

PROBABILIDAD Y APLICACIONES ESTADISTICAS

Paul L. Meyer

Departamento de Matemáticas
Washington State University

Versión en español por

Carlos Prado Campos

Departamento de Estadística
Instituto de Matemáticas
Universidad Católica de Chile

Con la colaboración de

Germán Ardila Cuéllar

Departamento de Matemáticas y Estadística
Universidad Nacional de Colombia



ADDISON-WESLEY IBEROAMERICANA

Argentina • Brasil • Chile • Colombia • Ecuador • España
Estados Unidos • México • Perú • Puerto Rico • Venezuela

Prefacio a la primera edición

Este texto está destinado para un curso de un semestre o para dos cursos trimestrales de introducción a la teoría de la probabilidad y algunas de sus aplicaciones. El prerrequisito es un año de cálculo diferencial e integral. No se supone conocimiento previo de probabilidad o estadística. En la Washington State University, el curso para el cual este texto fue desarrollado ha sido enseñado durante varios años y, principalmente, a estudiantes graduándose en ingeniería o en ciencias naturales. La mayoría de tales estudiantes sólo pueden dedicar un semestre al estudio de esta materia. Sin embargo, como esos estudiantes están familiarizados con el cálculo, pueden empezar dicho estudio más allá del nivel estrictamente elemental.

Muchos temas matemáticos pueden ser presentados a diversos grados de dificultad, y esto es realmente cierto de la probabilidad. En este texto se pretende aprovechar la ventaja que suponen los conocimientos matemáticos del lector, sin sobreponerlos. Se usa en él un lenguaje matemático preciso, pero se tiene cuidado de no llegar a profundizar demasiado en detalles matemáticos innecesarios. Este no es ciertamente un «libro de cocina». Aunque se presentan y exponen varios conceptos de una manera informal, las definiciones y los teoremas son enunciados cuidadosamente. Si no es posible o deseable una demostración detallada de un teorema, al menos se da un bosquejo de las ideas más importantes. Una de las características distintivas de este texto son las «observaciones» que siguen a la mayoría de los teoremas y definiciones; en ellas, el resultado particular o el concepto presentado es examinado desde un punto de vista intuitivo.

Debido a la restricción autoimpuesta de escribir un texto relativamente breve sobre una materia que abarca una extensa área, hubo necesidad de seleccionar en relación con la inclusión o exclusión de ciertos temas. Parece ser que no hay manera obvia de resolver este problema. Ciertamente, yo no sostengo que, para algunos de los temas excluidos, no se podría haber encontrado sitio; ni pretendo que ningún material podía haber sido omitido. Sin embargo, se ha hecho hincapié en gran parte en las nociones fundamentales, presentadas con detalle considerable. Solamente el capítulo 11, sobre confiabilidad, puede ser considerado «artículo de lujo»; pero, aun aquí, creo que las nociones asociadas con problemas de confiabilidad son de interés para muchas personas. Además, los conceptos de confiabilidad son un medio excelente para ilustrar muchas de las ideas presentadas anteriormente en el libro.

Aun si se piensa que la extensión ha sido limitada por el tiempo disponible, se ha logrado una selección amplia y razonable de temas. Una ojeada al Índice general muestra de manera evidente que unas tres cuartas partes del texto trata de temas probabilísticos mientras la cuarta parte está dedicada a una exposición de inferencia estadística. Aunque no hay nada de extraordinario en esta división particular del énfasis entre probabilidad y estadística, creo que un conocimiento profundo de los principios básicos de la probabilidad es imperativo para una comprensión propia de los métodos estadísticos. Idealmente, un curso en probabilidad debería ser seguido por otro en teoría estadística y metodología; sin embargo, como indiqué anteriormente, la mayoría de los estudiantes que toman este curso no tienen tiempo para dos semestres de exposición con estas materias y, por tanto, me sentí obligado

a exponer al menos algunos de los aspectos más importantes en el área general de la inferencia estadística.

El éxito potencial de una presentación particular de la materia no debería ser juzgado solamente en función de las ideas específicas aprendidas y de las técnicas específicas adquiridas; el juicio final debe tener en cuenta también si el estudiante está bien preparado para continuar estudiando el tema bien por sí mismo o por medio de un curso formal adicional. Si se considera que este criterio es importante, se hace evidente que debiera insistirse en los conceptos básicos y en las técnicas fundamentales, relegando al mismo tiempo los métodos y temas muy especializados a un papel secundario. Esto también resultó ser un factor importante en la decisión sobre temas a incluir.

Es difícil exagerar la importancia de la teoría de la probabilidad. El modelo matemático apropiado para el estudio de un gran número de fenómenos observables es probabilístico en vez de determinístico. Además, el tema completo de la inferencia estadística está basado en consideraciones probabilísticas. Las técnicas estadísticas se cuentan entre algunas de las herramientas más importantes de científicos e ingenieros. Para poder utilizar esas técnicas inteligentemente es necesaria una profunda comprensión de los conceptos probabilísticos.

Se espera que, además de familiarizarse con muchos métodos específicos y conceptos, el lector desarrolle cierto criterio: pensar probabilísticamente sustituyendo preguntas tales como: «¿Durante cuánto tiempo funcionará este mecanismo?» por «¿Cuál es la probabilidad de que este mecanismo funcione durante más de cien horas?». En muchas situaciones, la segunda pregunta puede no sólo ser la más atinada sino, de hecho, la única pertinente.

Tradicionalmente, muchos de los conceptos importantes de la probabilidad han sido ilustrados con la ayuda de varios «juegos de azar»: lanzar monedas o dados, sacar cartas de una baraja, hacer girar una ruleta, etc. Aunque yo no he evitado por completo referirme a tales juegos, porque sirven para ilustrar bien nociones básicas, he intentando poner al estudiante con ilustraciones más pertinentes de las aplicaciones de la probabilidad: la emisión de partículas α de una fuente radiactiva, muestreo de lote, la duración de instrumentos electrónicos y los problemas asociados de mecanismos y confiabilidad del sistema, etc.

Estoy reacio a mencionar una de las características más importantes en cualquier texto de matemáticas: los problemas; y, sin embargo, puede valer la pena señalar que el trabajar con problemas debe ser considerado parte integrante del curso. Sólo mediante el acto personal de plantear y resolver los ejercicios puede el estudiante desarrollar una comprensión y apreciación de las ideas, así como familiarizarse con las técnicas pertinentes. Es por eso que más de 330 problemas se han incluido en el texto y las respuestas a más de la mitad de ellos figuran al final del libro.

Este libro ha sido escrito en una manera bastante consecutiva: la comprensión de la mayoría de los capítulos requiere familiaridad con los anteriores; sin embargo, es posible tratar superficialmente con los capítulos 10 y 11 si se está interesado, en particular, en dedicar más tiempo a las aplicaciones estadísticas examinadas en los capítulos 13 a 15.

Como debe suceder a quienquiera que escribe un texto, debo estar agradecido a muchas personas: a mis colegas, por muchas conversaciones estimulantes y útiles; a mis propios profesores, por el conocimiento del tema y su interés en él; a los revisores de las primeras versiones del manuscrito, por sus muchas sugerencias útiles y críticas; a Addison-Wesley Publishing Company, por su gran ayuda y cooperación desde las primeras etapas de este proyecto hasta su finalización; a la señorita Carol Sloan, por ser una mecanógrafa muy eficiente y activa; a D. Van Nostrand, Inc., The Free Press, Inc. y Macmillan Publishing Company, por sus permisos para reproducir las tablas 3, 6 y 1, respectivamente; a McGraw-Hill Book Company, Inc., Oxford University Press, Inc., Pergamon Press, Ltd. y Prentice-Hall, Inc., por sus permisos para incluir ciertos ejemplos en el texto; y, finalmente, a mi esposa,

no sólo por la paciencia mantenida durante el esfuerzo sino también «por dejarme» y llevarse a nuestros dos hijos a visitar a sus abuelos durante dos cruciales meses de verano, durante los cuales pude convertir nuestro hogar en un taller desordenado pero tranquilo, del cual emergió milagrosamente la última, final, versión de este libro.

Pullman, Washington

PAUL L. MEYER

Prefacio a la segunda edición

En vista del considerable número de comentarios favorables que he recibido tanto de estudiantes como de profesores que han utilizado la primera edición de este libro, se han hecho en él relativamente pocos cambios. Durante mi propio repetido uso del texto he encontrado que su organización básica y el nivel general de presentación (p. ej.: la mezcla de argumentos matemáticos rigurosos con presentaciones y ejemplos más informales) son los más apropiados para el tipo de estudiante que toma este curso.

Sin embargo, se han hecho varios cambios y adiciones. En primer lugar se hizo un esfuerzo para eliminar varias erratas de imprenta y otros errores que aparecieron en la primera edición. El autor está muy agradecido a los muchos lectores que, no sólo descubrieron algunos de ellos, sino que se interesaron lo suficiente como para indicármelos.

En segundo lugar se intentó hacer más claras las relaciones entre varias distribuciones de probabilidades, de modo que el estudiante pueda conseguir una comprensión mayor de cómo usar varios modelos probabilísticos para aproximarlos entre sí.

Finalmente, algunos problemas nuevos han sido añadidos a la ya larga lista incluida en la primera edición.

El autor desea agradecer nuevamente a Addison-Wesley Publishing Company su cooperación en todos los aspectos que condujeron a esta nueva edición.

Pullman, Washington

P. L. M.

Índice general

Capítulo 1 Introducción a la probabilidad

1.1	Modelos matemáticos	1
1.2	Introducción a los conjuntos	3
1.3	Ejemplos de experimentos no determinísticos	7
1.4	El espacio muestral	8
1.5	Sucesos	10
1.6	Frecuencia relativa	12
1.7	Noción básicas de probabilidad	13
1.8	Varias observaciones	16
	Problemas	18

Capítulo 2 Espacios muestrales finitos

2.1	El espacio muestral finito	21
2.2	Resultados igualmente probables	22
2.3	Métodos de enumeración	24
	Problemas	31

Capítulo 3 Probabilidad condicional e independencia

3.1	Probabilidad condicional	34
3.2	Teorema de Bayes	40
3.3	Sucesos independientes	42
3.4	Consideraciones esquemáticas; probabilidad condicional e independencia	48
	Problemas	50

Capítulo 4 Variables aleatorias unidimensionales

4.1	Noción general de una variable aleatoria	55
4.2	Variables aleatorias discretas	60
4.3	La distribución binomial	63
4.4	Variables aleatorias continuas	67
4.5	Función de distribución acumulativa	71
4.6	Distribuciones mixtas	75
4.7	Variables aleatorias distribuidas uniformemente	76
4.8	Una observación	77
	Problemas	78

Capítulo 5 Funciones de variables aleatorias	
5.1 Un ejemplo	83
5.2 Sucesos equivalentes	83
5.3 Variables aleatorias discretas	86
5.4 Variables aleatorias continuas	88
Problemas	93
Capítulo 6 Variables aleatorias bidimensionales y de mayor dimensión	
6.1 Variables aleatorias bidimensionales	95
6.2 Distribuciones de probabilidades marginales y condicionales	101
6.3 Variables aleatorias independientes	105
6.4 Funciones de una variable aleatoria	108
6.5 Distribuciones del producto y el cociente de variables aleatorias independientes	112
6.6 Variables aleatorias n-dimensionales	114
Problemas	117
Capítulo 7 Otras características de las variables aleatorias	
7.1 El valor esperado de una variable aleatoria	120
7.2 Esperanza de una función de una variable aleatoria	126
7.3 Variables aleatorias bidimensionales	131
7.4 Propiedades del valor esperado	132
7.5 La varianza de una variable aleatoria	138
7.6 Propiedades de la varianza de una variable aleatoria	140
7.7 Expresiones aproximadas para la esperanza y la varianza	143
7.8 Desigualdad de Chebyshev	146
7.9 El coeficiente de correlación	148
7.10 Esperanza condicional	152
7.11 Regresión del promedio	154
Problemas	158
Capítulo 8 La variable aleatoria de Poisson y otras variables aleatorias	
8.1 La distribución de Poisson	164
8.2 La distribución de Poisson como una aproximación a la distribución binomial	165
8.3 El proceso de Poisson	170
8.4 La distribución geométrica	175
8.5 La distribución de Pascal	178
8.6 Relación entre las distribuciones binomial y de Pascal	179
8.7 La distribución hipergeométrica	180
8.8 La distribución multinomial	181
Problemas	183
Capítulo 9 Algunas variables aleatorias continuas importantes	
9.1 Introducción	187
9.2 La distribución normal	187
9.3 Propiedades de la distribución normal	188
9.4 Tabulación de la distribución normal	191
9.5 La distribución exponencial	195

