciones estimables, y la característica de $X'X$ es r , en virtud del teorema 14-3, cada función estimable deberá ser una combinación lineal de las filas de $X'X\beta$. Este resultado se formula en el teorema siguiente.

Teorema 14-4.—*Sea $\lambda'\beta$ una función estimable; existe entonces un vector b , de dimensión $p \times 1$, tal que $b'X'X = \lambda'$, y $b'X'Y$ es el estimador insesgado lineal óptimo (mínima varianza) de $\lambda'\beta$ en caso B, y el estimador máximo-verosímil de $\lambda'\beta$ en el caso A.*

Este teorema resalta de nuevo la importancia de las ecuaciones normales, ya que el estimador de mínimos cuadrados (en el caso B) o el estimador máximo-verosímil (en el caso A), de cualquier función estimable, proviene de una solución de las ecuaciones normales. Así, en el caso B, $\lambda'\beta$ es una función estimable, y el estimador insesgado lineal óptimo será $b'X'Y$, en donde b' es un vector $1 \times p$ tal que $b'X'X = \lambda'$. Corrientemente no es necesario hallar b' , sino que se obtiene $b'X'Y$ mediante combinaciones lineales de los primeros miembros de las ecuaciones normales, hasta hallar el $\lambda'\beta$ deseado; la misma combinación lineal de los segundos miembros proporciona entonces el estimador insesgado lineal óptimo de $\lambda'\beta$. Idéntico procedimiento da en el caso A el estimador máximo-verosímil de $\lambda'\beta$. Otra manera de obtener el estimador máximo-verosímil (o el de mínimos cuadrados) de $\lambda'\beta$ consiste en hallar una solución $\hat{\beta}$ de las ecuaciones normales, y utilizar $\lambda'\hat{\beta}$ como estimador. Demostraremos que cualquier solución $\hat{\beta}$ de las ecuaciones norma-

les $X'X\hat{\beta} = X'Y$, da el mismo valor $\lambda'\hat{\beta}$ si $\lambda'\beta$ es estimable. Este resultado se enuncia formalmente como sigue:

Teorema 14-5.—Sea $\lambda'\beta$ una función estimable en el modelo $Y = X\beta + e$, y sean $\hat{\gamma}$ y $\hat{\alpha}$ dos vectores $p \times 1$ que satisfacen las ecuaciones normales; es decir, $X'X\hat{\gamma} = X'Y$ y $X'X\hat{\alpha} = X'Y$. Entonces

$$\lambda'\hat{\gamma} = \lambda'\hat{\alpha} \quad \text{y} \quad E(\lambda'\hat{\gamma}) = E(\lambda'\hat{\alpha}) = \lambda'\beta.$$

Demostración.—Por hipótesis, $\hat{\gamma}$ y $\hat{\alpha}$ son dos vectores $p \times 1$ que satisfacen las ecuaciones normales; es decir

$$X'X\hat{\gamma} = X'Y$$

y

$$X'X\hat{\alpha} = X'Y$$

Pero también por hipótesis hecha $\lambda'\beta$ es estimable y, por tanto, en virtud del teorema 14-4, existe un vector b' tal que $b'X'X = \lambda'$. Multiplicando a la izquierda por b' ambas ecuaciones, se obtiene:

$$b'X'X\hat{\gamma} = b'X'Y$$

$$b'X'X\hat{\alpha} = b'X'Y$$

$$\lambda'\hat{\gamma} = b'X'Y$$

$$\lambda'\hat{\alpha} = b'X'Y$$

y, por tanto,

$$\lambda'\hat{\gamma} = \lambda'\hat{\alpha}$$

Ahora bien,

$$E(\lambda'\hat{\gamma}) = E(\lambda'\hat{\alpha}) = E(b'X'Y) = b'X' \cdot E(Y) = b'X'X\beta = \lambda'\beta$$

con lo que la proposición queda demostrada.

Teorema 14-6.—Si $\lambda_1'\beta$, ..., $\lambda_s'\beta$ es un conjunto de s funciones estimables, cualquier combinación lineal de ellas es también estimable.

Este teorema es análogo al 13-5, y su demostración se propone como ejercicio al lector.

Ejemplo 14-2.—Consideremos el modelo del ejemplo 14-1. La matriz $X'X$ es, por (14-1-3),

$$X'X = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 6 & 2 & 2 & 2 \\ 2 & 2 & 0 & 0 \\ 2 & 0 & 2 & 0 \\ 2 & 0 & 0 & 2 \end{pmatrix}$$

El vector $X'Y$ es

$$X'Y = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{22} \\ y_{31} \\ y_{32} \end{pmatrix} = \begin{pmatrix} Y_{..} \\ Y_{1.} \\ Y_{2.} \\ Y_{3.} \end{pmatrix}$$

en donde

$$Y_{i.} = \sum_j y_{ij}; \quad Y_{..} = \sum_i \sum_j y_{ij}$$

Emplearemos también la notación

$$Y_{.j} = \sum_i y_{ij} \quad \bar{y}_{i.} = \frac{Y_{i.}}{n} \quad \bar{y}_{.j} = \frac{Y_{.j}}{t}$$

siendo n el recorrido del subíndice j , y t , el del subíndice i .

Las ecuaciones normales son

$$\begin{aligned} 6\hat{\mu} + 2\hat{\tau}_1 + 2\hat{\tau}_2 + 2\hat{\tau}_3 &= Y_{..} \\ 2\hat{\mu} + 2\hat{\tau}_1 &= Y_{1.} \\ 2\hat{\mu} + 2\hat{\tau}_2 &= Y_{2.} \\ 2\hat{\mu} + 2\hat{\tau}_3 &= Y_{3.} \end{aligned} \quad (5)$$

El teorema 14-1 dice que $E(y_{ij}) = \mu + \tau_i$ es estimable para todo i ; luego $\mu + \tau_1$, $\mu + \tau_2$ y $\mu + \tau_3$ son también estimables. Por el teorema 14-4, podemos formar combinaciones lineales con las filas de las ecuaciones normales (5), con lo que en el primer miembro se obtiene $\hat{\mu} + \hat{\tau}_i$; el estimador es la misma combinación lineal del segundo miembro. Vemos que si se divide por 2 la segunda ecuación de (5), resulta

$$\hat{\mu} + \hat{\tau}_1 = (1/2)Y_{1.} = \bar{y}_{1.}$$

Análogamente para la tercera ecuación,

$$\hat{\mu} + \hat{\tau}_2 = (1/2)Y_{2.} = \bar{y}_{2.}$$

y para la cuarta,

$$\hat{\mu} + \hat{\tau}_3 = (1/2)Y_{3.} = \bar{y}_{3.}$$

Luego el estimador MV (máximo verosímil) o MC (mínimos cuadrados) de $\mu + \tau_i$ es $\bar{y}_i$. Es fácil comprobar que

$$E(\bar{y}_i) = \mu + \tau_i.$$

Existen otros estimadores insesgados de $\mu + \tau_1$, p. ej.,

$$E(y_{11}) = \mu + \tau_1,$$

pero este estimador y_{11} no tiene todas las propiedades de que goza el estimador $\bar{y}_1$, ya que $\bar{y}_1$ proviene de las *ecuaciones normales*. Por el teorema 14-4, si existe un estimador insesgado para una combinación lineal de los parámetros, será posible obtener un estimador insesgado a partir de las *ecuaciones normales* y, por el teorema 14-4, este es el estimador insesgado *óptimo*. Así, debemos examinar el *modelo* para ver qué funciones son estimables, aunque utilizamos las *ecuaciones normales* para *obtener* el estimador. Puesto que la característica de X en (14-1-3) es 3, por el teorema 14-3, existen exactamente tres funciones estimables linealmente independientes; estas son $\mu + \tau_1$, $\mu + \tau_2$ y $\mu + \tau_3$, que pueden escribirse:

$$\begin{aligned}\mu + \tau_1 &= \lambda'_1 \beta = (1 \quad 1 \quad 0 \quad 0) \begin{pmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \tau_3 \end{pmatrix} \\ \mu + \tau_2 &= \lambda'_2 \beta = (1 \quad 0 \quad 1 \quad 0) \begin{pmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \tau_3 \end{pmatrix} \\ \mu + \tau_3 &= \lambda'_3 \beta = (1 \quad 0 \quad 0 \quad 1) \begin{pmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \tau_3 \end{pmatrix}\end{aligned}$$

siendo 3 su característica; por tanto, las tres funciones estimables son, en efecto, linealmente independientes y, en consecuencia, *todas* las funciones estimables serán combinaciones lineales de estas. El lector comprobará sin dificultad que cada uno de los dos vectores

$$\hat{\gamma} = \begin{pmatrix} \bar{y}_{..} \\ \bar{y}_{1.} - \bar{y}_{..} \\ \bar{y}_{2.} - \bar{y}_{..} \\ \bar{y}_{3.} - \bar{y}_{..} \end{pmatrix} \quad \hat{\alpha} = \begin{pmatrix} \bar{y}_{2.} \\ \bar{y}_{1.} - \bar{y}_{2.} \\ 0 \\ \bar{y}_{3.} - \bar{y}_{2.} \end{pmatrix}$$

satisface las ecuaciones normales (5). Ahora bien, $\lambda'_1\beta = \mu + \tau_1$ es estimable; luego, por el teorema 14-5, $\lambda'_1\hat{\gamma}$ será igual a $\lambda'_1\hat{\alpha}$, que, a su vez, debe coincidir con $\bar{y}_1$. El lector puede comprobar fácilmente que esto es cierto. En (14-1-2),

$$E(\mathbf{y}_{11} - \mathbf{y}_{21}) = \tau_1 - \tau_2$$

y, por tanto, de acuerdo con el teorema 14-1, $\tau_1 - \tau_2$ es estimable. En (5), si se resta la segunda ecuación de la tercera, previamente multiplicadas por $\frac{1}{2}$, se obtiene

$$\hat{\tau}_1 - \hat{\tau}_2 = (\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2)$$

como estimador de $\tau_1 - \tau_2$. Sumando la mitad de la segunda ecuación con la mitad de la tercera y restando la cuarta, resulta:

$$\hat{\tau}_1 + \hat{\tau}_2 - 2\hat{\tau}_3 = \bar{\mathbf{y}}_1 + \bar{\mathbf{y}}_2 - 2\bar{\mathbf{y}}_3$$

como estimador de $\tau_1 + \tau_2 - 2\tau_3$. Para estimar σ^2 , observemos que, por (3),

$$\hat{\sigma}^2 = (\mathbf{Y} - \mathbf{X}\hat{\beta})(\mathbf{Y} - \mathbf{X}\hat{\beta})/n,$$

en donde $\hat{\beta}$ satisface a las ecuaciones normales. Sean $\hat{\alpha}$ y $\hat{\gamma}$ dos soluciones cualesquiera de dichas ecuaciones; por el teorema 14-1, $\mathbf{X}\hat{\beta}$ formará un conjunto de funciones estimables, y por el 14-5, se tiene que $\mathbf{X}\hat{\gamma} = \mathbf{X}\hat{\alpha}$; es decir, $\mathbf{X}\hat{\beta}$ es la misma para cualquier vector $\hat{\beta}$ que satisfaga a las ecuaciones normales. Por tanto, $\mathbf{Y} - \mathbf{X}\hat{\beta}$ resulta independientemente de la solución de las ecuaciones normales que se utilice, y de ahí que $\hat{\sigma}^2$ sea también invariable. Puesto que el estimador insesgado de σ^2 es

$$(\mathbf{Y} - \mathbf{X}\hat{\beta})(\mathbf{Y} - \mathbf{X}\hat{\beta})/(n - r),$$

emplearemos este como $\hat{\sigma}^2$.

Enunciaremos el teorema siguiente, referente a la docimaña de diversas hipótesis, pero omitiremos su demostración.

Teorema 14-7.—Sean $\lambda'_1\beta$, $\lambda'_2\beta$, ..., $\lambda'_k\beta$, k funciones estimables linealmente independientes, y supongamos que se desea docimar la hipótesis H_0 de que todas ellas son simultáneamente iguales a cero en el modelo (1) (caso A). A partir de las ecuaciones normales

$$\mathbf{X}'\mathbf{X}\hat{\beta} = \mathbf{X}'\mathbf{Y} \quad (6)$$

hallamos $\hat{\beta}'\mathbf{X}'\mathbf{Y}$, en donde $\hat{\beta}$ es cualquier solución de (6). Sustituyendo

$$\lambda'_1\beta = \lambda'_2\beta = \dots = \lambda'_k\beta = 0$$

en el modelo $\mathbf{Y} = \mathbf{X}\beta + \mathbf{e}$, se obtiene el modelo reducido designado por $\mathbf{Y} = \mathbf{Z}\gamma + \mathbf{e}$. A partir de las ecuaciones normales para este modelo

$$\mathbf{Z}'\mathbf{Z}\tilde{\gamma} = \mathbf{Z}'\mathbf{Y} \quad (7)$$

hallamos $\tilde{\gamma}'\mathbf{Z}'\mathbf{Y}$, donde $\tilde{\gamma}$ representa cualquier solución de (7). Entonces, si la hipótesis H_0 es cierta,

$$u = \frac{(\hat{\beta}'\mathbf{X}'\mathbf{Y} - \tilde{\gamma}'\mathbf{Z}'\mathbf{Y})/k}{(\mathbf{Y}'\mathbf{Y} - \hat{\beta}'\mathbf{X}'\mathbf{Y})/(n-r)} \quad (8)$$

tiene la distribución F con k y $n-r$ grados de libertad. Se rechaza H_0 , con una probabilidad α de cometer un error de tipo I cuando $u > F_{\alpha}(k, n-r)$. Esto se deduce a partir de la razón de verosimilitud.

La demostración sigue la misma línea que en el caso del teorema 13-6, excepto que las matrices $\mathbf{X}'\mathbf{X}$ y $\mathbf{Z}'\mathbf{Z}$ carecen de inversas. Estamos ahora en disposición de docimar cualquier hipótesis que pueda formularse por medio de funciones estimables linealmente independientes.

Ejemplo 14-3.—Docimaremos la hipótesis $H_0: \tau_1 = \tau_2 = \tau_3$ en el ejemplo 14-1. Para utilizar el teorema 14-7, tenemos que ver en primer lugar si H_0 equivale a que ciertas funciones estimables linealmente independientes sean nulas. Sea $\lambda'_1\beta = \tau_1 - \tau_2$, de modo que $\lambda'_1 = (0, 1, -1, 0)$, y sea $\lambda'_2\beta = \tau_1 + \tau_2 - 2\tau_3$ de manera que $\lambda'_2 = (0, 1, 1, -2)$. Evidentemente, estas son funciones estimables linealmente independientes. Es también obvio que $\lambda'_1\beta = \lambda'_2\beta = 0$ si, y solamente si, $\tau_1 = \tau_2 = \tau_3$; por tanto, podrá emplearse el teorema 14-7 para docimar H_0 . Obsérvese que la hipótesis no es que las τ_i son iguales a cero, sino que todas ellas son iguales entre sí. Podemos escribir $H_0: \tau_1 = \tau_2 = \tau_3 = \tau^*$, en donde τ^* es desconocida.

Las ecuaciones normales son las (5), y una solución es

$$\hat{\beta} = \begin{pmatrix} \hat{\mu} \\ \hat{\tau}_1 \\ \hat{\tau}_2 \\ \hat{\tau}_3 \end{pmatrix} = \begin{pmatrix} \bar{y}_{..} \\ \bar{y}_{1.} - \bar{y}_{..} \\ \bar{y}_{2.} - \bar{y}_{..} \\ \bar{y}_{3.} - \bar{y}_{..} \end{pmatrix}$$

lo que da

$$\hat{\beta}'\mathbf{X}'\mathbf{Y} = (\bar{y}_{..})(\mathbf{Y}_{..}) + \sum_{i=1}^3 (\bar{y}_{i.} - \bar{y}_{..})\mathbf{Y}_{i.}$$

Hagamos

$$\tau_1 = \tau^*, \quad \tau_2 = \tau^*, \quad \tau_3 = \tau^*,$$

y el modelo reducido es

$$y_{ij} = \mu + \tau^* + e_{ij} = \mu^* + e_{ij}$$

y hay un único parámetro desconocido μ^* . La ecuación normal se halla determinando el valor de μ^* que hace mínima a

$$\sum_{ij} e_{ij}^2 = \sum_{ij} (y_{ij} - \mu^*)^2.$$

Evidentemente, es

$$6\bar{\mu}^* = Y_{..}$$

que constituye el conjunto reducido de ecuaciones normales designadas por $Z'Z\tilde{y} = Z'Y$ en el teorema 14-7. La solución es

$$\bar{\mu}^* = \bar{y}_{..} \quad \text{y} \quad \tilde{y}'Z'Y = \bar{\mu}^*Y_{..} = (\bar{y}_{..})(Y_{..}).$$

Luego

$$\hat{\beta}'X'Y - \tilde{y}'Z'Y = \sum_{i=1}^3 (\bar{y}_{i.} - \bar{y}_{..})Y_{i.}$$

que puede escribirse en la forma

$$\sum_{j=1}^2 \sum_{i=1}^3 (\bar{y}_{i.} - \bar{y}_{..})^2 \quad \text{o bien} \quad \sum_i \frac{Y_{i.}^2}{2} - \frac{Y_{..}^2}{6}.$$

La suma de cuadrados del error es

$$\begin{aligned} Y'Y - \hat{\beta}'X'Y &= \sum_i \sum_j y_{ij}^2 - \frac{Y_{..}^2}{6} - \sum_{ij} (\bar{y}_{i.} - \bar{y}_{..})^2 \\ &= \sum_i \sum_j (y_{ij} - \bar{y}_{..})^2 - \sum_{ij} (\bar{y}_{i.} - \bar{y}_{..})^2 = \sum_i \sum_j (y_{ij} - \bar{y}_{i.})^2 \end{aligned}$$

y la cantidad u ,

$$u = \frac{\sum_i \sum_j (\bar{y}_{i.} - \bar{y}_{..})^2 / 2}{\sum_i \sum_j (y_{ij} - \bar{y}_{i.})^2 / 3}$$

ya que $k=2$ (número de funciones estimables linealmente independientes en H_0) y $n-r=6-3=3$ (r es la característica de X). También

$$\hat{\sigma}^2 = \frac{\sum \sum (y_{ij} - \bar{y}_i.)^2}{3}$$

es un estimador insesgado de σ^2 , puesto que

$$\hat{\sigma}^2 = [1/(n-r)](Y - X\hat{\beta})'(Y - X\hat{\beta}).$$

Obsérvese que se verifica la identidad

$$\sum_{ij} y_{ij}^2 = \frac{Y_{..}^2}{6} + \sum_{ij} (\bar{y}_i. - \bar{y}_{..})^2 + \sum_{ij} (y_{ij} - \bar{y}_i.)^2$$

que puede escribirse así

$$\sum_{ij} y_{ij}^2 = \frac{Y_{..}^2}{6} + \left(\sum_{i=1}^3 \frac{Y_{i.}^2}{2} - \frac{Y_{..}^2}{6} \right) + \left(\sum_{ij} y_{ij}^2 - \sum_{j=1}^3 \frac{Y_{.j}^2}{2} \right)$$

Esta segunda identidad resulta útil en el cálculo de varias cantidades.

$\frac{Y_{..}^2}{6}$ se denomina suma de cuadrados para la media, y

$$\sum_{ij} (\bar{y}_i. - \bar{y}_{..})^2$$

se llama suma de cuadrados debidos a las τ_i (que frecuentemente reciben el nombre de tratamientos);

$$\sum_{ij} (y_{ij} - \bar{y}_i.)^2$$

se llama suma de cuadrados para el error, y $\sum y_{ij}^2$, suma de cuadrados total no corregida.

Estimación por intervalos.—Para definir un intervalo confidencial para cualquier función estimable, se precisa conocer la varianza de la función y una estimación insesgada independiente de esta varianza.

Teorema 14-8.—*Un intervalo confidencial para la función estimable $\lambda'\beta$, con coeficiente de confianza $1 - \alpha$, es*

$$\lambda'\hat{\beta} - t_{\alpha/2} \sqrt{\widehat{\text{var}}(\lambda'\hat{\beta})} < \lambda'\beta < \lambda'\hat{\beta} + t_{\alpha/2} \sqrt{\widehat{\text{var}}(\lambda'\hat{\beta})} \quad (9)$$

en donde $\widehat{\text{var}}(\lambda'\hat{\beta})$ es la estimación insesgada de la varianza de $\lambda'\hat{\beta}$.

Con objeto de hallar el intervalo confidencial para $\lambda'\beta$, hay que obtener primero la varianza de la cantidad $\lambda'\hat{\beta}$; esta es una constante multiplicada por σ^2 , y la designaremos por $a\sigma^2$. Entonces,

$$\widehat{\text{var}}(\lambda'\hat{\beta}) = a\hat{\sigma}^2,$$

con

$$\hat{\sigma}^2 = (\mathbf{Y} - \mathbf{X}\hat{\beta})'(\mathbf{Y} - \mathbf{X}\hat{\beta}) / (n - r).$$

Ejemplo 14-4.—Para definir un intervalo confidencial para $\tau_1 - \tau_2$ en el ejemplo 14-3, hagamos $\lambda'\hat{\beta} = \bar{y}_1 - \bar{y}_2$. La varianza es

$$\begin{aligned} E[(\bar{y}_1 - \bar{y}_2) - (\tau_1 - \tau_2)]^2 &= E[(\mu + \tau_1 + \bar{e}_i - \mu - \tau_2 - \bar{e}_2)] \\ &\quad - (\tau_1 - \tau_2)]^2 = E(\bar{e}_1 - \bar{e}_2)^2 = \sigma^2 \end{aligned}$$

en donde

$$\bar{e}_i = 1/2 \sum_{j=1}^2 e_{ij}$$

Entonces

$$\text{var}(\hat{\tau}_1 - \hat{\tau}_2) = \sigma^2$$

y, a partir del ejemplo 14-3,

$$\hat{\sigma}^2 = \sum_i \sum_j \frac{(y_{ij} - \bar{y}_i)^2}{3};$$

estos valores se sustituirán en (9) para hallar el intervalo confidencial.

Puesto que las ecuaciones normales desempeñan un papel sumamente importante en los modelos de diseño experimental, daremos una regla mediante la cual puedan hallarse a partir del modelo $y_{ij\dots m} = \mu_{ij\dots m} + e_{ij\dots m}$ sin necesidad de construir efectivamente la matriz X . Para llegar a ella, observemos que en el modelo $\mathbf{Y} = \mathbf{X}\beta + \mathbf{e}$, a cada parámetro de β corresponde una ecuación del conjunto de ecuaciones normales, ya que, para obtener dicho conjunto, se igualan a cero las derivadas respecto a cada uno de los

parámetros. Observemos también que la componente β_i aparece en el s -ésimo elemento del vector $X\beta$ (de dimensión $n \times 1$) si, y solamente si, la i -ésima columna de X tiene un 1 en la s -ésima fila. Es decir, hay n ecuaciones en el modelo $Y = X\beta + e$ y el hecho de que el parámetro β_i aparezca en ciertas ecuaciones depende de los ceros y unos de la i -ésima columna de X ; si hay un 1 en la s -ésima fila de la i -ésima columna de X , β_i figurará en la s -ésima ecuación. Para aclarar esto, examinemos la ecuación (14-1-3) y observemos que τ_1 se presenta únicamente en las dos primeras ecuaciones del modelo. Esto se comprueba en la ecuación (14-1-1), en donde el modelo se ha desarrollado con detalle. Análogamente, para hallar en qué ecuaciones aparece μ , observamos la primera columna de la matriz X , ya que μ es el primer elemento de β . Esta columna tiene un 1 en cada fila, luego μ aparece en todas las ecuaciones del modelo, según se comprueba en la ecuación (14-1-1).

Para hallar las ecuaciones normales $X'X\hat{\beta} = X'Y$, estudiaremos el vector $X'Y$. La i -ésima columna de X pertenece a la componente i -ésima de β ; por tanto, la i -ésima fila de X' corresponderá al i -ésimo elemento de β . Sea X_i la i -ésima columna de X ; $X_i'Y$ es entonces el i -ésimo elemento del vector $X'Y$, y puesto que X_i está formada por ceros y unos, la cantidad $X_i'Y$ es la suma de aquellos elementos de Y que corresponden a la presencia de un 1 en X_i . Pero es exactamente en la ecuación del modelo en que aparece β_i donde hay un 1 en X_i . Por tanto, la regla es: *El i -ésimo elemento de $X'Y$ es la suma de los elementos de Y extendida a aquellas ecuaciones del modelo $Y = X\beta + e$ donde aparece β_i .*

Refiriéndonos de nuevo a la sección 14-1, dado que μ aparece en cada una de las ecuaciones de (14-1-1), el primer elemento de $X'Y$ es Y_1 . Así mismo, τ_1 aparece solo en las dos primeras ecuaciones (o sea, en aquellas en que y_{ij} tiene el primer subíndice igual a 1); por tanto, el segundo elemento es

$$\sum_{j=1}^2 y_{1j} = Y_1.$$

Análogamente, para τ_2 el elemento es Y_2 , etc. Luego $X'Y$ puede construirse examinando el modelo escrito en la forma (14-1-2). También $E(X'Y) = X'X\beta$, y, en consecuencia, se obtiene el primer miembro de las ecuaciones normales tomando el valor esperado de $X'Y$.

Hemos dedicado esta sección a los métodos generales que se utilizan para la estimación y docimasia de los parámetros en modelos de diseño experimental. En algunas de las secciones siguientes ilustraremos el empleo de la teoría examinando ciertos casos importantes y especiales.

14-3. Modelo de clasificación simple.—Supongamos que cierta compañía desea determinar si existe alguna diferencia entre los distintos métodos de fabricación de cable de acero. Imaginemos que se examina una muestra de n_i trozos de cable fabricado por el proceso i , para cada uno de t procesos, y que la resistencia a la rotura del j -ésimo trozo de cable fabricado por el i -ésimo proceso es y_{ij} , siendo

$$y_{ij} = \mu + \tau_i + e_{ij} \quad j = 1, 2, \dots, n_i; \quad i = 1, 2, \dots, t \quad (1)$$

donde μ es una media total; τ_i , el efecto del i -ésimo proceso, y e_{ij} , un error aleatorio no observable, debido a variaciones incontrolables en los procesos de fabricación y medición. Veremos con cierto detalle cómo se hallan las ecuaciones normales para este modelo; μ aparece en todas ellas y, por tanto, el elemento de $X'Y$ correspondiente a μ es

$$\sum_{ij} y_{ij} = Y_{..}$$

τ_p aparece solo en aquellas ecuaciones cuyo primer subíndice es p ; es decir,

$$\sum_j y_{pj} = Y_{p.}$$

Luego

$$X'Y = \begin{pmatrix} Y_{..} \\ Y_{1.} \\ Y_{2.} \\ \vdots \\ Y_{t.} \end{pmatrix}$$

También

$$E[Y_{p.}] = E \left[\sum_j^{n_p} y_{pj} \right] = E \left[\sum_j (\mu + \tau_p + e_{pj}) \right] = n_p \mu + n_p \tau_p$$

y

$$E[Y_{..}] = N\mu + \sum_{p=1}^t n_p \tau_p$$

en donde

$$\sum_{p=1}^t n_p = N.$$

(en donde SCM significa suma de cuadrados para la media; SCE suma de cuadrados del error, etc.) y se construye una tabla de análisis de la varianza, análoga a la tabla 14-1.

TABLA 14-1. ANÁLISIS DE LA VARIANZA

Fuente	Grados de libertad	Suma de cuadrados	Cuadrado medio	F
Total	N	$\sum y_{ij}^2$		
Media	1	$Y^2/N = \text{SCM}$		
Tratamientos (τ).	$t - 1$	$\sum_{ij} (\bar{y}_{i.} - \bar{y}_{..})^2 = \text{SCT}$	$\text{SCT}/(t - 1)$	$\frac{\text{SCT}}{\text{SCE}} \cdot \frac{N - t}{t - 1}$
Error	$N - t$	$\sum_{ij} (y_{ij} - \bar{y}_{i.})^2 = \text{SCE}$	$\text{SCE}/(N - t)$	

La columna F da el estadístico u del teorema 14-7. Para definir un intervalo confidencial para cualquier contraste $\Sigma a_p \tau_p$, observemos que $\Sigma a_p \hat{\tau}_p = \Sigma a_p \bar{y}_p$, se distribuye normalmente con media $\Sigma a_p \tau_p$ y varianza $\sigma^2 \Sigma a_p^2/n$ y $\text{SCE}/(N - t) = \hat{\sigma}^2$. Estas cantidades pueden utilizarse en el teorema 14-8.

14-4. Modelo de clasificación doble.—Tal vez se haya observado que el experimento descrito en el ejemplo 14-1 tenía un diseño poco satisfactorio. El inconveniente se debía a la presencia de un factor extrínseco, la aptitud de los diversos obreros que han de intervenir necesariamente en el experimento. Si la producción de una máquina resulta relativamente grande, ¿se deberá ello a la máquina o a la superioridad del grupo particular de trabajadores que se le asignó? No hay modo de decirlo con dicho experimento. En el lenguaje del diseño de experimentos, los efectos debidos a máquinas y los debidos a grupos de trabajadores están completamente *confundidos*, y no hay modo de distinguir entre estos dos factores.

La dificultad desaparece si se diseña de nuevo el experimento considerándolo como bifactorial. Supongamos, p. ej., que intervienen solo cinco hombres en el experimento y que cada uno de ellos trabaja un día en cada una de las tres máquinas. El orden en que ha de trabajar un hombre dado en las tres máquinas se fijará aleatoriamente. Ahora, los datos quedarán clasificados en una

tabla de doble entrada, de acuerdo con ambos factores, que será de forma análoga a la incluida como tabla 14-2. Cuando se utiliza un experimento bifactorial para controlar un factor extrínseco, como sucede aquí, el diseño se conoce con el nombre de *diseño de bloques aleatorizados*. El factor interesante se compara en bloques (hombres, en este ejemplo) de modo que las condiciones de la comparación resulten homogéneas dentro de cada bloque, aunque difieran de un bloque a otro.

TABLA 14-2

		Máquina		
		1	2	3
Hombre	1	53	47	57
	2	56	50	63
	3	45	47	54
	4	52	47	57
	5	49	53	58

El modelo puede escribirse así:

$$y_{ij} = \mu + \alpha_i + \tau_j + e_{ij} \quad \begin{matrix} i = 1, 2, \dots, r \\ j = 1, 2, \dots, c \end{matrix} \quad (1)$$

en donde y_{ij} es la observación de la i -ésima fila y j -ésima columna; μ , una media general; α_i , el efecto del i -ésimo nivel del factor A ; τ_j , el efecto del j -ésimo nivel del factor B , y e_{ij} , un error aleatorio no observable que supondremos encuadrado dentro de los casos A o B .

Para hallar las ecuaciones normales, aplicaremos la regla de la sección 14-2. Puesto que μ aparece en *todas* las ecuaciones de (1), el término de $X'Y$ que corresponde a μ se suma respecto a *todos* los subíndices, lo que da $Y_{..}$. En cambio, α_p aparece solo en aquellas ecuaciones de (1) para las que el primer subíndice de y_{ij} es p ; es decir, y_{pi} , cuya suma da $Y_{p.}$. Análogamente, para τ_q el término de $X'Y$

es $Y_{.q}$, y así las ecuaciones normales (tomamos el valor esperado de $X'Y$) son:

$$\begin{aligned}\mu: & \quad rc\hat{\mu} + c\Sigma\hat{\alpha}_i + r\Sigma\hat{\tau}_j = Y_{..} \\ \alpha_p: & \quad c\hat{\mu} + c\hat{\alpha}_p + \Sigma\hat{\tau}_j = Y_{p.} \quad p=1, 2, \dots, r \\ \tau_q: & \quad r\hat{\mu} + \Sigma\hat{\alpha}_i + r\hat{\tau}_q = Y_{.q} \quad q=1, 2, \dots, c\end{aligned}\quad (2)$$

Hay c ecuaciones para τ , r para α y una sola para μ ; o sea, en total $r+c+1$ ecuaciones. Puesto que la suma de las ecuaciones para τ es igual a la correspondiente a μ , y lo mismo sucede con la suma de las ecuaciones para α , existirán al menos dos dependencias lineales entre las ecuaciones normales y, por tanto, la característica será menor o igual que

$$(r+c+1)-2=r+c-1$$

En modelos como este, se desea en general estimar contrastes de las α_p y de las τ_q . Multiplicando por a_p la ecuación correspondiente a α_p de (2), y sumando respecto a p , resulta:

$$\Sigma a_p c \hat{\mu} + \Sigma a_p c \hat{\alpha}_p + \Sigma a_p (\Sigma \hat{\tau}_j) = \Sigma a_p Y_{p.}$$

Si $\Sigma a_p = 0$, obtenemos

$$\Sigma a_p \hat{\alpha}_p = \Sigma a_p \bar{y}_{p.};$$

por tanto, el estimador máximo verosímil (y el de mínimos cuadrados) del contraste $\Sigma a_p \alpha_p$ es $\Sigma a_p \bar{y}_{p.}$. Análogamente, $\Sigma b_q \hat{\tau}_q = \Sigma b_q \bar{y}_{.q}$ si $\Sigma b_q = 0$. Puesto que cada contraste de las α_p y τ_q es estimable, representan $(r-1)+(c-1)$ funciones estimables linealmente independiente. Además, por la primera ecuación de (2), $rc\hat{\mu} + c\Sigma\hat{\alpha}_p + r\Sigma\tau_q$ es estimable y linealmente independiente de los contrastes de α_p y τ_q . Por consiguiente, existen por lo menos $r+c-1$ funciones estimables linealmente independientes, y, en virtud de lo anterior, se deduce que hay exactamente $r+c-1$ funciones estimables linealmente independientes. En consecuencia, la característica del sistema de ecuaciones normales es $r+c-1$.

Para calcular

$$\hat{\sigma}^2 = (Y - X\hat{\beta})'(Y - X\hat{\beta}) / (rc - r - c + 1),$$

necesitamos una solución cualquiera de las ecuaciones normales; una, que es inmediata, se obtiene haciendo

$$\Sigma \hat{\alpha}_p = \Sigma \hat{\tau}_q = 0.$$

Entonces,

$$\begin{aligned}\hat{\mu} &= \bar{y}_{..} \\ \hat{\alpha}_p &= \bar{y}_{p.} - \bar{y}_{..} & p = 1, 2, \dots, r \\ \hat{\tau}_q &= \bar{y}_{.q} - \bar{y}_{..} & q = 1, 2, \dots, c\end{aligned}$$

Luego

$$\begin{aligned}(Y - X\hat{\beta})(Y - X\hat{\beta})' &= Y'Y - \hat{\beta}'X'Y = \sum_{ij} y_{ij}^2 - \hat{\mu}Y_{..} - \sum_p \hat{\alpha}_p(Y_{p.}) \\ &- \sum_q \hat{\tau}_q(Y_{.q}) = \sum_{ij} y_{ij}^2 - \frac{Y_{..}^2}{rc} - \left(\sum_p \frac{Y_{p.}^2}{c} - \frac{Y_{..}^2}{rc} \right) - \left(\sum_q \frac{Y_{.q}^2}{r} - \frac{Y_{..}^2}{rc} \right) \\ &= \sum_{ij} (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2 = \text{suma de cuadrados del error}\end{aligned}$$

y

$$\hat{\sigma}^2 = \frac{\sum_{ij} (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2}{(r-1)(c-1)}$$

Para docimar la hipótesis $H_0: \alpha_1 = \alpha_2 = \dots = \alpha_r = \alpha$ hacemos $\alpha_i = \alpha$ en (1) y reemplazamos $\mu + \alpha$ por μ^* . El modelo reducido es

$$y_{ij} = \mu^* + \tau_j + e_{ij} \quad \begin{aligned} i &= 1, 2, \dots, r \\ j &= 1, 2, \dots, c \end{aligned}$$

y las ecuaciones normales son

$$\begin{aligned}\mu^*: \quad rc\tilde{\mu}^* + r\sum_j \tilde{\tau}_j &= Y_{..} \\ \tau_q: \quad r\tilde{\mu}^* + r\tilde{\tau}_q &= Y_{.q} \quad q = 1, 2, \dots, c\end{aligned}$$

Evidentemente, la característica es c , y una solución es

$$\tilde{\mu}^* = \bar{y}_{..} \quad \tilde{\tau}_q = \bar{y}_{.q} - \bar{y}_{..};$$

por tanto,

$$\tilde{y}'Z'Y = \tilde{\mu}^*Y_{..} + \sum_q \tilde{\tau}_q Y_{.q} = \frac{Y_{..}^2}{rc} + \left(\sum_q \frac{Y_{.q}^2}{r} - \frac{Y_{..}^2}{rc} \right)$$

y

$$\hat{\beta}'X'Y - \tilde{y}'Z'Y = \sum_p \frac{Y_{p.}^2}{c} - \frac{Y_{..}^2}{rc} = \sum_{ij} (\bar{y}_{i.} - \bar{y}_{..})^2$$

Por el teorema 14-7, se obtiene

$$u = \frac{\sum_{ij} (\bar{y}_{i.} - \bar{y}_{..})^2 / (r-1)}{\sum_{ij} (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2 / (r-1)(c-1)}$$

que se distribuye como F si H_0 es cierta, y se rechaza H_0 si $u > F_{\alpha}$.

Por un procedimiento semejante obtenemos una d cima de la hip tesis H_0 :

$$\tau_1 = \tau_2 = \dots = \tau_c$$

la cual es

$$u = \frac{\sum_{ij} (\bar{y}_{.j} - \bar{y}_{..})^2 / (c-1)}{\sum_{ij} (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2 / (r-1)(c-1)}$$

Es interesante observar la identidad

$$\sum_{ij} y_{ij}^2 = \frac{Y_{..}^2}{rc} + \sum_{ij} (\bar{y}_{i.} - \bar{y}_{..})^2 + \sum_{ij} (\bar{y}_{.j} - \bar{y}_{..})^2 + \sum_{ij} (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2,$$

la cual se dispone frecuentemente en forma de tabla, como se ha hecho en la tabla 14-3.

TABLA 14-3. TABLA DE AN LISIS DE LA VARIANZA
PARA UNA CLASIFICACI N DOBLE

Fuente	Grados de libertad	Suma de cuadrados	Cuadrado medio	F
Total	rc	$\sum y_{ij}^2$		
Media	1	$\frac{Y_{..}^2}{rc}$		
Clase A (α) ...	$r-1$	$\sum_{ij} (\bar{y}_{i.} - \bar{y}_{..})^2$	s_1^2	s_1^2/s_3^2
Clase B (τ) ...	$c-1$	$\sum_{ij} (\bar{y}_{.j} - \bar{y}_{..})^2$	s_2^2	s_2^2/s_3^2
Error	$(r-1)(c-1)$	$\sum_{ij} (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2$	s_3^2	

14-5. Otros modelos.—Tanto el modelo de clasificación simple, $y_{ij} = \mu + \tau_i + e_{ij}$, como el de clasificación doble, $y_{ij} = \mu + \alpha_i + \tau_j + e_{ij}$, son casos especiales de un modelo más general que puede escribirse en la forma,

$$y_{ij\dots t} = \mu_{ij\dots t} + e_{ij\dots t}$$

en donde $\mu_{ij\dots t}$ es una constante desconocida, y $e_{ij\dots t}$, una variable aleatoria. Este se conoce con el nombre de modelo I de Eisenhart.

Otros casos especiales son:

Clasificación triple (sin interacción):

$$y_{ijk} = \mu + \alpha_i + \gamma_j + \tau_k + e_{ijk}$$

Clasificación doble (con interacción):

$$y_{ijk} = \mu + \alpha_i + \tau_j + (\alpha\tau)_{ij} + e_{ijk}$$

No estudiaremos estos modelos, pero los métodos utilizados para estimar los parámetros y docimar hipótesis en estos y otros casos especiales son análogos a los explicados en las secciones anteriores.

PROBLEMAS

1. Sea un modelo simple dado por

$$y_{11} = \mu + \alpha_1 + e_{11}$$

$$y_{12} = \mu + \alpha_1 + e_{12}$$

$$y_{21} = \mu + \alpha_2 + e_{21}$$

$$y_{31} = \mu + \alpha_3 + e_{31}$$

a) Escribáanse la matriz X y el vector β , si dicho modelo se representa en forma matricial,

$$Y = X\beta + e.$$

b) Hállese $X'X$.

c) Demuéstrese que

$$X'Y = \begin{pmatrix} Y_{..} \\ Y_{1.} \\ y_{21} \\ y_{31} \end{pmatrix}$$

d) Escribáanse las ecuaciones normales.

e) Hállese la característica de $X'X$.

f) Obténgase la estimación máximo-verosímil (o la de mínimos cuadrados) de $\alpha_1 - \alpha_2$.

2. El modelo de clasificación simple describe los datos de la sección 14-1, en donde $n_i=2$ y $t=3$. Por el método descrito en la sección 14-3:

a) Compruébese que

$$X'Y = \begin{pmatrix} 314 \\ 103 \\ 89 \\ 122 \end{pmatrix}$$

b) Utilizando la ecuación (14-3-1) escríbanse explícitamente las ecuaciones normales.

c) Hállese una solución $\hat{\beta}$ de las ecuaciones normales y calcúlese

$$\hat{\sigma}^2 = \frac{(Y - X\hat{\beta})'(Y - X\hat{\beta})}{3}$$

d) Compruébese que

$$\hat{\tau}_1 - \hat{\tau}_2 = 7,0; \quad \hat{\tau}_3 - \hat{\tau}_1 = 9,5 \quad \text{y} \quad \hat{\tau}_3 - \hat{\tau}_2 = 16,5.$$

e) Calcúlense las cantidades de la tabla 14-1.

f) En la parte e), compruébese que $SCE/(N-t)$ es igual a la cantidad de la parte c).

g) Use la parte e) para docimar $H_0: \tau_1 = \tau_2 = \tau_3$ con una probabilidad de error de tipo I igual a 0,05.

h) ¿Es estimable $\tau_1/2 + \tau_2/3 + \tau_3/6 + \mu$?

3. En el problema anterior, hállese la estimación de $\text{var}(\hat{\tau}_1 - \hat{\tau}_2)$.

4. En el problema 2, utilícese el teorema 14-8 para definir un intervalo confidencial del 0,95 para $\tau_1 - \tau_2$.

5. Los datos siguientes satisfacen el modelo de clasificación doble

$$y_{ij} = \mu + \alpha_i + \theta_j + e_{ij} \quad \begin{matrix} i=1, 2, 3, 4 \\ j=1, 2, 3, 4, 5, 6 \end{matrix}$$

Factor A (θ)

		1	2	3	4	5	6	Total
Factor B (α)	1	21	17	56	59	41	51	
	2	20	19	61	62	46	55	
	3	24	23	54	54	39	50	
	4	14	18	56	55	42	48	
Total								

a) Escribáse el vector β de dimensión 11×1 .

b) Utilícense las ecuaciones (14-4-2) y escribáse las 11 ecuaciones normales con los valores de r , c , $Y_{i.}$, $Y_{.j}$, etc., correspondientes a los datos de la tabla.

6. En el problema anterior, hállese los valores de $\hat{\alpha}_1 - \hat{\alpha}_2$, de $\hat{\alpha}_1 + \hat{\alpha}_2 - 2\hat{\alpha}_3$ y de $\hat{\theta}_1 + 2\hat{\theta}_3 + \hat{\theta}_4 - 4\hat{\theta}_5$.

7. Calcúlense las cantidades de la tabla 14-3 utilizando los datos del problema 5.

8. En un modelo de clasificación doble,

$$y_{ij} = \mu + \alpha_i + \theta_j + e_{ij}, \quad i = 1, 2, \dots, r \text{ y } j = 1, 2, \dots, c,$$

demuéstrese, a partir de las ecuaciones normales (14-4-2), que

$$a) \quad \hat{\theta}_1 - \hat{\theta}_2 = \bar{y}_{.1} - \bar{y}_{.2}.$$

$$b) \quad \hat{\theta}_1 + \hat{\theta}_2 - 2\hat{\theta}_3 = \bar{y}_{.1} + \bar{y}_{.2} - 2\bar{y}_{.3}.$$

c) Los dos estimadores de las partes a) y b) están incorrelacionados cuando las variables aleatorias e_{ij} satisfacen las condiciones de los casos A o B.

9. Sea el modelo de clasificación doble dado en el problema 8, y supóngase que las variables aleatorias e_{ij} satisfacen las estipulaciones del caso A o las del caso B. Por la teoría de la sección 14-4, el contraste

$$\sum_{p=1}^c a_p \theta_p$$

se estima por $\sum a_p \bar{y}_{.p}$, y el

$$\sum_{k=1}^c b_k \theta_k,$$

por $\sum b_p \bar{y}_{.p}$.

a) Pruébese que

$$\text{var}(\sum a_p \bar{y}_{.p}) = \frac{\sigma^2}{r} \sum_{p=1}^c a_p^2.$$

b) Demuéstrese que $\sum a_p \bar{y}_{.p}$ y $\sum b_k \bar{y}_{.k}$ están incorrelacionadas si

$$\sum_{p=1}^c a_p b_p = 0.$$

10. En el problema 5, pruébese que $\theta_1 - \bar{\theta}$ es estimable.

$$\left(\bar{\theta} = 1/6 \sum_{i=1}^6 \theta_i \right)$$

Hállese el valor de la estimación insesgada lineal «óptima».

11. En el problema 5, defínase un intervalo confidencial del 0,95 para $\alpha_1 + \alpha_2 - 2\alpha_3$.

12. En el modelo de clasificación triple sin interacción

$$\begin{aligned} y_{ijk} &= \mu + \alpha_i + \theta_j + \tau_k + e_{ijk} \\ i &= 1, 2, \dots, a \\ j &= 1, 2, \dots, b \\ k &= 1, 2, \dots, c \end{aligned}$$

demuéstrese que las ecuaciones normales son

$$\begin{aligned} abc\hat{\mu} + bc\Sigma\hat{\alpha}_i + ac\Sigma\hat{\theta}_j + ab\Sigma\hat{\tau}_k &= Y_{..} \\ bc\hat{\mu} + bc\hat{\alpha}_p + c\Sigma\hat{\theta}_j + b\Sigma\hat{\tau}_k &= Y_{p..} & p=1, 2, \dots, a \\ ac\hat{\mu} + c\Sigma\hat{\alpha}_i + ac\hat{\theta}_q + a\Sigma\hat{\tau}_k &= Y_{.q.} & q=1, 2, \dots, b \\ ab\hat{\mu} + b\Sigma\hat{\alpha}_i + a\Sigma\hat{\theta}_j + ab\hat{\tau}_r &= Y_{...r} & r=1, 2, \dots, c \end{aligned}$$

13. Compruébese que una solución de las ecuaciones normales del problema 12 es

$$\begin{aligned} \hat{\beta}' &= (\hat{\mu}, \hat{\alpha}_1, \dots, \hat{\alpha}_a, \hat{\theta}_1, \dots, \hat{\theta}_b, \hat{\tau}_1, \dots, \hat{\tau}_c) \\ &= (\bar{y}_{...}, \bar{y}_{1..} - \bar{y}_{...}, \dots, \bar{y}_{a..} - \bar{y}_{...}, \bar{y}_{.1.} - \bar{y}_{...}, \dots, \bar{y}_{.b.} - \bar{y}_{...}, \bar{y}_{..1} - \bar{y}_{...}, \dots, \bar{y}_{..c} - \bar{y}_{...}) \end{aligned}$$

o sea,

$$\hat{\mu} = \bar{y}_{...}; \quad \hat{\alpha}_p = \bar{y}_{p..} - \bar{y}_{...}; \quad \hat{\theta}_q = \bar{y}_{.q.} - \bar{y}_{...}; \quad \hat{\tau}_r = \bar{y}_{...r} - \bar{y}_{...}.$$

14. En los problemas 12 y 13, demuéstrese que

$$\begin{aligned} \hat{\beta}'X'Y &= \hat{\mu}Y_{...} + \sum_{p=1}^a \hat{\alpha}_p Y_{p..} + \sum_{q=1}^b \hat{\theta}_q Y_{.q.} + \sum_{r=1}^c \hat{\tau}_r Y_{...r} \\ &= \bar{y}_{...} Y_{...} + \sum_{p=1}^a (\bar{y}_{p..} - \bar{y}_{...}) Y_{p..} + \sum_{q=1}^b (\bar{y}_{.q.} - \bar{y}_{...}) Y_{.q.} + \sum_{r=1}^c (\bar{y}_{...r} - \bar{y}_{...}) Y_{...r} \\ &= \frac{Y_{...}^2}{abc} + \left(\sum_{p=1}^a \frac{Y_{p..}^2}{bc} - \frac{Y_{...}^2}{abc} \right) + \left(\sum_{q=1}^b \frac{Y_{.q.}^2}{ac} - \frac{Y_{...}^2}{abc} \right) + \left(\sum_{r=1}^c \frac{Y_{...r}^2}{ab} - \frac{Y_{...}^2}{abc} \right) \\ &= \frac{Y_{...}^2}{abc} + \sum_{pqr} (\bar{y}_{p..} - \bar{y}_{...})^2 + \sum_{pqr} (\bar{y}_{.q.} - \bar{y}_{...})^2 + \sum_{pqr} (\bar{y}_{...r} - \bar{y}_{...})^2 \end{aligned}$$

15. Para docimar la hipótesis $\alpha_1 = \alpha_2 = \dots = \alpha_a = \alpha$ (desconocida), el modelo reducido es

$$y_{ijk} = \mu^* + \theta_j + \tau_k + e_{ijk},$$

en donde $\mu^* = \mu + \alpha$. Escribanse las ecuaciones normales $Z'Z\tilde{y} = Z'Y$ para

el modelo y pruébese (utilizando los símbolos del teorema 14-7) que una solución es

$$\begin{aligned}\tilde{\gamma} &= (\tilde{\mu}^*, \tilde{\theta}_1, \dots, \tilde{\theta}_b, \tilde{\tau}_1, \dots, \tilde{\tau}_c) \\ &= (\tilde{y}_{...}, \tilde{y}_{.1.} - \tilde{y}_{...}, \dots, \tilde{y}_{.b.} - \tilde{y}_{...}, \tilde{y}_{.1.} - \tilde{y}_{...}, \dots, \tilde{y}_{.c.} - \tilde{y}_{...})\end{aligned}$$

En otras palabras,

$$\tilde{\mu}^* = \tilde{y}_{...} \quad \tilde{\theta}_q = \tilde{y}_{.q.} - \tilde{y}_{...} \quad \tilde{\tau}_r = \tilde{y}_{.r.} - \tilde{y}_{...}$$

16. En el problema 15, hállese $\tilde{\gamma}'Z'Y$.

17. En el problema 16, demuéstrese que

$$\hat{\beta}'X'Y - \tilde{\gamma}'Z'Y = \sum_{ijk} (\tilde{y}_{i..} - \tilde{y}_{...})^2$$

18. En el problema 13, pruébese que, para docimar la hipótesis

$$H_0: \alpha_1 = \alpha_2 = \dots = \alpha_a$$

utilizando el teorema 14-7, la cantidad

$$u = \frac{\sum_{ijk} (\tilde{y}_{i..} - \tilde{y}_{...})^2 / (a-1)}{\sum_{ijk} (y_{ijk} - \tilde{y}_{i..} - \tilde{y}_{.j.} - \tilde{y}_{..k} + 2\tilde{y}_{...})^2 / T},$$

en donde $T = abc - a - b - c + 2$, se distribuye como F con $a-1$ y T grados de libertad cuando H_0 es cierta.

19. Hállese la cantidad u para docimar la hipótesis $H_0: \theta_1 = \theta_2 = \dots = \theta_c$ en el problema 12.

20. En el modelo de clasificación doble con interacción

$$y_{ijk} = \mu + \alpha_i + \theta_j + (\alpha\theta)_{ij} + e_{ijk}$$

fórmense las ecuaciones normales.

BIBLIOGRAFIA

1. ANDERSON, R. L., y T. A. BANCROFT: *Statistical Theory in Research*, McGraw-Hill Book Company, Inc., Nueva York, 1952.
2. BENNETT, C. A., y N. L. FRANKLIN: *Statistical Analysis in Chemistry and the Chemical Industry*, John Wiley & Sons, Inc., Nueva York, 1954.
3. COCHRAN, W. G., y G. M. COX: *Experimental Designs*, John Wiley & Sons, Inc., Nueva York, 1957.
4. DAVIES, O. L.: *Métodos estadísticos aplicados a la investigación y a la producción*, Aguilar, 2.^a ed., 1965, Madrid.

5. FEDERER, W. T.: *Experimental Design, Theory and Application*, The Macmillan Company, Nueva York, 1955.
6. GOULDEN, C. H.: *Methods of Statistical Analysis*, John Wiley & Sons, Inc., Nueva York, 1939.
7. GRAYBILL, F. A.: *An Introduction to Linear Statistical Models*, vol. I, McGraw-Hill Book Company, Inc., Nueva York, 1961.
8. KEMPTHORNE, O.: *Design and Analysis of Experiments*, John Wiley & Sons, Inc., Nueva York, 1952.
9. RAO, C. R.: *Advanced Statistical Methods in Biometric Research*, John Wiley & Sons, Inc., Nueva York, 1952.
10. SCHEFFÉ, H.: *The Analysis of Variance*, John Wiley & Sons, Inc., Nueva York, 1959.
11. SNEDECOR, G. W.: *Statistical Methods*, Iowa State College Press, Ames, Iowa, 1956.

CAPITULO 15

DOCIMAS SUCESIONALES DE HIPOTESIS

15-1. Análisis sucesionales.—El *análisis sucesional* se refiere a técnicas para la docimasia de hipótesis o la estimación de parámetros, cuando el tamaño de la muestra no está fijado de antemano, sino que se determina durante el curso del experimento mediante criterios que dependen de las observaciones que se van obteniendo.

En el capítulo 12 hemos considerado la dócima de $H_0: \theta = \theta_0$ contra $H_1: \theta = \theta_1$. Se ha demostrado que, para muestras $(x_1, x_2, \dots, x_n)$ de tamaño n , la dócima de la razón de verosimilitud hace mínimo el error de tipo II cuando se ha fijado el de tipo I. Así, si se elige α como error de tipo I, α determina un número A mediante la ecuación

$$\iint_{\lambda_n > A} \dots \int f_0(x_1)f_0(x_2) \dots f_0(x_n) dx_1 \dots dx_n = \alpha \quad (1)$$

en donde

$$\lambda_n = \prod_1^n \frac{f_1(x_i)}{f_0(x_i)} \quad (2)$$

y la región crítica para el rechazamiento de H_0 es

$$\lambda_n > A \quad (3)$$

Esta región hace mínima la probabilidad β (error de tipo II) de aceptar H_0 cuando es cierta H_1 . Escribiremos $f(x; \theta_i)$ en la forma más simple $f_i(x)$.

Supongamos que se desean fijar de antemano α y β . Si se conociese el tamaño de la muestra, podría hacerse lo siguiente: empezar por determinar A_n en función de n , mediante (1), y obtener después β como función de n

$$\beta_n = \iint_{\lambda_n < A_n} \dots \int f_1(x_1)f_1(x_2) \dots f_1(x_n) dx_1 \dots dx_n \quad (4)$$

por último, elegir n de modo que β_n tenga el valor deseado.

Imaginemos, además, que para, p. ej., $\alpha=0,01$ y $\beta=0,01$, y para funciones particulares f_0 y f_1 hubiésemos efectuado los cálculos necesarios, y hallado el valor 100 para n . Las siguientes consideraciones ponen de relieve el interés del análisis sucesional, tanto desde el punto de vista teórico como desde el práctico. Al obtener 100 observaciones para docimar H_0 , es posible que entre las primeras haya una o más que tengan valores tales que el eventual rechazamiento de H_0 esté fuera de duda, por lo que supondría una pérdida de tiempo efectuar las observaciones restantes.

En otros ejemplos, las primeras 20 observaciones, o las 30 ó 40 primeras, pueden proporcionar base suficiente, respecto a α y β , para aceptar o rechazar H_0 . En resumen, surge la posibilidad de construir la dócima de manera que permita terminar el muestreo en cualquiera de las observaciones, con lo cual se podrá docimar H_0 con los errores α y β prefijados, utilizando, sin embargo, menos de 100 observaciones, por término medio. Así sucede en efecto, aunque a primera vista parezca sorprendente, dado que la mejor dócima para un tamaño prefijado de la muestra requiere efectuar 100 observaciones. El ahorro en el número de estas suele ser importante; a veces, superior al 50%. Esto es, repitiendo las dócimas de H_0 frente a H_1 , para valores prefijados de ambos errores, pueden necesitarse 100 observaciones por dócima para tamaños dados de la muestra; pero, en el muestreo sucesional con los mismos errores, se necesitan solo, por término medio, 50 observaciones por dócima.

15-2. Construcción de dócimas sucesionales.—En esta sección se desarrollará la teoría de la docimasia sucesional para el caso de una hipótesis nula H_0 contra una sola alternativa H_1 . En

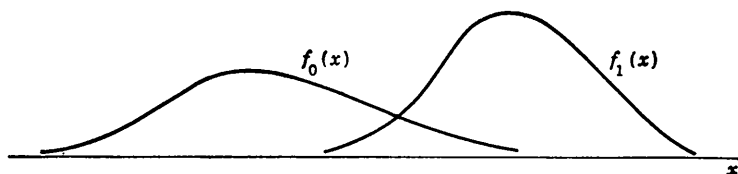


FIG. 15-1.

secciones posteriores del capítulo se pondrá en evidencia que este caso es útil en la aplicación de otros métodos a problemas prácticos. Designaremos por H_0 la hipótesis que supone una densidad $f_0(x)$ y por H_1 la correspondiente a $f_1(x)$. Véase figura 15-1. Las observaciones se representarán por $x_1, x_2, \dots$, en donde los sub-índices indican el orden en que se toman.

La d cima sucesional utiliza la raz n de verosimilitud

$$\lambda_m = \prod_{i=1}^m \frac{f_1(x_i)}{f_0(x_i)} \quad (1)$$

y dos n meros positivos A y B , siendo $A > 1$ y $B < 1$. Al efectuar observaciones, se calculan las razones $\lambda_1, \lambda_2, \lambda_3, \dots$, y se contin a realiz ndolas mientras

$$B < \lambda_m < A \quad (2)$$

Si para alg n m , λ_m es menor o igual a B , se acepta H_0 y la d cima queda terminada. Si λ_m resulta mayor o igual que A para cierto valor de m , se rechaza H_0 y queda completada la d cima. El procedimiento consiste, por tanto, en continuar el muestreo hasta que λ_m cae fuera del intervalo (2), en cuyo caso se da por terminado el muestreo.

La primera pregunta que se nos plantea es:  qu  impide que el muestreo prosiga indefinidamente? Se demuestra que esto no puede ocurrir; o sea, que la probabilidad de que el proceso termine, cualquiera que sea la distribuci n de x , vale 1. Sea

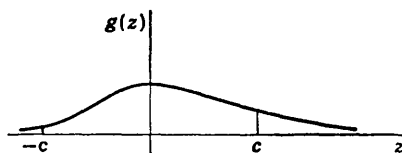
$$z = \log \left[\frac{f_1(x)}{f_0(x)} \right]; \quad (3)$$

z tendr  cierta distribuci n, $g(z)$ (con varianza positiva), determinada por la distribuci n de x [que no necesita ser precisamente $f_0(x)$ o $f_1(x)$]. La sucesi n de observaciones $x_1, x_2, \dots$ determina otra sucesi n de observaciones z , tal como $z_1, z_2, \dots$. La sucesi n de desigualdades (2) es

$$\log B < \sum_1^m z_i < \log A \quad (4)$$

en donde $\log B$ es negativo, y $\log A$, positivo. Sea $c = \log A - \log B$, y p , el  rea limitada por $g(z)$ entre $-c$ y c . Ahora bien: si cualquiera de las z_i cae fuera del intervalo $(-c, c)$, necesariamente una de las desigualdades (4) habr  quedado incumplida en esta o en alguna etapa anterior. Por tanto, si ha de verificarse (4) para todo valor de m , una al menos de las z_i deber  caer entre $-c$ y c . (Desde luego, las desigualdades pueden no cumplirse aunque todas las z sean interiores a dicho intervalo.) La probabilidad de que, para las primeras m observaciones, cada z_i caiga en el intervalo es

p^m , puesto que son independientes; esta probabilidad tiende a cero al aumentar m , ya que p es menor que uno. Por tanto, (4) no puede seguir siendo cierta con probabilidad 1. [En el caso de que $g(z)$ sea cero fuera del intervalo $(-c, c)$ cabría definir nuevas variables y_i , designando por y_1 la suma de las r primeras z , por y_2 la suma de



las r siguientes, y así sucesivamente, tomando r suficientemente grande para que la probabilidad de que un valor de y caiga fuera del intervalo $(-c, c)$ sea positiva.]

Volvamos ahora a la determinación de A y B . La probabilidad α de rechazar H_0 cuando es cierta se halla calculando la correspondiente a que λ_m exceda a A antes que resulte inferior a B . Es evidente que

$$\alpha = P(\lambda_1 \geq A) + P(B < \lambda_1 < A, \lambda_2 \geq A) \\ + P(B < \lambda_1 < A, B < \lambda_2 < A, \lambda_3 \geq A) + \dots \quad (5)$$

Análogamente, la probabilidad β de aceptar H_0 , cuando es cierta H_1 , será

$$\beta = P(\lambda_1 \leq B) + P(B < \lambda_1 < A, \lambda_2 \leq B) \\ + P(B < \lambda_1 < A, B < \lambda_2 < A, \lambda_3 \leq B) + \dots \quad (6)$$

Para dos densidades $f_0(x)$ y $f_1(x)$ dadas, pueden calcularse todas estas probabilidades utilizando $f_0(x)$ en (5) y $f_1(x)$ en (6). De esto se deduce que α y β son funciones conocidas de A y B ; por tanto, si α y β están prefijadas, A y B se determinan a partir de (5) y (6).

Como resulta fácil prever, la determinación efectiva de A y B a partir de tales expresiones supone la realización de cálculos complicados. En la práctica, nunca se determinan por tal procedimiento, porque se dispone de una aproximación muy sencilla y acurada. Las fórmulas aproximadas son

$$A \cong \frac{1 - \beta}{\alpha} \quad (7)$$

$$B \cong \frac{\beta}{1 - \alpha} \quad (8)$$

que se obtienen a partir de las consideraciones siguientes. Imaginemos que λ_m fuese una función continua de una variante continua m , de forma que λ_m pudiera representarse como una curva en función de m , y supongamos que se realizase la dócima moviéndose a lo largo del eje m hasta que λ_m fuese por primera vez igual a A o B . Esto es, la dócima se continúa mientras siga siendo cierta (2) y cesa cuando se verifica $\lambda_m = B$ (se acepta H_0), o $\lambda_m = A$ (se acepta H_1). En todos los puntos del espacio $(x_1, x_2, \dots)$ en que se acepta H_0 , la verosimilitud de H_1 , que designamos por L_1 , es exactamente igual a B veces la verosimilitud L_0 de H_0 , ya que en tales puntos $\lambda = L_1/L_0 = B$. Por consiguiente, la integral de L_1 en dichos puntos es exactamente igual a B veces la integral de L_0 en los mismos. Pero la primera integral es β , y la segunda es $1 - \alpha$ (probabilidad de aceptar H_0 cuando es cierta). De manera que tendríamos β exactamente igual a $B(1 - \alpha)$, y se verificaría (8). Por un razonamiento análogo en $\lambda_m = A$, (7) sería exactamente una igualdad. Puesto que el error al usar (7) y (8) es consecuencia de ser m discreta, cabría esperar que fuese pequeño, y, en efecto, se demuestra analíticamente que es muy pequeño cuando α y β son menores que $1/2$. Omitimos, sin embargo, los detalles de esta cuestión.

Las ecuaciones (7) y (8) hacen extraordinariamente simple la realización efectiva de una dócima sucesional. No es necesario desarrollar teoría alguna de distribuciones en el muestreo; basta elegir α y β arbitrariamente, calcular A y B y llevar a efecto la dócima sin más dilación. Estableceremos este resultado como una regla:

Regla 1.—Para realizar una dócima sucesional de la razón de verosimilitud de la hipótesis $H_0: \theta = \theta_0$ contra $H_1: \theta = \theta_1$ en la densidad $f(x; \theta)$, con probabilidades aproximadas α y β , respectivamente, para los errores de los tipos I y II, se calcula:

1. A , en donde $A = (1 - \beta)/\alpha$.
2. B , siendo $B = \beta/(1 - \alpha)$.
3. Se toma al azar una observación x_1 de $f(x; \theta)$ y se determina

$$\lambda_1 = \frac{f(x_1; \theta_1)}{f(x_1; \theta_0)}.$$

4. Si $\lambda_1 \leq B$, se acepta H_0 .
5. Si $\lambda_1 \geq A$, se rechaza H_0 .
6. Si $B < \lambda_1 < A$, se toma al azar otra observación de $f(x; \theta)$ y se halla el valor de

$$\lambda_2 = \frac{f(x_1; \theta_1)f(x_2; \theta_1)}{f(x_1; \theta_0)f(x_2; \theta_0)}$$

7. Se repiten 4 y 5 con λ_1 sustituida por λ_2 .
8. Se continúan tomando observaciones hasta que se satisfaga 4 ó 5 para un λ_m .

Ejemplo 15-1.—Supongamos que x se distribuye normalmente con varianza 1 y media θ , e imaginemos que se desea docimar la hipótesis $H_0: \theta=6$ contra la alternativa $H_1: \theta=8$, con una probabilidad 0,03 de error de tipo I, y una probabilidad de error de tipo II igual a 0,10. Aplicando la regla, haremos $\theta_0=6$; $\theta_1=8$; $\alpha=0,03$; $\beta=0,10$, y

$$f(x; \theta) = \frac{1}{\sqrt{2\pi}} e^{-(1/2)(x-\theta)^2}$$

También

$$A = \frac{1-\beta}{\alpha} = \frac{0,90}{0,03} = 30 \quad B = \frac{\beta}{1-\alpha} = \frac{0,10}{0,97} = 0,103$$

y

$$\lambda_m = \frac{f(x_1; \theta_1) \dots f(x_m; \theta_1)}{f(x_1; \theta_0) \dots f(x_m; \theta_0)} = \frac{\exp [-(1/2)\Sigma(x_i - 8)^2]}{\exp [-(1/2)\Sigma(x_i - 6)^2]}$$

$$\log \lambda_m = 2\Sigma x_i - 14m$$

Entonces, $\log B < \log \lambda_m < \log A$ es

$$7m + (1/2) \log 0,103 < \sum_{i=1}^m x_i < 7m + (1/2) \log 30$$

Puesto que las desigualdades $B < \lambda_m < A$ son equivalentes a las $\log B < \log \lambda_m < \log A$, no hay inconveniente en utilizarse estas últimas.

15-3. Funciones de potencia.—Supongamos una densidad $f(x; \theta)$, con un parámetro θ , y vamos a docimar la hipótesis $H_0: \theta = \theta_0$, frente a la hipótesis $H_1: \theta = \theta_1$. Interesa ver el comportamiento de la dócima para todos los valores posibles de θ , y, en particular, vamos a examinar la función de potencia de la dócima, $P(\theta)$, que es la probabilidad de rechazar θ_0 cuando el verdadero valor del parámetro es θ . Evidentemente,

$$P(\theta_0) = \alpha \tag{1}$$

$$P(\theta_1) = 1 - \beta \tag{2}$$

y (suponiendo, para fijar las ideas, que $\theta_0 < \theta_1$) parece natural esperar que la función de potencia tenga la forma aproximada de la curva de la figura 15-2.