9.6	Propiedades de la distribución exponencial	196
9.7	La distribución gama	199
9.8	Propiedades de la distribución gama	201
9.9	La distribución de χ^2 -cuadrado	203
9.10	Comparación entre varias distribuciones	205
9.11	La distribución normal bivariada	206
9.12	Distribuciones truncadas	208
	Problemas	212
Capítulo 10 La función generadora de momentos		
10.1	Introducción	217
10.2	La función generadora de momentos	218
10.3	Ejemplos de funciones generadoras de momentos	219
10.4	Propiedades de la función generadora de momentos	221
10.5	Propiedades reproductivas	225
10.6	Sucesiones de variables aleatorias	229
10.7	Nota final	230
	Problemas	230
Capítulo 11 Aplicaciones a la teoría de la confiabilidad		
11.1	Conceptos básicos	233
11.2	La ley normal de fallas	236
11.3	La ley exponencial de fallas	237
11.4	La ley exponencial de fallas y la distribución de Poisson	240
11.5	La ley de fallas de Weisbull	242
11.6	Confiabilidad de los sistemas	243
	Problemas	247
Capítulo 12 Sumas de variables aleatorias		
12.1	Introducción	252
12.2	La ley de los grandes números	252
12.3	Aproximación normal a la distribución binomial	255
12.4	El teorema del límite central	259
12.5	Otras distribuciones aproximadas por la distribución normal: Poisson, Pascal y gama	264
12.6	La distribución de la suma de un número finito de variables aleatorias	265
	Problemas	270
Capítulo 13 Muestras y distribuciones muestrales		
13.1	Introducción	273
13.2	Muestras aleatorias	274
13.3	Estadígrafos	277
13.4	Algunos estadígrafos importantes	277
13.5	La transformación integral	284
	Problemas	287
Capítulo 14 Estimación de los parámetros		
14.1	Introducción	290
14.2	Criterios para estimadores	291
14.3	Algunos ejemplos	294

14.4	Estimadores de máxima verosimilitud	298
14.5	El método de los mínimos cuadrados	307
14.6	El coeficiente de correlación	310
14.7	Intervalos de confianza	311
14.8	La distribución <i>t</i> de Student	313
14.9	Más sobre los intervalos de confianza	315
	Problemas	319
Capítulo 15 Docimasia de hipótesis		
15.1	Introducción	324
15.2	Formulación general: distribución normal con varianza conocida	329
15.3	Ejemplos adicionales	333
15.4	Dócima para la bondad de ajuste	336
	Problemas	343
Referencias		
Apéndice		
Respuestas a problemas seleccionados		
Índice de materias		

Introducción a la probabilidad

1.1 Modelos matemáticos

En este capítulo se tratará el tipo de fenómeno de que nos ocuparemos en este libro. Además, formularemos un modelo matemático que nos servirá para investigar, en forma bastante precisa, este fenómeno.

Al principio es muy importante distinguir entre el fenómeno observable en sí mismo y el modelo matemático para dicho fenómeno. Evidentemente, no influimos en manera alguna sobre lo que observamos; sin embargo, al elegir un modelo, sí podemos aplicar nuestro juicio crítico. Esto ha sido muy bien expresado por el Profesor J. Neyman, quien escribió*:

«Cada vez que utilizamos las matemáticas con el objeto de estudiar fenómenos observables es indispensable empezar por construir un modelo matemático (determinístico o probabilístico) para estos fenómenos. Necesariamente, este modelo debe simplificar las cosas y permitir la omisión de ciertos detalles. El éxito del modelo depende de si los detalles que se omitieron tienen o no importancia en el desarrollo de los fenómenos estudiados. La solución del problema matemático puede ser correcta y aún así estar muy en desacuerdo con los datos observados, debido sencillamente a que no estaba probada la validez de las suposiciones básicas que se hicieron. Corrientemente, es bastante difícil afirmar con certeza si un modelo matemático es adecuado o no, *antes* de obtener algunos datos, mediante la observación. Para verificar la validez del modelo, debemos *deducir* un cierto número de consecuencias del mismo y luego comparar con las observaciones esos resultados *predichos*».

Tendremos presentes las ideas anteriores mientras estemos considerando algunos fenómenos obtenidos en la observación y los modelos apropiados para su descripción. Examinemos primero lo que podría llamarse adecuadamente *un modelo determinístico*. Así designamos al modelo que estipula que las condiciones bajo las cuales se verifica un experimento determinan el resultado del mismo. Por ejemplo, si colocamos una batería en un circuito simple, el modelo matemático que posiblemente describiría el flujo observable de corriente sería $I = E/R$, que es la ley de Ohm. El modelo predice el valor de I tan pronto como se dan

* University of California Publications in Statistics. Vol. I, University of California Press, 1954.

E y *R*. En otras palabras, si se repitiese el experimento anterior cierto número de veces, empleando cada vez el mismo circuito (esto es, manteniendo fijos *E* y *R*) posiblemente hubiéramos esperado observar el mismo valor de *I*. Cualquier desviación que pudiese ocurrir sería tan pequeña que la mayor parte de los objetivos de la descripción anterior (que es el modelo) se cumplirían. La realidad es que la batería, el alambre y el amperímetro utilizados para generar y medir la corriente y nuestra destreza para usar los instrumentos de medida, determinan el resultado de cada repetición. (Hay ciertos factores que pueden muy bien ser distintos de repetición en repetición que, sin embargo, no afectarán el resultado de una manera notable. Por ejemplo, la temperatura y la humedad en el laboratorio, o bien la altura de la persona que lee el amperímetro se puede considerar, con razón, que no tienen influencia en el resultado).

Hay muchos ejemplos de «experimentos» en la naturaleza para los cuales los modelos determinísticos son apropiados. Por ejemplo, las leyes gravitacionales describen muy exactamente lo que sucede a un cuerpo que cae bajo ciertas condiciones. Las leyes de Kepler nos indican el comportamiento de los planetas. En cada caso, el modelo señala que las condiciones en las cuales se verifican ciertos fenómenos determinan el valor de ciertas variables observables: la magnitud de la velocidad, el área recorrida durante cierto período de tiempo, etc. Estas cifras aparecen en muchas de las fórmulas con las cuales estamos familiarizados. Por ejemplo, sabemos que bajo ciertas condiciones la distancia recorrida (verticalmente sobre el suelo) por un objeto está dada por: $s = -16t^2 = v_0 t$, en donde v_0 es la velocidad inicial y t es el tiempo empleado. Lo que queremos destacar no es la forma particular de la ecuación anterior (que es cuadrática), sino el hecho de que hay una relación definida entre t y s , que determina únicamente la cantidad del primer miembro de la ecuación, si se dan las del segundo miembro.

En muchos casos el modelo matemático determinístico descrito anteriormente es suficiente. Sin embargo, hay también muchos fenómenos que necesitan un modelo matemático distinto para su investigación. Esos son los que llamaremos modelos *no determinísticos* o *probabilísticos*. (Otro término muy usado es modelo *estocástico*). Posteriormente, en este capítulo consideraremos en forma muy precisa cómo se pueden describir tales modelos probabilísticos. De momento consideraremos unos pocos ejemplos.

Supongamos que tenemos un pedazo de material radiactivo que emite partículas α . Con la ayuda de un dispositivo para medir, podríamos registrar el número de partículas emitidas durante un determinado intervalo de tiempo. Es evidente que no podemos predecir exactamente el número de partículas emitidas, aunque sepamos la forma exacta, la dimensión, la composición química y la masa del objeto que se considera. Así no parece haber un modelo determinístico razonable que nos indique el número de partículas emitidas, digamos n , como una función de varias características propias de la fuente de radiactividad. En su lugar, debemos considerar un modelo probabilístico.

A manera de otro ejemplo consideremos la siguiente situación meteorológica. Deseamos determinar cuánta lluvia caerá debido a una tormenta que pasa por una zona específica. Los instrumentos para registrar la cantidad de lluvia están

listos. Las observaciones meteorológicas pueden darnos mucha información del frente de mal tiempo que se aproxima: la presión barométrica en diversos puntos, los cambios de presión, la velocidad del viento, el origen y la dirección de la tormenta, y además otros datos tomados a gran altura. Pero esta información tan valiosa como es para predecir de modo muy general la forma de la precipitación (débil, regular, intensa), sencillamente no permite indicar con mucha exactitud cuánta lluvia caerá. De nuevo, estamos considerando un fenómeno que por sí mismo no se presta a un tratamiento determinístico. Un modelo probabilístico describe la situación con mayor exactitud.

En principio podríamos indicar cuánta lluvia cayó, si la teoría se hubiera desarrollado (lo que no se hizo). Por lo tanto, usamos un modelo probabilístico. En el ejemplo relacionado con la desintegración radiactiva, debemos usar un modelo probabilístico aún en principio.

Aun a riesgo de adelantarnos a discutir un concepto que se definirá más tarde, indiquemos simplemente que en un modelo determinístico se supone que el resultado real (sea numérico o de otra especie) está determinado por las condiciones bajo las cuales se efectúa el experimento o procedimiento. En un modelo no determinístico, sin embargo, las condiciones experimentales determinan solamente el comportamiento probabilístico (más específicamente, la distribución probabilística) de los resultados observables.

En otras palabras, en un modelo determinístico utilizamos «consideraciones físicas» para predecir el resultado, mientras que en un modelo probabilístico usamos la misma clase de consideraciones que para especificar una distribución de probabilidades.

1.2 Introducción a los conjuntos

Con el fin de discutir los conceptos básicos del modelo probabilístico que deseamos desarrollar, será muy conveniente tener presentes algunas ideas y conceptos de la teoría matemática de conjuntos. Este tema es muy extenso y se ha escrito mucho acerca de él. Sin embargo, aquí sólo necesitaremos algunas nociones básicas.

Un *conjunto* es una colección de objetos. Comúnmente los conjuntos se designan con letras mayúsculas A , B , etc. para describir qué objetos están contenidos en el conjunto A , se dispone de tres métodos.

(a) Podemos anotar los elementos de A . Por ejemplo, $A = \{1, 2, 3, 4\}$ indica el conjunto que contiene los enteros positivos 1, 2, 3 y 4.

(b) Podemos describir al conjunto A con palabras. Por ejemplo, podríamos decir que A está formado por todos los números reales entre 0 y 1, inclusive.

(c) Para describir el conjunto anterior, simplemente podemos escribir $A = \{x \mid 0 \leq x \leq 1\}$; es decir, A es el conjunto de todas las x , en donde x es un número real comprendido entre 0 y 1, inclusive.

Los objetos que forman la colección del conjunto A se llaman *miembros* o *elementos* de A . Cuando « a » es un elemento de A escribimos $a \in A$ y cuando « a » no es un elemento de A escribimos $a \notin A$.

Hay dos conjuntos especiales que a menudo son de interés. En la mayor parte de los problemas, estamos interesados en el estudio de un conjunto definido de objetos, y no de otros, por ejemplo, en todos los números reales, en todos los artículos que salen de una línea de producción durante un período de 24 horas, etc. Definimos el *conjunto universal* como el conjunto de todos los objetos que se consideran. Corrientemente este conjunto se designa U .

Otro conjunto que se debe destacar especialmente, puede aparecer como sigue. Supongamos que se describe el conjunto A como el conjunto de todos los números *reales* x que satisfacen la ecuación $x^2 + 1 = 0$. Evidentemente sabemos que no pueden existir tales números. ¡El conjunto A no contiene ningún elemento! Esta situación ocurre tan a menudo que justifica la introducción de un nombre especial para tal conjunto. Por lo tanto, definimos el conjunto *nulo* o *vacio* como el conjunto que no contiene elementos. Generalmente este conjunto se designa \emptyset .

Puede suceder que dados dos conjuntos A y B un elemento de A es también un elemento de B . En tal caso se dice que A es un *subconjunto* de B y se escribe $A \subset B$. Se da una interpretación semejante a $B \subset A$. Decimos que dos conjuntos son el mismo $A = B$, si y sólo si $A \subset B$ y $B \subset A$. Esto es, dos conjuntos son iguales si y sólo si contienen los mismos elementos.