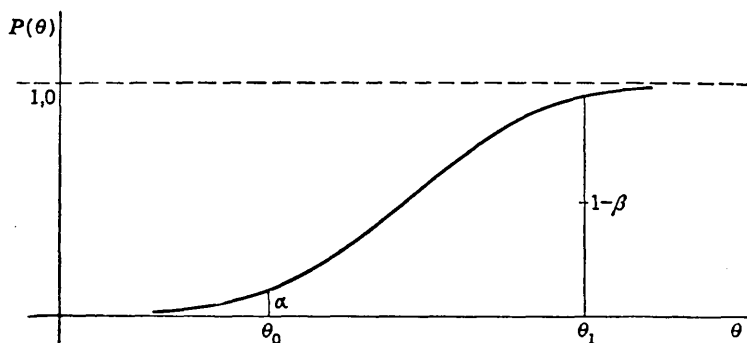


FIG. 15-2.

El modo directo de calcular $P(\theta)$ consiste en sumar las probabilidades de que H_0 se rechace en cada observación. Así,

$$P(\theta) = P(\lambda_1 \geq A) + P(B < \lambda_1 < A, \lambda_2 \geq A) \\ + P(B < \lambda_1 < A, B < \lambda_2 < A, \lambda_3 \geq A) + \dots \quad (3)$$

en donde, p. ej.,

$$P(B < \lambda_1 < A, \lambda_2 \geq A) = \int_R \int f(x_1; \theta) f(x_2; \theta) dx_1 dx_2 \quad (4)$$

y la integral doble se extiende a la región R del plano x_1, x_2 , definida por las desigualdades

$$B < \frac{f(x_1; \theta_1)}{f(x_1; \theta_0)} < A \quad \frac{f(x_1; \theta_1) f(x_2; \theta_1)}{f(x_1; \theta_0) f(x_2; \theta_0)} \geq A \quad (5)$$

Este procedimiento de determinar la función de potencia resulta, cuando menos, laborioso, y suele ser lo suficientemente molesto como para que nadie lo utilice en la práctica.

Con el fin de evitar el uso de (3) se ha ideado un procedimiento muy ingenioso. Este método supone que f satisface ciertas condiciones de regularidad y lo expondremos sin dar una demostración formal de su fundamento, sino esbozando solo las líneas generales de aquella. El razonamiento requiere, en primer lugar, la

existencia de un número h , distinto de cero, tal que

$$g(x; \theta) = \left[\frac{f(x; \theta_1)}{f(x; \theta_0)} \right]^h f(x; \theta_0) \quad (6)$$

sea una densidad; esto es, un número h tal que

$$\int_{-\infty}^{\infty} g(x; \theta) dx = 1 \quad (7)$$

Claro es que $h=0$ hace que $g(x; \theta)$ sea una densidad, ya que $f(x; \theta)$ lo es; para demostrar que existe tal valor de h , consideramos el valor esperado de $[f(x; \theta_1)/f(x; \theta_0)]^u$ como función de u [o sea, $\phi(u)$]:

$$\phi(u) = \int_{-\infty}^{\infty} \left[\frac{f(x; \theta_1)}{f(x; \theta_0)} \right]^u f(x; \theta_0) dx \quad (8)$$

Evidentemente, $\phi(u)$ es siempre positiva, y además $\phi(0)=1$; también podemos afirmar que $\phi(u)$ se hace infinita cuando u tiende a infinito, en sentido positivo o negativo. Puesto que $f(x; \theta_1)$ y $f(x; \theta_0)$ difieren, habrá un intervalo (o conjunto de intervalos) donde su razón sea mayor que la unidad. En tales intervalos, el integrando crece al aumentar u , y $\phi(u) \rightarrow \infty$ cuando $u \rightarrow \infty$. Análogamente, habrá intervalos donde la razón inversa sea superior a la unidad, y el integrando crecerá para valores negativos grandes de u . Esto basta para probar la existencia de h . [Claro es que $\phi(u)$ puede tener un mínimo en $u=0$, en cuyo caso no existiría h ; pero esto ocurrirá solo para valores particulares de θ , y no en general.] En lo que afecta a nuestro razonamiento, puede haber varios valores de u para los cuales sea $\phi(u)=1$. En realidad, hay solamente uno, pues la forma de $\phi(u)$ es la indicada en la figura 15-3; no obstante, el mínimo puede hallarse a la izquierda del origen, de modo que h sería negativo. Luego, en general, existe un valor de h distinto de cero, y tal que $\phi(h)=1$; por consiguiente, (6) es una densidad.

Se construye ahora una dócima sucesional de la hipótesis nula H_0 , según la cual la densidad es $f(x; \theta)$, frente a la hipótesis alternativa H'_1 de que fuese $g(x; \theta)$. Es obvio que en este caso la hipótesis nula es verdadera, porque así lo hemos supuesto. Como límites para la razón de verosimilitud se toman A^h y B^h . Por tanto, la dócima continúa mientras se verifique que

$$B^h < \frac{g(x_1; \theta)g(x_2; \theta) \dots g(x_m; \theta)}{f(x_1; \theta)f(x_2; \theta) \dots f(x_m; \theta)} < A^h \quad (9)$$

y cesa cuando la razón es igual a o cae fuera de estos límites. Suponemos ahora que h es positivo; en caso contrario, A y B se permutan entre sí. En virtud de (6), la dócima definida por (9) equivale exactamente a la dócima sucesional considerada al principio; o sea, (9) equivale a

$$B < \frac{f(x_1; \theta_1)f(x_2; \theta_1) \dots f(x_m; \theta_1)}{f(x_1; \theta_0)f(x_2; \theta_0) \dots f(x_m; \theta_0)} < A \quad (10)$$

Así, el rechazamiento de H_0 supone rechazar también H'_0 . Se calcula inmediatamente la probabilidad de que sea rechazada H'_0 cuando

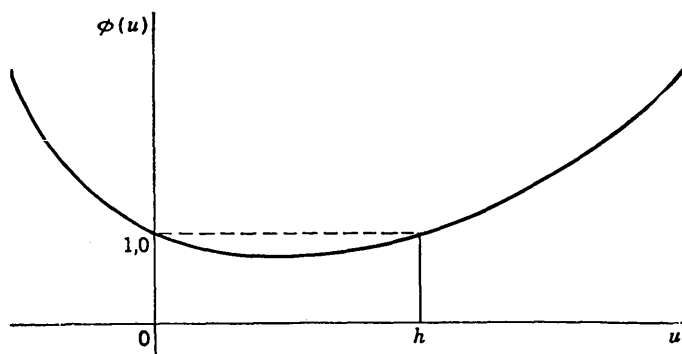


FIG. 15-3.

es cierta [esto es, cuando la densidad es $f(x; \theta)$]; luego a $f(x; \theta)$ le corresponde $P(\theta)$. H'_0 se rechazará cuando es cierta con probabilidad α' , y se aceptará cuando H'_1 es cierta con probabilidad β' donde, de acuerdo con (15-2-7) y (15-2-8),

$$A^h \cong \frac{1 - \beta'}{\alpha'} \quad (11)$$

$$B^h \cong \frac{\beta'}{1 - \alpha'} \quad (12)$$

La resolución de este sistema da

$$\alpha' = P(\theta) \cong \frac{1 - B^h}{A^h - B^h} \quad (13)$$

En resumen:

Regla 2.—Para obtener la función de potencia aproximada de la dócima sucesional de razón de probabilidad, se procede del modo siguiente:

1. Se halla $\phi(u)$ definida por (8) para un θ^* particular.
2. Se hace $\phi(u)=1$ y se resuelve respecto a u .
3. La raíz no nula así obtenida es el número h de (13).
4. La ordenada de la función de potencia en θ^* viene dada por (13):

$$P(\theta^*) \cong \frac{1 - B^h}{A^h - B^h}$$

Ejemplo 15-2.—Como ilustración, consideremos la hipótesis nula de que la media de una distribución normal es μ_0 , frente a la alternativa de que tal media sea μ_1 (siendo $\mu_1 < \mu_0$), suponiendo conocida la varianza σ^2 . Deseamos hallar la probabilidad $P(\mu)$ de rechazar μ_0 cuando la media verdadera es μ . La función $\phi(u)$ es

$$\phi(u) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-[(x-\mu)^2/2\sigma^2]} \left(\frac{e^{-[(x-\mu_1)^2/2\sigma^2]}}{e^{-[(x-\mu_0)^2/2\sigma^2]}} \right)^u dx \quad (14)$$

La integral se calcula sin dificultad, y haciendo $\phi(u)=1$, la resolución de esta ecuación respecto a u nos dice que una de las raíces es $u=0$, y la otra

$$h = \frac{\mu_1 + \mu_0 - 2\mu}{\mu_1 - \mu_0} \quad (15)$$

Si se sustituye esta expresión de h en (13), se obtiene una fórmula explícita para $P(\mu)$ en función de μ .

15-4. Tamaño muestral medio.—El tamaño n de la muestra en la docimasia sucesional es una variable aleatoria, cuya distribución viene dada por la función $p(n)$, la cual se determina a partir de la densidad verdadera, $f(x; \theta)$. Así, pues,

$$p(1) = P(\lambda_1 \leq B) + P(\lambda_1 \geq A) \quad (1)$$

$$p(2) = P(B < \lambda_1 < A, \lambda_2 \leq B) + P(B < \lambda_1 < A, \lambda_2 \geq A) \quad (2)$$

y así sucesivamente; las probabilidades del segundo miembro están expresadas por integrales como las de la ecuación (15-3-4). En esta sección hallaremos una expresión aproximada para el tamaño muestral esperado $E(n)$, viendo a continuación hasta qué punto resulta posible ahorrar observaciones mediante el uso de métodos sucesionales.

Sea

$$z = \log \frac{f(x; \theta_1)}{f(x; \theta_0)} \quad (3)$$

y sea n el menor entero para el cual $z_1 + z_2 + \dots + z_n = Z_n^*$ no satisfice a

$$\log B < Z_n^* < \log A \quad (4)$$

Mostraremos que el valor esperado de la variante Z_n^* , que depende de las variables aleatorias z y de la variante n , es

$$E(Z_n^*) = E(n)E(z) \quad (5)$$

Para verlo, supongamos que N es un valor muy grande, pero prefijado, de n , y prescindamos de la parte de la distribución de n situada a la derecha de N . Se consigue que el error resultante sea tan pequeño como queramos, tomando N suficientemente grande. Puesto que N es fijo, se deduce que

$$E(Z_N^*) = NE(z) \quad (6)$$

La variante Z_N^* puede escribirse en la forma

$$Z_N^* = Z_n^* + W_n \quad (7)$$

definiendo otra variante W_n , y, en virtud de (6),

$$E(Z_n^* + W_n) = NE(z) \quad (8)$$

El inconveniente con que se tropieza al intentar obtener directamente (5) se debe a que el recorrido de z_i depende de si $i \leq n$ ó $i > n$. En este último caso, $E(z_i) = E(z)$; pero cuando $i < n$, el recorrido de z_i viene restringido por (4). Ahora bien, en (8) la variante W_n consta de valores z con $i > n$, de modo que el valor esperado de cada z en W_n es $E(z)$. Así, pues,

$$E(W_n) = E(z)E(N - n) \quad (9)$$

donde el último factor del segundo miembro depende solo de la distribución de n . Combinando (8) y (9), resulta

$$NE(z) = E(Z_n^*) + E(W_n) \quad (10)$$

$$= E(Z_n^*) + E(z)[N - E(n)] \quad (11)$$

que es la misma que (5); despejando $E(n)$, se obtiene

$$E(n) = \frac{E(Z_n^*)}{E(z)} \quad (12)$$

Esta expresión permite deducir una fórmula aproximada sencilla para el tamaño esperado de la muestra. La variante Z_n^* toma solo valores superiores a $\log A$ o inferiores a $\log B$. Si se prescinde de la cuantía en que Z_n^* excede a $\log A$ o no llega a $\log B$, cabe decir que Z_n^* toma únicamente dos valores, a saber: $\log A$ y $\log B$. Cuando la distribución verdadera es $f(x; \theta)$, la probabilidad de que Z_n^* tome el valor $\log A$ es $P(\theta)$, mientras que la correspondiente al valor $\log B$ es $1 - P(\theta)$. Luego

$$E(Z_n^*) \cong P(\theta) \log A + [1 - P(\theta)] \log B \quad (13)$$

que en unión de (12) nos lleva al siguiente enunciado.

Teorema 15-1.—*El tamaño muestral medio aproximado de la dócima secuencial de la razón de verosimilitud es*

$$E(n) \cong \frac{P(\theta) \log A + [1 - P(\theta)] \log B}{E(z)} \quad (14)$$

Este resultado permite comparar dócimas sucesionales con dócimas de tamaño muestral fijo.

Ejemplo 15-3.—Como aclaración consideremos la dócima de la hipótesis $\mu=0$ frente a la de $\mu=1$, para una población normal con varianza unitaria. Elegiremos $\alpha=0,01$ y $\beta=0,01$; entonces, (15-2-7) y (15-2-8) dan $A=99$ y $B=1/99$; supongamos, además, que el valor paramétrico verdadero es cero, de modo que $P(\theta)$ en (14) sea precisamente 0,01. Necesitamos también calcular el valor esperado de

$$z = \log \frac{e^{-[(x-1)^2/2]}}{e^{-(x^2/2)}} = x - \frac{1}{2} \quad (15)$$

que vale $-1/2$ en la distribución verdadera. Así, pues,

$$\begin{aligned} E(n) &\cong \frac{0,01 \log 99 + 0,99 \log 1/99}{-1/2} \\ &\cong 1,96 \log 99 \cong 9 \end{aligned} \quad (16)$$

Para obtener el mismo control de ambos errores, para un tamaño determinado de la muestra, recordemos que la mejor dócima se obtiene eligiendo un número c y aceptando o rechazando $\mu=0$, según que $\bar{x}$ sea menor o mayor que c . La probabilidad α de rechazar H_0 (suponiendo $\mu=0$) es,

$$\alpha = \sqrt{\frac{n}{2\pi}} \int_c^\infty e^{-(n/2)\bar{x}^2} d\bar{x} = \frac{1}{\sqrt{2\pi}} \int_{\sqrt{nc}}^\infty e^{-t^2/2} dt$$

de modo que para $\alpha=0,01$,

$$\sqrt{n}c = 2,326 \quad (17)$$

La probabilidad β de aceptar H_0 , suponiendo cierta H_1 ($\mu=1$), es

$$\beta = \sqrt{\frac{n}{2\pi}} \int_{-\infty}^c e^{-(n/2)(\bar{x}-1)^2} d\bar{x} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\sqrt{n}(c-1)} e^{-(t^2/2)} dt$$

de manera que para $\beta=0,01$,

$$\sqrt{n}(c-1) = -2,326 \quad (18)$$

Resolviendo las ecuaciones (17) y (18), obtenemos para n el valor 22. Así, pues, en décimas repetidas de la hipótesis en cuestión, el procedimiento sucesional requeriría solo un promedio de $\frac{9}{22}$ ó 41% de las observaciones necesarias en el procedimiento de tamaño muestral fijo.

15-5. Inspección por muestreo.—Una aplicación de particular importancia de la docimasia sucesional se presenta en la inspección de artículos manufacturados. Algunos clientes importantes, tales como cadenas de establecimientos, fábricas de montaje, organismos gubernamentales, etc., suelen contratar suministros periódicos de artículos en grandes grupos denominados *lotes*. En el contrato se estipulan determinadas condiciones para los artículos en cuestión, estableciendo además que los artículos serán inspeccionados total o parcialmente para asegurarse de que solo una pequeña proporción de los mismos deja de cumplir las condiciones exigidas por el contrato. Por lo general, los artículos defectuosos no tienen tanta importancia como para compensar el gasto de una inspección detallada y completa, y se usa una inspección por muestreo, de forma que el proveedor inspecciona una muestra de los artículos del lote y estima la proporción defectuosa del mismo. Si la calidad del lote resulta aparentemente satisfactoria, se hace entrega de este; en otro caso, puede venderse a un consumidor menos exigente o al consumidor en cuestión a precio más bajo, o bien se inspecciona por completo (si la inspección no es destructiva), retirando los artículos defectuosos. Cuando ha de utilizarse la inspección por muestreo, el contrato suele especificar el procedimiento efectivo de tomar las muestras. El proveedor no asegura que la proporción de objetos defectuosos en los lotes examinados sea inferior a cierto límite; lo que garantiza es que únicamente presentará a examen lotes que hayan pasado determinada décima de inspección por muestreo.

El plan más sencillo de inspección es el llamado *plan de muestreo simple*. Se inspecciona una muestra de tamaño n y se acepta el lote como satisfactorio si el número de objetos defectuosos es inferior o igual a un número dado c ; en caso contrario, se rechaza el lote. La probabilidad de aceptar un lote con dicho plan depende, por supuesto, de la proporción de artículos defectuosos en el lote. La densidad que da el número x de defectuosos es

$$g(x) = \frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}} \quad (1)$$

en donde N es el tamaño del lote, y M , el número de objetos defectuosos contenido en el mismo. Esta distribución es incómoda de manejar, y puesto que n suele ser muy pequeño con relación a N , es costumbre dar como aproximación la binomial

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x} \quad (2)$$

en donde $p = M/N$ es la proporción de defectuosos en el lote.

El desarrollo de un plan de inspección por muestreo puede representarse mediante la curva *característica operante*, que se reduce a la representación de la probabilidad de aceptar el lote expresada en función de p . Esta probabilidad para el plan de muestreo simple es

$$L(p) = \sum_{x=0}^c g(x) \cong \sum_{x=0}^c f(x) \quad (3)$$

utilizando la aproximación binomial, como haremos en esta sección y en la siguiente. Una curva característica operante puede verse en la figura 15-4. Si, p. ej., se desea aceptar todos los lotes con un 6% o menos de defectuosos y rechazar los lotes con más de un 6%, la característica operante real será la representada en la figura 15-4. Pero esto solo podría realizarse mediante una inspección completa. La inspección por muestreo rechazará necesariamente algunos de los lotes aceptables y aceptará algunos de los que debieran rechazarse. Cuanto más detallado sea el muestreo que se va a realizar, más se acercará la característica operante a su forma ideal. La extensión efectiva del muestreo en un ejemplo dado depende, por supuesto, de varios factores económicos asocia-

dos con el problema particular de que se trate, factores tales como los costes de producción y de inspección por cada objeto, la diferencia entre los valores de mercado de los lotes aceptados y rechazados, etc.

Cabe considerar los planes de la inspección por muestreo como procedimientos para la docimasia de hipótesis. Así, el plan de muestreo simple es precisamente el procedimiento que debería

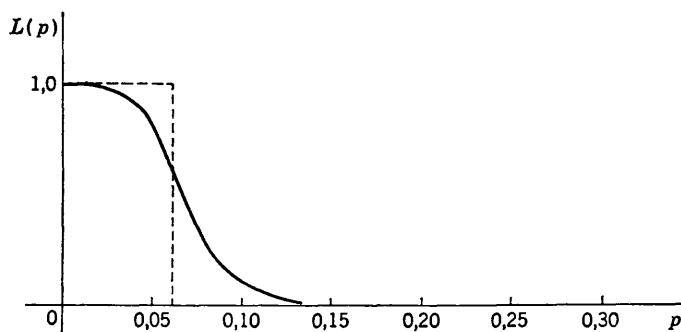


FIG. 15-4.

usarse para docimar la hipótesis nula, según la cual el parámetro p de una distribución binomial tiene el valor p_0 , frente a las alternativas $p > p_0$.

Otros planes de inspección algo más complicados son los de *muestreo doble*. Se examina una pequeña muestra de tamaño n_1 y se acepta o rechaza el lote basándose en la misma. Pero en los casos dudosos, se examina una segunda muestra de tamaño n_2 antes de clasificar definitivamente el lote en una u otra clase. El procedimiento consta de las siguientes fases:

- 1) Se examina una muestra de tamaño n_1 .
- 2) Si x_1 (número de objetos defectuosos en n_1) $\leq c_1$, se acepta el lote.
- 3) Si $x_1 \geq c_2$, se rechaza el lote.
- 4) Si $c_1 < x_1 < c_2$, se examina una segunda muestra de tamaño n_2 .
- 5) Si $x_1 + x_2 \leq c_3$, se acepta el lote.
- 6) Si $x_1 + x_2 > c_3$, se rechaza el lote.

Este procedimiento contiene el germen de la idea sucesional. Es preferible al muestreo simple en el sentido siguiente: Dado un

plan de muestreo simple con tamaño muestral n y un plan de muestreo doble con tamaño muestral medio n , resulta posible aproximar más la característica operante ideal con el segundo de los planes. En otras palabras, para una característica operante dada, el muestreo doble requiere, por término medio, menos observaciones que el muestreo simple.

15-6. Inspección por muestreo sucesional.—Supongamos que se trata de lotes grandes, de modo que el error cometido al emplear la distribución binomial carece de importancia práctica. Imaginemos, además, que el proceso de producción del fabricante, si todo marcha bien, produce alrededor de un 2% de defectuosos, y que el plan de distribución por muestreo es tal que se acepta la mayor parte de los lotes con menos de un 3% de defectuosos y se rechaza la mayoría de aquellos que contienen más de un 3% de defectuosos. La situación corriente es esta: un proveedor que se haya comprometido a suministrar una calidad superior a la que resulta de su proceso de producción, obtendrá poca ventaja si lleva a efecto la inspección por muestreo.

Al establecer un plan sucesional, debe empezarse por expresar la dócima basada en una hipótesis nula con una sola alternativa. Así, en el caso presente podría docimarse la hipótesis nula $p_0=0,025$ frente a la alternativa $p_1=0,04$, aceptando el lote siempre que se acepte la hipótesis nula. En general, se eligen dos valores p_0 y p_1 , y dos probabilidades α y β , para los errores de Tipo I y Tipo II. Así, se dispone de dos puntos de la característica operante: $(p_0, 1-\alpha)$ y (p_1, β) . Podría considerarse un plan de inspección que fuese muy crítico en el punto $p=0,03$, eligiendo, p. ej., los dos puntos $(0,029; 0,999)$ y $(0,031; 0,001)$; pero al hacer esto, habría que realizar un muestreo considerable. La elección efectiva de uno de estos dos puntos depende de consideraciones económicas.

Las observaciones individuales y_i tienen la densidad

$$f(y) = p^y(1-p)^{1-y} \quad (1)$$

y, representando $\sum_1^n y_i$ por x_n , la razón de verosimilitud es

$$\lambda_n = \frac{p_1^{x_n}(1-p_1)^{n-x_n}}{p_0^{x_n}(1-p_0)^{n-x_n}} \quad (2)$$

Se toman observaciones hasta que se verifica $\lambda_n \leq B$, en cuyo caso se acepta el lote, o bien $\lambda_n \geq A$, y entonces se rechaza; A y B se calculan a partir de (15-2-7) y (15-2-8).

Para obtener la característica operante se empieza por hallar $\phi(u)$, que es

$$\phi(u) = E \left[\frac{p_1^y (1-p_1)^{1-y}}{p_0^y (1-p_0)^{1-y}} \right]^u \quad (3)$$

$$\begin{aligned} &= \sum_{y=0}^1 p^y (1-p)^{1-y} \left(\frac{p_1}{p_0} \right)^{uy} \left(\frac{1-p_1}{1-p_0} \right)^{u(1-y)} \\ &= p \left(\frac{p_1}{p_0} \right)^u + (1-p) \left(\frac{1-p_1}{1-p_0} \right)^u \end{aligned} \quad (4)$$

El número h de la sección 15-3 es la raíz no nula de $\phi(u)=1$, de modo que h está definido por

$$p \left(\frac{p_1}{p_0} \right)^h + (1-p) \left(\frac{1-p_1}{1-p_0} \right)^h = 1 \quad (5)$$

Esta ecuación, en unión de

$$L(p) = \frac{A^h - 1}{A^h - B^h} \quad (6)$$

[obtenida restando de 1 los dos miembros de (15-3-13)], determina la función característica operante. Como la obtención de h a partir de (5) constituye un cálculo laborioso, se determinan puntos de la curva eligiendo valores arbitrarios para h y calculando los valores correspondientes de p y $L(p)$ a partir de (5) y (6).

A menudo suele obtenerse una evaluación suficiente de la característica operante a partir de cinco puntos de la curva, que se calculan con facilidad:

$$L(0) = 1 \quad (7)$$

$$L(1) = 0 \quad (8)$$

$$L(p_0) = 1 - \alpha \quad (9)$$

$$L(p_1) = \beta \quad (10)$$

$$L(p') = \frac{\log A}{\log A - \log B} \quad (11)$$

en donde

$$p' = \frac{\log [(1-p_0)/(1-p_1)]}{\log (p_1/p_0) - \log [(1-p_1)/(1-p_0)]} \quad (12)$$

El quinto punto $[p', L(p')]$ está situado entre p_0 y p_1 y corresponde a $h=0$. Las fórmulas (11) y (12) se obtienen haciendo tender h

a cero en (5) y (6), las cuales se transforman en expresiones indeterminadas al sustituir $h=0$.

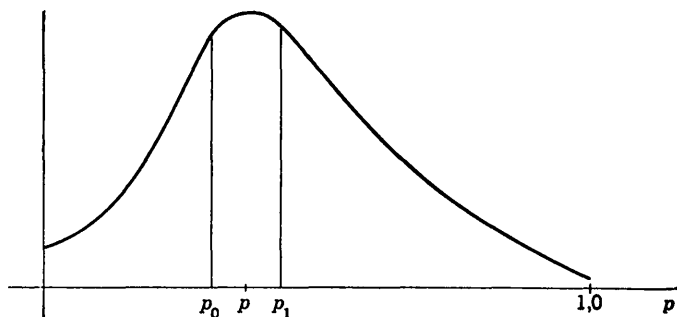


FIG. 15-5.

La curva del tamaño muestral medio puede representarse fácilmente una vez representada $L(p)$. Con referencia a la ecuación (15-4-14), la ordenada de esta curva (Fig. 15-5) viene dada por

$$E(n) \cong \frac{[1 - L(p)] \log A + L(p) \log B}{p \log (p_1/p_0) + (1 - p) \log [(1 - p_1)/(1 - p_0)]} \quad (13)$$

en donde se ha sustituido $P(p)$ por $1 - L(p)$ y

$$E(\bar{x}) = E \left[\log \frac{p_1^x (1 - p_1)^{1-x}}{p_0^x (1 - p_0)^{1-x}} \right] \quad (14)$$

$$= p \log \frac{p_1}{p_0} + (1 - p) \log \frac{1 - p_1}{1 - p_0} \quad (15)$$

El valor máximo de $E(n)$ se alcanza en un punto muy próximo al p' dado por (12). En dicho punto, (13) es:

$$\frac{\log A \log B}{\log (p_1/p_0) \log [(1 - p_1)/(1 - p_0)]} \quad (16)$$

Este es, aproximadamente, el tamaño muestral de promedio máximo, y se presenta cuando la proporción defectuosa verdadera toma el valor dado por (12).

15-7. Dócima sucesional para la media de una población normal.—Como último ejemplo de docimasia sucesional daremos, sin demostración, la dócima bilateral de la hipótesis nula H_0 , se-

gún la cual la media de una población normal tiene el valor μ_0 . Se supone que la varianza es 1. Es necesario construir la dócima a partir de una sola alternativa H_1 . Si interesara una dócima unilateral, p. ej., frente a las alternativas $\mu > \mu_0$, nos limitaríamos a elegir como alternativa cierto valor arbitrario μ_1 (mayor que μ_0). Pero tal alternativa no serviría para la dócima bilateral, puesto que la función de potencia tiende a cero cuando μ se desplaza hacia la izquierda.

Se pueden utilizar varios artificios para eludir esta dificultad. Quizá el más satisfactorio consista en efectuar una ligera alteración de la función de la razón de verosimilitud. Bajo la hipótesis H_0 la densidad es

$$f(x; \mu_0) = \frac{1}{\sqrt{2\pi}\sigma} e^{-[(x-\mu_0)^2/2]} \quad (1)$$

En lugar de un parámetro alternativo $\mu \neq \mu_0$, utilizamos un parámetro $\delta_1 > 0$ y formamos la razón

$$\lambda_n = \frac{{}^{1/2} \prod_{i=1}^n f(x_i; \mu_0 + \delta_1) + {}^{1/2} \prod_{i=1}^n f(x_i; \mu_0 - \delta_1)}{\prod_{i=1}^n f(x_i; \mu_0)} \quad (2)$$

en donde los productos se toman sobre las observaciones sucesivas, indicadas por $i=1, 2, \dots, n$. Es evidente que esta razón λ_n se comporta del modo que deseamos. En la hipótesis nula, su numerador y denominador son equivalentes. Si la alternativa μ está realmente a la izquierda de μ_0 , el segundo término del numerador llega a dominar al denominador, y si se halla a la derecha, dominará el primero.

La realización de la dócima está ahora de acuerdo con el procedimiento usual. Se eligen probabilidades α y β para los dos tipos de error y se calculan A y B a partir de (15-2-7) y (15-2-8). Si se desea una dócima muy sensible, se elegirán δ_1 , α y β pequeñas. Se continuarán las observaciones hasta que λ_n exceda a A o sea menor que B . (Véase Fig. 15-6.)

Cuando se desconoce la varianza, se dispone de varias dócimas; la mayor parte de estas utilizan funciones ponderales de un tipo u otro. Tal vez la dócima más sencilla sea la basada en la distribución t . Si designamos por $g_n(t; \mu)$ la densidad de t , con n

grados de libertad, siendo μ la media de la población normal, podemos definir

$$\lambda_n = \frac{g_n(t; \mu_1)}{g_n(t; \mu_0)} \quad (3)$$

siendo $n=2, 3, 4, \dots$ Aunque esta función no es del mismo tipo que las consideradas hasta aquí (porque el numerador y el denominador no son productos de funciones correspondientes a va-

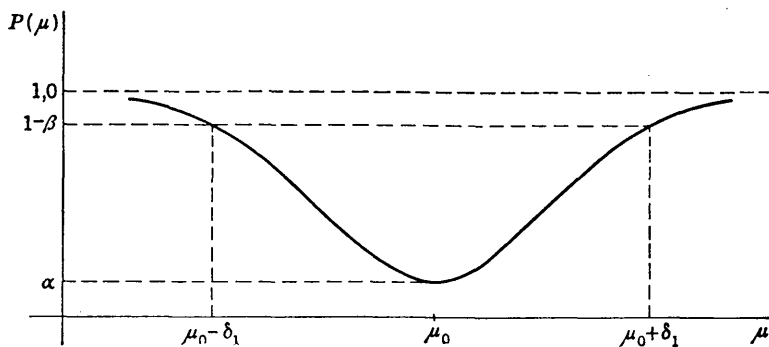


FIG. 15-6.

riantes independientes), se demuestra que la dódima tiene fin y que (15-2-7) y (15-2-8) determinan A y B como anteriormente.

El criterio (3) se refiere, desde luego, a la dódima unilateral de $H_0: \mu = \mu_0$ frente a una alternativa $H_1: \mu_1 > \mu_0$. Para una dódima bilateral de μ_0 , cabe utilizar

$$\lambda_n = \frac{1/2 g_n(t; \mu_0 + \delta_1) + 1/2 g_n(t; \mu_0 - \delta_1)}{g_n(t; \mu_0)} \quad (4)$$

en donde δ_1 tiene el mismo significado que antes.

PROBLEMAS

1. Efectúese una dódima sucesional de la hipótesis nula $p=0,45$ frente a la alternativa $p=0,30$. Supongamos que p se refiere a la probabilidad de obtener una cara al lanzar una moneda y que se efectúa la dódima lanzándola, y empleando $\alpha=0,10$ y $\beta=0,10$. Los cálculos se simplifican resolviendo $\log \lambda_n = B$ y $\log \lambda_n = A$, para x_n (número de caras en n tiradas), obteniendo así, como funciones lineales de n , los números de aceptación y de rechazamiento.

2. Demuéstrese que la ecuación (15-3-13) es correcta cuando h es negativo.

3. Suponiendo que un lote es de tamaño N , con M defectuosos, ¿cuál es la expresión exacta para la función característica operante?

4. Demuéstrese que la razón 9/22 obtenida al final de la sección 15-4 depende solamente de los valores de α y β y no de los tamaños de σ^2 , μ_0 y μ_1 .

5. Compárese el tamaño muestral sucesional promedio con el tamaño muestral fijo, para la dócima unilateral de la media de una población normal, cuando $\alpha=0,01$, $\beta=0,05$, y es cierta la hipótesis alternativa ($\sigma=1$).

6. Demuéstrese que la dócima unilateral para la media de una población normal de varianza conocida puede efectuarse representando las dos rectas

$$y = \frac{\sigma^2}{\mu_1 - \mu_0} \log B + \frac{\mu_0 + \mu_1}{2} n$$

$$y = \frac{\sigma^2}{\mu_1 - \mu_0} \log A + \frac{\mu_0 + \mu_1}{2} n$$

en el plano n, y ; a continuación se representa $\sum_{i=1}^n x_i$ en función de n a medida que se hacen las observaciones. La dócima termina cuando se atraviesa una de las rectas.

7. Con referencia al problema 6, sea $c = (\mu_0 + \mu_1)/2$, y representemos por b y a las dos constantes de las ecuaciones que allí aparecen; esto es,

$$a = \frac{\sigma^2 \log A}{\mu_1 - \mu_0} \quad b = \frac{\sigma^2 \log B}{\mu_1 - \mu_0}$$

Demuéstrese que la función de potencia de la dócima puede expresarse en la forma siguiente:

$$P(\mu) \cong \frac{1 - e^{2(c-\mu)b/\sigma^2}}{e^{2(c-\mu)a/\sigma^2} - e^{2(c-\mu)b/\sigma^2}}$$

8. Con referencia a los problemas 6 y 7, demuéstrese que la expresión del tamaño muestral promedio puede expresarse por

$$E(n) \cong \frac{b + P(\mu)(b - a)}{\mu - c}$$

9. Compruébense las ecuaciones (15-6-11) y (15-6-12).

10. Represéntese la función de potencia y la función del tamaño muestral promedio para la dócima del problema 1.

11. Represéntese la función de potencia y la función de tamaño muestral promedio, para la dócima de la hipótesis según la cual la media de una población normal es cero, frente a la alternativa de que es igual a 1, haciendo $\sigma^2=1$, $\alpha=0,01$, $\beta=0,05$.