Las dos propiedades siguientes del conjunto nulo y del conjunto universal son inmediatas.

(a) Para cualquier conjunto A , se tiene $\emptyset \subset A$.

(b) Una vez que el conjunto universal se ha acordado, entonces para cualquier conjunto A considerado que está en U , tenemos $A \subset U$.

EJEMPLO 1.1. Suponga que U = todos los números reales, $A = \{x \mid x^2 + 2x - 3 = 0\}$, $B = \{x \mid (x - 2)(x^2 + 2x - 3) = 0\}$, y $C = \{x \mid x = -3, 1, 2\}$. Entonces $A \subset B$ y $B = C$.

Ahora consideraremos la idea importante de *combinar* conjuntos dados con el fin de formar un nuevo conjunto. Se consideran dos operaciones básicas. Estas son paralelas, en ciertos aspectos, a las operaciones de suma y multiplicación de números. Supongamos que A y B son dos conjuntos. Definamos C como la *unión* de A y B (algunas veces llamada la suma de A y de B) de la manera siguiente:

$$C = \{x \mid x \in A \text{ o } x \in B \text{ (o ambos)}\}.$$

Escribimos $C = A \cup B$. Así C está formado por elementos que están en A , o en B , o en ambos.

Definimos D como la *intersección* de A y B (algunas veces designado como el producto de A y B) como sigue:

$$D = \{x \mid x \in A \text{ y } x \in B\}.$$

Escribamos esto como $D = A \cap B$. Es así como D posee todos los elementos que están en A y en B .

Finalmente presentamos la idea del *complemento* de un conjunto A como sigue: el conjunto designado por \bar{A} , formado por todos los elementos que *no*

están en A (sino en el conjunto universal U) se llama el complemento de A . Esto es, $\bar{A} = \{x \mid x \notin A\}$.

Se puede usar con mucha ventaja un recurso gráfico conocido como *gráfico de Venn* cuando se combinan conjuntos de la manera indicada anteriormente. En cada uno de los gráficos de la figura 1.1, la región sombreada representa el conjunto considerado.

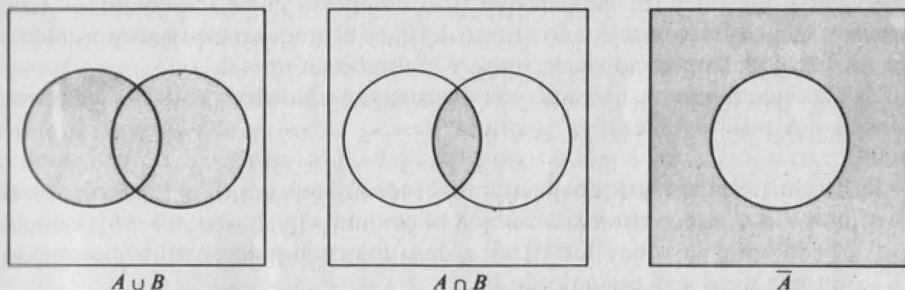


FIGURA 1.1

EJEMPLO 1.2. Supóngase que $U = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$; $A = \{1, 2, 3, 4\}$, $B = \{3, 4, 5, 6\}$. Hallamos que $\bar{A} = \{5, 6, 7, 8, 9, 10\}$, $A \cup B = \{1, 2, 3, 4, 5, 6\}$ y $A \cap B = \{3, 4\}$. Nótese que al describir un conjunto (tal como $A \cup B$) anotamos cada elemento exactamente una vez.

Las operaciones anteriores de unión e intersección definidas justamente para dos conjuntos pueden extenderse de una manera obvia para cualquier número finito de conjuntos. Así definimos $A \cup B \cup C$ como $A \cup (B \cup C)$ o $(A \cup B) \cup C$, que es el mismo, como fácilmente se puede verificar. De igual manera, definimos $A \cap B \cap C$ como $A \cap (B \cap C)$ o $(A \cap B) \cap C$ que, también, puede verificarse que son iguales. Y es evidente que podemos continuar esas construcciones de conjuntos nuevos con *cualquier* número *finito* de conjuntos dados.

Afirmábamos que ciertos conjuntos eran lo mismo, por ejemplo $A \cap (B \cap C)$ y $(A \cap B) \cap C$. Resulta que hay un número de tales conjuntos *equivalentes*, algunos de los cuales se indican más adelante. Si recordamos que dos conjuntos son iguales siempre que contengan los mismos elementos, es fácil verificar que los enunciados establecidos son verdaderos. El lector debe convencerse por sí mismo con ayuda de los diagramas de Venn.

- | | | |
|--|---|-------|
| (a) $A \cup B = B \cup A,$ | (b) $A \cap B = B \cap A,$ | (1.1) |
| (c) $A \cup (B \cup C) = (A \cup B) \cup C,$ | (d) $A \cap (B \cap C) = (A \cap B) \cap C$ | |

Nos referimos a (a) y (b) como las propiedades *comutativas*, y (c) y (d) como las propiedades *asociativas*.

Hay otros conjuntos *identidades* que contienen unión, intersección y complementación. Los más importantes de estos se indican a continuación. En cada caso, su validez puede verificarse con ayuda de un diagrama de Venn.

- (e) $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$,
 (f) $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$,
 (g) $A \cap \emptyset = \emptyset$,
 (h) $\overline{A \cup \emptyset} = A$, (i) $\overline{(A \cup B)} = \overline{A} \cap \overline{B}$,
 (j) $\overline{(A \cap B)} = \overline{A} \cup \overline{B}$, (k) $\overline{\overline{A}} = A$.

Observamos que (g) y (h) indican que \emptyset se comporta entre los conjuntos (con respecto a las operaciones \cup e \cap) como lo hace el número cero entre números (con respecto a las operaciones de suma y multiplicación).

Para lo que sigue se necesita una construcción adicional de un conjunto, dados dos (o más) conjuntos.

Definición. Sean A y B dos conjuntos. Indicaremos como el *producto cartesiano* de A y B escrito como $A \times B$ al conjunto $\{(a, b), a \in A, b \in B\}$, esto es, el conjunto de todos los pares ordenados en donde el primer elemento se toma de A y el segundo de B .

EJEMPLO 1.3. Sea $A = \{1, 2, 3\}$; $B = \{1, 2, 3, 4\}$.

Entonces $A \times B = \{(1, 1), (1, 2), \dots, (1, 4), (2, 1), \dots, (2, 4), (3, 1), \dots, (3, 4)\}$.

Observación: en general $A \times B \neq B \times A$.

La noción anterior puede extenderse como sigue: A_1, \dots, A_n son conjuntos, entonces $A_1 \times A_2 \times \dots \times A_n = \{(a_1, a_2, \dots, a_n), a_i \in A_i\}$, esto es el conjunto de todos los n -tuplos ordenados.

Un caso especialmente importante aparece cuando tomamos el producto cartesiano de un conjunto consigo mismo, esto es, $A \times A$ o $A \times A \times A$. Ejemplos así aparecen cuando nos relacionamos con el plano euclidiano, $R \times R$, en donde R es el conjunto de todos los números reales y el espacio euclidiano tridimensional se representa $R \times R \times R$.

El *número de elementos* en un conjunto nos será de mucha utilidad. Si hay un número finito de elementos en A , digamos a_1, a_2, \dots, a_n , decimos que A es *finito*. Si hay un número infinito de elementos en A que pueden ponerse en una *correspondencia uno-a-uno* con los enteros positivos, decimos que A es *contable* o *infinito numerable*. (Se puede demostrar por ejemplo, que el conjunto de todos los números racionales es infinito pero contable). Finalmente debemos considerar el caso de un conjunto infinito no numerable. Tales conjuntos contienen un número infinito de elementos que no pueden ser enumerados. Se puede demostrar, por ejemplo, que para dos números reales cualesquiera $b > a$, el conjunto $A = \{x \mid a \leq x \leq b\}$ tiene un número no numerable de elementos. Puesto que debemos asociar con cada número real un punto sobre la recta de los números reales, lo anterior expresa que cualquier intervalo (no degenerado) contiene más de un número contable de puntos.

Los conceptos presentados anteriormente, aunque representan sólo un breve bosquejo de la teoría de conjuntos, son suficientes para nuestro propósito: describir con rigor y precisión considerables, las ideas básicas de la teoría de la probabilidad.

1.3 Ejemplos de experimentos no determinísticos

Estamos ahora listos para discutir lo que entendemos por un experimento «aleatorio» o «no determinístico». (Más precisamente, daremos ejemplos de fenómenos para los cuales los modelos no determinísticos son apropiados. Esta es una distinción que el lector deberá mantener presente. Así nos referiremos frecuentemente a experimentos no determinísticos o aleatorios, cuando en realidad estamos hablando de un *modelo* no determinístico para un experimento). No pretendemos dar una definición precisa de diccionario para este concepto. En su lugar, daremos numerosos ejemplos que la ilustran.

- E_1 : Se lanza un dado y se observa el número que aparece en la cara superior.
- E_2 : Se lanza una moneda cuatro veces y se cuenta el número total de caras obtenidas.
- E_3 : Se lanza una moneda cuatro veces y se observa la sucesión de caras y sellos obtenidos.
- E_4 : Se fabrican artículos en una línea de producción y se cuenta el número de artículos defectuosos producidos en un período de 24 horas.
- E_5 : El ala de un aeroplano se arma con un gran número de remaches. Se cuenta el número de remaches defectuosos.
- E_6 : Se fabrica una bombilla. Luego se prueba su duración poniéndola en un portálámparas y se anota el tiempo transcurrido (en horas) hasta que se quema.
- E_7 : En un lote de 10 artículos hay 3 defectuosos. Se elige un artículo después de otro (sin sustituir el artículo elegido) hasta que se obtiene el último artículo defectuoso. Se cuenta el número total de artículos sacados del lote.
- E_8 : Se fabrican artículos hasta producir 10 no defectuosos. Se cuenta el número total de artículos manufacturados.
- E_9 : Se lanza un proyectil. Después de un tiempo determinado t , se anotan las tres componentes de la velocidad v_x , v_y , y v_z .
- E_{10} : Se observa un proyectil recién lanzado en tiempos, t_1, t_2, \dots, t_n . En cada oportunidad se anota la altura del proyectil sobre el suelo.
- E_{11} : Medir la resistencia a la tensión de una barra de acero.
- E_{12} : De una urna que contiene sólo esferas negras, se escoge una esfera y se anota su color.
- E_{13} : Un termógrafo anota la temperatura, continuamente, en un período de 24 horas. En un sitio y en una fecha señalados, «leer» dicho termógrafo.
- E_{14} : En la situación descrita en E_{13} , se anotan x e y , las temperaturas mínima y máxima del período de 24 horas considerado.

¿Qué tienen en común los experimentos anteriores? Los siguientes aspectos son importantes para nuestra descripción de un *experimento aleatorio*.

(a) Es posible repetir cada experimento indefinidamente sin cambiar esencialmente las condiciones.

(b) Aunque en general no podemos indicar cuál será un resultado *particular*, podemos describir el conjunto de *todos* los resultados *posibles* del experimento.

(c) A medida que el experimento se repite, los resultados individuales parecen ocurrir en forma caprichosa. Sin embargo, como el experimento se repite un *gran* número de veces, aparece un modelo definido de regularidad. Esta regularidad hace posible la construcción de un modelo matemático preciso con el cual analizamos el experimento. Tendremos más que decir acerca de la naturaleza e importancia de esta regularidad más adelante. Por el momento, el lector necesita pensar solamente en lanzamientos repetidos de una moneda regular. Aunque las caras y sellos aparecerán, sucesivamente, de una manera casi arbitraria, es bien conocido el hecho empírico de que, después de un gran número de lanzamientos, la proporción de caras y sellos será aproximadamente igual.

Debe notarse que todos los experimentos descritos anteriormente satisfacen estas características generales. (Naturalmente, la última característica mencionada solamente se puede verificar por experimentación: dejaremos a la intuición del lector creer que si el experimento se repitiese un gran número de veces la regularidad mencionada sería evidente. Por ejemplo, si se probase un número de bombillas del mismo fabricante, posiblemente el número de bombillas quemadas en más de 100 horas, digamos, podría ser predicha con bastante exactitud). Nótese que el experimento E_{12} tiene la peculiaridad de que sólo es posible un resultado. En general tales experimentos no serán de interés, por el hecho de que no sabemos qué resultado particular ocurrirá cuando se realice un experimento y qué lo hace interesante para nosotros.

Observación: al describir los diversos experimentos, hemos especificado no sólo el procedimiento que se realiza sino que también lo que estamos interesados en observar (ver, por ejemplo, la diferencia entre E_2 y E_3 mencionados previamente). Este es un punto muy importante al cual nos referiremos más adelante cuando discutamos las variables aleatorias. Por el momento, observemos simplemente que, como una consecuencia de un solo procedimiento experimental o la ocurrencia de un solo fenómeno, se pudieron calcular *varios* valores numéricos diferentes. Por ejemplo, si se escoge una persona entre un gran grupo de personas (y la elección propiamente dicha se hace según el procedimiento experimental indicado previamente), podríamos estar interesados en la altura, peso, ingreso anual, número de niños, etc., de la persona. Naturalmente, en la mayoría de los casos sabemos, antes de comenzar nuestro experimento, las características numéricas que nos interesan.

1.4 El espacio muestral

Definición. Con cada experimento ε del tipo que consideramos, definimos el *espacio muestral* como el conjunto de todos los resultados posibles de ε .

Usualmente designamos este conjunto como S . (En nuestro contexto, S representa el conjunto universal descrito previamente).

Consideraremos cada uno de los experimentos anteriores y describamos el espacio muestral de cada uno. El espacio muestral S_i se referirá al experimento E_i .

- $S_1: \{1, 2, 3, 4, 5, 6\}.$
- $S_2: \{0, 1, 2, 3, 4\}.$
- $S_3: \{\text{Todas las sucesiones posibles de la forma } a_1, a_2, a_3, a_4 \text{ en donde cada } a_i = C \text{ o } S \text{ según si aparece cara o sello en el } i\text{-ésimo lanzamiento}\}.$
- $S_4: \{0, 1, 2, \dots, N\}, \text{ en donde } N \text{ es el número máximo que pudo ser producido en 24 horas.}$
- $S_5: \{0, 1, 2, \dots, M\}, \text{ en donde } M \text{ es el número de remaches instalados.}$
- $S_6: \{t | t \geq 0\}.$
- $S_7: \{3, 4, 5, 6, 7, 8, 9, 10\}.$
- $S_8: \{10, 11, 12\}.$
- $S_9: \{v_x, v_y, v_z | v_x, v_y, v_z \text{ números reales}\}.$
- $S_{10}: \{h_1, \dots, h_n | h_i \geq 0, i = 1, 2, \dots, n\}.$
- $S_{11}: \{S | S \geq 0\}.$
- $S_{12}: \{\text{esfera negra}\}.$
- $S_{13}: \text{Este espacio muestral es el más importante de los que aquí consideramos. Debemos suponer prácticamente que la temperatura en cierta localidad específica nunca puede subir o bajar con relación a ciertos valores, digamos } M \text{ y } m. \text{ Fuera de esta restricción, debemos admitir la posibilidad de que aparezca cualquier gráfico con determinadas características. Posiblemente el gráfico no tenga saltos (esto es, representará una función continua). Además el gráfico tendrá ciertas características de suavidad que pueden resumirse matemáticamente al decir que el gráfico representa una función diferenciable. Así podemos enunciar finalmente que el espacio muestral es } \{f | f \text{ una función diferenciable, que satisface } m \leq f(t) \leq M, \text{ para todo } t\}.$
- $S_{14}: \{(x, y) | m \leq x \leq y \leq M\}. \text{ Es decir, } S_{14} \text{ consta de todos los puntos que están sobre y en un triángulo en el plano bi-dimensional } x, y.$

(En este libro no nos preocuparemos por los espacios muestrales de la complejidad encontrada en S_{13} . Sin embargo, tales espacios muestrales aparecen, pero para su estudio se necesitan matemáticas más avanzadas que las que presuponemos.)

A fin de describir un espacio muestral asociado con un experimento, debemos tener una idea muy clara de lo que medimos u observamos. Por tanto, deberíamos hablar de «un» espacio muestral asociado con un experimento en vez de «el» espacio muestral. A este respecto obsérvese la diferencia entre S_2 y S_3 .

Nótese también que el resultado de un experimento no necesita ser un número. Por ejemplo, en E_3 cada resultado es una sucesión de caras y sellos. En E_9 y E_{10} cada resultado consiste en un vector, mientras que en E_{13} consiste en una función.

Será importante discutir de nuevo el *número* de resultados de un espacio muestral. Surgen tres posibilidades: el espacio muestral puede ser finito, infinito numerable, o infinito no numerable. Refiriéndonos a los ejemplos anteriores notemos que $S_1, S_2, S_3, S_4, S_5, S_7$, y S_{12} son finitos, S_8 es infinito numerable, y $S_6, S_9, S_{10}, S_{11}, S_{13}, S_{14}$ son infinitos no numerables.

A esta altura podría ser útil comentar la diferencia entre un espacio muestral matemáticamente «idealizado» y uno realizable experimentalmente. Para este propósito, consideremos el experimento E_6 y su espacio muestral asociado S_6 . Es evidente que cuando anotamos el tiempo total t durante el cual funciona una bombilla, somos «victimas» de la precisión de nuestros instrumentos de medida. Supongamos que tenemos un instrumento que es capaz de anotar el tiempo con dos cifras decimales, por ejemplo, 16,43 horas. Con esta restricción impuesta nuestro espacio muestral llega a ser *infinito numerable*: $\{0,0; 0,01; 0,02, \dots\}$. Aun más, es muy realista suponer que ninguna bombilla puede durar posiblemente más de H horas, donde H podría ser un número muy grande. Así, parece que si somos completamente realistas en la descripción de este espacio muestral, estamos considerando un espacio muestral *finito*: $\{0,0; 0,01; 0,02, \dots, H\}$. El número total del resultado sería $(H/0,01) + 1$, que sería un número muy grande aun si H es moderadamente grande, por ejemplo, $H = 100$. Resultaría matemáticamente más simple y conveniente, suponer que *todos* los valores de $t \geq 0$ son resultados posibles y por tanto considerar el espacio muestral S_6 como se definió originalmente.

1.5 Sucesos

Otra noción básica es el concepto de un *suceso*. Un suceso A (respecto a un espacio muestral particular S asociado con un experimento ε) es simplemente un conjunto de resultados posibles. En la terminología de conjuntos, un suceso es un *subconjunto* del espacio muestral S . En vista de nuestra discusión previa, esto significa que S mismo es un suceso y también lo es el conjunto vacío \emptyset . Cualquier resultado individual puede también ser considerado como un suceso.

Los siguientes son ejemplos de sucesos. Otra vez, nos referimos a los experimentos anotados anteriormente: A_i se referirá a un suceso asociado con el experimento E_i .

- A_1 : Un número par ocurre; esto es, $A_1 = \{2, 4, 6\}$.
- A_2 : {2}; es decir, ocurren dos caras.
- A_3 : {CCCC, CCCS, CCSC, CSCC, SCCC}; es decir, salen más caras que sellos.
- A_4 : {0}; es decir, todos los artículos fueron no defectuosos.
- A_5 : {3, 4, ..., M }; es decir, más de dos remaches fueron defectuosos.
- A_6 : $\{t|t < 3\}$; es decir, la bombilla se quema en menos de tres horas.
- A_{14} : $\{(x, y)|y = x + 20\}$; es decir, el máximo es 20° mayor que el mínimo.

Cuando el espacio muestral S es finito o infinito numerable, *todo* subconjunto se puede considerar **como** un suceso. [Es un ejercicio fácil de verificar, y que haremos en breve, si S tiene n elementos, hay exactamente 2^n subconjuntos (sucesos).] Sin embargo, si S es infinito no numerable, aparece una dificultad teórica. Resulta que cualquier subconjunto concebible *no* se puede considerar como un suceso. Por razones que escapan al nivel de esta presentación, ciertos subconjuntos «no admisibles» deben ser excluidos. Afortunadamente tales conjuntos no admisibles

no aparecen realmente en las aplicaciones y, por tanto, no nos interesarán aquí. En lo que sigue se supondrá tácitamente que cada vez que mencionemos un suceso será de la clase que nos está permitido considerar.

Podemos usar ahora los diversos métodos para combinar conjuntos (es decir, sucesos) y obtener los nuevos conjuntos (es decir, sucesos) que presentamos al comienzo.

(a) Si A y B son sucesos, $A \cup B$ es el suceso que ocurre si y sólo si A o B (o ambos) ocurren.

(b) Si A y B son sucesos, $A \cap B$ es el suceso que ocurre si y sólo si A y B ocurren.

(c) Si A es un suceso, \bar{A} es el suceso que ocurre si y sólo si A no ocurre.

(d) Si A_1, \dots, A_n es cualquier colección finita de sucesos, entonces $\cup_{i=1}^n A_i$ es el suceso que ocurre si y sólo si *al menos* uno de los sucesos A_i ocurre.

(e) Si A_1, \dots, A_n es cualquier colección finita de sucesos, entonces $\cap_{i=1}^n A_i$ es el suceso que ocurre si y sólo si *todos* los sucesos A_i ocurren.

(f) Si A_1, \dots, A_n es cualquier colección infinita (numerable) de sucesos, entonces $\cup_{i=1}^{\infty} A_i$ es el suceso que ocurre si y sólo si *a lo menos* uno de los sucesos A_i ocurre.

(g) Si A_1, \dots, A_n, \dots es cualquier colección infinita (numerable) de sucesos, entonces $\cap_{i=1}^{\infty} A_i$ es el suceso que ocurre si y sólo si *todos* los sucesos A_i ocurren.

(h) Supóngase que S representa el espacio muestral asociado con un experimento ε y realiza ε dos veces. Entonces $S \times S$ se puede utilizar para representar todos los resultados de esas dos repeticiones. Es decir $(s_1, s_2) \in S \times S$ significa que s_1 resultó cuando se realizó ε la primera vez y s_2 cuando ε se realizó la segunda vez.

(i) Evidentemente, el ejemplo h se puede generalizar. Consideremos n repeticiones de un experimento ε cuyo espacio muestral es S . Entonces $S \times S \times \dots \times S = \{(s_1, s_2, \dots, s_n), s_i \in S, i = 1, \dots, n\}$ representa el conjunto de todos los resultados posibles cuando ε se realiza n veces. En cierto sentido, $S \times S \times \dots \times S$ es un espacio muestral en sí mismo, o sea el espacio muestral asociado con n repeticiones de ε .

Definición. Se dice que dos sucesos, A y B , son mutuamente excluyentes si no pueden ocurrir juntos. Expresamos esto escribiendo $A \cap B = \emptyset$; es decir, la intersección de A y B es el conjunto vacío.

EJEMPLO 1.4. Se prueba un artefacto electrónico y se anota el tiempo total de uso, digamos t . Supongamos $\{t|t \geq 0\}$.

Sean A , B , y C tres sucesos definidos como sigue:

$$A = \{t|t < 100\}; \quad B = \{t|50 \leq t \leq 200\}; \quad C = \{t|t > 150\}.$$

Entonces

$$A \cup B = \{t|t \leq 200\}; \quad A \cap B = \{t|50 \leq t \leq 100\};$$

$$B \cup C = \{t|t \geq 50\}; \quad B \cap C = \{t|150 < t \leq 200\}; \quad A \cap C = \emptyset;$$

$$A \cup C = \{t|t < 100 \text{ o } t > 150\}; \quad \bar{A} = \{t|t \geq 100\}; \quad \bar{C} = \{t|t \leq 150\}.$$

Como se discutió en la sección anterior, una de las características básicas del concepto de «experimento» es que no sabemos qué resultado particular se obtendrá al realizar el experimento. En otras palabras, si A es un suceso asociado con un experimento, no podemos indicar con certeza que A ocurrirá o no. Por lo tanto, llega a ser muy importante tratar de asociar un número con el suceso A que medirá, de alguna manera, la posibilidad de que el suceso A ocurra. Esta tarea nos conduce a la teoría de probabilidades.

1.6 Frecuencia relativa

Para motivar el enfoque adoptado como solución del problema anterior, consideremos el procedimiento siguiente. Supóngase que repetimos n veces el experimento ε y sean A y B dos sucesos asociados con ε . Sean n_A y n_B el número respectivo de veces que el suceso A y el suceso B ocurrieron en las n repeticiones.