12. Hállense fórmulas para la función de potencia y del tamaño muestral promedio, para décimas sucesionales relativas a la media de una distribución de Poisson.

13. Supongamos que un proceso de producción da lugar a lotes de tamaño N con M defectuosos, de modo que M tenga una distribución binomial. Demuéstrese que una muestra de tamaño n (con x defectuosos) no puede facilitar información sobre la proporción de defectuosos en los $N-n$ ítems restantes de un lote.

14. Supongamos lotes que se rechazan bajo un procedimiento de inspección por muestreo sucesional, y que se inspeccionan después por completo, reemplazando los elementos defectuosos por buenos, lo cual constituye una práctica corriente. Sea p la proporción de defectuosos en los lotes originales. ¿Cuál será la proporción media de defectuosos en todos los lotes entregados, contando los inspeccionados por completo y los sometidos al plan de muestreo? Esta función de p se denomina función de la *calidad media resultante*; el máximo de esta función se denomina el *límite de la calidad media resultante*. Se trata de hacer un ligero croquis que exprese el aspecto general de la función.

15. Con referencia a la situación descrita en el problema 14, hállese el porcentaje medio de elementos inspeccionados como función de p , contando los lotes pasados y los inspeccionados por completo. Hágase un ligero croquis que exponga el aspecto general de la función.

16. Se supone una distribución uniforme $f(x)=1/\theta$ con recorrido $0 < x < \theta$. Se trata de discutir la décima sucesional de $\theta=\theta_0$ frente a $\theta=\theta_1$, siendo $\theta_0 < \theta_1$. Obsérvese que algunas de las fórmulas generales pueden no ser aplicables a este caso.

17. Utilizando un argumento análogo al empleado para (15-4-5), Wald demostró que

$$E\{e^{Z_n t}[\phi(t)]^{-n}\}=1$$

siendo $\phi(t)$ la función generatriz de momentos de z , esto es, $\phi(t)=E(e^{zt})$, extendida la esperanza E a la distribución conjunta de las z y de la variable aleatoria n . Esto se denomina identidad fundamental del análisis sucesional. Utilícese esta identidad para obtener (15-4-5).

18. Demuéstrese, mediante la identidad del problema 17, que

$$E(n)=\frac{E(Z_n^{*2})}{E(z^2)}$$

cuando $E(z)=0$.

19. Utilícese el resultado del problema 18 para obtener (15-6-16).

20. Utilícese el resultado del problema 18 para demostrar que el tamaño muestral promedio máximo para décimas unilaterales relativas a la media de una población normal es aproximadamente igual a $-ab/\sigma^2$, en donde a y b se definen como en el problema 7. Hay que suponer, sin demostrarlo, que el máximo corresponde a $h=0$.

BIBLIOGRAFIA

1. BLACKWELL, D., y M. A. GIRSCHICK: *Theory of Games and Statistical Decisions*, John Wiley & Sons, Inc., Nueva York, 1954.
2. BRUNK, H.: *An Introduction to Mathematical Statistics*, Ginn & Company, Boston, 1960.
3. FRASER, D. A. S.: *Statistics—An Introduction*, John Wiley & Sons, Inc., Nueva York, 1959.
4. FREEMAN, H., M. FRIEDMAN, F. MOSTELLER y W. WALLIS: *Sampling Inspection*, McGraw-Hill Book Company, Inc., Nueva York, 1948.
5. HOEL, P. G.: *Introduction to Mathematical Statistics*, 2.^a ed., John Wiley & Sons, Inc., Nueva York, 1956.
6. WALD, A.: *Sequential Analysis*, John Wiley & Sons, Inc., Nueva York, 1947.

CAPITULO 16

METODOS NO PARAMETRICOS

16-1. Introducción.—El importante lugar que corresponde a la distribución normal en la teoría estadística está bien justificado por el teorema central del límite. Sin embargo, con frecuencia no se conoce si la distribución básica es tal que sea aplicable dicho teorema o si la aproximación a la distribución normal es lo suficientemente buena como para que los intervalos confidenciales resultantes y las dójimas de hipótesis basadas en la teoría normal tengan la acuracidad deseada. Así, p. ej., si se toman repetidas muestras de tamaño n de una población con densidad normal, y se halla un intervalo confidencial del 0,95 para la media, la interpretación frecuencial será: Si se toman repetidas muestras de esta población y para cada muestra aleatoria se obtiene un intervalo confidencial del 95%, al cabo de un gran número de pruebas el 95% de esos intervalos contendrán la media de la densidad. Si el muestreo se hace para una densidad distinta de la normal, en lugar de haber el 95% de intervalos que contienen la muestra, el porcentaje podrá ser el 99%, el 90% o cualquier otro. Si este porcentaje está próximo al 95% (p. ej., 93% o 97%), corrientemente el experimentador se dará por satisfecho. Pero si el porcentaje se desvía mucho del deseado, probablemente el experimentador no quedará satisfecho. En los casos en que se sabe que los métodos usuales, basados en el supuesto de una densidad normal, no son aplicables, se precisa recurrir a otro método. Si se conoce la distribución básica (no necesariamente la normal), pueden deducirse dójimas de hipótesis e intervalos confidenciales basados en esa distribución. En muchos casos, el experimentador no conoce la forma de la distribución básica y necesita técnicas estadísticas cuya aplicación sea independiente de la forma de la densidad. Estas técnicas se denominan métodos *no paramétricos*, o métodos *a libre distribución*.

El término *no paramétricos* proviene de consideraciones sobre la docimasia de hipótesis (Cap. 12). Al formar, p. ej., la razón de verosimilitud, se opera con un espacio paramétrico, el cual define una familia de distribuciones al variar en él los parámetros de la forma funcional de la distribución. Los métodos que desarrollaremos en este capítulo no utilizan formas funcionales ni parámetros

de tales formas. Se aplican a familias muy extensas de distribuciones en vez de a familias caracterizadas por una forma funcional particular. También se usa a veces el término *a libre distribución*, que, análogamente, indica que los métodos no dependen de la forma funcional de las funciones de distribución.

Hasta ahora, al representar una muestra por

$$x_1, x_2, \dots, x_n,$$

el símbolo x_1 se refería a la primera observación, el x_2 a la segunda, y así sucesivamente. En este capítulo, a menos que se indique otra cosa, la x_i será la i -ésima observación ordenada; es decir, el símbolo x_1 representará la menor de las n observaciones, x_2 la más pequeña después de la primera, y así sucesivamente, siendo x_n la mayor. Así, p. ej., para una muestra de cuatro observaciones, 2, -4, -1, 1, x_1 representará la segunda observación, x_2 , la tercera, x_3 la cuarta y x_4 la primera. Los métodos no paramétricos se basan en gran parte en estas observaciones ordenadas, o *estadígrafos de orden*.

Los métodos que vamos a exponer se aplican tanto a variantes continuas como discretas, pero nos ocuparemos casi exclusivamente del caso continuo, indicando en ocasiones las modificaciones que sería necesario hacer cuando se trate de variantes discretas. Por tanto, salvo que explícitamente se diga otra cosa, *en este capítulo se supondrá que las variables aleatorias básicas son continuas*.

16-2. Una distribución básica.—Toda la estructura de los métodos no paramétricos descansa en una sencilla propiedad de los estadígrafos de orden. *La distribución del área comprendida bajo la función de densidad, entre dos observaciones ordenadas, es independiente de la forma de dicha función.* Para demostrar esto, supongamos que z representa una variable aleatoria continua con densidad

$$f(z) \quad -\infty < z < \infty \quad (1)$$

Sea $z_1, z_2, \dots, z_n$ una muestra aleatoria procedente de $f(z)$, y $x_1, x_2, \dots, x_n$ la correspondiente muestra ordenada. La densidad conjunta de las x_i es (Sec. 10-9)

$$q(x_1, x_2, \dots, x_n) = n! f(x_1) f(x_2) \dots f(x_n) \\ -\infty < x_1 < x_2 < \dots < x_n < \infty \quad (2)$$

en donde $f(x_i)$ es la densidad (1).

Utilizaremos u_i para designar el área de $f(z)$ que queda a la izquierda de la i -ésima observación ordenada x_i ; es decir,

$$u_i = \int_{-\infty}^{x_i} f(z) dz = F(x_i) \quad i = 1, 2, \dots, n \quad (3)$$

siendo $F(z)$ la función acumulativa de $f(z)$.

Teorema 16-1.—La densidad conjunta de las variables aleatorias $u_1, u_2, \dots, u_n$ en donde u_i está definida por (3), es

$$g(u_1, u_2, \dots, u_n) = n! \quad 0 < u_1 < u_2 < \dots < u_n < 1 \quad (4)$$

y no depende de $f(z)$.

Demostración.—Utilícese la $q(x_1, \dots, x_n)$ de (2) y transfórmese en u_i mediante (3).

La densidad $g(u_1, \dots, u_n)$ permite hallar la distribución de cualquier conjunto de áreas situadas bajo $f(z)$ y entre pares de observaciones ordenadas. Así, p. ej., cabe que deseemos hallar la densidad del área comprendida bajo $f(z)$ entre las observaciones ordenadas mayor y menor, o sea entre x_n y x_1 .

Teorema 16-2.—La densidad de la variable aleatoria v , que es el área comprendida por debajo de $f(z)$, entre la mayor y la menor de las observaciones de una muestra de tamaño n procedente de $f(z)$, es la densidad beta:

$$p(v) = n(n-1)v^{n-2}(1-v) \quad 0 < v < 1 \quad (5)$$

Demostración.—Se tiene que

$$v = F(x_n) - F(x_1) = u_n - u_1 \quad (6)$$

La densidad conjunta de las u_i es la dada en (4), que integraremos respecto a $u_2, u_3, \dots, u_{n-1}$ para obtener la densidad conjunta de u_1 y u_n . Así,

$$h(u_1, u_n) = \int_{u_1}^{u_n} \dots \int_{u_1}^{u_4} \int_{u_1}^{u_3} n! du_2 du_3 \dots du_{n-1} \quad (7)$$

$$= n! \frac{(u_n - u_1)^{n-2}}{(n-2)!} \quad 0 < u_1 < u_n < 1 \quad (8)$$

y la densidad de u_1 y v será

$$k(u_1, v) = n(n-1)v^{n-2} \quad 0 < u_1 < (1-v) < 1 \quad (9)$$

Integrando respecto a u_1 , se obtiene la densidad deseada.

El siguiente teorema es más general.

Teorema 16-3.—*Supongamos que la variable aleatoria W_{rs} es el área comprendida bajo $f(z)$ entre x_r y x_s ($r < s$); entonces, la densidad de W_{rs} será:*

$$f(W_{rs}) = \frac{n!}{(s-r-1)!(n-s+r)!} (W_{rs})^{s-r-1} (1-W_{rs})^{n-s+r} \quad 0 < W_{rs} < 1 \quad (10)$$

La demostración es análoga a la del teorema 16-2 y se propone como ejercicio al lector.

El valor esperado de u_i es

$$\begin{aligned} E(u_i) &= \int_0^1 \dots \int_0^{u_2} \int_0^{u_2} n! u_i du_1 du_2 \dots du_n \\ &= \frac{i}{n+1} \end{aligned} \quad (11)$$

Por tanto, el valor esperado del área situada bajo $f(z)$ entre dos observaciones sucesivas será

$$E(u_i) - E(u_{i-1}) = \frac{1}{n+1} \quad (12)$$

Así, por término medio, las n observaciones ordenadas dividen la superficie situada bajo $f(z)$ en $n+1$ partes iguales, cada una de área $1/(n+1)$.

16-3. Posición y dispersión.—En el caso paramétrico hemos utilizado la media y la desviación estándar poblacionales como medidas de posición y dispersión, pero métodos no paramétricos utilizan otras medidas. Como centro de la población se toma la mediana ν , que es el punto que divide en dos partes iguales el área limitada por la función de densidad. Se define, pues, ν mediante

$$1/2 = \int_{-\infty}^{\nu} f(z) dz = F(\nu) \quad (1)$$

en donde $f(z)$ es la densidad y $F(z)$ la distribución acumulativa. La mediana suele designarse por $\xi_{0,50}$ y se utiliza una notación análoga para otros puntos porcentuales; p. ej.,

$$F(\xi_{0,15}) = 0,15 \quad (2)$$

define el punto $\xi_{0,15}$, correspondiente al 15% de la población.

Como medida de dispersión se emplea la distancia entre dos puntos porcentuales; p. ej., una medida de dispersión de uso frecuente es

$$\tau_{0.50} = \xi_{0.75} - \xi_{0.25} \quad (3)$$

la cual recibe el nombre de *recorrido del 50%*, o *recorrido intercuartílico*. Pero también se usan a menudo muchos otros recorridos, como el recorrido del 90%, $\tau_{0.90} = \xi_{0.95} - \xi_{0.05}$, o el recorrido del 33 $\frac{1}{3}$ %, $\tau_{1/3} = \xi_{2/3} - \xi_{1/3}$.

Estimación puntual.—La mediana poblacional ν se estima mediante la mediana muestral $\tilde{x}$, que coincide con la observación media cuando el tamaño de la muestra es impar, o es la media aritmética de las dos observaciones medias si aquel es par. Se tiene así.

$$\tilde{x} = x_{k+1} \quad \text{si } n = 2k + 1 \quad (4)$$

$$= (1/2)(x_k + x_{k+1}) \quad \text{si } n = 2k \quad (5)$$

La mediana muestral $\tilde{x}$ no es de ordinario una estimación insesgada de ν , aun en el caso de que n sea impar, ya que $E[F(x)] = F(\nu)$ no significa que se haya de verificar $E(\tilde{x}) = \nu$. Sin embargo, el sesgo suele ser pequeño y tiende a cero al aumentar el tamaño de la muestra.

Para estimar puntos porcentuales, las mismas x_i proporcionan estimaciones de los puntos del $100i/(n+1)$ por ciento. Para otros valores puede recurrirse a la interpolación lineal. Así, para estimar $\xi_{0.25}$ de una muestra de tamaño $n=10$, observemos que x_2 estima el punto $2/11$ y x_3 el punto $3/11$; por tanto, utilizaremos como estimación

$$\tilde{x}_{0.25} = x_2 + \frac{0.25 - 2/11}{1/11} (x_3 - x_2) \quad (6)$$

Dadas estimaciones de los puntos porcentuales, se podrán evidentemente estimar los diversos recorridos.

Intervalos confidenciales.—Mediante la distribución binomial es fácil construir un intervalo confidencial para ν . La probabilidad de que una observación caiga a la derecha o a la izquierda de ν es, en uno y otro caso, igual a $1/2$. La probabilidad de que haya exactamente i observaciones a la izquierda de ν es, precisamente,

$$\binom{n}{i} \left(\frac{1}{2}\right)^n \quad (7)$$

Así, pues, la probabilidad de que x_r , estadígrafo r -ésimo, exceda a ν será

$$P(x_r > \nu) = \sum_{i=0}^{r-1} \binom{n}{i} \left(\frac{1}{2}\right)^n \quad (8)$$

y, análogamente,

$$P(x_s < \nu) = \sum_{i=s}^n \binom{n}{i} \left(\frac{1}{2}\right)^n \quad (9)$$

Si ahora suponemos $s > r$, sumando (8) y (9), y restando ambos miembros de la unidad, se obtiene

$$P(x_r < \nu < x_s) = \sum_{i=r}^{s-1} \binom{n}{i} \left(\frac{1}{2}\right)^n \quad (10)$$

que proporciona un intervalo confidencial para ν . Ordinariamente, se toma para s el valor $n-r+1$, de modo que se utilizan las observaciones r -ésimas, en orden de magnitud, a partir de los extremos.

Teorema 16-4.—Sea $x_1, x_2, \dots, x_n$ una muestra ordenada de una población con densidad $f(z)$; la probabilidad de que el intervalo aleatorio $x_r < z < x_{n-r+1}$ cubra la mediana ν es

$$\sum_{i=r}^{n-r} \binom{n}{i} \left(\frac{1}{2}\right)^n \quad (11)$$

Si se desea un intervalo confidencial para la mediana ν , con coeficiente de confianza $1-\alpha$, se elige el entero r de tal modo que (11) esté lo más próximo posible a $1-\alpha$, o bien de manera que la cantidad (11) sea mayor que $1-\alpha$. Si r' es el máximo entero tal que (11) sea mayor o igual que $1-\alpha$, el intervalo confidencial es

$$P(x_{r'} < \nu < x_{n-r'}) \geq 1-\alpha$$

Así, para una muestra de tamaño (6), se obtendrían como intervalos confidenciales para ν

$$P(x_1 < \nu < x_6) = 1 - (1/2)^6 - (1/2)^6 = 31/32 = 0,97 \quad (12)$$

y

$$P(x_2 < \nu < x_5) = 1 - 14/(2^6) = 50/64 = 0,78 \quad (13)$$

Si estuviéramos interesados en un intervalo del 95%, emplearíamos probablemente (12). Si se deseara, p. ej., un intervalo del 90%, se podría utilizar una combinación aleatoria de ambos intervalos.

Si el tamaño muestral es pequeño, se dispone solo de algunos niveles de confianza; en particular, si $n=2$, solo hay un intervalo confidencial del 50%, dado por

$$P(x_1 < \nu < x_2) = 0,50 \quad (13)$$

Para tamaños muestrales moderados, se calcula directamente la suma binomial de (10) o se obtiene por medio de tablas. Para valores grandes de n podría utilizarse la aproximación normal a la binomial. Como el índice i de (7) es aproximadamente normal, con media $n/2$ y desviación estándar $\sqrt{n/2}$ para valores grandes de n , se obtiene, p. ej., un intervalo confidencial del 95%, contando $1,96 \sqrt{n/2}$ observaciones a la izquierda y a la derecha de la mediana muestral.

Se emplea una técnica análoga en el cálculo de intervalos confidenciales para los puntos porcentuales. Si ξ_p es el punto porcentual 100 p de la distribución, el mismo razonamiento utilizado para deducir (10) prueba que

$$P(x_r < \xi_p < x_s) = \sum_{i=r}^{s-1} \binom{n}{i} p^i (1-p)^{n-i} \quad (14)$$

Así, para una muestra de tamaño seis, se obtiene un posible intervalo confidencial para el punto del 25%, mediante

$$P(x_1 < \xi_{0,25} < x_4) = \sum_{i=1}^3 \binom{6}{i} \left(\frac{1}{4}\right)^i \left(\frac{3}{4}\right)^{6-i} \approx 0,78 \quad (15)$$

Un límite superior del 96% para $\xi_{0,25}$ viene dado por

$$P(\xi_{0,25} < x_4) = \sum_{i=0}^3 \binom{6}{i} \left(\frac{1}{4}\right)^i \left(\frac{3}{4}\right)^{6-i} \approx 0,96 \quad (16)$$

Docimasia de hipótesis.—Para docimar la hipótesis nula $\nu = \nu_0$ frente a alternativas $\nu > \nu_0$, se utiliza la relación (8), eligiendo de antemano un entero r de modo que la probabilidad de un error de Tipo I se aproxime lo más posible al valor deseado. Así, para una muestra de tamaño seis, la probabilidad de un error de Tipo I

puede hacerse aproximadamente igual a 0,11 eligiendo como entero $r=2$.

Si después de obtener la muestra se verifica que $x_2 < \nu_0$, se acepta la hipótesis nula, y si $x_2 > \nu_0$, se rechaza. Del mismo modo se construyen dúcimas bilaterales de $\nu = \nu_0$. La dúcima bilateral equivale, evidentemente, a la construcción de un intervalo confidencial para ν , y a aceptar o rechazar $\nu = \nu_0$ según que el intervalo confidencial cubra o no a ν_0 . Mediante el mismo proceso pueden efectuarse dúcimas relativas a un punto porcentual ξ_p , por medio de las probabilidades p y $(1-p)$ en lugar de $1/2$ y $1/2$.

Resulta, pues, evidente que los métodos no paramétricos, además de ser muy generales por no requerir hipótesis alguna relativa a la forma de la función de distribución, son también extraordinariamente sencillos. No se necesitan análisis complejos ni se requiere una teoría de la distribución; la distribución binomial simple proporciona todo lo necesario para la estimación y docimasia de hipótesis si se trata de una sola población. El único inconveniente estriba en la escasez de niveles de confianza cuando el tamaño de la muestra es muy pequeño.

Digamos ahora unas palabras sobre el caso de las distribuciones discretas. Hasta aquí hemos supuesto que se trataba de distribuciones continuas; cuando estas son discretas, habrá que sustituir por desigualdades las igualdades obtenidas en esta sección para dúcimas e intervalos confidenciales. Así, p. ej., (10) será ahora

$$P(x_r < \nu < x_s) \geq \sum_r^{s-1} \binom{n}{i} \left(\frac{1}{2}\right)^n \quad (17)$$

La razón de la desigualdad se debe al hecho de que ciertas observaciones pueden duplicarse. Supongamos, p. ej., que quisiéramos estimar ν para una distribución discreta, utilizando una muestra de tamaño seis y un intervalo confidencial del 78%, dado por x_2 y x_5 . A veces las dos observaciones más pequeñas x_1 y x_2 serán iguales, de modo que el intervalo (x_2, x_5) es equivalente al (x_1, x_5) y corresponde, por tanto, a una probabilidad superior a 0,78. Cabe también que suceda algo análogo con el límite superior; si x_5 y x_6 son iguales, el intervalo (x_2, x_5) será equivalente al (x_2, x_6) ; hasta puede suceder en algún caso que coincida con el intervalo (x_1, x_6) , correspondiendo así al nivel del 97% en vez de al citado, del 78%.

16-4. Comparación de dos poblaciones.—Se han establecido muchos métodos no paramétricos para docimar si la distribución de dos poblaciones es una misma. Nos referiremos solo a dos de estos, y al final de esta sección deduciremos un intervalo confi-

dencial para la diferencia entre dos medianas poblacionales. En primer lugar, vamos a dar un resultado sencillo relativo a la distribución de las disposiciones de dos conjuntos de observaciones procedentes de la misma población.

Sea $x_1, x_2, \dots, x_{n_1}$ una muestra ordenada de una población con densidad $f(x)$, e $y_1, y_2, \dots, y_{n_2}$ otra muestra ordenada procedente de la misma población. Supongamos combinadas ambas muestras y dispuestas en orden de magnitud; p. ej.,

$$y_1, x_1, x_2, y_2, x_3, y_3, y_4, y_5, x_4, \dots \quad (1)$$

Deseamos hallar la probabilidad de obtener una disposición determinada de este tipo.

Si se transforman las x en u y las y en v mediante la relación (16-2-3), la densidad conjunta de las u y las v es

$$g(u_1, u_2, \dots, u_{n_1}, v_1, v_2, \dots, v_{n_2}) = n_1! n_2! \quad (2)$$

La probabilidad de una disposición concreta, como la (1), se deduce integrando (2) en la región definida por

$$0 < v_1 < u_1 < u_2 < v_2 < u_3 < \dots < 1 \quad (3)$$

Esto es, se integra v_1 de 0 a u_1 , después u_1 de 0 a u_2 , etc. Se ve fácilmente que el valor de la integral es $n_1! n_2! / (n_1 + n_2)!$, o sea

$$1 / \binom{n_1 + n_2}{n_1}.$$

Puesto que hay exactamente $\binom{n_1 + n_2}{n_1}$ disposiciones de n_1 valores x y n_2 valores y , se deduce que todas las disposiciones de las x y de las y son igualmente verosímiles.

Décima de rachas.—Volvemos ahora a la cuestión de la dõcimasia de la hipótesis nula de que dos muestras proceden de la misma población. Las observaciones de ambas muestras se representarán como antes por x e y . Se combinan las dos series de observaciones como en (1) y se cuenta el número d de rachas o runflas. Una racha es una sucesión de letras de la misma clase limitada por letras de clase distinta. Así, (1) comienza por una racha de una y ; a continuación viene una racha de dos x , luego otra de una y , y así sucesivamente; en esta expresión hay seis rachas. Es evidente que si las dos muestras proceden de la misma población, las x y las y estarán en general bien mezcladas y d será grande. Si ambas poblaciones están muy apartadas, de modo que no se

solapen sus recorridos, d valdrá solo dos y, en general, las diferencias entre ambas poblaciones tenderán a reducir d . Así, pues, ambas poblaciones pueden tener la misma media o mediana, pero si la población x está muy concentrada, mientras que la y está muy dispersa, habrá una larga racha de y en cada extremo de la muestra combinada y, por tanto, una tendencia a reducir d . La d6cima se efectúa observando el n6mero total de rachas en la muestra combinada, aceptando la hip6tesis nula si d es mayor que un n6mero determinado d_0 , o rechaz6ndola si $d \leq d_0$. Trataremos de determinar la distribuci6n de d en la hip6tesis nula, de modo que podamos establecer d_0 para un nivel dado de significaci6n.

Hemos visto que, en la hip6tesis nula, todas las $\binom{n_1+n_2}{n_1}$ disposiciones de n_1 valores x y n_2 valores y son igualmente veros6miles. Ahora es necesario contar todas aquellas que tienen exactamente un n6mero de rachas igual a d . Supongamos que d es par e igual a $2k$; habr6 k rachas de x y k de y . Para obtener k rachas de x deben dividirse los n_1 valores x en k grupos, y contar todas las permutaciones de los k n6meros en cada grupo. En resumen, contar todas las particiones ordenadas k de n_1 , excluyendo aquellas partes que sean iguales a cero. Esto se consigue f6cilmente utilizando la funci6n generatriz definida en el cap6tulo 2, para contar las maneras de obtener un total determinado con un conjunto de dados. El n6mero buscado es el coeficiente de t^n en

$$(t + t^2 + t^3 + \dots)^k = \left(\frac{t}{1-t} \right)^k \quad (4)$$

$$= t^k \sum_{i=0}^{\infty} \binom{k-1+i}{k-1} t^i \quad (5)$$

que es $\binom{n_1-1}{k-1}$. An6logamente, hay $\binom{n_2-1}{k-1}$ particiones (de k partes) de n_2 , excluyendo las partes iguales a cero. Cualquier partici6n de las x puede combinarse de dos maneras con cualquiera de las y para formar una sucesi6n como (1); la primera partici6n x o la primera y puede ponerse al principio de la sucesi6n. Se obtiene as6 la densidad para valores pares de d :

$$h(d) = 2 \frac{\binom{n_1-1}{k-1} \binom{n_2-1}{k-1}}{\binom{n_1+n_2}{n_1}} \quad k = \frac{d}{2} \quad (6)$$

y, por un razonamiento análogo, para valores impares de d resulta:

$$h(d) = \frac{\binom{n_1-1}{k} \binom{n_2-1}{k-1} + \binom{n_1-1}{k-1} \binom{n_2-1}{k}}{\binom{n_1+n_2}{n_1}} \quad k = \frac{d-1}{2} \quad (7)$$

Para docimar la hipótesis nula en cuestión, con una probabilidad α para el error de Tipo I, se halla el entero d_0 de modo que (con la mayor aproximación posible)

$$\sum_{d=0}^{d_0} h(d) = \alpha \quad (8)$$

y se rechaza dicha hipótesis si la d observada no excede a d_0 .

Los cálculos que supone (8) son en general muy laboriosos, a menos que n_1 y n_2 sean pequeños. La distribución de d se aproxima a la normal para muestras grandes, y, desde el punto de vista práctico, esta aproximación suele ser suficiente cuando n_1 y n_2 son mayores que 10. La media y la varianza de d son:

$$E(d) = \frac{2n_1n_2}{n_1+n_2} + 1 \quad (9)$$

$$\sigma_d^2 = \frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1+n_2)^2(n_1+n_2-1)} \quad (10)$$

y si hacemos

$$n_1+n_2=n \quad n_1=n\alpha \quad n_2=n\beta \quad (11)$$

las expresiones aproximadas de estos momentos toman, para valores grandes de n ,

$$E(d) \cong 2n\alpha\beta \quad (12)$$

$$\sigma_d^2 \cong 4n\alpha^2\beta^2 \quad (13)$$

La normalidad de $h(d)$ para muestras grandes se comprueba utilizando la fórmula de Stirling para calcular las factoriales de (6), sustituyendo los valores de d en función de t , definida por

$$t = \frac{d - 2n\alpha\beta}{2\alpha\beta\sqrt{n}} \quad (14)$$

y viendo que el logaritmo de la expresión resultante tiende a

$$-\log \sqrt{2\pi} - \frac{1}{2}t^2 \quad (15)$$

cuando n tiende a infinito. Omitimos los detalles de la demostración. Utilizando este resultado se determinaría d_0 para la docimasia de la hipótesis nula con un nivel del 0,05, p. ej., haciendo el segundo miembro de (14) igual a $-1,645$ y hallando el correspondiente valor de d .

La dócima de rachas es muy sensible a las diferencias, tanto en forma como en situación, entre las dos distribuciones. No obstante, en la práctica no suelen interesar las diferencias de forma, sino solo las de posición. Es decir, la hipótesis nula que interesa docimar es la de la igualdad de las medianas poblacionales: $\nu_1 = \nu_2$. La siguiente dócima de $f_1(x) = f_2(y)$ es especialmente sensible a las diferencias de posición, y apenas lo es a las diferencias de forma.

Dócima de medianas.—Sea, como antes, una muestra ordenada $x_1, x_2, \dots, x_{n_1}$ de $f_1(x)$, y otra muestra $y_1, y_2, \dots, y_{n_2}$ de $f_2(y)$, y sea $z_1, z_2, \dots, z_{n_1+n_2}$ la muestra ordenada que resulta al combinar las dos anteriores. La dócima de la hipótesis nula $f_1(x) = f_2(x) = f(x)$ consiste en hallar la mediana $\tilde{z}$ de la muestra combinada, en contar luego

el número de x (m_1 , p. ej.) que superan a $\tilde{z}$, y el número de y (m_2) que también la superan. Si la hipótesis nula es verdadera, es de esperar que m_1 sea aproximadamente igual a $n_1/2$ y m_2 , a su vez, aproximadamente igual a $n_2/2$. Es obvio que esta dócima será sensible a las diferencias en posición entre $f_1(x)$ y $f_2(x)$, pero no a las diferencias de forma. Por tanto, si $f_1(x)$ y $f_2(x)$ tienen la misma mediana cabe esperar que la hipótesis nula sea aceptada en general, aunque sus formas sean muy distintas.

Para hacer esta dócima se necesita conocer, en la hipótesis nula, la distribución de m_1 y m_2 . Se demuestra que esta densidad es

$$g(m_1, m_2) = \frac{\binom{n_1}{m_1} \binom{n_2}{m_2}}{\binom{n_1+n_2}{a}} \quad (16)$$

en donde

$$a = \frac{n_1 + n_2 + 1}{2}$$

si $n_1 + n_2$ es impar y

$$a = \frac{n_1 + n_2}{2}$$

si $n_1 + n_2$ es par. Observemos que esta expresión es precisamente la distribución de las frecuencias de casillas en una tabla de contingen-

cia 2×2 con todos los totales marginales fijos, cuando existe independencia. La tabla de contingencia es

m_1	m_2	$n_1 + n_2 - a$
$n_1 - m_1$	$n_2 - m_2$	a
n_1	n_2	

en donde los totales marginales están a la derecha y debajo de la parte cerrada de la tabla. La hipótesis nula se docima utilizando, bien el criterio λ , o bien el criterio ji cuadrado. Si $n_1 + n_2$ es pequeño, se empleará (16) para calcular las probabilidades exactas, en vez de utilizar la probabilidad aproximada dada por la distribución ji cuadrado con un grado de libertad. Esta aproximación es bastante aceptable si tanto n_1 como n_2 son mayores que 10.

Intervalos confidenciales.—Para obtener intervalos confidenciales exactos para la diferencia entre las medianas de dos poblaciones, es necesario suponer que las distribuciones difieren solo en posición. Si $x_1, x_2, \dots, x_{n_1}$ es una muestra de una población de mediana ν_1 , e $y_1, y_2, \dots, y_{n_2}$ una muestra de otra población de mediana ν_2 , supondremos que las variantes

$$u_i = x_i - \nu_1 \quad \text{y} \quad v_i = y_i - \nu_2$$

tienen la misma densidad, $f(u)$, con mediana cero. Luego tanto las u como las v son dos muestras de la misma población. Eligiendo dos números r y s , la probabilidad de que u_r exceda a v_s se calcula del modo siguiente:

$$P(u_r > v_s) = \int_{-\infty}^{\infty} s \binom{n_2}{s} [F(v_s)]^{s-1} [1 - F(v_s)]^{n_2-s} \sum_{i=0}^{r-1} \binom{n_1}{i} [F(v_s)]^i [1 - F(v_s)]^{n_1-i} f(v_s) dv_s \quad (17)$$

$$= \sum_{i=0}^{r-1} \frac{s \binom{n_2}{s} \binom{n_1}{i}}{(s+i) \binom{n_1+n_2}{s+i}} \quad (18)$$

$$= \sum_{i=0}^{r-1} \frac{\binom{s+i-1}{s-1} \binom{n_1+n_2-s-i}{n_2-s}}{\binom{n_1+n_2}{n_1}} \quad (19)$$

Análogamente,

$$P(u_{r'} < v_{s'}) = \sum_{i=r'}^{n_1} \frac{\binom{s'+i-1}{s'-1} \binom{n_1+n_2-s'-i}{n_2-s'}}{\binom{n_1+n_2}{n_1}} \quad (20)$$

Si suponemos ahora que $r < s$, $r' > s'$, y que $v_2 > v_1$, se tendrá

$$P(y_{s'} - x_{r'} < v_2 - v_1 < y_s - x_r) = P(u_r < v_s \text{ y } u_{r'} < v_{s'}) \quad (21)$$

$$= 1 - P(u_r > v_s) - P(u_{r'} < v_{s'}) \quad (22)$$

y el primer miembro de esta relación proporciona un intervalo de confianza para $v_2 - v_1$ con un nivel que puede calcularse mediante (19) y (20). El intervalo confidencial proporciona una tercera dódima de la hipótesis nula de ser iguales ambas distribuciones. Si este intervalo no incluye el cero, la hipótesis se rechaza.