Definición. $f_A = n_A/n$ se llama la *frecuencia relativa* del suceso A en las n repeticiones de ε . La frecuencia relativa f_A tiene las siguientes propiedades importantes, que son verificables fácilmente.

- (1) $0 \leq f_A \leq 1$.
- (2) $f_A = 1$ si y sólo si A ocurre *cada vez* en las n repeticiones.
- (3) $f_A = 0$ si y sólo si A *nunca* ocurre en las n repeticiones.
- (4) Si A y B son dos sucesos que se excluyen mutuamente y si $f_{A \cup B}$ es la frecuencia relativa asociada al suceso $A \cup B$, entonces $f_{A \cup B} = f_A + f_B$.
- (5) f_A , basada en las n repeticiones del experimento y considerada para una función de n , «converge» en cierto sentido probabilístico a $P(A)$ cuando $n \rightarrow \infty$.

Observación: la propiedad (5) anterior obviamente está indicada de una manera vaga en este momento. Sólo posteriormente (Sección 12.2) podremos precisar más esta idea. Por ahora indiquemos simplemente que la Propiedad (5) encierra la notación bastante intuitiva de que la frecuencia relativa basada en un número creciente de observaciones tiende a «estabilizarse» en la proximidad de un valor definitivo. Esto *no* es lo mismo que el concepto corriente de convergencia que se encuentra en otra parte en matemáticas. De hecho, como se indicó aquí, esta *no* es del todo una conclusión matemática, sino simplemente una realidad empírica.

La mayor parte de nosotros intuitivamente estamos conscientes de este fenómeno de estabilización aunque puede ser que nunca lo hayamos verificado. Hacerlo requiere una cantidad considerable de tiempo y paciencia, ya que requiere un gran número de repeticiones de un experimento. Sin embargo, algunas veces podemos ser observadores inocentes de este fenómeno como lo ilustra el siguiente ejemplo.

EJEMPLO 1.5. Supóngase que estamos parados en una acera y nos fijamos en dos losas de cemento adyacentes. Imaginemos que empieza a llover de tal manera que en realidad podemos distinguir unas gotas de otras y les seguimos

la pista para averiguar si caen en una losa o en otra. Continuamos observando las gotas individuales y anotamos su punto de impacto. Simbolizando la i -ésima gota por X_i , en donde $X_i = 1$ si la gota cae en una losa y 0 si cae en la otra, podríamos observar una sucesión tal como 1, 1, 0, 1, 0, 0, 0, 1, 0, 0, 1. Ahora está claro que no podemos predecir en donde caerá la gota particular. (Nuestro experimento consiste en una especie de situación meteorológica que causa la caída de las gotas de lluvia.) Si calculamos la frecuencia relativa del suceso $A = \{\text{la gota cae en la losa}\}$, entonces la sucesión anterior de resultados da origen a las frecuencias relativas siguientes (con base en la observación de 1, 2, 3, ..., gotas): $\frac{1}{1}, \frac{1}{2}, \frac{3}{4}, \frac{3}{5}, \frac{3}{6}, \frac{3}{7}, \frac{4}{8}, \frac{4}{9}, \frac{4}{10}, \frac{5}{11}, \dots$. Esos valores muestran un grado considerable de variación, especialmente al comienzo. Intuitivamente es claro que si se continua indefinidamente el experimento anterior, esas frecuencias relativas se estabilizarian próximas al valor $\frac{1}{2}$. Porque tenemos toda la razón para creer que después de que haya transcurrido cierto tiempo las dos losas estarían igualmente mojadas.

Esta propiedad de estabilidad de la frecuencia relativa es aún una noción bastante intuitiva y sólo podremos precisarla matemáticamente más tarde. Lo importante de esta propiedad es que si se realiza un experimento un gran número de veces, la frecuencia relativa con que ocurre un suceso A tiende a variar menos y menos cuando el número de repeticiones aumenta. A esta característica se le designa como regularidad estadística.

También hemos sido algo imprecisos en nuestra definición de experimento. Exactamente, ¿cuándo un procedimiento o mecanismo es un experimento en nuestra opinión, susceptible de ser estudiado matemáticamente mediante un modelo no determinístico? Previamente indicamos que debe ser posible efectuar un experimento repetidamente sin cambiar las mismas condiciones esenciales. Podemos agregar ahora otro requisito. Cuando el experimento se realiza repetidamente debe presentar la regularidad estadística a que nos referimos anteriormente. Más adelante discutiremos un teorema (llamado la ley de los grandes números) que muestra que la regularidad estadística es de hecho una *consecuencia* de la primera condición: la repetición.

1.7 Nociones básicas de probabilidad

Volvamos ahora al problema propuesto anteriormente: asignar un número a cada suceso A que medirá la posibilidad de que A ocurra cuando el experimento se realiza. Un enfoque *possible* podría ser el siguiente: repetir el experimento un gran número de veces, calcular la frecuencia relativa f_A , y usar este número. Cuando recordamos las propiedades f_A , es claro que este número *da* una indicación muy definida de qué posibilidad existe de que A ocurra. Aun más, como sabemos que el experimento se repite más y más veces, la frecuencia relativa se estabiliza cerca de algún número, digamos p . Sin embargo, hay dos objeciones serias a este enfoque. (a) No está claro cómo de grande debe ser n antes de que conozcamos el número. ¿1000? ¿2000? ¿10.000? (b) Una vez que se ha descrito completamente el experimento y se ha especificado el suceso A , el número que buscamos no debe

depender del experimentador o de una racha de suerte en particular con la que él experimenta. (Por ejemplo, es posible que con una moneda perfectamente balanceada que se lanzó 10 veces, resulten 9 caras y 1 sello. La frecuencia relativa del suceso $A = \{\text{salen caras}\}$ es así igual a $\frac{9}{10}$. Aunque es posible que en los 10 lanzamientos siguientes el modelo de caras y sellos pueda estar invertido.) Lo que queremos es un medio de obtener tal número sin recurrir a la experimentación. Naturalmente, para que el número estipulado sea significativo, cualquier experimento debería dar una frecuencia relativa "cercana" al valor estipulado, en especial si el número de repeticiones en las cuales se calculó la frecuencia relativa es muy grande. Procederemos formalmente como sigue.

Definición. Sea ε un experimento. Sea S un espacio muestral asociado con ε . Con cada suceso A asociamos un número real, designado por $P(A)$ y llamado *la probabilidad de que A satisfaga* las siguientes propiedades.

- (1) $0 \leq P(A) \leq 1$.
- (2) $P(S) = 1$.
- (3) Si A y B son sucesos que se excluyen mutuamente, $P(A \cup B) = P(A) + P(B)$.
- (4) Si $A_1, A_2, \dots, A_n, \dots$ son sucesos que se excluyen mutuamente de par en par, entonces

$$P(\cup_{i=1}^{\infty} A_i) = P(A_1) + P(A_2) + \dots + P(A_n) + \dots \quad (1.3)$$

Observemos que de la Propiedad 3 se *deduce* de inmediato que para cualquier n finito,

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i).$$

La Propiedad 4 *no* sigue; sin embargo, cuando consideremos el espacio muestral idealizado, esta condición será necesaria y por tanto se incluye aquí.

La elección de la lista de propiedades de las probabilidades está obviamente motivada por las características correspondientes de las frecuencias relativas. La propiedad anteriormente mencionada como regularidad estadística se ligará con esta definición de probabilidad más tarde. Por el momento indicamos sólo que demostraremos que los valores de $P(A)$ y f_A están próximos uno al otro (en cierto sentido), si f_A se basa en un gran número de repeticiones. Este hecho es el que justifica el empleo de $P(A)$ para medir la probabilidad de que A ocurra.

Por el momento no sabemos *como* calcular $P(A)$. Solamente hemos anotado algunas propiedades generales que posee $P(A)$. El lector tendrá que tener paciencia un poco más (hasta el próximo capítulo) antes de que aprenda *como* calcular $P(A)$. Antes de volver a este tema indiquemos y probemos varias consecuencias relativas a $P(A)$ que se deducen de las condiciones, y que no dependen de *como* calculamos $P(A)$ en realidad.

Teorema 1.1. Si \emptyset es el conjunto vacío, entonces $P(\emptyset) = 0$.

Demostración: podemos escribir, para cualquier suceso A , $A = A \cup \emptyset$. Puesto que A y \emptyset son mutuamente excluyentes, se deduce de la Propiedad 3 que $P(A) = P(A \cup \emptyset) = P(A) + P(\emptyset)$. A partir de esto la conclusión del teorema es inmediata.

Observación: tendremos ocasión de ver más tarde que el recíproco del teorema anterior no es verdadero. Esto es, si $P(A) = 0$, en general no podemos concluir que $A = \emptyset$, porque hay situaciones en que asignamos probabilidad cero a un suceso que *puede* ocurrir.

Teorema 1.2. Si \bar{A} es el suceso complementario de A , entonces

$$P(A) = 1 - P(\bar{A}) \quad (1.4)$$

Demostración: podemos escribir $S = A \cup \bar{A}$ y, usando las Propiedades 2 y 3, obtenemos $1 = P(A) + P(\bar{A})$.

Observación: este es un resultado muy útil, porque indica que cada vez que deseamos calcular $P(A)$ podemos calcular $P(\bar{A})$ en su lugar y obtener el resultado deseado por sustacción. Veremos después que en muchos problemas es más fácil calcular $P(\bar{A})$ que $P(A)$.

Teorema 1.3. Si A y B son dos sucesos cualesquiera, entonces

$$P(A \cup B) = P(A) + P(B) - P(A \cap B). \quad (1.5)$$

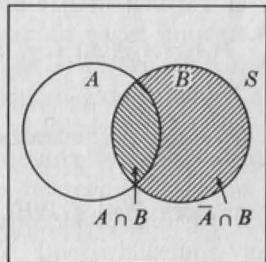


FIGURA 1.2

Demostración: la idea de esta demostración es descomponer $A \cup B$ y B en sucesos que se excluyen mutuamente y luego aplicar la Propiedad 3. (Ver el diagrama de Venn en la figura 1.2.)

Así escribimos

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

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

Por lo tanto

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

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

Sustrayendo la segunda ecuación de la primera

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

y por tanto se obtiene el resultado.

Observación: este teorema representa una extensión obvia de la Propiedad 3, porque si $A \cap B = \emptyset$, obtenemos de lo anterior el enunciado de la Propiedad 3.

Teorema 1.4. Si A , B , y C son tres sucesos cualesquiera entonces

$$\begin{aligned} P(A \cup B \cup C) &= P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) \\ &\quad - P(B \cap C) + P(A \cap B \cap C). \end{aligned} \tag{1.6}$$

Demostración: la demostración consiste en escribir $A \cup B \cup C$ como $(A \cup B) \cup C$ y aplicar el resultado del teorema anterior. Dejamos los detalles al lector.

Observación: una extensión obvia del teorema anterior se sugiere por sí misma. Sean A_1, \dots, A_k k sucesos cualesquiera. Entonces

$$\begin{aligned} P(A_1 \cup A_2 \cup \dots \cup A_k) &= \sum_{i=1}^k P(A_i) - \sum_{i < j = 2}^k P(A_i \cap A_j) \\ &\quad + \sum_{i < j < r = 3}^k P(A_i \cap A_j \cap A_r) + \dots + (-1)^{k-1} P(A_1 \cap A_2 \cap \dots \cap A_k). \end{aligned} \tag{1.7}$$

Este resultado se puede establecer fácilmente por inducción matemática.

Teorema 1.5. Si $A \subset B$, entonces $P(A) \leq P(B)$.

Demostración: podemos descomponer B en dos sucesos que se excluyen mutuamente como sigue: $B = A \cup (B \cap \bar{A})$. Por lo tanto $P(B) = P(A) + P(B \cap \bar{A}) \geq P(A)$, puesto que $P(B \cap \bar{A}) \geq 0$ según la Propiedad 1.

Observación: este resultado es intuitivamente atractivo. Porque dice que si B debe ocurrir cada vez que A ocurre, entonces B es al menos tan probable como A .

1.8 Varias observaciones

(a) Cabe aquí una advertencia. De la discusión previa se podría inferir (incorrectamente) que cuando elegimos un modelo probabilístico para la descripción de algún fenómeno observable descartamos todas las relaciones determinísti-

cas. Nada puede estar más lejos de la verdad. Mantenemos todavía el hecho de que, por ejemplo, la ley de Ohm $I = E/R$ es válida en ciertas circunstancias. La diferencia será de interpretación. En vez de decir que la relación anterior determina I para E y R dados, reconoceremos que E (y/o) R pueden variar de una manera aleatoria e imprecisa y que por lo tanto I variará también de una manera aleatoria. Para E y R dados, todavía I se determina por la relación anterior. Lo importante es que cuando adoptamos un modelo probabilístico para la descripción de un circuito, consideraremos la posibilidad que E y R pueden variar de una manera imprecisa que sólo se puede describir probabilísticamente. Así, puesto que será de importancia considerar sólo la *probabilidad* de que E y R tomen ciertos valores, llega a ser significativo hablar sólo de la probabilidad de que I tome ciertos valores.

(b) La elección entre adoptar un modelo determinístico o probabilístico puede ser difícil de hacer algunas veces. Puede depender de lo intrincado de nuestra técnica de medida y la precisión asociada. Por ejemplo, si las medidas precisas son tan difíciles de obtener que las lecturas repetidas de la misma cantidad produzcan resultados variables, un modelo probabilístico es indudablemente más adecuado para describir la situación.

(c) Señalaremos brevemente que bajo ciertas circunstancias, estamos en posición de hacer hipótesis adicionales acerca de la conducta probabilística de nuestros resultados experimentales que nos conducirán a un método para evaluar las probabilidades básicas. La elección de esas hipótesis adicionales puede estar basada en consideraciones físicas del experimento (ciertas propiedades de simetría por ejemplo), evidencia empírica, o en algunos casos, simplemente un juicio personal basado en una experiencia previa con una situación similar. La frecuencia relativa puede jugar un papel importante en nuestra decisión acerca de una asignación numérica de $P(A)$. Sin embargo, es importante darse cuenta de que cualquier suposición que hagamos acerca de $P(A)$ debe ser tal que se satisfagan los axiomas del (1) al (4) de la Definición (1.3).

(d) En el transcurso del desarrollo de las ideas básicas de la teoría de probabilidades, haremos algunas referencias a ciertas analogías de la mecánica. La primera de ellas puede ser apropiada aquí. En mecánica, asignamos la masa a cada cuerpo B , digamos $m(B)$. Luego hacemos varios cálculos y llegamos a diversas conclusiones acerca de la conducta de B y su relación con otros cuerpos, muchos de los cuales implican su masa $m(B)$. El hecho de que en realidad tengamos que recurrir a alguna aproximación para obtener $m(B)$ para un cuerpo específico no disminuye la utilidad del concepto de masa. De igual manera, establecemos para cada suceso A , asociado con el espacio muestral de un experimento, un número $P(A)$ llamado la probabilidad de A y que satisface los axiomas básicos. En realidad al calcular $P(A)$ para un suceso específico, tenemos que hacer hipótesis adicionales o bien obtener una aproximación basada en la evidencia empírica.

(e) Es muy importante darnos cuenta de que hemos *postulado* la existencia del número $P(A)$ y que hemos postulado ciertas propiedades que este número posee. La validez de las diversas consecuencias (teoremas) derivadas de esos pos-

tulados de ninguna manera depende de cómo obtenemos un valor numérico para $P(A)$. Es vital que este punto esté claro. Por ejemplo, hemos supuesto que $P(A \cup B) = P(A) + P(B)$. A fin de usar esta relación para la *evaluación* actual de $P(A \cup B)$, debemos conocer el valor de $P(A)$ y de $P(B)$. Discutiremos brevemente como, bajo ciertas circunstancias, debemos hacer suposiciones adicionales que conduzcan a un método para evaluar esas probabilidades. Si estas (u otras) suposiciones no están garantizadas, debemos recurrir a la experimentación para aproximar el valor de $P(A)$ de los datos reales. La frecuencia relativa f_A desempeñará un papel importante en esto, y de hecho, se usará como una aproximación de $P(A)$.

Sin embargo, es importante tener presente que f_A y $P(A)$ no son lo mismo, que simplemente usamos f_A como aproximación de $P(A)$, y que cada vez que nos referimos a $P(A)$ nos estamos refiriendo al valor postulado. Si identificamos f_A con $P(A)$ debemos verificar que simplemente sustituimos un valor postulado por uno aproximado obtenido experimentalmente. Qué tan buena o mala pueda ser esta aproximación, de ninguna manera influye en la estructura lógica de nuestro modelo. Aunque el fenómeno que pretende representar el modelo se consideró al construirlo, nos hemos separado del fenómeno mismo (temporalmente al menos), cuando entramos en el dominio del modelo.

PROBLEMAS

1.1. Supóngase que el conjunto universal consta de los enteros positivos de 1 a 10. Sean $A = \{2, 3, 4\}$, $B = \{3, 4, 5\}$, y $C = \{5, 6, 7\}$. Anote los elementos de los siguientes conjuntos.

- (a) $\bar{A} \cap B$ (b) $\bar{A} \cup B$ (c) $\overline{A \cap \bar{B}}$ (d) $\overline{A \cap (\bar{B} \cap C)}$ (e) $\overline{A \cap (\bar{B} \cup C)}$

1.2. Supóngase que el conjunto universal U está dado por $U = \{x | 0 \leq x \leq 2\}$. Sean los conjuntos A y B definidos como sigue: $A = \{x | \frac{1}{2} < x \leq 1\}$ y $B = \{x | \frac{1}{4} \leq x < \frac{3}{2}\}$. Describa los conjuntos siguientes:

- (a) $\overline{A \cup B}$ (b) $A \cup \bar{B}$ (c) $\overline{A \cap B}$ (d) $\bar{A} \cap B$

1.3. ¿Cuáles de las siguientes relaciones son verdaderas?

- (a) $(A \cup B) \cap (A \cup C) = A \cup (B \cap C)$ (b) $(A \cup B) = (A \cap \bar{B}) \cup B$
 (c) $\bar{A} \cap B = A \cup B$ (d) $(\overline{A \cup B}) \cap C = \bar{A} \cap \bar{B} \cap \bar{C}$
 (e) $(A \cap B) \cap (\bar{B} \cap C) = \emptyset$

1.4. Supóngase que el conjunto universal consta de todos los puntos (x, y) cuyas coordenadas son enteros y quedan dentro o sobre el contorno del cuadrado acotado por las rectas $x = 0$, $y = 0$, $x = 6$, $y = 6$. Indique los elementos de los conjuntos siguientes.

- (a) $A = \{(x, y) | x^2 + y^2 \leq 6\}$ (b) $B = \{(x, y) | y \leq x^2\}$
 (c) $C = \{(x, y) | x \leq y^2\}$ (d) $B \cap C$ (e) $(B \cup A) \cap C$

1.5. Use los diagramas de Venn para establecer las siguientes relaciones.

- (a) $A \subset B$ y $B \subset C$ implican que $A \subset C$ (b) $A \subset B$ implica que $A = A \cap B$
 (c) $A \subset B$ implica que $\bar{B} \subset \bar{A}$ (d) $A \subset B$ implica que $A \cup C \subset B \cup C$
 (e) $A \cap B = \emptyset$ y $C \subset A$ implica que $B \cap C = \emptyset$

1.6. Los artículos provenientes de una línea de producción se clasifican defectuosos (D) o no defectuosos (N). Se observan los artículos y se anota su condición. Este proceso se continúa hasta que se produzcan dos artículos defectuosos consecutivos o se hayan verificado cuatro artículos, cualesquiera que ocurra primero. Describir un espacio muestral para este experimento.

1.7. (a) Una caja con N bombillas tiene r ($r < N$) unidades con filamentos rotos. Estas se prueban una por una, hasta que se encuentra una defectuosa. Describir un espacio muestral para este experimento.

(b) Supóngase que las bombillas anteriores se prueban una por una, hasta que se prueban todas las defectuosas. Describir el espacio muestral para este experimento.

1.8. Considérense cuatro objetos, a, b, c , y d . Supóngase que el orden en el cual se anotan esos objetos representa el resultado de un experimento. Sean A y B los sucesos definidos como sigue: $A = \{a$ está en el primer lugar $\}; B = \{b$ está en el segundo lugar $\}$.

- (a) Anote todos los elementos del espacio muestral.
- (b) Anote todos los elementos de los sucesos $A \cap B$ y $A \cup B$.

1.9. Un lote contiene artículos que pesan 5, 10, 15, ..., 50 libras. Supóngase que al menos dos artículos de cada peso se encuentran allí. Se eligen dos artículos del lote. Identifíquese por X el peso del primer artículo elegido y por Y el peso del segundo artículo. Así el par de números (X, Y) representa un solo resultado del experimento. Usando el plano XY , indíquese el espacio muestral y los sucesos siguientes.

- (a) $\{X = Y\}$
- (b) $\{Y > X\}$
- (c) El segundo artículo pesa el doble del primero.
- (d) El primer artículo pesa 10 libras menos que el segundo.
- (e) El promedio de peso de los dos artículos es menor de 30 libras.

1.10. En un período de 24 horas, en un momento X , un interruptor se pone en la posición "encendido". Posteriormente, en un momento Y (todavía en el mismo período de 24 horas) el interruptor se pone en la posición "apagado". Supóngase que X y Y se miden en horas en el eje de tiempo con el comienzo del período como origen. El resultado del experimento consta del par de números (X, Y) .

- (a) Describa el espacio muestral.
- (b) Describa y dibuje los siguientes sucesos en el plano XY .
 - (i) El circuito funciona durante una hora o menos.
 - (ii) El circuito funciona en el tiempo z en donde z es algún intervalo durante el período dado de 24 horas.
 - (iii) El circuito empieza a funcionar antes del tiempo t_1 y deja de funcionar después del tiempo t_2 (en donde otra vez $t_1 < t_2$ son dos intervalos de tiempo durante el período especificado).
 - (iv) El circuito funciona el doble de lo que será interrumpido.

1.11. Sean A , B , y C tres sucesos asociados con un experimento. Exprese las siguientes proposiciones verbales en notación de conjuntos.

- (a) Al menos uno de los sucesos ocurre.
- (b) Exactamente uno de los sucesos ocurre.
- (c) Exactamente dos de los sucesos ocurren.
- (d) No ocurren más de dos sucesos simultáneamente.

1.12. Demuestre el teorema 1.4.

1.13. (a) Demostrar que para dos sucesos cualesquiera, A_1 y A_2 , tenemos $P(A_1 \cup A_2) \leq P(A_1) + P(A_2)$.

(b) Demuestre que para n sucesos cualesquiera A_1, \dots, A_n , tenemos

$$P(A_1 \cup \dots \cup A_n) \leq P(A_1) + \dots + P(A_n).$$

[Consejo: Use una inducción matemática. El resultado que se indica en (b) se llama desigualdad de Boole.]

1.14. El teorema 1.3 trata de la probabilidad de que *al menos uno* de los sucesos A o B ocurra. La proposición siguiente trata de la probabilidad de que *exactamente* uno de los sucesos A o B ocurra.

Demostrar que $[P(A \cap \bar{B}) \cup (B \cap \bar{A})] = P(A) + P(B) - 2P(A \cap B)$.

1.15. Cierta tipo de motor eléctrico falla por obstrucción de los cojinetes, por combustión del embobinado o por desgaste de las escobillas. Supóngase que la probabilidad de la obstrucción es el doble de la de combustión, la cual es cuatro veces más probable que la inutilización de las escobillas. ¿Cuál es la probabilidad de que el fallo sea por cada uno de esos tres mecanismos?

1.16. Supóngase que A y B son sucesos para los cuales $P(A) = x$, $P(B) = y$, y $P(A \cap B) = z$. Expresar cada una de las probabilidades siguientes en términos de x , y , y z .

- (a) $P(\bar{A} \cup \bar{B})$ (b) $P(\bar{A} \cap B)$ (c) $P(\bar{A} \cup B)$ (d) $P(\bar{A} \cap \bar{B})$

1.17. Supóngase que A , B , y C son sucesos tales que $P(A) = P(B) = P(C) = \frac{1}{4}$, $P(A \cap B) = P(C \cap B) = 0$, y $P(A \cap C) = \frac{1}{8}$. Calcular la probabilidad de que al menos uno de los sucesos A , B , o C ocurra.