Bosquejaremos ahora una aproximación para muestras grandes que suele utilizarse cuanto tanto n_1 como n_2 son mayores que 10. Como la suma (19) vale 1 si se extiende a todo el recorrido i , podemos considerar la expresión como una densidad para la variante i y hallar la aproximación normal a dicha función. La suma puede sustituirse por la integral de la función. La media y la varianza de i son

$$E(i) = \frac{sn_1}{n_2 + 1} \quad (23)$$

$$\sigma_i^2 = \frac{sn_1}{n_2 + 1} \left[\frac{(s+1)(n_2+3)}{n_2+2} + \frac{(s+1)n_1}{n_2+2} - (2s+1) - \frac{sn_1}{n_2+1} \right] \quad (24)$$

y sus valores aproximados para valores grandes de n_1 y n_2 se obtienen haciendo

$$n_1 + n_2 = n \quad n_1 = n\alpha \quad n_2 = n\beta \quad s = \gamma n_2 = \beta\gamma n \quad (25)$$

y conservando solo los términos que comprenden la máxima potencia de n . Los resultados son

$$E(i) \cong n\alpha\gamma \quad (26)$$

$$\sigma_i^2 \cong n\alpha\gamma \frac{1-\gamma}{\beta} \quad (27)$$

La normalidad de i en muestras grandes se demuestra del mis-

mo modo que se indicó para d en (6). Un valor aproximado de suma que aparece en (19) es

$$\int_{-\infty}^A \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt \quad (28)$$

en donde

$$A = \frac{(r - 1 + 1/2) - n\alpha\gamma}{\sqrt{n\alpha\gamma(1-\gamma)}/\beta} \quad (29)$$

Dada s , se elegiría A para obtener el nivel de probabilidad deseado (p. ej., 1,96 para que la probabilidad fuese 0,025) y se resolvería la ecuación respecto a r .

Se plantea el problema de cómo elegir s ; evidentemente s debe ser superior a $n_2/2$, y r , inferior a $n_1/2$. Cabe, p. ej., hacer iguales ambas diferencias; pero es de esperar que el intervalo confidencial sea más corto al igualar las diferencias en escala estándar. El número de observaciones x inferior a ν_1 tiene una distribución aproximadamente normal, con media $n_1/2$ y desviación estándar $\sqrt{n_1}/2$. Análogamente, el número de observaciones y , superior a ν_2 , tiene una distribución aproximadamente normal, con media $n_2/2$ y desviación estándar $\sqrt{n_2}/2$. Se determina después s de modo que

$$\frac{(n_1/2) - r}{\sqrt{n_1}} = \frac{s - (n_2/2)}{\sqrt{n_2}} \quad (30)$$

Sustituyendo en esta relación los valores de n_1 , n_2 y s dados por (25), se resuelve la ecuación para hallar el valor de r , y se iguala después el resultado a la solución de r en (29). Se obtiene

$$\gamma \cong \frac{1}{2} + \frac{A}{2\sqrt{n}(\sqrt{\alpha\beta} + \beta)} \quad (31)$$

despreciando los términos con potencias superiores a $1/\sqrt{n}$; con los símbolos originales, resulta la expresión

$$s \cong \frac{n_2}{2} + \frac{A\sqrt{n_2}\sqrt{n_1+n_2}}{2(\sqrt{n_1} + \sqrt{n_2})} \quad (32)$$

y de (30)

$$r \cong \frac{n_1}{2} - \frac{A\sqrt{n_1}\sqrt{n_1+n_2}}{2(\sqrt{n_1} + \sqrt{n_2})} \quad (33)$$

Análogamente, podrá verse que se obtienen valores adecuados de r' y s' en (22), cambiando los signos en (32) y (33).

La técnica expuesta para hallar intervalos confidenciales para la diferencia de dos medianas puede utilizarse cualesquiera que sean los parámetros de posición, tales como medias o cuartilas.

16-5. Límites de tolerancia.—Una máquina automática de una fábrica de cojinetes de bolas se supone que produce cojinetes de 0,25 pulg de diámetro. Desde un punto de vista técnico, los cojinetes se consideran aceptables si su diámetro se halla comprendido entre los límites 0,249 pulg y 0,251 pulg. La producción es comprobada con regularidad midiendo diariamente el diámetro de una muestra aleatoria de cojinetes, y calculando límites estadísticos de tolerancia, L_1 y L_2 , a partir de las muestras. Si L_1 cae por encima de 0,249 y L_2 por debajo de 0,251, se acepta la producción. ¿Qué tamaño deberá tener la muestra, con el fin de que sea posible asegurar con probabilidad del 90% que los límites estadísticos de tolerancia contendrán el 99%, al menos, de la población de diámetros? Para los problemas de esta clase existe una solución no paramétrica sencilla.

Planteemos la cuestión en términos más generales: sea $f(z)$ una densidad y, por medio de una muestra de n valores, se desea determinar dos números L_1 y L_2 tales que, p. ej., el 0,99 del área de $f(z)$ quede comprendida entre L_1 y L_2 . Basándose en la muestra no se puede tener la *certeza* de que el 0,99 del área de $f(z)$ esté entre L_1 y L_2 , pero sí es posible determinar la probabilidad de que esto ocurra.

En otras palabras, se pretende hallar dos funciones, $L_1 = L_1(z_1, \dots, z_n)$, $L_2 = L_2(z_1, \dots, z_n)$, de la muestra aleatoria $z_1, \dots, z_n$, tales que la probabilidad de que

$$\int_{L_1}^{L_2} f(z) dz > \beta \quad (1)$$

sea igual a $1 - \alpha$.

O, expresado de otra manera, deseamos tener la probabilidad $1 - \alpha$ de que el área de $f(z)$ comprendida entre L_1 y L_2 sea mayor o igual que β . Los límites L_1 y L_2 se denominan *límites de tolerancia*. Hay multitud de funciones L_1 y L_2 que podrían utilizarse, pero aquí elegiremos $L_1 = x_1$ (la más pequeña de las n observaciones) y $L_2 = x_n$ (la observación mayor). Entonces, la densidad del área entre x_1 y x_n está dada por el teorema 16-2; es decir, si

$$v = \int_{x_1}^{x_n} f(z) dz \quad (2)$$

la densidad de v es

$$p(v) = n(n-1)v^{n-2}(1-v) \quad 0 < v < 1 \quad (3)$$

Ejemplo 16-1.—Supóngase que se desea determinar el tamaño que debe tener una muestra para asegurar, con probabilidad 0,90, que el 99% al menos de la producción diaria de cojinetes tiene sus diámetros comprendidos entre las observaciones menor y mayor. Las cantidades son $1-\alpha=0,90$, $\beta=0,99$, y se pretende determinar n de modo que

$$P(v > \beta) = 1 - \alpha$$

en donde la densidad v está dada por (3). Obtenemos

$$1 - \alpha = P(v > \beta) = \int_{\beta}^1 n(n-1)v^{n-2}(1-v) dv = 1 - n\beta^{n-1} + (n-1)\beta^n$$

Si sustituimos α y β , resulta la ecuación

$$0,90 = 1 - n(0,99)^{n-1} + (n-1)(0,99)^n$$

cuya resolución permite determinar n .

16-6. Dócima de rangos para dos muestras.—Una interesante dócima no paramétrica para dos muestras ha sido descrita por Wilcoxon y estudiada por Mann y Whitney. Dadas dos muestras aleatorias, $x_1, x_2, \dots, x_n$ e $y_1, y_2, \dots, y_m$, de poblaciones con distribuciones acumulativas continuas $F(x)$ y $G(y)$, se ordenan las $m+n$ observaciones muestrales en orden ascendente y se reemplaza la más pequeña por 1, la siguiente por 2, y así sucesivamente, hasta reemplazar la mayor por $m+n$. Estos enteros se denominan rangos de las observaciones. El criterio de la dócima es

$$T = \text{suma de rangos de observaciones de } y$$

y si T es significativamente grande o pequeña, se rechaza la hipótesis nula de que $F=G$. El problema de Teoría estadística consiste en hallar la distribución de T de modo que permita decir qué valores de T son significativos. Tal problema plantea grandes dificultades.

Sin embargo, Mann y Whitney han calculado la distribución para valores pequeños de m y n ; han hallado los momentos de T en general; han conseguido probar que la distribución de T es aproximadamente normal para m y n grandes, y han demostrado que la aproximación normal resulta suficientemente acurada si m y n

son mayores que 7. Nos contentaremos aquí con indicar cómo se obtienen los momentos de T mediante los métodos de dichos autores.

En lugar de T , usan el criterio U , definido como el número total de veces que las y preceden a las x cuando las dos muestras se disponen en orden. Así, como ejemplo sencillo, si la disposición es

$$xyxyxx$$

será $U=5$, ya que la primera y precede a tres x , y la segunda, a dos. En este ejemplo, $T=2+4=6$. Entre U y T existe la sencilla relación lineal

$$U = mn + \frac{m(m+1)}{2} - T \quad (1)$$

de modo que los resultados para U se transforman inmediatamente en resultados para T .

Sea ahora $N_{n,m}(U)$ el número de disposiciones en las cuales y precede a x exactamente U veces; omitiendo la última letra de tales disposiciones, hallamos

$$N_{n,m}(U) = N_{n-1,m}(U-m) + N_{n,m-1}(U) \quad (2)$$

en donde el primer término del segundo miembro procede de cuando la letra omitida es una x , y el segundo de cuando dicha letra es una y . En la hipótesis nula, el número total de disposiciones es $(m+n)!/m!n!$ [es decir, el número de maneras de seleccionar m posiciones para las y entre las $(m+n)$ posibles]; por tanto, la probabilidad de obtener U puede escribirse

$$p_{n,m}(U) = \frac{N_{n,m}(U)}{(m+n)!/m!n!} \quad (3)$$

y, utilizando esta definición en (2), se obtiene la relación

$$p_{n,m}(U) = \frac{n}{m+n} p_{n-1,m}(U-m) + \frac{m}{m+n} p_{n,m-1}(U) \quad (4)$$

Si designamos por $E_{n,m}(U)$ la esperanza de U para muestras de tamaños n y m , resulta que

$$E_{n,m}(U) = \sum_U U p_{n,m}(U) \quad (5)$$

y, multiplicando (4) por U y sumando respecto a U , se tiene

$$\begin{aligned} E_{n,m}(U) &= \frac{n}{m+n} \sum_U U p_{n-1,m}(U-m) + \frac{m}{m+n} \sum_U U p_{n,m-1}(U) \\ &= \frac{n}{m+n} \left[\sum_U (U-m) p_{n-1,m}(U-m) + m \sum_U p_{n-1,m}(U-m) \right] \\ &\quad + \frac{m}{m+n} E_{n,m-1}(U) \\ &= \frac{n}{m+n} [E_{n-1,m}(U) + m] + \frac{m}{m+n} E_{n,m-1}(U) \end{aligned}$$

o bien

$$(m+n)E_{n,m}(U) = nE_{n-1,m}(U) + mE_{n,m-1}(U) + mn \quad (6)$$

Esta última relación puede considerarse como una ecuación en diferencias que define $E_{n,m}(U)$ como función de m y n ; según la teoría de tales ecuaciones, su solución es un polinomio en m y n ,

$$E_{n,m}(U) = amn + bm + cn + d \quad (7)$$

Al sustituir (7) en (6) e igualar los coeficientes de los términos en m y n correspondientes en ambos miembros, se encuentra que $a = 1/2$ y que $b = c = d = 0$, de modo que

$$E_{n,m}(U) = \frac{mn}{2} \quad (8)$$

Para obtener el segundo momento de U , se multiplicaría (4) por U^2 y se sumaría respecto a U , obteniéndose, después de simplificar y tener en cuenta (8),

$$E_{n,m}(U^2) = \frac{n}{m+n} E_{n-1,m}(U^2) + \frac{m}{m+n} E_{n,m-1}(U^2) + \frac{m^2 n^2}{m+n} \quad (9)$$

que es también una ecuación en diferencias. La solución viene dada por un polinomio cuadrático en m y n . Sustituyendo tal polinomio e igualando coeficientes, se halla que el segundo momento es

$$E_{n,m}(U^2) = \frac{m^2 n^2}{4} + \frac{mn(m+n+1)}{12} \quad (10)$$

Teniendo en cuenta (8), la varianza de U es el segundo término del segundo miembro de (10). Utilizando estos resultados, junto

con la ecuación (1), se obtiene que la media y la varianza de T son

$$E(T) = \frac{m(m+n+1)}{2} \quad (11)$$

$$\sigma_T^2 = \frac{mn(m+n+1)}{12} \quad (12)$$

Con estos resultados puede aplicarse la d6cima de rangos, utilizando la aproximaci6n normal, cuando m y n son mayores que 7.

16-7. Eficiencias asint6ticas y d6cima de aleatorizaci6n.—

Un problema que surge en conexi6n con las d6cimas no param6tricas es si se perder6 gran parte de la eficiencia al usar tales d6cimas en lugar de las est6ndar, basadas en la teor6a normal. No existe posibilidad de responder a esta pregunta sin conocer las distribuciones poblacionales con que se opera. Sin embargo, resulta instructivo comparar algunas de las d6cimas no param6tricas con la d6cima t , en la hip6tesis de que la poblaci6n en cuesti6n es normal. Haremos tal comparaci6n para la d6cima de rangos, suponiendo que x e y tienen las densidades

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \quad y \quad g(y) = \frac{1}{\sqrt{2\pi}} e^{-(y-\theta)^2/2} \quad (1)$$

Sus formas acumulativas se designar6n por $F(x)$ y $G(y)$.

En general, es posible demostrar (aunque no lo hagamos aqu6) que, bajo condiciones completamente generales, la funci6n de potencia para docimar si un par6metro θ toma el valor θ_0 tiene, en el caso de muestras grandes, forma parab6lica en las proximidades del valor verdadero θ_0 . M6s concretamente, si es S_n el criterio para docimar θ en muestras de tama6o n , si S_n tiene media μ_n y varianza σ_n^2 , y si la distribuci6n de S_n es aproximadamente normal para tama6os muestrales n grandes, la ecuaci6n de la funci6n de potencia en las proximidades de θ_0 es

$$\beta_n(\theta) = \alpha + \frac{k\phi(k)}{\sigma_n^2} \left(\frac{d\mu_n}{d\theta_0} \right)^2 (\theta - \theta_0)^2 \quad (2)$$

en donde $\beta_n(\theta)$ es la funci6n de potencia, α el nivel de significaci6n de la d6cima, k el coeficiente de σ_n para la d6cima (determinado por α) y

$$\phi(k) = \frac{1}{\sqrt{2\pi}} e^{-k^2/2}$$

La forma de la parábola está determinada por el coeficiente de $(\theta - \theta_0)^2$, y tanto más potente será la dócima cuanto mayor sea este coeficiente. Supongamos ahora que se tiene otro estadígrafo S_n^* para la docimasia de la misma hipótesis y que su distribución es aproximadamente normal para muestras grandes con media μ^* y varianza σ_n^{*2} . Su función de potencia en las proximidades de θ_0 tendría una ecuación semejante a la (2), excepto que μ_n y σ_n vendrían reemplazados por μ_n^* y σ_n^* . Se obtiene una medida de la potencia relativa de las dos dócimas formando la razón de los coeficientes de $(\theta - \theta_0)^2$. La *eficiencia asintótica relativa* de S_n respecto a S_n^* (EAR) se define como el límite, cuando n tiende a infinito, de la razón:

$$\text{EAR}(S_n \text{ respecto a } S_n^*) = \lim_{n \rightarrow \infty} \frac{1/\sigma_n^2 (d\mu_n/d\theta_0)^2}{1/\sigma_n^{*2} (d\mu_n^*/d\theta_0)^2} \quad (3)$$

Nos proponemos ahora calcular esta razón, cuando S_n^* es la dócima t , y S_n , la dócima de rangos.

Para obtener la derivada de μ_n respecto a θ , es necesario hallar la esperanza del criterio U de la dócima de rangos bajo la hipótesis alternativa [$F \neq G$, es decir $\theta \neq 0$ en (1)]. Sea

$$\begin{aligned} w_{ij} &= 0 & \text{si } x_i < y_i \\ &= 1 & \text{si } x_i \geq y_i \end{aligned}$$

(x_i se refiere a la i -ésima observación x , no al i -ésimo valor ordenado de las x ; análogamente para y_i)

$$\begin{aligned} E(w_{ij}) &= P(x_i \geq y_i) \\ &= \int_{-\infty}^{\infty} [1 - F(v)] g(v) dv \\ &= \int_{-\infty}^{\infty} \left(\int_v^{\infty} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du \right) \frac{1}{\sqrt{2\pi}} e^{-(v-\theta)^2/2} dv \end{aligned}$$

y

$$\frac{dE(w_{ij})}{d\theta} = \int_{-\infty}^{\infty} \int_v^{\infty} \frac{1}{2\pi} e^{-u^2/2} (v - \theta) e^{-(v-\theta)^2/2} du dv$$

Poniendo $\theta = \theta_0 = 0$

$$\frac{dE(w_{ij})}{d\theta_0} = \int_{-\infty}^{\infty} \int_v^{\infty} \frac{v}{2\pi} e^{-(v^2+u^2)/2} du dv = -\frac{1}{2\sqrt{\pi}} \quad (4)$$

Por tanto, puesto que

$$\mu_n = E(U) = E \sum w_{ij} = mnE(w_{ij})$$

se deduce que

$$\frac{d\mu_n}{d\theta_0} = -\frac{mn}{2\sqrt{\pi}} \quad (5)$$

En la hipótesis nula

$$\sigma_n^2 = \frac{mn(m+n+1)}{12} \quad (6)$$

Para la dódima t

$$\mu_n^* = \theta \quad (7)$$

$$\sigma_n^{*2} = \frac{1}{m} + \frac{1}{n} \quad (8)$$

$$\frac{d\mu_n^*}{d\theta_0} = 1 \quad (9)$$

Sustituyendo (5), (6), (8) y (9) en (3), hallamos

$$\begin{aligned} \text{EAR}(S_n \text{ respecto a } S_n^*) &= \lim \frac{[12mn(m+n+1)](mn/2\sqrt{\pi})^2}{\frac{1}{1/m + 1/n}} \\ &= \frac{3}{\pi} \end{aligned} \quad (10)$$

que es, aproximadamente, el 95%. Esto significa que si se utiliza la dódima de rangos (en lugar de la dódima t) cuando la población se distribuye normalmente, solo se pierde alrededor del 5% de eficiencia para muestras grandes. La pérdida es, como se ve, pequeña cuando se sospecha que es aplicable la hipótesis de normalidad. Sin embargo, este cálculo nada nos dice acerca de las eficiencias relativas de las dos dódimas en el caso de que las distribuciones no coincidan con la normal. Cabe esperar que unas veces resulte preferible una dódima y que otras suceda lo contrario, según cuál sea la distribución. En la decisión de la dódima que va a utilizarse es, naturalmente, importante saber que los niveles de significación de la dódima de rangos no dependen de la forma de la distribución.

Diremos, incidentalmente, que la dódima de la mediana, dada en la sección 16-3 para dos muestras, tiene una eficiencia relativa

de $2/\pi$ respecto a la d6cima t , cuando la distribuci6n de la poblaci6n es normal.

Una cuesti6n relacionada con esto es la de saber si existir6 una d6cima no param6trica igualmente eficiente que la d6cima t para poblaciones normales. Lehmann y Stein han demostrado que tal d6cima existe. Es la llamada d6cima de aleatorizaci6n. Suponiendo varianzas iguales, esta d6cima se aplica como sigue: Se calculan las medias $\bar{x}$ e $\bar{y}$ de las dos muestras $x_1, \dots, x_n$ e $y_1, \dots, y_m$ y la diferencia $d = \bar{x} - \bar{y}$. A continuaci6n se dividen de todas las maneras posibles las $m+n$ observaciones en dos grupos de tama1os m y n [hay $(m+n)!/m!n!$ formas de hacer esto] y se calculan las diferencias en orden ascendente. Si la diferencia d cae entre las $2^{1/2}\%$ diferencias m6s peque1as, o entre las $2^{1/2}\%$ diferencias mayores, se rechaza la hip6tesis nula al nivel de significaci6n del 5%. Esta d6cima de aleatorizaci6n no param6trica es bastante importante desde el punto de vista te6rico, aunque su inter6s pr6ctico es menor a causa de la excesiva cantidad de c6lculo que frecuentemente requiere su aplicaci6n.

PROBLEMAS

1. H6llese la funci6n de densidad de $u=F(x_r)$, en donde x_r es la observaci6n ordenada r -6sima en una muestra de tama1o n de una poblaci6n con funci6n de distribuci6n acumulativa $F(x)$.

2. Obt6ngase la densidad dada en la ecuaci6n (16-2-10) utilizando (16-2-4).

3. Obt6ngase (16-2-10) mediante un razonamiento geom6trico, considerando dividido el eje x en cinco intervalos, seg6n se indica. Se considera que la muestra procede de una poblaci6n polinomial con cinco catego-



r6as que tenga probabilidades $F(y-\Delta y/2)$, $f(y)\Delta y$, $F(z-\Delta z/2)-F(y+\Delta y/2)$, $f(z)\Delta z$, $1-F(z+\Delta z/2)$, de manera que $r-1$ observaciones caigan en la primera categor6a, una en la segunda, etc. La densidad de x es $f(x)$ y la funci6n de distribuci6n acumulativa $F(x)$.

4. Util6cese el m6todo geom6trico del problema 3 para hallar la distribuci6n conjunta de u (6rea comprendida entre x_q y x_r) y v (6rea entre x_s y x_t), siendo $q < r < s < t$.

5. Demu6strese que el valor esperado de la observaci6n mayor de una muestra de tama1o dos, procedente de una poblaci6n normal de media cero y varianza unidad, es $1/\sqrt{\pi}$ y, por tanto, que para la poblaci6n normal general, el valor esperado es $\mu + (\sigma/\sqrt{\pi})$.

6. Si (x, y) es una observaci6n de una poblaci6n normal bivalente

con medias cero, varianzas unitarias y correlación ρ , demuéstrese que el valor esperado del mayor de los valores x e y es $\sqrt{(1-\rho)/\pi}$.

7. Obténgase la ecuación (16-4-7).

8. Compruébense las ecuaciones (16-4-9) y (16-4-10).

9. Demuéstrese que la variable t , definida por la ecuación (16-4-14), tiene una distribución aproximadamente normal para valores grandes de n .

10. Compruébense las ecuaciones (16-4-23) y (16-4-24).

11. Compruébese la ecuación (16-4-31).

12. Si $x_1, x_2, \dots, x_n$ es una muestra ordenada de una población cuya distribución acumulativa es $F(x)$, hállese la densidad de

$$u = \frac{[F(x_n) - F(x_2)]}{[F(x_n) - F(x_1)]}$$

13. La vida activa x , en horas, de los átomos radiactivos tiene la densidad $(1/\theta)e^{-x/\theta}$. Para estimar θ a partir de un determinado tipo de átomo se observa una muestra de n átomos, terminando la experiencia cuando haya dejado de existir el átomo r -ésimo; esto es, se trata de no esperar a que todos los átomos cesen en su actividad, sino hasta que r de ellos (r es un número elegido de antemano) hayan cesado. Los datos consisten en r medidas $x_1, x_2, \dots, x_r$ y $n-r$ medidas, de las cuales solo se sabe que exceden a x_r . Hállese la estimación máximo-verosímil de θ y demuéstrese que tiene la distribución ji cuadrado. Obsérvese que la verosimilitud contiene el factor $[1 - F(x_r)]^{n-r}$, en donde $F(x)$ es la distribución acumulativa.

14. Con referencia al problema 13, ¿debe empezarse con átomos de nueva actividad o basta empezar con átomos que hayan sido ya activos para diversas duraciones (y sigan siendo activos)?

15. Si x tiene una distribución uniforme entre $\theta - 1/2$ y $\theta + 1/2$, hállese la distribución de la mediana $\tilde{x}$ para muestras de tamaño $2k+1$.

16. Con referencia al problema 15, hállese la densidad de

$$z = (x_1 + x_{2k+1})/2$$

¿Cuál es mejor estimador de θ , z o $\tilde{x}$?

17. Demuéstrese que la mediana muestral es un estimador consistente de la mediana poblacional del problema 15.

18. Hemos visto que la media muestral para una distribución de varianza infinita (como la distribución de Cauchy) no es necesariamente un estimador consistente de la media poblacional. ¿Es la mediana muestral un estimador consistente de la mediana poblacional?

19. Si una población tiene la densidad

$$\begin{aligned} f(x) &= 1/2 e^{-(x-\theta)} & x > \theta \\ f(x) &= 1/2 e^{(x-\theta)} & x \leq \theta \end{aligned}$$

hállese la estimación máximo-verosímil de θ para muestras de tamaño n .

20. Una medida corriente de la asociación de dos variantes x e y es la *correlación por rangos* o *correlación de Spearman*. Se colocan ordenados por rangos los valores de x y se sustituyen las observaciones por sus rangos; análogamente, se sustituyen por sus rangos las observaciones y . Así, para muestras de tamaño n , p. ej., se tendría:

x	1	2	3	...	n
y	7	4	11	...	6

Utilizando estos rangos en pares, se calcula la correlación ordinaria

$$S = \frac{\sum (X_i - \bar{X})(y_i - \bar{y})}{\sqrt{\sum (X_i - \bar{X})^2 \sum (y_i - \bar{y})^2}} = 1 - \frac{6 \sum d_i^2}{n^3 - n}$$

en donde las mayúsculas representan los rangos, y $d_i = X_i - y_i$. Compruébese que la relación dada es cierta.

$$\left[\text{NOTA: } \sum_1^n i^2 = n(n+1)(2n+1)/6 \right]$$

21. Demuéstrese que la distribución de S en el problema 20 es independiente de la forma de las distribuciones de x e y , suponiendo que estas sean independientes y, por tanto, que S sea un criterio no paramétrico para la docimasia de la hipótesis nula de independencia.

22. Demuéstrese que la media y la varianza de S (Probl. 21) en la hipótesis de independencia, son cero y $1/(n-1)$, respectivamente. Para ello, demuéstrese que S puede expresarse en la forma

$$S = \frac{12}{n^3 - n} \left[Q - \frac{n(n+1)^2}{4} \right]$$

en donde $Q = \sum i y_i$ (sustituyendo X_i por i), y obsérvese que el coeficiente de $\prod_1^n u_i$ en

$$\phi(t) = \frac{1}{n!} \prod_{i=1}^n \left(\sum_{j=1}^n u_i t^{ij} \right)$$

es una función generatriz de momentos factoriales para Q .

23. Hállese n en el ejemplo 16-1.

24. Utilícese la dódima de rangos para docimar al nivel 5% la hipótesis nula de que estas dos muestras provienen de la misma población:

x	1,3	1,4	1,4	1,5	1,7	1,9	1,9
y	1,6	1,8	2,0	2,1	2,1	2,2	2,3

25. Utilícese la dódima de aleatorización en lugar de la de rangos en el problema 24.

26. Dedúzcase la relación entre U y T dada por la ecuación (16-6-1).

27. Complétense los detalles que implica la obtención de las ecuaciones (16-6-9) y (16-6-10) a partir de la (16-6-4).

BIBLIOGRAFIA

1. FRASER, D. A. S.: *Nonparametric Methods in Statistics*, John Wiley & Sons, Inc., Nueva York, 1957.
2. HOEFFDING, W.: «Large sample power of tests based on permutations of observations», *Annals of Mathematical Statistics*, vol. 23 (1952), páginas 169-192.
3. LEHMANN, E. L., y C. STEIN: «On some theory of nonparametric hypotheses», *Annals of Mathematical Statistics*, vol. 20 (1949), págs. 28-46.
4. LEHMANN, E. L.: *Testing Statistical Hypotheses*, John Wiley & Sons, Inc., Nueva York, 1959.
5. MANN, H. B., y D. R. WHITNEY: «On a test of whether one of two random variables is stochastically larger than the other», *Annals of Mathematical Statistics*, vol. 18 (1947), págs. 50-61.
6. MOOD, A. M.: «On the asymptotic efficiency of certain nonparametric two-sample tests», *Annals of Mathematical Statistics*, vol. 25 (1954), páginas 514-522.
7. SIEGEL, S.: *Nonparametric Statistics for the Behavioral Sciences*, McGraw-Hill Book Company, Inc., Nueva York, 1956.
8. WILCOXON, F.: «Individual comparisons by ranking methods», *Biometrics*, vol. 1 (1945), págs. 80-83.
9. WILKS, S. S.: «Order statistics», *Bulletin of the American Mathematical Society*, vol. 54 (1948), págs. 6-50.

TABLAS

DESCRIPCION DE LAS TABLAS

I. Ordenadas de la función de densidad normal.—Esta tabla da los valores de

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

para valores de x entre 0 y 4, a intervalos de 0,01. Naturalmente, para los valores negativos de x se hace uso de la propiedad $f(-x) = f(x)$.

II. Distribución normal acumulativa.—Da los valores de

$$F(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$

para valores de x comprendidos entre 0 y 3,5, a intervalos de 0,01. Para valores negativos de x se utiliza la relación $F(-x) = 1 - F(x)$. En la parte inferior de la tabla principal se dan los valores de x que corresponden a unos cuantos valores elegidos de F .

III. Distribución ji cuadrado acumulativa.—Esta tabla da los valores de u correspondientes a unos pocos valores seleccionados de $F(u)$, donde

$$F(u) = \int_0^u \frac{x^{(n-2)/2} e^{-x/2} dx}{2^{n/2} [(n-2)/2]!}$$

para n , número de grados de libertad, igual a 1, 2, ..., 30. Para valores mayores de n resulta completamente acurada una aproximación normal. La cantidad $\sqrt{2u} - \sqrt{2n-1}$ tiene una distribución casi normal, con media cero y varianza unidad. En consecuencia, u_α , punto α de la distribución, puede calcularse mediante

$$u_\alpha = (1/2) (x_\alpha + \sqrt{2n-1})^2$$

donde x_α es el punto α de la distribución normal acumulativa.

Como ejemplo, calculemos el valor 0,95 de u para $n=30$ grados de libertad:

$$\begin{aligned} u_{0,95} &= \frac{1}{2}(1,645 + \sqrt{59})^2 \\ &= 43,5 \end{aligned}$$

con error inferior al 1%.

IV. Distribución acumulativa de «Student».—Esta tabla da los valores de t correspondientes a unos pocos valores seleccionados de

$$F(t) = \int_{-\infty}^t \frac{\left(\frac{n-1}{2}\right)!}{\left(\frac{2}{n-2}\right)! \sqrt{\pi n} \left(1 + \frac{x^2}{n}\right)^{(n+1)/2}} dx$$

con $n=1, 2, \dots, 30, 40, 60, 120, \infty$. Puesto que la densidad es simétrica respecto a t , resulta que $F(-t)=1-F(t)$. No debe interpolarse linealmente entre grados de libertad, sino entre sus recíprocos, si se desea obtener un error inferior a una unidad de la última cifra. Como ejemplo, calculemos el valor 0,975 para 40 grados de libertad. Los valores para 30 y 60 son 2,042 y 2,000. Utilizando los valores recíprocos de n , el valor interpolado es

$$2,042 - \frac{1/30 - 1/40}{1/30 - 1/60} (2,042 - 2,000) = 2,021$$

que es el valor correcto. Por interpolación lineal se hubiera obtenido 2,028.

V. Distribución acumulativa F .—Esta tabla da los valores de F que corresponden a cinco valores de

$$G(F) = \int_0^F \frac{\left(\frac{m+n-2}{2}\right)! m^{m/2} n^{n/2} x^{(m-2)/2} (n+mx)^{-(m+n)/2} dx}{\left(\frac{m-2}{2}\right)! \left(\frac{n-2}{2}\right)!}$$

para valores elegidos de m y n ; m es el número de grados de libertad en el numerador de F , y n el correspondiente al denominador. La tabla da también los valores que corresponden a $G=0,10, 0,05, 0,025, 0,01$ y $0,005$, debido a que F_{1-a} para m y n grados de libertad es el recíproco de F_a para n y m grados de libertad. Así,

para $G=0,05$, con 3 y 6 grados de libertad, se tiene

$$F_{0,95}(3, 6) = \frac{1}{F_{0,05}(6, 3)} = \frac{1}{8,94} = 0,112$$

La interpolación debe hacerse respecto a los recíprocos de m y n , como en la tabla IV, si se desea exactitud.

VI a VIII. Recorrido studentizado.—Sean $x_1, x_2, \dots, x_n$ normales e independientes con media μ y varianza σ^2 , y definamos el recorrido R por

$$R = \max_i x_i - \min_i x_i$$

Sea $\nu s^2/\sigma^2$ una variante ji cuadrado con ν grados de libertad, independiente de las x_i , y sea $q = R/s$. Las tablas VI a VIII dan los valores de q correspondientes a

$$\int_{q\alpha}^{\infty} f(q) dq = \alpha$$

para $\alpha=0,01$; 0,05 y 0,10 para varios de n y ν .