1.18. Una instalación consta de dos calderas y un motor. Sea A el suceso de que el motor está en buenas condiciones, mientras que los sucesos B_k ($k = 1, 2$) son los sucesos de que la k -ésima caldera esté en buenas condiciones. El suceso C es que la instalación pueda funcionar. Si la instalación funciona cada vez que el motor y al menos una caldera funcione, exprese C y \bar{C} en función de A y de los sucesos B_i .

1.19. Un mecanismo tiene dos tipos de repuestos, digamos I y II. Suponga que hay dos del tipo I y tres del tipo II. Definir los sucesos A_k , $k = 1, 2$, y B_j , $j = 1, 2, 3$ como sigue: A_k : la k -ésima unidad del tipo I está funcionando correctamente; B_j : la j -ésima unidad del tipo II está funcionando correctamente. Finalmente C representa el suceso: el mecanismo funciona. Dado que el mecanismo funciona si al menos una unidad del tipo I y dos unidades del tipo II funcionan, exprese el suceso C en función de los A_k y los B_j .

Espacios muestrales finitos

2.1 El espacio muestral finito

En este capítulo nos ocuparemos sólo de los experimentos para los cuales el espacio muestral S consta de un número *finito* de elementos. Es decir, suponemos que S se puede escribir como $S = \{a_1, a_2, \dots, a_k\}$. Si nos referimos a los ejemplos de espacios muestrales de la sección 1.4, observamos que $S_1, S_2, S_3, S_4, S_5, S_7$ y S_{12} son todos finitos.

A fin de caracterizar $P(A)$ en este modelo consideraremos primero el suceso que está constituido por un *solo resultado*, llamado algunas veces un suceso *elemental*, digamos $A = \{a_i\}$. Procedemos como sigue.

A cada uno de los sucesos elementales $\{a_i\}$ asignamos un número p_i , llamado la probabilidad de $\{a_i\}$, que satisface las condiciones siguientes:

- (a) $p_i \geq 0, \quad i = 1, 2, \dots, k,$
- (b) $p_1 + p_2 + \dots + p_k = 1.$

Puesto que $\{a_i\}$ es un suceso, estas condiciones deben estar de acuerdo con las postuladas para las probabilidades de sucesos en general, como se hizo en la ecuación (1.3). Es muy sencillo verificar que es así.

A continuación, supongamos que un suceso A está constituido por r resultados, $1 \leq r \leq k$, digamos

$$A = \{a_{j_1}, a_{j_2}, \dots, a_{j_r}\},$$

en donde j_1, j_2, \dots, j_r representa cualquier índice r de $1, 2, \dots, k$. Por lo tanto, se deduce de la ecuación (1.3), Propiedad 4, que

$$P(A) = p_{j_1} + p_{j_2} + \dots + p_{j_r}. \quad (2.1)$$

Para resumir: la asignación de probabilidades p_i a cada uno de los sucesos elementales $\{a_i\}$, sujeto a las condiciones anteriores (a) y (b), determina de un modo único $P(A)$ para cada uno de los sucesos $A \subset S$, en donde $P(A)$ está dado por la ecuación (2.1).

A fin de evaluar las p_j , se deben hacer algunas suposiciones respecto a los resultados individuales.

EJEMPLO 2.1. Supongamos que sólo son posibles tres resultados en un experimento, digamos a_1, a_2 y a_3 . Además, supongamos que la ocurrencia de a_1 es dos veces más probable que la de a_2 , la cual, a su vez, es dos veces más probable que a_3 .

Por tanto, $p_1 = 2p_2$ y $p_2 = 2p_3$. Puesto que $p_1 + p_2 + p_3 = 1$, tenemos que $4p_3 + 2p_3 + p_3 = 1$, lo que finalmente da

$$p_3 = \frac{1}{7}, \quad p_2 = \frac{2}{7}, \quad y \quad p_1 = \frac{4}{7}.$$

2.2 Resultados igualmente probables

La suposición que más comúnmente se hace para espacios muestrales finitos es que todos los resultados son igualmente probables. De ninguna manera esta suposición puede darse como un hecho; debe justificarse cuidadosamente. Hay muchos experimentos para los cuales se garantiza tal suposición, pero también hay muchas situaciones experimentales en las cuales sería un error hacer tal suposición. Por ejemplo, sería muy poco realista suponer que es tan probable no recibir llamadas telefónicas en una central entre la 1 a.m. y 2 a.m. como entre las 5 p.m. y las 6 p.m.

Si los k resultados son igualmente probables, se deduce que cada $p_i = 1/k$. Porque la condición $p_1 + \dots + p_k = 1$ llega a ser $kp_i = 1$ para todo i . De esto se deduce que para cualquier suceso A que conste de r resultados, tenemos

$$P(A) = r/k.$$

Este método de evaluar $P(A)$ a menudo se indica como sigue:

$$P(A) = \frac{\text{número de maneras en que } \varepsilon \text{ puede ocurrir favorable a } A}{\text{número total de maneras en que } \varepsilon \text{ puede ocurrir}}.$$

Es importante comprender que la expresión anterior de $P(A)$ es sólo una consecuencia de la suposición de que todos los resultados son igualmente probables y sólo es aplicable cuando se satisface esta suposición. Indudablemente no sirve como una definición general de probabilidad.

EJEMPLO 2.2. Se lanza un dado y se supone que todos los resultados son igualmente probables. El suceso A ocurre si y sólo si aparece un número mayor que 4. Esto es, $A = \{5,6\}$. Por lo tanto $P(A) = \frac{1}{6} + \frac{1}{6} = \frac{2}{6}$.

EJEMPLO 2.3. Se lanza una moneda normal. Sea A el suceso: {aparece una cara}. Para evaluar $P(A)$ un análisis del problema podría ser de la manera siguiente. El espacio muestral es $S = \{0, 1, 2\}$, en donde cada uno de los resultados representa un número de caras que ocurren. Por lo tanto $P(A) = \frac{1}{3}$! Este análisis es obviamente incorrecto, puesto que en el espacio muestral considerado anteriormente, todos los resultados *no* son igualmente probables. A fin de aplicar el método anterior deberíamos considerar en su lugar, el espacio muestral $S' = \{CC, CS, SC, SS\}$. En este espacio muestral todos los resultados son igualmente probables y, por lo tanto, obtenemos para la solución correcta a nuestro problema $P(A) = \frac{2}{4} = \frac{1}{2}$.

Podríamos emplear correctamente el espacio muestral S como sigue: los resultados 0 y 2 son igualmente posibles, mientras que el resultado 1 es probablemente el doble de cada uno de los otros. Por lo tanto, $P(A) = \frac{1}{2}$, lo cual concuerda con la respuesta anterior.

Este ejemplo ilustra dos puntos. Primero, debemos estar completamente seguros de que todos los resultados que se pueden suponer son igualmente probables antes de utilizar el procedimiento anterior. Segundo, a menudo podemos reducir el problema a uno, en el cual todos los resultados *son* igualmente posibles, mediante una elección apropiada del espacio muestral. Cada vez que sea posible se debe hacer esto, puesto que generalmente simplifica los cálculos. Se tratará este punto nuevamente en ejemplos subsecuentes.

Muy a menudo la manera como se realiza un experimento determina si los resultados son o no igualmente probables. Por ejemplo, supongamos que vamos a escoger un perno de una caja que contiene tres de tamaño diferente. Si elegimos el perno acercándonos a la caja y sacando el primero que tocamos, es obvio que el perno más grande tendrá una probabilidad de ser escogido, mayor que la de los otros dos. Sin embargo, si rotulamos cuidadosamente cada perno con un número, lo escribimos en una etiqueta, y escogemos una de ellas, podemos tratar de asegurar que cada perno, de hecho, tiene la misma probabilidad de ser escogido. Así quizás tengamos que afrontar considerables dificultades a fin de asegurarnos que la suposición matemática de resultados igualmente probables es de hecho apropiada.

En ejemplos anteriores y muchos otros subsecuentes, nos interesa la elección al azar de uno o más objetos de una colección dada. Definamos más precisamente esta noción. Supongamos que tenemos N objetos, digamos a_1, a_2, \dots, a_N .

(a) *Escoger al azar un objeto* de los N , significa que cada uno de los objetos tiene la misma probabilidad de ser escogido. Esto es,

$$\text{Prob}(\text{elegir } a_i) = 1/N, \quad i = 1, 2, \dots, N.$$

(b) *Escoger al azar dos objetos* entre N objetos significa que *cada uno de los pares* de objetos (sin considerar el orden) tiene la misma probabilidad de ser escogido que cualquier otro par. Por ejemplo, si debemos elegir dos objetos al azar de (a_1, a_2, a_3, a_4) , y luego obtener a_1 y a_2 es tan probable como obtener a_2 y a_3 , etc. Esta afirmación nos lleva inmediatamente a la cuestión de *cuántos pares* diferentes hay. Supongamos que hay K de tales pares. Entonces la probabilidad de cada par sería $1/K$. Muy pronto aprenderemos a calcular K .

(c) *Escoger al azar n objetos* ($n \leq N$) entre N objetos significa que *cada n -tuplo*, $a_{i_1}, a_{i_2}, \dots, a_{i_n}$, tiene tantas probabilidades de ser escogido como cualquier otro n -tuplo.

Observación: anteriormente ya sugerimos que se debe tener mucho cuidado en la parte experimental para asegurar que la suposición matemática de «escoger al azar» se cumpla.

2.3 Métodos de enumeración

Tenemos que hacer un alto a fin de aprender cómo enumerar. Nuevamente consideraremos la forma anterior de $P(A)$ llamada $P(A) = r/k$, en donde k es el número total de maneras en que ε puede ocurrir mientras que r es igual al número de maneras en que A puede ocurrir. En los ejemplos presentados hasta ahora, se encontró poca dificultad para calcular r y k . Pero es necesario considerar situaciones sólo ligeramente más complicadas para apreciar la necesidad de contar sistemáticamente o de procedimientos de enumeración.

EJEMPLO 2.4. Un lote de cien artículos contiene 20 defectuosos y 80 no defectuosos. Se eligen diez artículos al azar, sin sustituir un artículo antes que sea elegido el próximo. ¿Cuál es la probabilidad de que exactamente la mitad de los artículos escogidos sean defectuosos?

Para analizar este problema, consideremos el siguiente espacio muestral S . Cada uno de los elementos de S consta de diez artículos posibles del lote, digamos $(i_1, i_2, \dots, i_{10})$. ¿Cuántos hay de tales resultados? Y entre estos resultados, ¿cuántos tienen la característica de que exactamente la mitad sean defectuosos? Evidentemente necesitamos saber contestar tales interrogantes para resolver el problema propuesto. Muchos problemas similares dan origen a preguntas análogas. En las próximas secciones presentaremos algunos procedimientos sistemáticos de enumeración.

A. Principio de multiplicación. Supongamos que un procedimiento, designado como 1 puede hacerse de n_1 maneras. Supongamos que un segundo procedimiento designado como 2, se puede hacer de n_2 maneras. También supongamos que cada una de las maneras de efectuar 1 puede ser seguida por cualquiera de las maneras de efectuar 2. Entonces el procedimiento que consta de 1 seguido por 2 se puede hacer de $n_1 n_2$ maneras.

Para indicar la validez de este principio es más sencillo considerar el siguiente enfoque esquemático. Consideraremos un punto P y dos rectas L_1 y L_2 . El procedimiento 1 consiste en ir de P a L_1 mientras que el procedimiento 2 consiste en ir de L_1 a L_2 . La figura 2.1 indica cómo se obtiene el resultado final.

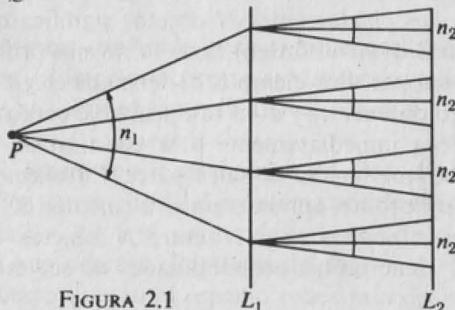


FIGURA 2.1

Observación: obviamente este principio puede extenderse a cualquier número de procedimientos. Si hay k procedimientos y el i -ésimo procedimiento se puede hacer de n_i maneras, $i = 1, 2, \dots, k$, entonces el procedimiento que consiste en 1, seguido por 2, ..., seguido por el procedimiento k puede hacerse en $n_1 n_2 \cdots n_k$ maneras.