TABLA II. DISTRIBUCIÓN NORMAL ACUMULATIVA

$$F(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$

x	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3,0	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990
3,1	0,9990	0,9991	0,9991	0,9991	0,9992	0,9992	0,9992	0,9992	0,9993	0,9993
3,2	0,9993	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995	0,9995
3,3	0,9995	0,9995	0,9995	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9997
3,4	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9998

x ...	1,282	1,645	1,960	2,326	2,576	3,090	3,291	3,891	4,417
$F(x)$...	0,90	0,95	0,975	0,99	0,995	0,999	0,9995	0,99995	0,999995
$2[1-F(x)]$...	0,20	0,10	0,05	0,02	0,01	0,002	0,001	0,0001	0,00001

TABLA III. DISTRIBUCIÓN II CUADRADO ACUMULATIVA *

$$F(u) = \int_0^u \frac{x^{(n-2)/2} e^{-x/2} dx}{2^{n/2} [n-2]!}$$

F n	0,005	0,010	0,025	0,050	0,100	0,250	0,500	0,750	0,900	0,950	0,975	0,990	0,995
1	0,0393	0,0157	0,00982	0,00393	0,0158	0,102	0,455	1,32	2,71	3,84	5,02	6,63	7,88
2	0,0100	0,0201	0,0506	0,103	0,211	0,575	1,39	2,77	4,61	5,99	7,38	9,21	10,6
3	0,0717	0,115	0,216	0,352	0,584	1,21	2,37	4,11	6,25	7,81	9,35	11,3	12,8
4	0,207	0,297	0,484	0,711	1,06	1,92	3,36	5,39	7,78	9,49	11,1	13,3	14,9
5	0,412	0,534	0,831	1,15	1,61	2,67	4,55	6,63	9,24	11,1	12,8	15,1	16,7
6	0,676	0,872	1,24	1,64	2,20	3,45	5,35	7,84	10,6	12,6	14,4	16,8	18,5
7	0,989	1,24	1,69	2,17	2,83	4,25	6,35	9,04	12,0	14,1	16,0	18,5	20,3
8	1,34	1,65	2,18	2,73	3,49	5,07	7,34	10,2	13,4	15,5	17,5	20,1	22,0
9	1,73	2,09	2,70	3,33	4,17	5,90	8,34	11,4	14,7	16,9	19,0	21,7	23,6
10	2,16	2,56	3,25	3,94	4,87	6,74	9,34	12,5	16,0	18,3	20,5	23,2	25,2
11	2,60	3,05	3,82	4,57	5,58	7,58	10,3	13,7	17,3	19,7	21,9	24,7	26,8
12	3,07	3,57	4,40	5,23	6,30	8,44	11,3	14,8	18,5	21,0	23,3	26,2	28,3
13	3,57	4,11	5,01	5,89	7,04	9,30	12,3	16,0	19,8	22,4	24,7	27,7	29,8
14	4,07	4,66	5,63	6,57	7,79	10,2	13,3	17,1	21,1	23,7	26,1	29,1	31,3
15	4,60	5,23	6,26	7,26	8,55	11,0	14,3	18,2	22,3	25,0	27,5	30,6	32,8
16	5,14	5,81	6,91	7,96	9,31	11,9	15,3	19,4	23,5	26,3	28,8	32,0	34,3
17	5,70	6,41	7,56	8,67	10,1	12,8	16,3	20,5	24,8	27,6	30,2	33,4	35,7
18	6,26	7,01	8,23	9,39	10,9	13,7	17,3	21,6	26,0	28,9	31,5	34,8	37,2
19	6,84	7,63	8,91	10,1	11,7	14,6	18,3	22,7	27,2	30,1	32,9	36,2	38,6
20	7,43	8,26	9,59	10,9	12,4	15,5	19,3	23,8	28,4	31,4	34,2	37,6	40,0
21	8,03	8,90	10,3	11,6	13,2	16,3	20,3	24,9	29,6	32,7	35,5	38,9	41,4
22	8,64	9,54	11,0	12,3	14,0	17,2	21,3	26,0	30,8	33,9	36,8	40,3	42,8
23	9,26	10,2	11,7	13,1	14,8	18,1	22,3	27,1	32,0	35,2	38,1	41,6	44,2
24	9,89	10,9	12,4	13,8	15,7	19,0	23,3	28,2	33,2	36,4	39,4	43,0	45,6
25	10,5	11,5	13,1	14,6	16,5	19,9	24,3	29,3	34,4	37,7	40,6	44,3	46,9
26	11,2	12,2	13,8	15,4	17,3	20,8	25,3	30,4	35,6	38,9	41,9	45,6	48,3
27	11,8	12,9	14,6	16,2	18,1	21,7	26,3	31,5	36,7	40,1	43,2	47,0	49,6
28	12,5	13,6	15,3	16,9	18,9	22,7	27,3	32,6	37,9	41,3	44,5	48,3	51,0
29	13,1	14,3	16,0	17,7	19,8	23,6	28,3	33,7	39,1	42,6	45,7	49,6	52,3
30	13,8	15,0	16,8	18,5	20,6	24,5	29,3	34,8	40,3	43,8	47,0	50,9	53,7

* Esta tabla se ha resumido de «Tables of percentage points of the incomplete beta function and of the chi-square distribution», *Biometrika*, vol. 32 (1941). Se publica con permiso del autor, Catherine M. Thompson, y del editor de *Biometrika*.

TABLA IV. DISTRIBUCIÓN ACUMULATIVA DE «STUDENT» *

$$F(t) = \int_{-\infty}^t \frac{\left(\frac{n-1}{2}\right)!}{\left(\frac{n-2}{2}\right)! \sqrt{\pi n} \left(1 + \frac{x^2}{n}\right)^{(n+1)/2}} dx$$

$n \backslash F$	0,75	0,90	0,95	0,975	0,99	0,995	0,9995
1	1,000	3,078	6,314	12,706	31,821	63,657	636,619
2	0,816	1,886	2,920	4,303	6,965	9,925	31,598
3	0,765	1,638	2,353	3,182	4,541	5,841	12,941
4	0,741	1,533	2,132	2,776	3,747	4,604	8,610
5	0,727	1,476	2,015	2,571	3,365	4,032	6,859
6	0,718	1,440	1,943	2,447	3,143	3,707	5,959
7	0,711	1,415	1,895	2,365	2,998	3,499	5,405
8	0,706	1,397	1,860	2,306	2,896	3,355	5,041
9	0,703	1,383	1,833	2,262	2,821	3,250	4,781
10	0,700	1,372	1,812	2,228	2,764	3,169	4,587
11	0,697	1,363	1,796	2,201	2,718	3,106	4,437
12	0,695	1,356	1,782	2,179	2,681	3,055	4,318
13	0,694	1,350	1,771	2,160	2,650	3,012	4,221
14	0,692	1,345	1,761	2,145	2,624	2,977	4,140
15	0,691	1,341	1,753	2,131	2,602	2,947	4,073
16	0,690	1,337	1,746	2,120	2,583	2,921	4,015
17	0,689	1,333	1,740	2,110	2,567	2,898	3,965
18	0,688	1,330	1,734	2,101	2,552	2,878	3,922
19	0,688	1,328	1,729	2,093	2,539	2,861	3,883
20	0,687	1,325	1,725	2,086	2,528	2,845	3,850
21	0,686	1,323	1,721	2,080	2,518	2,831	3,819
22	0,686	1,321	1,717	2,074	2,508	2,819	3,792
23	0,685	1,319	1,714	2,069	2,500	2,807	3,767
24	0,685	1,318	1,711	2,064	2,492	2,797	3,745
25	0,684	1,316	1,708	2,060	2,485	2,787	3,725
26	0,684	1,315	1,706	2,056	2,479	2,779	3,707
27	0,684	1,314	1,703	2,052	2,473	2,771	3,690
28	0,683	1,313	1,701	2,048	2,467	2,763	3,674
29	0,683	1,311	1,699	2,045	2,462	2,756	3,659
30	0,683	1,310	1,697	2,042	2,457	2,750	3,646
40	0,681	1,303	1,684	2,021	2,423	2,704	3,551
60	0,679	1,296	1,671	2,000	2,390	2,666	3,460
120	0,677	1,289	1,658	1,980	2,358	2,617	3,373
∞	0,674	1,282	1,645	1,960	2,326	2,576	3,291

* Esta tabla es un resumen de *Statistical Tables* de R. A. Fisher y Frank Yates, publicado por Oliver & Boyd, Ltd., Edimburgo y Londres, 1938. Se reproduce con permiso de los autores y editores. Hay edición española, Aguilar, S. A. de Ediciones.

TABLA V. DISTRIBUCIÓN F ACUMULATIVA * (m GRADOS DE LIBERTAD EN EL NUMERADOR; n EN EL DENOMINADOR)

$$G(F) = \int_0^F \frac{[(m+n-2)/2]! m^{m/2} n^{n/2} x^{(m-2)/2} (n+mx)^{-(m+n)/2} dx}{[(m-2)/2]! [(n-2)/2]!}$$

G	n	m	1	2	3	4	5	6	7	8	9	10	12	15	20	30	60	120	∞
0.90	1		39.9	49.5	53.6	55.8	57.2	58.2	58.9	59.4	59.9	60.2	60.7	61.2	61.7	62.3	62.8	63.1	63.3
0.95			161	200	216	225	230	234	237	239	241	242	246	248	248	250	252	253	254
0.975			648	800	864	900	922	937	948	957	963	969	977	985	993	1000	1010	1010	1020
0.99			4050	5000	5400	5620	5760	5860	5930	5980	6020	6060	6110	6160	6210	6260	6310	6340	6370
0.995			16200	20000	21600	22500	23100	23400	23700	23900	24100	24200	24400	24600	24800	25000	25200	25400	25500
0.90	2		8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38	9.39	9.41	9.42	9.44	9.46	9.47	9.48	9.49
0.95			18.5	19.0	19.2	19.2	19.3	19.3	19.4	19.4	19.4	19.4	19.4	19.4	19.5	19.5	19.5	19.5	19.5
0.975			38.5	39.0	39.2	39.2	39.3	39.3	39.4	39.4	39.4	39.4	39.4	39.4	39.4	39.5	39.5	39.5	39.5
0.99			98.5	99.0	99.2	99.2	99.3	99.3	99.4	99.4	99.4	99.4	99.4	99.4	99.4	99.5	99.5	99.5	99.5
0.995			199	199	199	199	199	199	199	199	199	199	199	199	199	199	199	199	199
0.90	3		5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24	5.23	5.22	5.20	5.18	5.17	5.15	5.14	5.13
0.95			10.1	9.95	9.88	9.82	9.78	9.74	9.71	9.68	9.65	9.62	9.59	9.56	9.53	9.50	9.47	9.45	9.43
0.975			17.4	16.0	15.4	15.1	14.9	14.7	14.6	14.5	14.4	14.4	14.3	14.3	14.2	14.1	14.0	13.9	13.9
0.99			34.1	30.8	29.5	28.7	28.2	27.9	27.7	27.5	27.3	27.2	27.1	26.9	26.7	26.5	26.3	26.2	26.1
0.995			55.6	49.8	47.5	46.2	45.4	44.8	44.4	44.1	43.9	43.7	43.4	43.1	42.8	42.5	42.1	42.0	41.8
0.90	4		4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.93	3.92	3.90	3.87	3.84	3.82	3.79	3.78	3.76
0.95			7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.75	5.69	5.66	5.63
0.975			12.2	10.6	9.98	9.60	9.36	9.20	9.07	8.98	8.90	8.84	8.75	8.66	8.56	8.46	8.36	8.31	8.26
0.99			21.2	18.0	16.7	16.0	15.5	15.2	15.0	14.8	14.7	14.5	14.4	14.2	14.0	13.8	13.7	13.6	13.5
0.995			31.3	26.3	24.3	23.2	22.5	22.0	21.6	21.4	21.1	21.0	20.7	20.4	20.2	19.9	19.6	19.5	19.3
0.90	5		4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32	3.30	3.27	3.24	3.21	3.17	3.14	3.12	3.11
0.95			6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.50	4.43	4.40	4.37
0.975			10.3	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68	6.62	6.52	6.43	6.33	6.23	6.12	6.07	6.02
0.99			16.3	13.3	12.1	11.4	11.0	10.7	10.5	10.3	10.2	10.1	9.89	9.72	9.55	9.38	9.20	9.11	9.02
0.995			22.8	18.3	16.5	15.6	14.9	14.5	14.2	14.0	13.8	13.6	13.4	13.1	12.9	12.7	12.4	12.3	12.1
0.90	6		3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96	2.94	2.90	2.87	2.84	2.80	2.76	2.74	2.72
0.95			5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.81	3.74	3.70	3.67
0.975			8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52	5.46	5.37	5.27	5.17	5.07	4.96	4.90	4.85
0.99			13.7	10.9	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.72	7.56	7.40	7.23	7.06	6.97	6.88
0.995			18.6	14.5	12.9	12.0	11.5	11.1	10.8	10.6	10.4	10.2	10.0	9.81	9.59	9.36	9.12	9.00	8.88
0.90	7		3.26	3.07	2.92	2.86	2.82	2.83	2.78	2.75	2.72	2.70	2.67	2.63	2.59	2.56	2.51	2.49	2.47
0.95			5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.38	3.30	3.27	3.23
0.975			8.07	6.52	5.89	5.52	5.29	5.12	4.99	4.90	4.82	4.76	4.67	4.57	4.47	4.36	4.25	4.20	4.14
0.99			12.2	9.53	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.47	6.31	6.16	5.99	5.82	5.74	5.65
0.995			16.2	12.4	10.9	10.1	9.52	9.16	8.89	8.68	8.51	8.38	8.18	7.97	7.75	7.53	7.31	7.19	7.08
0.90	8		3.11	2.92	2.81	2.76	2.73	2.67	2.62	2.59	2.56	2.54	2.50	2.46	2.42	2.38	2.34	2.31	2.29
0.95			5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.08	3.01	2.97	2.93
0.975			7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36	4.30	4.20	4.10	4.00	3.89	3.78	3.73	3.67
0.99			11.5	8.66	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.67	5.52	5.36	5.20	5.03	4.95	4.86
0.995			14.7	11.0	9.60	8.81	8.30	7.95	7.69	7.50	7.34	7.21	7.01	6.81	6.61	6.40	6.18	6.06	5.95

0.90	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44	2.42	2.38	2.34	2.30	2.25	2.21	2.18	2.16
0.95	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.86	2.79	2.75	2.71
0.975	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03	3.96	3.87	3.77	3.67	3.56	3.48	3.40	3.33
0.99	10.6	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.11	4.96	4.81	4.65	4.54	4.43	4.31
0.995	13.6	10.1	8.72	7.96	7.47	7.13	6.88	6.69	6.54	6.42	6.23	6.03	5.83	5.62	5.41	5.30	5.19
0.90	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35	2.32	2.28	2.24	2.20	2.15	2.11	2.08	2.06
0.95	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.84	2.77	2.70	2.62	2.58	2.54
0.975	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78	3.72	3.62	3.52	3.42	3.31	3.20	3.14	3.08
0.99	10.0	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.71	4.56	4.41	4.25	4.08	4.00	3.91
0.995	12.8	9.83	8.68	7.94	7.34	6.87	6.54	6.30	6.12	5.97	5.85	5.67	5.57	5.07	4.86	4.75	4.64
0.90	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21	2.19	2.15	2.10	2.06	2.01	1.96	1.93	1.90
0.95	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.47	2.38	2.34	2.30
0.975	6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44	3.37	3.28	3.18	3.07	2.96	2.85	2.79	2.72
0.99	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.16	4.01	3.86	3.70	3.54	3.46	3.36
0.995	11.8	8.51	7.23	6.52	6.07	5.76	5.52	5.35	5.20	5.09	4.91	4.72	4.53	4.33	4.12	4.01	3.90
0.90	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09	2.06	2.02	1.97	1.92	1.87	1.82	1.79	1.76
0.95	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.25	2.16	2.11	2.07
0.975	6.20	4.77	4.15	3.80	3.58	3.41	3.29	3.20	3.12	3.06	2.96	2.86	2.76	2.64	2.52	2.46	2.40
0.99	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.67	3.52	3.37	3.21	3.05	2.96	2.87
0.995	10.8	7.70	6.48	5.80	5.37	5.07	4.85	4.67	4.54	4.42	4.25	4.07	3.88	3.69	3.48	3.37	3.26
0.90	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96	1.94	1.89	1.84	1.79	1.74	1.68	1.64	1.61
0.95	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.04	1.95	1.90	1.84
0.975	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84	2.77	2.68	2.57	2.46	2.35	2.22	2.16	2.09
0.99	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.23	3.09	2.94	2.78	2.61	2.52	2.42
0.995	9.94	6.99	5.82	5.17	4.76	4.47	4.26	4.09	3.96	3.85	3.68	3.50	3.32	3.12	2.92	2.81	2.69
0.90	2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85	1.82	1.77	1.72	1.67	1.61	1.54	1.50	1.46
0.95	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.84	1.74	1.68	1.62
0.975	5.57	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.57	2.51	2.41	2.31	2.20	2.07	1.94	1.87	1.79
0.99	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.84	2.70	2.55	2.39	2.21	2.11	2.01
0.995	9.18	6.35	5.24	4.62	4.23	3.95	3.74	3.58	3.45	3.34	3.18	3.01	2.82	2.63	2.42	2.30	2.18
0.90	2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74	1.71	1.66	1.60	1.54	1.48	1.40	1.35	1.29
0.95	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.84	1.74	1.68	1.62
0.975	5.29	3.93	3.34	3.01	2.79	2.63	2.51	2.41	2.33	2.27	2.17	2.06	1.94	1.82	1.67	1.58	1.48
0.99	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.50	2.35	2.20	2.03	1.84	1.73	1.60
0.995	8.49	5.80	4.73	4.14	3.76	3.49	3.29	3.13	3.01	2.90	2.74	2.57	2.39	2.19	1.96	1.83	1.69
0.90	2.75	2.35	2.13	1.99	1.90	1.82	1.77	1.72	1.68	1.65	1.60	1.54	1.48	1.41	1.32	1.26	1.19
0.95	3.92	3.00	2.60	2.37	2.21	2.10	2.02	2.02	1.96	1.91	1.83	1.75	1.66	1.55	1.43	1.35	1.25
0.975	5.15	3.80	3.23	2.89	2.67	2.52	2.39	2.30	2.22	2.16	2.05	1.94	1.82	1.69	1.53	1.43	1.31
0.99	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.34	2.19	2.03	1.86	1.66	1.53	1.38
0.995	8.18	5.54	4.50	3.92	3.55	3.28	3.09	2.93	2.81	2.71	2.54	2.37	2.19	1.98	1.75	1.61	1.43
0.90	2.71	2.30	2.08	1.94	1.85	1.77	1.72	1.67	1.63	1.60	1.55	1.49	1.42	1.34	1.24	1.17	1.00
0.95	3.84	3.00	2.60	2.37	2.21	2.10	2.02	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.46	1.32	1.22
0.975	5.02	3.69	3.12	2.79	2.57	2.41	2.29	2.29	2.11	2.05	1.94	1.83	1.71	1.57	1.39	1.27	1.00
0.99	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.18	2.04	1.88	1.70	1.47	1.32	1.00
0.995	7.88	5.30	4.28	3.72	3.35	3.09	2.90	2.74	2.62	2.52	2.36	2.19	2.00	1.79	1.53	1.36	1.00

* Esta tabla es un resumen de «Tables of percentage points of the inverted beta distribution», *Biometrika*, vol. 33 (1943). Se reproduce con permiso de los autores, Maxine Merrington y Catherine M. Thompson, y del editor de *Biometrika*.

TABLA VI. PUNTOS PORCENTUALES SUPERIORES DEL 1% DEL RECORRIDO STUDENTIZADO *

Las cifras son $q_{0.01}$ con $P(q < q_{0.01}) = 0.99$

$\frac{n}{v}$	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	90.03	135.0	164.3	185.6	202.2	215.8	227.2	237.0	245.6	253.2	260.0	266.2	271.8	277.0	281.8	286.3	290.4	294.3	298.0
2	10.04	19.02	22.29	24.72	26.63	28.20	29.53	30.68	31.69	32.59	33.40	34.13	34.81	35.43	36.00	36.53	37.03	37.50	37.95
3	8.26	10.62	12.17	13.33	14.24	15.00	15.64	16.20	16.69	17.13	17.53	17.89	18.22	18.52	18.81	19.07	19.32	19.55	19.77
4	6.51	8.12	9.17	9.96	10.58	11.10	11.55	11.93	12.27	12.57	12.84	13.09	13.32	13.53	13.73	13.91	14.08	14.24	14.40
5	5.70	6.98	7.80	8.42	8.91	9.32	9.67	9.97	10.24	10.48	10.70	10.89	11.08	11.24	11.40	11.55	11.68	11.81	11.93
6	5.24	6.33	7.03	7.56	7.97	8.32	8.61	8.87	9.10	9.30	9.48	9.65	9.81	9.95	10.08	10.21	10.32	10.43	10.54
7	4.95	5.92	6.54	7.01	7.37	7.68	7.94	8.17	8.36	8.55	8.71	8.86	9.00	9.12	9.24	9.35	9.46	9.55	9.65
8	4.75	5.64	6.20	6.62	6.96	7.24	7.47	7.68	7.86	8.03	8.18	8.31	8.44	8.55	8.66	8.76	8.85	8.94	9.03
9	4.60	5.43	5.96	6.35	6.66	6.91	7.13	7.33	7.49	7.65	7.78	7.91	8.03	8.13	8.23	8.33	8.41	8.49	8.57
10	4.48	5.27	5.77	6.14	6.43	6.67	6.87	7.05	7.21	7.36	7.49	7.60	7.71	7.81	7.91	7.99	8.08	8.15	8.23
11	4.39	5.15	5.62	5.97	6.25	6.48	6.67	6.84	6.99	7.13	7.25	7.36	7.46	7.56	7.65	7.73	7.81	7.88	7.95
12	4.32	5.05	5.50	5.84	6.10	6.32	6.51	6.67	6.81	6.94	7.06	7.17	7.26	7.36	7.44	7.52	7.59	7.66	7.73
13	4.26	4.96	5.40	5.73	5.98	6.19	6.37	6.53	6.67	6.79	6.90	7.01	7.10	7.19	7.27	7.35	7.42	7.48	7.55
14	4.21	4.89	5.32	5.63	5.88	6.08	6.26	6.41	6.54	6.66	6.77	6.87	6.96	7.05	7.13	7.20	7.27	7.33	7.39
15	4.17	4.84	5.25	5.56	5.80	5.99	6.16	6.31	6.44	6.55	6.66	6.76	6.84	6.93	7.00	7.07	7.14	7.20	7.26
16	4.13	4.79	5.19	5.49	5.72	5.92	6.08	6.22	6.35	6.46	6.56	6.66	6.74	6.82	6.90	6.97	7.03	7.09	7.15
17	4.10	4.74	5.14	5.43	5.66	5.85	6.01	6.15	6.27	6.38	6.48	6.57	6.66	6.73	6.81	6.87	6.94	7.00	7.05
18	4.07	4.70	5.09	5.38	5.60	5.79	5.94	6.08	6.20	6.31	6.41	6.50	6.58	6.65	6.73	6.79	6.85	6.91	6.97
19	4.05	4.67	5.05	5.33	5.55	5.73	5.89	6.02	6.14	6.25	6.34	6.43	6.51	6.58	6.65	6.72	6.78	6.84	6.89
20	4.02	4.64	5.02	5.29	5.51	5.69	5.84	5.97	6.09	6.19	6.28	6.37	6.45	6.52	6.59	6.65	6.71	6.77	6.82
24	3.96	4.55	4.91	5.17	5.37	5.54	5.69	5.81	5.92	6.02	6.11	6.19	6.26	6.33	6.39	6.45	6.51	6.56	6.61
30	3.89	4.45	4.80	5.05	5.24	5.40	5.54	5.65	5.76	5.85	5.93	6.01	6.08	6.14	6.20	6.26	6.31	6.36	6.41
40	3.82	4.37	4.70	4.93	5.11	5.26	5.39	5.50	5.60	5.69	5.76	5.83	5.90	5.96	6.02	6.07	6.12	6.16	6.21
60	3.76	4.28	4.59	4.82	4.99	5.13	5.25	5.36	5.45	5.53	5.60	5.67	5.73	5.78	5.84	5.89	5.93	5.97	6.01
120	3.70	4.20	4.50	4.71	4.87	5.01	5.12	5.21	5.30	5.37	5.44	5.50	5.56	5.61	5.66	5.71	5.75	5.79	5.83
∞	3.64	4.12	4.40	4.60	4.76	4.88	4.99	5.08	5.16	5.23	5.29	5.35	5.40	5.45	5.49	5.54	5.57	5.61	5.65

* De E. S. Pearson y H. O. Hartley, *Biometrika Tables for Statisticians*, vol. 1, págs. 176-77, publicada por The Biometrika Trustees, Cambridge University Press, Londres, 1954. Reproducida con permiso de los autores y editores. Han sido incorporadas algunas correcciones de ± 1 en la última cifra, proporcionadas por el doctor Leon Harter.

TABLA VII. PUNTOS PORCENTUALES SUPERIORES DEL 5 % DEL RECORRIDO STUDENTIZADO *

Las cifras son $q_{0,05}$ con $P(q < q_{0,05}) = 0,95$

$\frac{n}{v}$	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	17,97	26,98	32,82	37,08	40,41	43,12	45,40	47,36	49,07	50,59	51,96	53,20	54,33	55,36	56,32	57,22	58,04	58,83	59,56
2	6,08	8,33	9,80	10,88	11,74	12,44	13,03	13,54	13,99	14,39	14,75	15,08	15,38	15,65	15,91	16,14	16,37	16,57	16,77
3	4,50	5,91	6,82	7,50	8,04	8,48	8,85	9,18	9,46	9,72	9,95	10,15	10,35	10,52	10,69	10,84	10,98	11,11	11,24
4	3,93	5,04	5,76	6,29	6,71	7,05	7,35	7,60	7,83	8,03	8,21	8,37	8,52	8,66	8,79	8,91	9,03	9,13	9,23
5	3,64	4,60	5,22	5,67	6,03	6,33	6,58	6,80	6,99	7,17	7,32	7,47	7,60	7,72	7,83	7,93	8,03	8,12	8,21
6	3,46	4,34	4,90	5,30	5,63	5,90	6,12	6,32	6,49	6,65	6,79	6,92	7,03	7,14	7,24	7,34	7,43	7,51	7,59
7	3,34	4,16	4,68	5,06	5,36	5,61	5,82	6,00	6,16	6,30	6,43	6,55	6,66	6,76	6,85	6,94	7,02	7,10	7,17
8	3,26	4,04	4,53	4,89	5,17	5,40	5,60	5,77	5,92	6,05	6,18	6,29	6,39	6,48	6,57	6,65	6,73	6,80	6,87
9	3,20	3,95	4,41	4,76	5,02	5,24	5,43	5,59	5,74	5,87	5,98	6,09	6,19	6,28	6,36	6,44	6,51	6,58	6,64
10	3,15	3,88	4,33	4,65	4,91	5,12	5,30	5,46	5,60	5,72	5,83	5,93	6,03	6,11	6,19	6,27	6,34	6,40	6,47
11	3,11	3,82	4,26	4,57	4,82	5,03	5,20	5,35	5,49	5,61	5,71	5,81	5,90	5,98	6,06	6,13	6,20	6,27	6,33
12	3,08	3,77	4,20	4,51	4,75	4,95	5,12	5,27	5,39	5,51	5,61	5,71	5,80	5,88	5,95	6,02	6,09	6,15	6,21
13	3,06	3,73	4,15	4,45	4,69	4,88	5,05	5,19	5,32	5,43	5,53	5,63	5,71	5,79	5,86	5,93	5,99	6,05	6,11
14	3,03	3,70	4,11	4,41	4,64	4,83	4,99	5,13	5,25	5,36	5,46	5,55	5,64	5,71	5,79	5,85	5,91	5,97	6,03
15	3,01	3,67	4,08	4,37	4,59	4,78	4,94	5,08	5,20	5,31	5,40	5,49	5,57	5,65	5,72	5,78	5,85	5,90	5,96
16	3,00	3,65	4,05	4,33	4,56	4,74	4,90	5,03	5,15	5,26	5,35	5,44	5,52	5,59	5,66	5,73	5,79	5,84	5,90
17	2,98	3,63	4,02	4,30	4,52	4,70	4,86	4,99	5,11	5,21	5,31	5,39	5,47	5,54	5,61	5,67	5,73	5,79	5,84
18	2,97	3,61	4,00	4,28	4,49	4,67	4,82	4,96	5,07	5,17	5,27	5,35	5,43	5,50	5,57	5,63	5,69	5,74	5,79
19	2,96	3,59	3,98	4,25	4,47	4,65	4,79	4,92	5,04	5,14	5,23	5,31	5,39	5,46	5,53	5,59	5,65	5,70	5,75
20	2,95	3,58	3,96	4,23	4,45	4,62	4,77	4,90	5,01	5,11	5,20	5,28	5,36	5,43	5,49	5,55	5,61	5,66	5,71
24	2,92	3,53	3,90	4,17	4,37	4,54	4,68	4,81	4,92	5,01	5,10	5,18	5,25	5,32	5,38	5,44	5,49	5,55	5,59
30	2,89	3,49	3,85	4,10	4,30	4,46	4,60	4,72	4,82	4,92	5,00	5,08	5,15	5,21	5,27	5,33	5,38	5,43	5,47
40	2,86	3,44	3,79	4,04	4,23	4,39	4,52	4,63	4,73	4,82	4,90	4,98	5,04	5,11	5,16	5,22	5,27	5,31	5,36
60	2,83	3,40	3,74	3,98	4,16	4,31	4,44	4,55	4,65	4,73	4,81	4,88	4,94	5,00	5,06	5,11	5,15	5,20	5,24
120	2,80	3,36	3,68	3,92	4,10	4,24	4,36	4,47	4,56	4,64	4,71	4,78	4,84	4,90	4,95	5,00	5,04	5,09	5,13
∞	2,77	3,31	3,63	3,86	4,03	4,17	4,29	4,39	4,47	4,55	4,62	4,68	4,74	4,80	4,85	4,89	4,93	4,97	5,01

* De E. S. Pearson y H. O. Hartley, *Biometrika Tables for Statisticians*, vol. 1, págs. 176-77, publicada por The Biometrika Trustees, Cambridge University Press, Londres, 1954. Reproducida con permiso de los autores y editores. Han sido incorporadas algunas correcciones de ± 1 en la última cifra, proporcionadas por el doctor James Pachares.

TABLA VIII. PUNTOS PORCENTUALES SUPERIORES DEL 10% DEL RECORRIDO STUDENTIZADO *

Las cifras son $q_{0,10}$, con $P(q < q_{0,10}) = 0,90$

$\frac{n}{v}$	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	8,93	13,44	16,36	18,49	20,15	21,51	22,64	23,62	24,48	25,24	25,92	26,54	27,10	27,62	28,10	28,54	28,96	29,35	29,71
2	4,13	5,73	6,77	7,54	8,14	8,63	9,05	9,41	9,72	10,01	10,26	10,49	10,70	10,89	11,07	11,24	11,39	11,54	11,68
3	3,33	4,47	5,20	5,74	6,16	6,51	6,81	7,06	7,29	7,49	7,67	7,83	7,98	8,12	8,25	8,37	8,48	8,58	8,68
4	3,01	3,98	4,59	5,03	5,39	5,68	5,93	6,14	6,33	6,49	6,65	6,78	6,91	7,02	7,13	7,23	7,33	7,41	7,58
5	2,85	3,72	4,26	4,66	4,98	5,24	5,46	5,65	5,82	5,97	6,10	6,22	6,34	6,44	6,54	6,63	6,71	6,79	6,86
6	2,75	3,56	4,07	4,44	4,73	4,97	5,17	5,34	5,50	5,64	5,76	5,87	5,98	6,07	6,16	6,25	6,32	6,40	6,47
7	2,68	3,45	3,93	4,28	4,55	4,78	4,97	5,14	5,28	5,41	5,53	5,64	5,74	5,83	5,91	5,99	6,06	6,13	6,19
8	2,63	3,37	3,83	4,17	4,43	4,65	4,83	4,99	5,13	5,25	5,36	5,46	5,56	5,64	5,72	5,80	5,87	5,93	6,00
9	2,59	3,32	3,76	4,08	4,34	4,54	4,72	4,87	5,01	5,13	5,23	5,33	5,42	5,51	5,58	5,66	5,72	5,79	5,85
10	2,56	3,27	3,70	4,02	4,26	4,47	4,64	4,78	4,91	5,03	5,13	5,23	5,32	5,40	5,47	5,54	5,61	5,67	5,73
11	2,54	3,23	3,66	3,96	4,20	4,40	4,57	4,71	4,84	4,95	5,05	5,15	5,23	5,31	5,38	5,45	5,51	5,57	5,63
12	2,52	3,20	3,62	3,92	4,16	4,35	4,51	4,65	4,78	4,89	4,99	5,08	5,16	5,24	5,31	5,37	5,44	5,49	5,55
13	2,50	3,18	3,59	3,88	4,12	4,30	4,46	4,60	4,72	4,83	4,93	5,02	5,10	5,18	5,25	5,31	5,37	5,43	5,48
14	2,49	3,16	3,56	3,85	4,08	4,27	4,42	4,56	4,68	4,79	4,88	4,97	5,05	5,12	5,19	5,26	5,32	5,37	5,43
15	2,48	3,14	3,54	3,83	4,05	4,23	4,39	4,52	4,64	4,75	4,84	4,93	5,01	5,08	5,15	5,21	5,27	5,32	5,38
16	2,47	3,12	3,52	3,80	4,03	4,21	4,36	4,49	4,61	4,71	4,81	4,89	4,97	5,04	5,11	5,17	5,23	5,28	5,33
17	2,46	3,11	3,50	3,78	4,00	4,18	4,33	4,46	4,58	4,68	4,77	4,86	4,93	5,01	5,07	5,13	5,19	5,24	5,30
18	2,45	3,10	3,49	3,77	3,98	4,16	4,31	4,44	4,55	4,65	4,75	4,83	4,90	4,98	5,04	5,10	5,16	5,21	5,26
19	2,45	3,09	3,47	3,75	3,97	4,14	4,29	4,42	4,53	4,63	4,72	4,80	4,88	4,95	5,01	5,07	5,13	5,18	5,23
20	2,44	3,08	3,46	3,74	3,95	4,12	4,27	4,40	4,51	4,61	4,70	4,78	4,85	4,92	4,99	5,05	5,10	5,16	5,20
24	2,42	3,05	3,42	3,69	3,90	4,07	4,21	4,34	4,44	4,54	4,63	4,71	4,78	4,85	4,91	4,97	5,02	5,07	5,12
30	2,40	3,02	3,39	3,65	3,85	4,02	4,16	4,28	4,38	4,47	4,56	4,64	4,71	4,77	4,83	4,89	4,94	4,99	5,03
40	2,38	2,99	3,35	3,60	3,80	3,96	4,10	4,21	4,32	4,41	4,49	4,56	4,63	4,69	4,75	4,81	4,86	4,90	4,95
60	2,36	2,96	3,31	3,56	3,75	3,91	4,04	4,16	4,25	4,34	4,42	4,49	4,56	4,62	4,67	4,73	4,78	4,82	4,86
120	2,34	2,93	3,28	3,52	3,71	3,86	3,99	4,10	4,19	4,28	4,35	4,42	4,48	4,54	4,60	4,65	4,69	4,74	4,78
∞	2,33	2,90	3,24	3,48	3,66	3,81	3,93	4,04	4,13	4,21	4,28	4,35	4,41	4,47	4,52	4,57	4,61	4,65	4,69

* De James Pachares, «Tables of the upper 10% points of the Studentized ranges, *Biometrika*, vol. 46, partes 3 y 4 (1959), págs. 461-66. Reproducción con permiso del autor y de Biometrika Trustees.

**SOLUCIONES DE LOS PROBLEMAS
PROPUESTOS AL FINAL DE LOS CAPITULOS**

SOLUCIONES DE LOS PROBLEMAS PROPUESTOS AL FINAL DE LOS CAPITULOS

CAPITULO 2

1. a) $\{x: x=5\}$. d) S .
b) $\{x: x=8, 10\}$. e) $\emptyset$.
c) S . f) $\emptyset$.
2. a) $\{x: 0 \leq x \leq 10\}$. d) $\{x: 0 \leq x < 7\}$.
b) $\{x: 5 < x \leq 20\}$. e) $\{x: 5 < x < 7, 15 < x \leq 20\}$.
c) $\{x: 0 \leq x < 3\}$. f) $\{x: 0 \leq x < 3, 10 < x \leq 20\}$.
3. a) $\{(x, y): x \geq 0, y > 5\}$.
b) $\{(x, y): x \geq 0, 0 \leq y < 5 \text{ e } y > 10\}$.
c) $\{(x, y): x \geq 0, 0 \leq y < 5\}$. e) $\{(x, y): x \geq 0, y > 10\}$.
d) $\{(x, y): x \geq 0, y \geq 0, y \neq 5\}$. f) $\{(x, y): x \geq 0, 5 < y \leq 10\}$.
g) $\{(x, y): x \geq 0, 0 \leq y \leq 5 \text{ e } y > 10\}$.
4. a) O bien ocurre $\bar{A}$ u ocurre B , o bien ocurren $\bar{A}$ y B .
b) O bien no ocurre A o no ocurre B , o bien no ocurren ni A ni B .
c) Ocurren A y B .
d) No ocurre A y ocurre B .
e) No ocurren ni A ni B .
f) Ocurre S .
5. $S = \{(R, R), (R, V), (V, R), (V, V)\}$.
 $P[(R, R)] = 9/25$; $P[(R, V)] = P[(V, R)] = 6/25$; $P[(V, V)] = 4/25$.
Este particular espacio muestral contiene cuatro puntos.
6. $4/10$. 7. $1/2$. 8. $1/24$. 9. $1/8$; $1/2$.
10. $1/10$. 11. 20; 25. 12. 48; 30. 13. 84.
14. 18. 15. 63. 16. 60. 17. 112.
18. $\binom{12}{7} \binom{10}{4}$. 19. 1545. 20. 676 000. 21. 54.
22. 91. 23. $1/6$. 24. $1/17$.
25. $\binom{4}{2} \binom{48}{3} / \binom{52}{5}$; $\left[\binom{4}{2} \binom{48}{3} + \binom{4}{3} \binom{48}{2} + \binom{48}{1} \right] / \binom{52}{5}$.

26. $4 / \binom{52}{13}$. 27. 35/108. 29. $n! / n_1! n_2! \dots n_k! m!$.
30. $m/(m+n)$. 31. 5/324. 32. 35/648. 33. 3/216.
34. 1/5. 35. $\binom{n}{m} (k-1)^{n-m} / k^n$. 40. 0.
41. -840. 42. 20/27. 43. 155/45.
44. $(13!)^2/26!$; $(13!)^4/(8!)^2(5!)^2 26!$.
45. $\frac{13!}{418!} \binom{4}{2} 4^4 / \binom{52}{6}$; $\frac{13! 4^2}{9!(2!)^2} \binom{4}{2}^2 / \binom{52}{6}$. 46. 0,0633.
47. 3/80. 48. 49/90. 49. 53/99. 50. 5/13.
51. $1 - \frac{\binom{33}{13}}{\binom{39}{13}}$; $1 - \frac{\binom{33}{13}}{\binom{39}{13}} - 6 \frac{\binom{33}{12}}{\binom{39}{13}}$
52. $2 \binom{5}{3} \binom{21}{10} / \binom{26}{13}$. 53. $2 \binom{23}{12} / \binom{26}{13}$; 1/2.
54. 5/11. 55. 1/2. 56. 244/495. 57. 3/4; 1/3.
58. a) Amarillo liso: 1/4; amarillo rugoso: 1/4; verde liso: 1/4; verde rugoso: 1/4.
 b) Amarillo rugoso: 1/8; amarillo liso: 3/8; verde liso: 3/8; verde rugoso: 1/8.
 c) Amarillo liso: 9/16; amarillo rugoso: 3/16; verde liso: 3/16; verde rugoso: 1/16.
59. q^2 ; $2q^2/(1+q)$. La proporción de albinos tendería a q .
60. 1/5. 61. 4/7. 62. 45/109.
63. 35 por ciento; 800/35 por ciento. 64. Aproximadamente 0,35.
65. Aproximadamente 0,062.

CAPITULO 3

1. $\binom{13}{x} \binom{39}{5-x} / \binom{52}{5}$, $0 \leq x \leq 5$;
- $$\sum_{x' < x} \binom{13}{x'} \binom{39}{5-x'} / \binom{52}{5}, \quad 0 \leq x \leq 5.$$

$$2. \quad \binom{10}{x} \left(\frac{1}{4}\right)^x \left(\frac{3}{4}\right)^{10-x}, \quad 0 \leq x \leq 10;$$

$$\sum_{x' < x} \binom{10}{x'} \left(\frac{1}{4}\right)^{x'} \left(\frac{3}{4}\right)^{10-x'}, \quad 0 \leq x \leq 10.$$

$$3. \quad \left(\frac{1}{2}\right)^x, \quad x=1, 2, 3, \dots$$

$$4. \quad \left(\frac{1}{6}\right), \quad 1 \leq x \leq 6; \quad \sum_{x' \leq x} \left(\frac{1}{6}\right), \quad 1 \leq x \leq 6.$$

$$5. \quad \frac{x-1}{36}, \quad 2 \leq x \leq 7, \quad \frac{13-x}{36}, \quad 7 \leq x \leq 12.$$

$$6. \quad \binom{13}{1} \binom{39}{x-1} / \binom{x}{1} \binom{52}{x}, \quad 1 \leq x \leq 40.$$

$$7. \quad \binom{6}{x} \left(\frac{1}{3}\right)^x \left(\frac{2}{3}\right)^{6-x}, \quad 0 \leq x \leq 6.$$

$$8. \quad \binom{n}{x} \binom{m}{k-x} / \binom{m+n}{k}, \quad \text{siempre que } 0 < k \leq m+n; \text{ si } k \leq n, \text{ y } k \leq m, \\ 0 \leq x \leq k; \text{ si } k \leq n \text{ y } k \geq m, k-m \leq x \leq k; \text{ si } k \geq n \text{ y } k \geq m, k-m \leq x \leq n; \\ \text{si } k \geq n \text{ y } k \leq m, 0 \leq x \leq n.$$

$$9. \quad \frac{n!}{x!y!z!(n-x-y-z)!} \left(\frac{1}{8}\right)^x \left(\frac{3}{8}\right)^y \left(\frac{3}{8}\right)^z \left(\frac{1}{8}\right)^{n-x-y-z} \\ 0 \leq x+y+z \leq n.$$

$$10. \quad \binom{60}{x} (0,01)^x (0,99)^{60-x}, \quad 0 \leq x \leq 60.$$

$$11. \quad 5 \binom{8}{5} \binom{12}{x-5} / x \binom{20}{x}, \quad 5 \leq x \leq 17.$$

$$12. \quad \binom{4}{x} \binom{4}{y} \binom{44}{6-x-y} / \binom{52}{6}, \quad 0 \leq x \leq 4, \quad 0 \leq y \leq 4, \quad x+y \leq 6.$$

$$14. \quad \binom{4}{x} \binom{48}{6-x} / \binom{52}{6}, \quad 0 \leq x \leq 4.$$

$$15. \quad \sum_{i=60}^{100} \binom{2500}{i} \binom{2500}{100-i} / \binom{5000}{100} \cong \sum_{x=60}^{100} \binom{100}{x} \left(\frac{1}{2}\right)^{100}.$$

$$16. \quad 1 - \sum_{x=0}^{20} \binom{200}{x} (0,05)^x (0,95)^{200-x}.$$

17. $1 - (0,99)^{10} \cong 0,096.$

18. 28.

19. $\left(\frac{1}{2}\right)^x, x=1, 2, 3, \dots$

20. $\frac{3}{2^{x+1}}, x=1, 3, 5, \dots$

21. $\sum_{x=1}^5 \binom{10}{x} \left(\frac{1}{3}\right)^x \left(\frac{2}{3}\right)^{10-x}$

22. $4/e^2.$

23. $\sum_{x=0}^{14} \frac{e^{-20} (20)^x}{x!}$

24. $1 - \sum_{x=0}^3 \frac{e^{-1/2} (1/2)^x}{x!}$

25. $1 - \sum_{x=0}^{20} \frac{e^{-10} (10)^x}{x!}$

26. $1 - \sum_{x=1}^5 \left(\frac{1}{6}\right) \left(\frac{5}{6}\right)^{x-1}.$

27. $\sum_{x=0}^2 \sum_{y=0}^2 \frac{10!}{x! y! (10-x-y)!} \left(\frac{25}{36}\right)^x \left(\frac{1}{36}\right)^y \left(\frac{10}{36}\right)^{10-x-y}.$

28. $\sum_{x=0}^4 \frac{10!}{x! (4-x)! 6!} \left(\frac{25}{36}\right)^y \left(\frac{1}{36}\right)^{4-x} \left(\frac{10}{36}\right)^6 = \binom{10}{4} \left(\frac{26}{36}\right)^4 \left(\frac{10}{36}\right)^6;$

$$\sum_{t=2}^4 \binom{10}{t} \left(\frac{26}{36}\right)^t \left(\frac{10}{36}\right)^{10-t}$$

29. $\sum_{y=0}^6 \sum_{x=2y}^{20-y} \frac{20!}{x! y! (20-x-y)!} \left(\frac{1}{3}\right)^x \left(\frac{1}{6}\right)^y \left(\frac{1}{2}\right)^{20-x-y}.$

30. $\sum_{y=0}^4 \sum_{x=y+1}^{10-y} \binom{13}{x} \binom{13}{y} \binom{26}{10-x-y} / \binom{52}{10}.$

31. 9/81, 12/81, 16/81, 12/81, 10/81, 6/81, 3/81, 1/81 para 1, 2, 3, ..., 9, respectivamente.

32. $\binom{4}{x} \binom{44}{7-x-y} / \binom{48}{7-y}, 0 \leq x \leq \min(4, 7-y).$

33. $\frac{(n-y)!}{x! z! (n-x-y-z)!} \left(\frac{1}{8}\right)^x \left(\frac{3}{8}\right)^z \left(\frac{1}{8}\right)^{n-x-y-z} / \left(\frac{5}{8}\right)^{n-y},$
 $y=0, 1, \dots, n, \quad x=0, 1, \dots, n, \quad x+z \leq n-y.$

$$34. \sum_{x=0}^1 \sum_{y=0}^{3-2x} f(x, y).$$

$$35. f(0, 5) + f(3, 4) + f(4, 3) + f(5, 0).$$

$$36. \sum_{x=0}^3 f(x|1 \leq y \leq 6).$$

$$37. \sum_{x=0}^m \sum_{y=0}^i f(x, y), \text{ donde } i \text{ es el mayor entero inferior a } \frac{a+x}{2}$$

$$38. \sum_{z=0}^{m-2} \sum_{y=z+1}^{m-1} \sum_{x=y+1}^m f(x, y, z).$$

$$39. f(2|3).$$

$$40. \sum_{x=0}^5 f(x, 5-x|3).$$

$$41. \sum_{x=0}^3 \sum_{y=0}^4 \sum_{z=5}^n \sum_{w=6}^m f(x, y, z, w).$$

$$42. \sum_{x=a}^b \sum_{y=0}^m f(x, y, y) / \sum_{x=0}^m \sum_{y=0}^m f(x, y, y).$$

$$43. \sum_{x=1}^m \sum_{y=0}^i \sum_{z=0}^{x-1} f(x, y, z) / \sum_{x=1}^m \sum_{y=0}^m \sum_{z=0}^{x-1} f(x, y, z), \text{ donde } i \text{ es el mayor entero inferior a } x/2.$$

CAPITULO 4

1. a) 1.
b) 1.
c) 1.
d) 0.
e) 1/2.

- f) 1.
g) 1/2.
h) 0.
i) 1.
j) 1/2.

- k) 0.
l) 0.
m) 0.
n) 0.
o) 1/2.
p) 1/2.

2. a) 1.
b) 1.
c) 1.
d) 0.
e) 1-1/e.

- f) 1.
g) 1-1/e.
h) 0.
i) 1.
j) 1/e.

- k) 0.
l) 0.
m) 0.
n) 0.
o) 1-1/e.
p) 1/e.

3. $1/4$; $3/16$; $7/12$.
4. $3x(2-x)/4$, $0 < x < 2$, $b^2(3-b)/4 - a^2(3-a)/4$; 1.
5. $2^{-1/4}$; $(0,95)^{1/4}$. 6. $9/16$; $27/64$. 7. $(0,1)^{1/4}$.
8. $8/27$; $1/27$. 9. $1 - \int_{0,242}^{0,258} f(x) dx$. 10. $1 - (0,36)^3$.
11. Más pesado. 12. $[1 - (0,01)^{1/5}]$ millares. 13. $\left(\frac{1 - e^{-4}}{1 - e^{-9}} \right)^5$.
14. a) $1/4$; b) $1/2$; c) $1/2$; d) $3/4$; e) $2/3$;
f) $\pi/16$; g) 0; h) $1/4$; i) $1/8$.
15. $1/e$; $e^{-a}(1+a) - e^{-b}(1+b)$; $3/4$. 16. Aprox. 1,68.
17. $1 - (2/e - 1/e^2)^3$. 18. 0,96. 19. $1 - 100/x$, $x > 100$.
20. $(1 - e^{-x})(1 - e^{-y})$. 21. e^{-x} . 22. e^{-x} ; $1 - 1/e$.
23. a) $1 - (1+x)^{2-n} - (1+y)^{2-n} + (1+x+y)^{2-n}$; $x > 0$; $y > 0$.
b) $(n-2)(1+x)^{1-n}$, $x > 0$.
c) $1 - (1+x)^{2-n}$, $x > 0$.
d) $(n-1)(1+x)^{n-1}(1+x+y)^{-n}$, $x > 0$, $y > 0$.
24. $2(1-x-y)/(1-y)^2$; $0 < x < 1-y$. 25. $2x/(1-y^2)$; $0 < y < x < 1$.
26. $47/64$. 27. $1/6$.
28. $2/(1+x)^2$; $x > 1$. 29. $n+1$, $0 < y < x^n$, $0 < x < 1$.
30. $1/3$; $1 < y < 4$. 31. e^{-y} ; $0 < y < \infty$.
32. $h(z) = \begin{cases} z & 0 < z < 1 \\ 2-z & 1 < z < 2 \end{cases}$ 33. $4ze^{-2z}$, $0 < z < \infty$.
34. $2z^3 e^{-z^2}$, $0 < z < \infty$. 35. 1, $0 < u < 1$, $0 < v < 1$.
36. $3(1-z^2)$, $0 < z < 1$. 37. $1/2\sqrt{y}$; $0 < y < 1$.
38. 1, $0 < z < 1$. 39. $h(u, v) = e^{-u}$, $0 < v < u < \infty$; e^{-v} , $0 < v < \infty$.
40. $27u^2 e^{-3u}/2$, $0 < u < \infty$. 41. $2x$, $0 < x < 1$.
42. $\frac{1}{2\sqrt{a(y-b)}} f[\sqrt{(y-b)/a}]$, $y > b$. 43. $f[u^{-1}(y)] \left[d \frac{u^{-1}(y)}{dy} \right]$.
44. $1/2$. 45. $2/3$. 46. 32, 33, 34.

CAPITULO 5

1. 1,33 dólares.
2. $5/3$.
3. 2.
4. p ; pq .
5. np ; npq .
6. a .
7. $1/4$; $1/48$.
8. $1/6$; $1/45$.
9. $1/2$; $3/4$.
12. $(2 - \sqrt{2})/2$.
13. $a/(a-t)$, $t < a$; $1/a$; $1/a^2$.
14. $(pt+q)^n$; $n(n-1)(n-2)p^3 + 3n(n-1)p^2 + np$.
15. $4r/(r+1)$.
16. $a^2/(a-t)^2$, $t < a$; a^2ue^{-au} .
18. $m(t)e^{-t\mu}$.
20. $(r-1)!/a^r$.
24. 0.
25. Si; $\mu_r' = [(-1)^r/r!]$ multiplicado por la derivada de orden r en $t=0$.
26. $\frac{a^2}{(a-t_2)(a-t_1)}$, $t_1, t_2 < a$; $2/a^3$.
28. a) $(26-9x)/3(3-x)$; b) $(78-28x)/3(3-x)$; c) $3x^2-18x+26$.
29. $\binom{n}{y} q^y (1-q)^{n-y}$.
30. $\binom{n-y}{x} p^{*n-x-y} (1-p^*)^x$, $p^* = \frac{1-p-q}{1-q}$, $y=0, 1, 2, \dots, n$,
 $x=0, 1, 2, \dots, n$, $x \leq n-y$; $(n-y)p^*$.
33. $\frac{26x-9x^2}{3(3-x)}$

CAPITULO 6

1. 0 si $x < \alpha$; $\frac{x-\alpha}{\beta-\alpha}$ si $\alpha < x < \beta$; 1 si $x > \beta$.
2. $y=(x-1)^2/4$; $0,0025 < y < 0,9025$.
4. 0,9938; 0,9544.
5. 1,96; 1,645; 2,576; 1,645.
6. $e^{a^2/2}$.
7. Cero para r impar; $r! \sigma^r / \left(\frac{r}{2}\right)! 2^{r/2}$ para r par.
8. $e^{-1/4/\sqrt{\pi}}$; $1/2$; $1/2$.
9. 0,8821.
10. $3\sqrt{2\pi}$.
12. $(\alpha+1)(\alpha+2)(\alpha+3)\beta^3$.
13. $(1-2t)^{-n/2}$ $t < 1/2$; n ; $2n$.
14. 5,991.
15. $(\alpha+r)! \beta^r / \alpha!$

$$18. (\alpha+1)/(\alpha+\beta+2); \frac{\alpha+1}{\alpha+\beta+2} \left(\frac{\alpha+2}{\alpha+\beta+3} - \frac{\alpha+1}{\alpha+\beta+2} \right).$$

$$20. \mu.$$

$$21. a) g(y|x) = e^{-\frac{x}{2\sigma^2}(y-\mu)^2} / \sqrt{2\pi\sigma^2/x}, \quad x > 0, \quad -\infty < y < \infty.$$

$$b) m(x, y) = \frac{x^{(m-1)/2} e^{-\frac{x}{2}[1+(y-\mu)^2/\sigma^2]}}{\sigma \sqrt{\pi} \Gamma\left(\frac{m}{2}\right) 2^{(m+1)/2}}, \quad x > 0, \quad -\infty < y < \infty.$$

$$22. a) \frac{\sigma^m \Gamma\left(\frac{m+1}{2}\right) [\sigma^2 + (y-\mu)^2]^{-(m+1)/2}}{\sqrt{\pi} \Gamma\left(\frac{m}{2}\right)} \quad -\infty < y < \infty.$$

$$b) \sigma^2/(m-2).$$

25. Los dos primeros son iguales a μ y σ^2 ; los restantes son nulos.

$$26. 10.$$

$$27. (\alpha+t)! \beta!/\alpha!$$

$$28. \mu = \frac{2^{1/4}}{\sqrt{\pi}} (-1/4)!; \quad \sigma^2 = \frac{\sqrt{2}}{\pi} [1 - (-1/4)!^2].$$

$$29. (1-t)^{-1}, \quad t < 1; \quad e^{-x}, \quad x > 0.$$

$$30. -\log(1-x).$$

$$31. -2 \log(1-x).$$

$$32. n(x; 0, 2).$$

$$34. 0,242.$$

$$35. 0,9999+.$$

40. Resolviendo $F(x) = G(u)$ respecto a u .

CAPITULO 7

3. Población muestreada: Los transistores fabricados en la planta A.

Población objetivo: Todos los transistores fabricados por la compañía.

4. Población objetivo: Todos los transistores fabricados en las plantas B y C.

5. a) (0, 1), (1, 1), (1, 0), (0, 0). b) (1, 1), (2, 1), (1, 0), (0, 0).

6. $g(y, x_2) = p^y q^{2-y}$ o

$y=0, 1, 2$

$x_2=0, 1$

$x_2 \backslash y$	y		
	0	1	2
0	q^2	pq	0
1	0	pq	p^2

$$7. \begin{array}{|c|c|c|c|} \hline y & 0 & 1 & 2 \\ \hline P(y) & q^2 & 2pq & p^2 \\ \hline \end{array} \quad \text{o} \quad \binom{2}{y} p^y q^{2-y}, \quad y=0, 1, 2.$$

8. $3/4$.

9. $3/4$.

10. $\left(\frac{2}{3}\right)^{\sum x_i} \left(\frac{1}{3}\right)^{9-\sum x_i}; \quad \left(\frac{9}{y}\right) \left(\frac{2}{3}\right)^y \left(\frac{1}{3}\right)^{9-y}$.

11. $2/3; 16/81$.

13. $(n-1)\sigma^2/n$.

14. $\bar{x}$.

16. 0.

18. 250.

19. Supóngase normalidad; 68.

20. $\pm 0,633; \pm 0,329$.

21. 13.

22. 136.

23. 62.

24. 3375.

25. 0,106.

26. $3; 3/4$.

27. 1.

30. $p(1-p)(1-2p)/n^2$.

31. $0,0001 te^{-0,01t}, t > 0$.

32. $(N+1)/2$.

CAPITULO 8

1. $5\mu^2/3$.

2. $\{\hat{\mu}: -\infty < \hat{\mu} < \infty\}; \{\mu: -\infty < \mu < \infty\}$.

3. $\frac{15\mu^2}{8}; \frac{3\mu^2}{2}$.

4. $R(d_3; \mu) \leq R(d_1; \mu) \leq R(d_2; \mu)$.

6. $9/8; 9/10; 4/5$.

7. La mitad de la mayor observación; $2\bar{x}$.

8. 1,1; 0,403.

9. $-1 - \frac{n}{\sum \log x_i}; (1-2\bar{x})/(\bar{x}-1)$.

10. $\bar{x}/(\alpha+1)$; sf; sf.

11. $\bar{x}$; sf; sf.

12. $\frac{1}{n} \sum (x_i - \mu)^2$; sf; no.

13. $x^2/(\alpha+1)$.

14. No.

17. $\bar{x} + 1,645 \sqrt{\sum (x_i - \bar{x})^2/n}; \bar{x} + 1,645 \frac{\Gamma\left(\frac{n-1}{2}\right) \sqrt{\sum (x_i - \bar{x})^2}}{\Gamma\left(\frac{n}{2}\right) \sqrt{2}}$.

19. $2(\bar{x}-1)/\bar{x}$; no.

20. $[3(x_1 + x_2) + \sqrt{9x_1^2 - 14x_1x_2 + 9x_2^2}]/4$; no; $3(x_1 + x_2)/2$.

21. $\hat{a}/3$.

22. $x/(n-x)$, donde x es el número de bolas negras extraídas.

23. $1/\bar{x}$. 24. a) $6\hat{\mu} + 4\hat{\sigma}^2$; b) $\hat{\mu}^2 - \sigma^2/n - 5\hat{\sigma}^2$.
25. a) $3\bar{x}$; b) $5\bar{x} - 1$. 26. 0,9; 0,06; 0,03. 27. 0,96; 0,93.
28. $\bar{x}_1 - \bar{x}_2$; n_1 es el entero más próximo a $n\sigma_1/(\sigma_1 + \sigma_2)$.
29. $\hat{a} = (\bar{x}_1 + \bar{x}_2 + \bar{x}_3 + \bar{x}_4)/4$;
 $\hat{b} = (\bar{x}_1 + \bar{x}_2 - \bar{x}_3 - \bar{x}_4)/4$;
 $\hat{c} = (\bar{x}_1 - \bar{x}_2 + \bar{x}_3 - \bar{x}_4)/4$; $\hat{\sigma}^2 = \sum_{i,j} (x_{ij} - \hat{\mu}_i)^2/4n$, donde
 $\hat{\mu}_1 = \hat{a} + \hat{b} + \hat{c}$, $\hat{\mu}_2 = \hat{a} + \hat{b} - \hat{c}$, etc.
30. No; $\sum_i \left(\frac{x_i}{\sigma_i^2} \right) / \sum_i \left(\frac{1}{\sigma_i^2} \right)$. 31. No. 32. 130.
33. $\frac{1}{150\pi} \exp \left[-\frac{1}{450} (9x^2 + 10\mu^2 - 18x\mu - 200\mu + 10\,000) \right]$
 $-\infty < x < \infty$; $-\infty < \mu < \infty$. $\frac{1}{10\sqrt{5}} \exp \left[-\frac{1}{500} (x - 100)^2 \right]$
 $-\infty < x < \infty$.
35. 127. 36. 1/2.
37. $q(x, \theta) = 6\theta^{x+1}(1-\theta)^{2-x}$ $x=0, 1$; $0 \leq \theta \leq 1$;
 $k(x) = \frac{\Gamma(x+2)\Gamma(3-x)}{4}$ $x=0, 1$.
38. $\frac{24\theta^{x+1}(1-\theta)^{2-x}}{\Gamma(x+2)\Gamma(3-x)}$ $x=0, 1$; $0 \leq \theta \leq 1$.
39. $\frac{2}{\Gamma(x+2)} \left[\theta^2 \Gamma(x+2) - \frac{2}{5} \theta \Gamma(x+3) + \frac{1}{30} \Gamma(x+4) \right]$,
 $x=0, 1$; $0 \leq \theta \leq 1$.
40. $\frac{\Gamma(x+3)}{5\Gamma(x+2)} = \frac{x+2}{5}$ 41. 1/2.
43. $\frac{\lambda \sum x_i e^{-(n+1)\lambda} (n+1) \sum x_i + 1}{\Gamma(\sum x_i + 1)}$, $x_i=0, 1, \dots$; $i=1, 2, \dots, n$; $0 < \lambda < \infty$.
44. $(\sum x_i + 1)/(n+1)$.

CAPITULO 9

3. $\mu_x = -1$; $\mu_y = 2$; $V = \begin{pmatrix} 3/5 & 3/5 \\ 3/5 & 18/5 \end{pmatrix}$; $\rho = \sqrt{6}/6$.

4. $e^{-t_1 + 2t_2 + 1/10[3t_1^2 + 6t_1t_2 + 18t_2^2]}$.

5. $e^{-(t_1^2\sigma_x^2 + 2\rho t_1t_2\sigma_x\sigma_y + t_2^2\sigma_y^2)}$.

6. $\begin{pmatrix} 2/5 & -1/5 & 0 & 0 \\ -1/5 & 3/5 & 0 & 0 \\ 0 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 1/4 \end{pmatrix}$.

7. $V = \begin{pmatrix} 2/3 & 1/3 & 1/6 \\ 1/3 & 7/6 & 1/12 \\ 1/6 & 1/12 & 7/24 \end{pmatrix}$.

8. $n(y_2; 0; 7/6)$.

9. $n\left(\frac{2y_3 + y_2}{4}; \frac{1}{2}\right)$.

11. a) $1/5$.

b) 0 .

c) 0 .

d) 2 .

e) 0 .

f) $(6, 2y_4 - 7, 0)'$.

13. $R = \begin{pmatrix} \frac{1}{\sigma_x^2(1-\rho^2)} & \frac{-\rho}{\sigma_x\sigma_y(1-\rho^2)} \\ \frac{-\rho}{\sigma_x\sigma_y(1-\rho^2)} & \frac{1}{\sigma_y^2(1-\rho^2)} \end{pmatrix}$

15. $\hat{\mu}_x = \frac{1}{n} \sum_{i=1}^n x_i$; $\hat{\mu}_y = \frac{1}{n} \sum_{i=1}^n y_i$; $\sigma_x^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$;

$$\sigma_y^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2;$$

$$\rho = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$$

16. $\hat{\mu}_{y_2|y_1} = -0,06 + 2,40y_1$.

17. $E(x) = \frac{1}{n} \sum_{i=1}^k \mu_i$; $\text{var } x = \frac{1}{n^2} \sum_{i,j} \sigma_{ij}$.

19. $\sum_{ij} \sigma_{ij} \alpha_i b_j / \sqrt{\left(\sum_{i,j} \alpha_i \alpha_j \sigma_{ij}\right) \left(\sum_{i,j} b_i b_j \sigma_{ij}\right)}$

26. $k\rho/(k\rho + 1 - \rho)$.

CAPITULO 10

2. $2/9\sqrt{u}$, $0 < u < 1$.
 $(1 + \sqrt{u})/9\sqrt{u}$, $1 < u < 4$.
3. $4\bar{x}$, $0 < \bar{x} < 1/2$.
 $4(1 - \bar{x})$, $1/2 < \bar{x} < 1$.
5. $n(\bar{x}; -2, 4/n)$.
8. k ; $2k$.
9. $n/(n-2)$; $2n^2(n+m-2)/m(n-4)(n-2)^2$.
10. 6,94.
15. $3u^2/(1+u^4)$, $u > 0$.
16. $-\log u$, $0 < u < 1$.
18. Ji cuadrado, Cauchy.
20. Bivariante normal con
vector media = $\begin{pmatrix} k\mu \\ (n-r+1)\mu \end{pmatrix}$
y

$$V = \begin{pmatrix} k\sigma^2 & \sigma^2(k-r+1) \\ \sigma^2(k-r+1) & (n-r+1)\sigma^2 \end{pmatrix}$$
23. $x^2 + y^2 - 2xy = 593$.
24. $(n-1)\sigma^2/n$; $2(n-1)\sigma^4/n^2$.
26. Las $x_i = \begin{pmatrix} y_{1i} \\ y_{2i} \end{pmatrix}$ tienen distribuciones independientes y normales (μ, V) , con $\mu' = (\mu_1, \mu_2)$.
Si $a_i = (a_{1i}, a_{2i})$, $i = 1, 2, \dots, k$, $u = \sum_{i=1}^k a_i' x_i$ está distribuida normalmente, con $E(u) = \sum_{i=1}^k a_i' \mu$ y $\text{var } u = \sum_i a_i' V a_i$.
27. $(\sqrt{n(\alpha+1)}/\beta\sqrt{2\pi}) \exp[-n(\alpha+1)(\hat{\beta}-\beta)^2/2\beta^2]$, $\hat{\beta} > 0$.
28. $\sqrt{n/2\pi\mu} \exp[-n(\hat{\mu}-\mu)^2/2\mu]$, $\hat{\mu} > 0$.
29.
$$\exp \left[-1/2 \sum_1^{k-1} \sum_1^{k-1} \frac{1}{n} \left(\frac{\delta_{ij}}{p_i} + \frac{1}{p_k} \right) (n_i - np_i)(n_j - np_j) \right] /$$

$$\sqrt{(2\pi n)^{k-1} p_1 p_2 \dots p_k}; \quad n_i > 0, \quad \sigma_{ii} = np_i(1-p_i); \quad \sigma_{ij} = -np_i p_j.$$
31. $\prod_{i=1}^k p_i$.

$$33. \left(\frac{(n_1-1)(n_2-1)\sigma_3^4}{(n_3-1)^2\sigma_1^2\sigma_2^2} \right) \left(\frac{n_1+n_2+n_3-5}{2} \right) ! y^{(n_1-3)/2} z^{(n_2-3)/2}$$

$$\prod_{i=1}^3 \left(\frac{n_i-3}{2} \right) ! (1+y+z)^{(n_1+n_2+n_3-3)/2}$$

$$y > 0, z > 0, \text{ donde } y = (n_1-1)\sigma_3^2 u / (n_3-1)\sigma_1^2$$

$$z = (n_2-1)\sigma_3^2 v / (n_3-1)\sigma_2^2.$$

$$34. \frac{\left(\frac{(n_1-1)\sigma_2^2}{(n_2-1)^{3/2}\sigma_1^2} \right) \left(\frac{n_1+n_2-3}{2} \right) ! y^{(n_1-3)/2}}{\sqrt{\pi} \prod_{i=1}^2 \left(\frac{n_i-3}{2} \right) ! (1+y+z^2)^{(n_1+n_2-1)/2}}, \quad y > 0, -\infty < z < \infty.$$

$$\text{donde } y = (n_1-1)\sigma_2^2 v / (n_2-1)\sigma_1^2$$

$$z = u / \sqrt{n_1-1}.$$

$$35. \frac{e^{-R^2/4}}{\sqrt{\pi}} \quad 0 < R < \infty, \quad R = y_2 - y_1.$$

$$36. a) 3e^{-3y_1} \quad 0 < y_1 < \infty. \quad b) 1/3.$$

$$37. \frac{2^{3/2} \sqrt{m} \Gamma\left(\frac{m+1}{2}\right)}{\Gamma\left(\frac{m}{2}\right) \pi}$$

$$38. 3/4.$$

$$39. 2e^{-w}, \quad w = y_1 + y_2, \quad w > z > 0.$$

$$40. 1 - (1/2)^n.$$

$$41. ne^{-ny_1}, \quad y_1 > 0.$$

$$42. 1/(n+1).$$

CAPITULO 11

$$1. -2,09 < \mu < 2,84; \quad -1,90 < \mu < 2,65.$$

$$2. 468,3 < \mu < 647,7; \quad 439,15.$$

$$3. 2200 < \sigma^2 < 29\,367. \quad 49,90 < \sigma < 171,4.$$

$$5. 22,4 < \sigma^2 < 91,4.$$

$$6. x' < \theta < x'/(1-\gamma)^{1/n}.$$

$$7. 2,9 < \theta < 6,13.$$

$$8. -1,46 < \mu_1 - \mu_2 < 10,04.$$

$$9. 4,3 < \theta < 10,8; \text{ en este caso no se ha obtenido la longitud mínima.}$$

$$11. 0,18 < \alpha < 1,02.$$

12. $0,98 < \mu = 1/\alpha < 5,56$; $0,96 < \sigma^2 = 1/\alpha^2 < 30,8$; 90.
13. $0,017 < p < 0,865$. 14. $0,397 < p < 0,478$; $0,374 < p < 0,501$.
15. $\{(p_1, p_2): 3890(p_1 - 0,6)^2 + 6680(p_1 - 0,6)(p_2 - 0,3) + 4460(p_2 - 0,3)^2 < 1\}$.
20. 3,82. 21. $\text{Var } \hat{\mu} = \sigma^2/n$; $\text{var } \hat{\sigma} = \sigma^2/2n$.
22. $\text{Var } (\hat{\mu} + k\hat{\sigma}) = \frac{\sigma^2}{2n} [k^2 + 2]$; $3,67 < \alpha < 3,97$.
23. Utilícense los teoremas 10-3 a y 10-4.
26. $P[\chi^2(19) < 95/t^2_{\alpha/2}]$, suponiendo un intervalo confidencial del $(1-\alpha)100\%$.
27. Distribución t : $\text{intervalo} = \frac{\sqrt{2} \Gamma\left(\frac{n}{2}\right) \sigma}{\Gamma\left(\frac{n-1}{2}\right) \sqrt{n-1}}$;
 Distribución normal: $\text{intervalo} = \frac{3,92 \sigma}{\sqrt{n}}$
29. 310.
33. $-0,607 < \mu_2 - \mu_1 < 2,647$
 $-0,447 < \mu_3 - \mu_1 < 2,807$
 $-1,467 < \mu_3 - \mu_2 < 1,787$
 $-2,647 < \mu_1 - \mu_2 < 0,607$
 $-2,807 < \mu_1 - \mu_3 < 0,447$
 $-1,787 < \mu_2 - \mu_3 < 1,467$.

CAPITULO 12

2. $A = \{a: a = a_1, a_2\}$, donde a_1 es la decisión «se acepta H_1 , $\mu = 6$ », y a_2 es la decisión «se acepta H_2 , $\mu = 7$ ».
2. $\Omega = \{\mu: -\infty < \mu < \infty\}$.
3. Adóptese la acción a_1 si $\bar{x} < 2\pi$; adóptese la acción a_2 si $\bar{x} \geq 2\pi$.
4. Sí; no. 5. 0,079; 0,5.

		Estado de la naturaleza	
		$\mu = 6$	$\mu = 7$
Acción:	a_1	0,921	0,50
	a_2	0,079	0,50

7. $\delta_1(d; \theta) = 0,079$; $\delta_2(d; \theta) = 0,5$.

8. $R(d; \mu=6)$ $R(d; \mu=7)$ 9. d_1, d_2, d_3 y d_4 dominan a d_5 .

d_1	1	0,16
d_2	0,617	0,309
d_3	0,32	0,5
d_4	0,134	0,692
d_5	1,38	0,692

10. $R(d_c; 6) = 2P(x \geq c | \mu=6) = 2\{1 - N(c-6)\}$.
 $R(d_c; 7) = P(x < c | \mu=7) = N(c-7)$.

11. excelente bueno mediano

d_1	a_1	a_1	a_1
d_2	a_1	a_2	a_2
d_3	a_1	a_2	a_1
d_4	a_1	a_1	a_2
d_5	a_2	a_2	a_1
d_6	a_2	a_1	a_2
d_7	a_2	a_1	a_1
d_8	a_2	a_2	a_2

12. $R(d_1; \theta_1)=0$; $R(d_1; \theta_2)=5$.
 $R(d_2; \theta_1)=3/2$; $R(d_2; \theta_2)=5/9$.
 $R(d_3; \theta_1)=1$; $R(d_3; \theta_2)=5/2$.
 $R(d_4; \theta_1)=1/2$; $R(d_4; \theta_2)=55/18$.
 $R(d_5; \theta_1)=5/2$; $R(d_5; \theta_2)=35/18$.
 $R(d_6; \theta_1)=2$; $R(d_6; \theta_2)=5/2$.
 $R(d_7; \theta_1)=3/2$; $R(d_7; \theta_2)=40/9$.
 $R(d_8; \theta_1)=3$; $R(d_8; \theta_2)=0$.
13. $R(d_e; \theta_1)=1$; $R(d_e; \theta_2)=65/36$.
14. $R(d_p; \theta_1)=(2p+1)/2$; $R(d_p; \theta_2)=(55-45p)/18$.
15. $R(d_1; \mu < 0) = \mu^2[1 - N(2-\mu)]$, $\mu < 0$.
 $R(d_1; \mu \geq 0) = \mu^2 N(2-\mu)$, $\mu \geq 0$.
 $R(d_2; \mu < 0) = \mu^2[1 - N(3-\mu)]$, $\mu < 0$.
 $R(d_2; \mu \geq 0) = \mu^2 N(3-\mu)$, $\mu \geq 0$.
17. Acéptese H_0 .

18. Rechácese; $\bar{x}_2 - \bar{x}_1 = 3,25$ es exterior a $(-3,2, 3,2)$.
19. Rechácese; $\bar{x} = 1267$ cae fuera de $(1253,1, 1266,9)$.
21. Por debajo del hiperplano $n\bar{x} = [20n + 2\sigma^2 \log k]/4$.
24. Acéptese H_0 ($t = 1,58$; 8 g. de l.).
25. No ($\chi^2 = 10$, 3 g. de l.).
26. No puede rechazarse ($F = 1,85$).
29. Rechácese H_0 ($\chi^2 = 9,56$, 2 g. de l.).
30. Acéptese H_0 ($-2 \log \lambda = 1,3$; 1 g. de l.).
31. No puede rechazarse.
33.
$$\lambda = \frac{[\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2 - (\sum (x_i - \bar{x})(y_i - \bar{y}))^2]^{n/2}}{[\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2]^{n/2}}$$
34. Si se verifica H_0 , $\sum_{i=1}^n (x_i - \hat{x})^2 / \sigma_i^2$ tiene una distribución ji cuadrado con $n-1$ grados de libertad, siendo $\hat{x} = \Sigma \left(\frac{x_i}{\sigma_i^2} \right) / \Sigma \left(\frac{1}{\sigma_i^2} \right)$.
35. Sí al nivel de 0,05.
36. Sí.
37. No al nivel de 0,05.
38. Acéptese H_0 ($\chi^2 = 8,96$, $-2 \log \lambda = 8,8$, 5 g. de l.).
39. Sí ($\chi^2 = 1,44$, $-2 \log \lambda = 1,4$, 2 g. de l.).
40. Sí ($-2 \log \lambda = 0,81$, 1 g. de l.).
41. $n_i \cdot n_{.j} / n$; $n_i \cdot n_{.j} (n - n_{.j}) / (n^2 (n - 1))$.
43. La misma dócima.
44. Sí ($\chi^2 = 0,92$; 1 g. de l.).
45. Sí ($\chi^2 = 1,90$; 2 g. de l.).
46. Rechácese H_0 .
47. $\chi^2 = 135$; 4 g. de l.,
48. $\chi^2 = 7,57$; 2 g. de l.
49. El mismo criterio que cuando no se fijan los totales marginales; $(r-1)(s-1)$ grados de libertad.
50. Resuélvase las ecuaciones de máxima verosimilitud (por iteración o aproximaciones sucesivas) respecto a p y q , calculando a continuación χ^2 ó $-2 \log \lambda$; 1 g. de l.
51.
$$\lambda = \left(\prod_1^r n_{i..}^{n_{i..}} \right) \left(\prod_1^s n_{.j.}^{n_{.j.}} \right) \left(\prod_1^t n_{..k}^{n_{..k}} \right) / n^{2n} \prod_{i,j,k} n_{ijk}^{n_{ijk}};$$

 $rst - r - s - t + 2$ g. de l.

52. $\chi^2=970$ (4 g. de l.), $\chi^2=548$ (3 g. de l.).
53. $\lambda = \left(\prod_1^r n_{i..}^{n_{i..}} \right) \left(\prod_{j,k=1}^{s,t} n_{.jk}^{n_{.jk}} \right) / n^n \prod_{i,j,k} n_{ijk}^{n_{ijk}}$ con $(r-1)(st-1)$ g. de l., igual que cuando no se fijan los totales.
54. La misma que en el problema 41. 55. $2/33$.

CAPITULO 13

1. $X'X = \begin{pmatrix} 8 & -1,3 \\ -1,3 & 13,95 \end{pmatrix}$; $X'Y = (5,1, 43,6)'$; $\beta' = (1,16, 3,23)$.
2. $1,90 < \beta_2 < 4,56$.
3. $\frac{\sigma^2}{\sum (x_i - \bar{x})^2} \left[\frac{\sum x_i^2}{n} + x_0^2 - 2x_0\bar{x} \right]$.
4. Normal con media $\alpha + \beta x_0$; la varianza es la misma que en el problema 3.
6. $2,07 < \alpha + \beta < 6,71$.
7. Rechácese; véase el intervalo confidencial en el problema 2.
11. $\beta_0=14,71$; $\beta_1=-1,20$; $\beta_2=-0,46$; $1,2 < \sigma^2 < 11,7$.
12. $-2,05 < \beta_1 < -0,35$.
13. No puede rechazarse H_0 : $t=-1,0$, 7 g. de l.
14. Rechácese; $t=-3,6$.
15. $\beta_0=12,1$; $\beta_1=-1,06$; $\beta_2=0,02$; acéptese, $t=0,013$.
17. $\beta_0=7,9$; $\beta_1=-0,88$; $\beta_2=0,02$.
18. $E(y_1|y_2) = \mu_1 + \frac{\rho\sigma_{y_1}}{\sigma_{y_2}} (y_2 - \mu_2)$;
 $E(y_2|y_1) = \mu_2 + \frac{\rho\sigma_{y_2}}{\sigma_{y_1}} (y_1 - \mu_1)$; sí.
19. $E(y_1|y_2, y_3) = \mu_1 - \frac{r_{12}}{r_{11}} (y_2 - \mu_2) - \frac{r_{13}}{r_{11}} (y_3 - \mu_3)$.
 r_{ij} es el elemento ij -ésimo de $R = V^{-1}$.
22. Rechácese; $t=5,96$, 6 g. de l. 23. $0,63 < \rho < 0,986$.
25. $1,27 \sigma$.

26. a) $\gamma_1' = (\beta_1, \beta_2)$; $X_1' = \begin{pmatrix} 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 7 & 4 & 44 & 6 & 4 & 2 & 1 & 1 & 1 & 0 \end{pmatrix}$.

b) $\gamma_2' = (\beta_0)$; $X_2' = (1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1)$.

27. $10\hat{\beta}_0 + 45\hat{\beta}_1 + 70\hat{\beta}_2 = 79$;

$45\hat{\beta}_0 + 285\hat{\beta}_1 + 157\hat{\beta}_2 = 283$; $\beta'X'Y = 687,76$;

$70\hat{\beta}_0 + 157\hat{\beta}_1 + 2060\hat{\beta}_2 = 691,7$.

28. $10\tilde{\beta}_0 = 79$; $\tilde{\gamma}_2'X_2'Y = 624,1$.

30. La solución no es única; cualquier recta que determine cuatro segmentos verticales de longitud 2, centrados en los cuatro puntos del plano x - y es una estimación máximo-verosímil. La recta obtenida por mínimos cuadrados es $y = 0,85 - 0,73x$.

31. $E(y|x) = (n-x)p^*$; $p^* = (1-p-q)/(1-q)$.

CAPITULO 14

1. a) $X = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}$; b) $X'X = \begin{pmatrix} 4 & 2 & 1 & 1 \\ 2 & 2 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix} \beta = \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix}$

d) $4\hat{\mu} + 2\hat{\alpha}_1 + \hat{\alpha}_2 + \hat{\alpha}_3 = Y_{..}$

e) 3;

f) $Y_{1.}/2 - 2\gamma_{21}$.

$2\hat{\mu} + 2\hat{\alpha}_1 = Y_{1.}$

$\hat{\mu} + \hat{\alpha}_2 = \gamma_{21}$

$\hat{\mu} + \hat{\alpha}_3 = \gamma_{31}$

2. b) $\mu: 6\hat{\mu} + 2(\hat{\tau}_1 + \hat{\tau}_2 + \hat{\tau}_3) = 314$

$\tau_1: 2\hat{\mu} + 2\hat{\tau}_1 = 103$

$\tau_2: 2\hat{\mu} + 2\hat{\tau}_2 = 89$

$\tau_3: 2\hat{\mu} + 2\hat{\tau}_3 = 122$.

c) $\beta' = (y_{..}, y_{1.} - y_{..}, y_{2.} - y_{..}, y_{3.} - y_{..})$

$= (52,33, -0,83, -7,83, 8,67)$.

$\hat{\sigma}^2 = 123$.

e)

Fuente	Grados de libertad	Suma de cuadrados	Cuadrado medio
Total	6	17 076,00	
Media	1	16 432,67	
Tratamientos	2	274,33	137,16
Error	3	369,00	123,00

g) No puede rechazarse; $F=1,12$.

h) Sí; la estimación es $y_1/2 + y_2/3 + y_3/6$.

3. $\sigma^2=123$.

4. $-28,29 < \tau_1 - \tau_2 < 42,29$.

5. a) $\beta' = (\mu, \alpha_1, \alpha_2, \alpha_3, \alpha_4, \theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6)$.

b) $\mu: 24\hat{\mu} + 6\Sigma\hat{\alpha}_i + 4\Sigma\hat{\theta}_j = 985$

$\alpha_1: 6\hat{\mu} + 6\hat{\alpha}_1 + \Sigma\hat{\theta}_j = 245$

$\alpha_2: 6\hat{\mu} + 6\hat{\alpha}_2 + \Sigma\hat{\theta}_j = 263$

$\alpha_3: 6\hat{\mu} + 6\hat{\alpha}_3 + \Sigma\hat{\theta}_j = 244$

$\alpha_4: 6\hat{\mu} + 6\hat{\alpha}_4 + \Sigma\hat{\theta}_j = 233$

$\theta_1: 4\hat{\mu} + \Sigma\hat{\alpha}_i + 4\hat{\theta}_1 = 79$

$\theta_2: 4\hat{\mu} + \Sigma\hat{\alpha}_i + 4\hat{\theta}_2 = 77$

$\theta_3: 4\hat{\mu} + \Sigma\hat{\alpha}_i + 4\hat{\theta}_3 = 227$

$\theta_4: 4\hat{\mu} + \Sigma\hat{\alpha}_i + 4\hat{\theta}_4 = 230$

$\theta_5: 4\hat{\mu} + \Sigma\hat{\alpha}_i + 4\hat{\theta}_5 = 168$

$\theta_6: 4\hat{\mu} + \Sigma\hat{\alpha}_i + 4\hat{\theta}_6 = 204$

6. $-3,0; 3,32; -13,25$.

7.

Fuente	Grados de libertad	Suma de cuadrados	Cuadrado medio
Total	24	46 803,00	
Media	1	40 426,04	
Clase A (α)	3	77,13	25,71
Clase B (θ)	5	6 183,71	1236,74
Error	15	116,12	7,74

10. $-21,25$.

11. $-2,60 < \alpha_1 + \alpha_2 - 2\alpha_3 < 9,24$.

15. $\mu^*: abc\bar{\mu}^* + ac\Sigma\bar{\theta}_j + ab\Sigma\bar{\tau}_k = Y_{...}$

$\theta_q: ac\bar{\mu}^* + ac\bar{\theta}_q + a\Sigma\bar{\tau}_k = Y_{.q} \quad q=1, 2, \dots, b$

$\tau_r: ab\bar{\mu}^* + a\Sigma\bar{\theta}_j + ab\bar{\tau}_r = Y_{..r} \quad r=1, 2, \dots, c$

16. $\frac{Y^2}{abc} + \sum_{ijk} (\bar{y}_{.j} - \bar{y}_{...})^2 + \sum_{ijk} (\bar{y}_{..k} - \bar{y}_{...})^2$

$$\sum_{ijk} (\bar{y}_{.j} - \bar{y}_{...})^2 / (b-1)$$

19. Si se verifica H_0 , $u =$ _____ se distri-

$$\sum_{ijk} (y_{ijk} - \bar{y}_{i..} - \bar{y}_{.j.} - \bar{y}_{..k} + 2\bar{y}_{...})^2 / T$$

buye como F con $b-1$ y T grados de libertad, siendo

$$T = abc - a - b - c + 2.$$

20. $\mu: ab\hat{\mu} + b\hat{\alpha} + a\hat{\theta} + (\hat{\alpha\theta})_{..} = Y_{..}$
 $\alpha_p: b\hat{\mu} + b\hat{\alpha}_p + \hat{\theta} + (\hat{\alpha\theta})_{p.} = Y_{p.} \quad p=1, 2, \dots, a$
 $\theta_q: a\hat{\mu} + \hat{\alpha} + a\hat{\theta}_q + (\hat{\alpha\theta})_{.q} = Y_{.q} \quad q=1, 2, \dots, b$
 $(\alpha\theta)_{pq}: \hat{\mu} + \hat{\alpha}_p + \hat{\theta}_q + (\hat{\alpha\theta})_{pq} = Y_{pq} \quad \begin{matrix} p=1, 2, \dots, a \\ q=1, 2, \dots, b. \end{matrix}$

CAPITULO 15

3. $L\left(\frac{M}{N}\right) = \sum_{x=0}^c \binom{M}{x} \binom{N-M}{n-x} / \binom{N}{n}.$
12. $P(\theta) = (1-B^h)/(A^h-B^h)$, siendo h una raíz no nula de
 $\theta \left(\frac{\theta_1}{\theta_0}\right)^h - h(\theta_1 - \theta_0) = \theta.$
14. $pL(p).$
15. $100 \left[\frac{E(N)}{N} L(p) + 1 - L(p) \right]$, siendo N el tamaño del lote.

CAPITULO 16

1. $r \binom{n}{r} u^{r-1} (1-u)^r, \quad 0 < u < 1.$
4. $\frac{n! u^{r-q-1} v^{t-s-1} (1-u-v)^{n+q+s-r-t}}{(r-q-1)! (t-s-1)! (n+q+s-r-t)!} \quad 0 < u < 1, \quad 0 < v < 1.$
12. $(n-2)u^{n-3}, \quad 0 < u < 1.$
13. $\hat{\theta} = \sum_{i=1}^r \frac{y_i}{r} + \frac{(n-r)y_1}{r}$ 14. No.
15. $\frac{(2k+1)!}{k!^2} \left[\frac{1}{4} - (\bar{x} - \theta)^2 \right]^k \quad \theta - 1/2 < \bar{x} < \theta + 1/2.$
16. $(2k+1) 2^{2k} (z - \theta + 1/2)^{2k}$ para $\theta - 1/2 < z < \theta$;
 $(2k+1) 2^{2k} (z - \theta - 1/2)^{2k}$ para $\theta < z < \theta + 1/2$;
 $\text{var } z < \text{var } \bar{x}$ para cualquier $k.$
18. Sí.
19. $\bar{x}$ para n impar; cualquier valor comprendido entre la pareja central de observaciones si n es par.
23. 392. 24. Rechácese. 25. Rechácese.

INDICE ALFABETICO

INDICE ALFABETICO

- Acumulativas, distribuciones, 66, 99.
 Aleatorias, variables, 52, 53, 253.
 Aleatorio, muestreo, 163.
 Aleatorización, dócima de, 489.
 Algebra de matrices, 234.
 Análisis de la varianza:
 en regresión lineal, 406.
 experimentación sobre el:
 doble, 433, 437.
 simple, 431, 433.
 triple, 438.
 interacción en el, 438.
 modelos de diseño experimental
 en el, 425.
- Bayes:
 estimadores de, 215, 221.
 estrategia de, 328.
- Beta, distribución, 148.
- Binomial:
 distribución, 75.
 aproximación normal a la, 176.
 con parámetro p , intervalos con-
 fidenciales de la, 302.
 negativa, 83.
 teorema, 30.
- Bondad del ajuste, dócima de la,
 356.
- Calidad media resultante, 465.
- Característica operante, curva, 457.
- Cauchy, distribución de, 151, 156.
- Clasificación:
 doble, 438, 440.
 triple, 438, 441.
- Coeficiente confidencial experimen-
 tativo, 309.
- Combinaciones, 23.
- Compuestas, hipótesis, 335.
- Condicional:
 distribución, 69, 106.
 continua, 106.
 discreta, 70.
- normal:
 bivariante, 233.
 multivariante, 245.
- esperanza, 134.
 probabilidad, 40, 47, 48.
- Confidencial:
 experimental, coeficiente, 309.
 intervalo (véase *Intervalos confi-
 denciales*).
 región (véase *Regiones confiden-
 ciales*).
- Conjunta, distribución, 53, 65, 97.
- Conjuntos convexos, 329.
- Consistencia, 199.
- Continuas, distribuciones, 88.
- Correlación, 250, 412.
 múltiple, 252.
 parcial, 246.
 por rangos de Spearman, 491.
- Cuadráticas, formas, 235, 239.
- Cuadrático medio, error, 192, 202.
- Cumulantes, 157.
- Curva característica operante, 457.
- Deductiva, inferencia, 160.
- Deformación, 126.
- Densidad:
 completa, 151, 158, 205.
 funciones de, 53, 65, 92.
- Desviación estándar o típica, 124.
- Diferencias entre medias, 350.
- Discretas, distribuciones, 61.
- Discriminación, 389.
- Diseño experimental, modelos de,
 415.
- Distribución:
 acumulativa, 66, 99.
 beta, 148.
 binomial (véase *Binomial, distribu-
 ción*).
 condicional (véase *Condicional,
 distribución*).
 conjunta, 53, 65, 97.

- continua, 88.
 - de Cauchy, 151, 156.
 - de estadísticos ordinales, 276.
 - de la razón de varianza, 265.
 - de Poisson, 81, 157, 268.
 - deducida de otra, 110, 254.
 - discreta, 61.
 - F*, 265.
 - gamma, 145.
 - hipergeométrica, 83, 127.
 - ji cuadrado, 156, 259, 267, 292, 347, 354, 384.
 - logarítmico normal, 151.
 - marginal, 70, 104.
 - multivariante, 66, 96, 119, 133.
 - normal (véase *Normal, distribución*).
 - polinomial, 80, 273.
 - «*t*» de Student, 267.
 - uniforme, 83, 140, 212.
- Distribuciones en el muestreo, 253.
- de estadísticos ordinales, 276.
 - de estimadores máximo-verosímiles, 270.
 - de la diferencia entre dos medias, 350.
 - de la media:
 - en muestras grandes, 171.
 - en poblaciones:
 - binomiales, 268, 269.
 - de Poisson, 269.
 - normales, 261.
 - de la razón de varianza, 265.
 - de los coeficientes de regresión, 383, 384.
- Dóclimas:
- insesgadas, 343.
 - sucesionales, 444.
 - funciones de potencia de, 449.
 - identidad fundamental de las, 465.
 - para la media de una población normal, 448, 453, 461.
 - para una distribución binomial, 459.
 - tamaño muestral medio en las, 453.
 - uniformemente más potentes, 340.
- Docimasia de hipótesis, 317, 336.
- a libre distribución, 461.
 - de bondad del ajuste, 356.
 - de igualdad:
 - de dos medias, 350.
 - de dos varianzas, 355.
 - de independencia en tablas de contingencia, 359.
 - en muestras grandes, 347.
 - nulas, 343.
 - potencia de la, 337.
 - relativas a:
 - la media de una población normal, 344, 347.
 - la razón de verosimilitud, 329, 343.
 - la varianza de una distribución normal, 354.
 - sucesionales (véase *Dóclimas sucesionales*).
- Ecuaciones normales, 381, 399, 418, 423.
- Eficiencia, 200.
- relativa, 192.
- Error:
- cuadrático medio, 192, 202.
 - de tipo:
 - I, 321.
 - II, 321.
- Espacio:
- de acción, 185.
 - paramétrico, 185, 318.
- Estadísticos, 166.
- ordinales, 276.
 - suficientes, 193, 196, 205, 366.
- Estimable, función, 419.
- Estimadores:
- consistentes, 193, 199, 221.
 - de Bayes, 215, 221.
 - eficientes, 192, 193, 200, 221.
 - en una distribución normal multivariante, 250.
 - insesgados, 193, 198, 202.
 - de varianza mínima, 202, 205.
 - máximo-verosímiles (véase *Máximo-verosímiles, estimadores*).
 - OAN, 201, 221.

- obtenidos por el método de los momentos, 214.
Estrategia admisible, 328.
- F*, distribución, 265.
Formas cuadráticas, 235, 239.
Frecuencial, probabilidad, 13.
Funciones:
 de cuantía (véase *Distribución*).
 de decisión, 186.
 de distribución (véase *Distribución*).
 de pérdida, 187, 189, 190, 318, 319.
 medidas por el error cuadrático, 199.
 de potencia, 322.
 de dótimas sucesionales, 449.
 de regresión, 233, 379.
 de riesgo, 188, 190.
 de verosimilitud, 208, 329.
 estimables, 419.
 generatrices, 33.
 combinatorias, 33.
 de momentos (véase *Momentos*, *función generatriz de*).
- Gamma, distribución, 145.
Gauss-Markoff, teorema de, 392.
Generatrices, funciones, 33.
Grados de libertad, 260.
- Herencia mendeliana, 59.
Hipergeométrica, distribución, 83, 127.
Hipótesis:
 compuestas, 335.
 lineales, 378, 379, 384, 425.
 nulas, 343.
 simples, 324.
- Igualdad de dos varianzas, dócima de la, 355.
Independencia:
 de la media y varianza muestrales, 261.
 en sentido probabilístico, 51, 69, 107, 108.
- en tablas de contingencia, 359.
funcional, 69.
- Inferencia:
 deductiva, 160.
 inductiva, 160.
- Inesgadas, dótimas, 343.
Inesgados, estimadores, 193, 198, 202, 205.
Inspección por muestreo, 456-61.
- Interacción:
 en el análisis de la varianza, 438.
 en tablas de contingencia, 361.
- Intercuartílico, recorrido, 471.
Intervalo de predicción, 386.
Intervalos confidenciales, 285.
 método general para la obtención de, 295.
 múltiples, 307.
 para coeficientes de regresión, 384, 385.
 para el parámetro de una distribución binomial, 299.
 para la media de una distribución normal, 289.
 para la mediana, 472.
 para la varianza de una distribución normal, 291.
 para muestras grandes, 301.
 para poblaciones rectangulares, 312 313.
- Ji cuadrado:
 criterio, de Pearson, 358.
 distribución, 156, 259, 267, 292, 347, 354, 384.
 dócima:
 de la bondad del ajuste, 356.
 en tablas de contingencia, 358, 359.
- Ley de los grandes números, 169.
Libre distribución, métodos a, 467.
Límites de tolerancia, 482.
- Lineal:
 hipótesis, 378, 379, 384, 425.
 regresión, 378, 394.
- Logarítmico normal, distribución, 151.

- Marginal:
distribución:
continua, 104.
discreta, 70.
para la distribución normal multivariante, 232, 239.
probabilidad, 37.
- Matrices:
álgebra de, 234.
de varianzas y covarianzas, 242, 243.
inversas de las, 249.
- Máxima verosimilitud, principio de, 206.
- Máximo-verosímiles, estimadores, 206, 208, 210, 212.
distribución de los, en muestras grandes, 270.
propiedades de los, 213.
- Media:
de una población normal, dócima sucesional para la, 448, 453, 461.
distribución de la, 259.
dócimas relativas a la, 338, 349, 350.
intervalos confidenciales para la, 289.
muestral, 259.
varianza de la, 169.
- Mediana, 123.
- Mínimos cuadrados, 391, 424.
- Mixtos, momentos, 133, 231.
- Moda, 123.
- Modelo de clasificación simple, 431.
- Momentos, 122.
estimadores obtenidos por el método de los, 214.
factoriales, 130.
función generatriz de los, 132.
función generatriz de, 131, 231, 248, 381.
de algunas variantes, 249.
de distribuciones multivariantes, 133, 248.
de la distribución:
de Poisson, 132.
gamma, 148.
ji cuadrado, 260.
normal, 145, 173.
mixtos, 133, 231.
muestrales, 166, 167, 214.
problema de los, 134.
- Muestral:
discreto, espacio, 22.
media, 259.
varianza de la, 169.
- Muestrales, momentos, 166, 167, 214.
- Muestras:
aleatorias, 108, 163.
grandes:
distribución:
de estimadores en, 270.
de la media en, 172.
de la razón de verosimilitud en, 348, 349.
intervalos confidenciales para, 303.
regiones confidenciales para, 303.
- Muestreada, población, 163.
- Muestreo:
distribuciones en el (véase *Distribuciones en el muestreo*).
doble, 458.
inspección por, 456.
simple, 457.
sucesional, 459.
- Múltiple:
correlación, 252.
intervalo confidencial, 307.
- Multivariante, distribución, 66, 96, 119, 133.
- No paramétricos, métodos, 467.
correlación por rangos en los, 491.
dócima:
de aleatorización en los, 489.
de rachas en los, 475.
de rangos en los, 483.
eficiencia asintótica de los, 486.
estadígrafos de orden en los, 468.
estimación puntual en los, 471.
igualdad:
de distribuciones en los, 474, 478, 483.
de medianas en los, 476.
intervalos confidenciales en los, 471, 479.

- mediana en los, 470, 478.
puntos porcentuales en los, 471.
recorrido intercuartílico en los, 471.
- Normal:**
bivariante, distribución, 228, 271.
función generatriz de momentos de la, 231, 263.
distribución, 141, 264, 280, 379, 417.
de la media muestral, 259.
formas:
condicionales de la, 233, 245.
marginales de la, 232, 241.
función generatriz de momentos de la, 145, 173.
funciones de regresión de la, 233.
independencia de la media y varianza muestrales en la, 261, 264.
papel de la, 179.
ecuación, 381, 399, 418, 423.
multivariante, distribución, 234, 238, 263, 397.
estimadores en la, 250.
formas marginales y condicionales de la, 241, 245.
función generatriz de momentos de la, 133, 248.
- Nula, hipótesis, 343.**
- OAN, estimadores, 201, 221.**
- Objetivo, población, 162.**
- Paramétrico, espacio, 185, 318.**
- Parcial, correlación, 246.**
- Particiones, 32.**
- Pearson, criterio ji cuadrado de, 358.**
- Pérdida, función de, 187, 289, 190, 199, 318.**
- Permutaciones, 23.**
- Población:**
finita, muestreo a partir de una, 183.
muestreada, 163.
objetivo, 162.
- Poisson, distribución de, 81, 157, 268.**
- Polinomial:**
distribución, 80, 273.
teorema, 30.
- Polinomios ortogonales, 411.**
- Potencia:**
de una décima, 322, 337.
funciones de, 322.
de décimas sucesionales, 449.
- Predicción, 386.**
- Principio de máxima verosimilitud, 206.**
- Probabilidad:**
condicional, 40, 47, 48.
definición de la, 21.
frecuencial, 13.
función de, 21.
leyes de la, 21, 42, 45.
marginal, 37.
- Rachas, 475.**
- Razón de verosimilitud:**
décima de la, 328.
en muestras grandes, distribución de la, 348, 349.
- Recorrido:**
intercuartílico, 471.
«studentizado», 279.
- Región crítica, 337.**
- Regiones confidenciales:**
para coeficientes de regresión, 385.
para la media y la varianza, 293.
para muestras grandes, 303.
- Regresión, 378.**
coeficiente de, 379.
curva de, 380.
función de, 233, 378.
lineal, 379.
múltiple, 394.
normal, 379, 397.
- Riesgo, 188, 190, 319.**
- Sesgo, 198.**
- Simples, hipótesis, 324.**
- Spearman, correlación por rangos de, 491.**

- Stirling, fórmula de, 29.
Student, distribución de, 267.
«Studentizado», recorrido, 279.
Sucesional, análisis (véase *Dócximas sucesionales*).
Suma de cuadrados, 391, 401.

«t», distribución, 267.
Tablas de contingencia, 359, 360, 367, 479.
 dócximas de independencia en, 359.
 interacción en las, 361.
Tamaño muestral medio, 453.
Tchebysheff, desigualdad de, 170.
Teorema central del límite, 172.
Teoría de la decisión, 185, 318, 328.

Uniforme, distribución, 83, 140, 212.

Valores esperados, 118, 121, 134.
Variante, 54.
Varianza, 123, 126.
 análisis de (véase *Análisis de la varianza*).
 de la media muestral, 169.
 de una distribución normal, 264, 291.
 de una función lineal, 243, 250.
 estimador de, 210.
 mínima, estimadores insesgados de, 202, 205.
Varianzas y covarianzas, matrices de, 242, 243.