EJEMPLO 2.5. Un artículo manufacturado debe pasar por tres controles. En cada uno de los controles, se inspecciona una característica particular del artículo y se la anota de conformidad. En el primer control, hay tres mediciones posibles mientras que en cada uno de los dos últimos controles hay cuatro mediciones posibles. Por lo tanto hay $3 \cdot 4 \cdot 4 = 48$ maneras de anotar el artículo.

B. Principio de adición. Supongamos que un procedimiento, designado como 1, se puede hacer de n_1 maneras. Supongamos que un segundo procedimiento, designado como 2, se puede hacer de n_2 maneras. Supongamos además que *no* es posible que *ambos*, 1 y 2, se hagan juntos. Entonces el número de maneras como se puede hacer 1 o 2 es $n_1 + n_2$.

Usemos otra vez el enfoque esquemático para convencernos de la validez del principio de adición, como se indica en la figura 2.2.

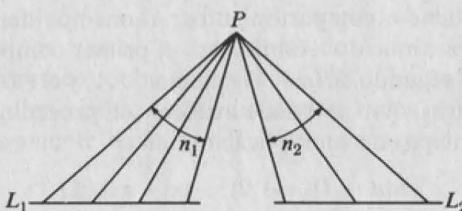


FIGURA 2.2

Observación: también este principio puede generalizarse como sigue: si hay k procedimientos y el i -ésimo procedimiento se puede hacer en n_i maneras, $i = 1, 2, \dots, k$, entonces el número de maneras como podemos hacer el procedimiento 1, el procedimiento 2 o ... o el procedimiento k está dado por $n_1 + n_2 + \dots + n_k$, suponiendo que los procedimientos no se pueden realizar conjuntamente.

EJEMPLO 2.6. Supongamos que proyectamos un viaje y debemos decidir entre el transporte por bus o tren. Si hay tres rutas para el bus y dos para el tren, entonces hay $3 + 2 = 5$ rutas disponibles para el viaje.

C. Permutaciones

(a) Supongamos que tenemos n objetos diferentes. ¿De cuántas maneras, digamos ${}_nP_n$, se pueden agrupar (permutar) estos objetos? Por ejemplo, si tenemos los objetos a, b , y c , podemos considerar las siguientes agrupaciones: abc, acb, bac, bca, cab y cba . Así la respuesta es 6. En general, consideraremos el esquema siguiente. Agrupar los n objetos es equivalente a ponerlos en una caja con n compartimentos, en algún orden específico.

1	2	.	.	.	n
---	---	---	---	---	-----

La primera casilla se puede llenar en una cualquiera de las n maneras, la segunda de cualquiera de las $(n - 1)$ maneras, ..., y la última casilla de sólo una manera. Por tanto, aplicando el principio de multiplicación anterior, vemos que

la caja se puede llenar de $n(n - 1)(n - 2) \cdots 1$ maneras. Este número ocurre tan a menudo en matemáticas que presentamos un nombre y un símbolo especiales para él.

Definición. Si n es un entero positivo, definimos $n! = (n)(n - 1)(n - 2) \cdots 1$ y lo llamamos n -factorial. También definimos $0! = 1$.

Así el número de permutaciones de n objetos diferentes está dado por

$${}_nP_n = n!$$

(b) Nuevamente consideremos n objetos diferentes. Esta vez deseamos escoger r de esos objetos, $0 \leq r \leq n$, y permutamos el r elegido. Indiquemos el número de maneras de hacerlo, por ${}_nP_r$. Recurrimos nuevamente al esquema anterior de llenar una caja que tiene n compartimientos; ahora nos detenemos después que se ha llenado el compartimiento r -ésimo. Así, el primer compartimiento puede llenarse de n maneras, el segundo de $(n - 1)$ maneras, ..., y el r -ésimo compartimiento de $n - (r - 1)$ maneras. Así se puede realizar el procedimiento completo; de nuevo usando el principio de multiplicación, de

$$n(n - 1)(n - 2) \cdots (n - r + 1)$$

maneras. Usando la notación factorial introducida anteriormente, podemos escribir

$${}_nP_r = \frac{n!}{(n - r)!}.$$

D. Combinaciones. Consideremos nuevamente n objetos diferentes. Esta vez estamos interesados en contar el número de maneras como podemos escoger r de esos n objetos sin considerar el orden. Por ejemplo, tenemos los objetos a, b, c , y d , y $r = 2$; deseamos contar ab, ac, ad, bc, bd , y cd . En otras palabras, no contamos ab y ba puesto que los mismos objetos están relacionados y sólo difiere el orden.

Para obtener el resultado general recordemos la fórmula derivada anteriormente: el número de maneras de elegir r objetos entre n y permutar los r elegidos es igual a $n!/(n - r)!$ Sea C el número de maneras de elegir r entre n , sin considerar el orden. (Esto es, el número buscado es C). Observe que una vez que se han escogido los r artículos, hay $r!$ maneras de permutarlos. Por tanto, aplicando nuevamente el principio de multiplicación junto con el resultado anterior, obtenemos

$$Cr! = \frac{n!}{(n - r)!}.$$

Así el número de maneras de elegir r entre n objetos diferentes, sin considerar el orden, está dado por

$$C = \frac{n!}{r!(n - r)!}.$$

* *N. del T.* Esta expresión se conoce también como arreglo o variación.

Este número aparece en muchos contextos en matemáticas y, por lo tanto, se emplea un símbolo especial para designarlo. Escribiremos

$$\frac{n!}{r!(n-r)!} = \binom{n}{r}.$$

Para nuestros propósitos, $\binom{n}{r}$ se define sólo si n es un entero positivo y si r es un entero $0 \leq r \leq n$. Sin embargo, podemos definir $\binom{n}{r}$ muy ampliamente para cualquier número real n y para cualquier entero no negativo r como sigue:

$$\binom{n}{r} = \frac{n(n-1)(n-2)\cdots(n-r+1)}{r!}.$$

Los números $\binom{n}{k}$ a menudo se llaman *coeficientes binomiales*, porque aparecen como coeficientes en el desarrollo de la expresión binomial $(a+b)^n$. Si n es un entero positivo, $(a+b)^n = (a+b)(a+b)\cdots(a+b)$. Cuando se efectúa la multiplicación, cada uno de los términos consta de el producto de k aes y $(n-k)$ bees, $k = 0, 1, 2, \dots, n$. ¿Cuántos términos serán de la forma $a^k b^{n-k}$? Contemos simplemente el número de maneras como podemos elegir k entre n aes, sin considerar el orden. Pero esto está precisamente dado por $\binom{n}{k}$. Así, tenemos lo que se conoce como *teorema del binomio*

$$(a+b)^n = \sum_{k=0}^n \binom{n}{k} a^k b^{n-k}. \quad (2.2)$$

Los números $\binom{n}{r}$ tienen muchas propiedades interesantes de las cuales sólo dos mencionaremos aquí. (A no ser que se indique otra cosa, suponemos que n es un entero positivo y r un entero, $0 \leq r \leq n$.)

$$(a) \quad \binom{n}{r} = \binom{n}{n-r},$$

$$(b) \quad \binom{n}{r} = \binom{n-1}{r-1} + \binom{n-1}{r}.$$

Es muy fácil verificar algebraicamente las dos identidades anteriores. Simplemente escribimos, en cada uno de los casos, el lado izquierdo y derecho de las identidades anteriores y notamos que son iguales.

Sin embargo, hay otro método de verificar esas identidades que hace uso de la interpretación que hemos dado a $\binom{n}{r}$, llamada el número de maneras de escoger r entre n objetos.

(a) Cuando escogemos r entre n objetos simultáneamente “dejamos atrás” $(n-r)$ objetos y, por tanto, escoger r de n es equivalente a escoger $(n-r)$ de n . Esta es precisamente la primera proposición que se debe verificar.

(b) Escojamos cualquiera de los n objetos, sea este el primero, a_1 . Al elegir r objetos, a_1 está incluido o excluido pero no las dos cosas. Por tanto al contar el número de maneras como podemos escoger r objetos, podemos aplicar el principio de adición tratado al comienzo.

Si a_1 está excluido, debemos escoger los r objetos que se desean de los $(n - 1)$ objetos restantes y hay $\binom{n-1}{r}$ maneras de hacer esto.

Si a_1 va a incluirse, sólo $(r - 1)$ objetos más deben escogerse de los restantes $(n - 1)$ objetos y esto se puede hacer de $\binom{n-1}{r-1}$ maneras. Así el número pedido es la suma de esos dos, lo que verifica la segunda identidad.

Observación: en lo expuesto anteriormente los coeficientes binomiales $\binom{n}{k}$ son significativos sólo si n y k son enteros no negativos con $0 \leq k \leq n$. Sin embargo, si escribimos

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n(n-1)\cdots(n-k+1)}{k!}$$

observamos que la última expresión es significativa si n es cualquier número real y k es cualquier entero no negativo. Así,

$$\binom{-3}{5} = \frac{(-3)(-4)\cdots(-7)}{5!},$$

y así sucesivamente.

Utilizando esta versión extendida de los coeficientes binomiales, podemos indicar la *forma generalizada del teorema del binomio*:

$$(1 + x)^n = \sum_{k=0}^{\infty} \binom{n}{k} x^k$$

Esta serie es significativa para cualquier n real y para todas las x tales como $|x| < 1$. Observamos que si n es un entero positivo, las series infinitas se reducen a un número finito de términos puesto que en ese caso $\binom{n}{k} = 0$ si $k > n$.

EJEMPLO 2.7. (a) ¿Cuántos comités de tres miembros se pueden elegir con ocho personas? Puesto que dos comités son el mismo si están formados por los mismos miembros (sin considerar en el orden en el cual fueron elegidos), tenemos $\binom{8}{3} = 56$ comités posibles.

(b) ¿Cuántas señales con tres banderas pueden obtenerse con ocho banderas diferentes? Este problema se parece mucho al anterior. Sin embargo, aquí el orden constituye una diferencia, y, por lo tanto, obtenemos $8!/5! = 336$ señales.

(c) Un grupo de ocho personas consta de cinco hombres y tres mujeres. ¿Cuántos comités que consten de dos hombres exactamente se pueden formar? Aquí debemos hacer dos cosas: escoger dos hombres (entre cinco) y escoger una mujer (entre tres). Así obtenemos el número pedido $\binom{5}{2} \cdot \binom{3}{1} = 30$ comités.

(d) Ahora podemos verificar una proposición formulada anteriormente, decíamos que el número de subconjuntos de un conjunto que tiene n elementos es 2^n (contando el conjunto vacío y el conjunto mismo). Simplemente clasificamos cada elemento con un uno o con un cero, sea que el elemento vaya a ser incluido, o excluido en el subconjunto. Hay dos maneras de clasificar cada uno de los elementos, y hay n elementos. Luego el principio de multiplicación nos dice que hay $2 \cdot 2 \cdot 2 \cdots 2 = 2^n$ clasificaciones posibles. Pero cada una de las clasificaciones en particular representa una elección de un subconjunto. Por ejemplo, $(1, 1, 0, 0, 0, \dots, 0)$ consistiría en el subconjunto formado por a_1 y a_2 pre-

cisamente. Nuevamente, $(1, 1, \dots, 1)$ representaría S y $(0, 0, \dots, 0)$ representaría al conjunto vacío.

(e) Podemos obtener el resultado anterior usando el principio de adición como sigue. Para obtener subconjuntos debemos escoger el conjunto vacío, los subconjuntos que constan de sólo un elemento, los que constan de sólo 2 elementos, ..., y el conjunto mismo que consta de todos los n elementos. Esto se puede hacer de

$$\binom{n}{0} + \binom{n}{1} + \binom{n}{2} + \dots + \binom{n}{n}$$

maneras. Sin embargo, la suma de esos coeficientes binomiales es simplemente el desarrollo de $(1 + 1)^n = 2^n$.

Ahora volvamos al ejemplo 2.4. De un lote que consta de 80 artículos buenos y 20 artículos defectuosos, escogemos 10 al azar (sin sustitución). El número de maneras de hacer esto es $\binom{100}{10}$. Por tanto la probabilidad de encontrar 5 artículos defectuosos y 5 no defectuosos entre los 10 elegidos está dada por

$$\frac{\binom{20}{5} \binom{80}{5}}{\binom{100}{10}}$$

Mediante logaritmos de factoriales (que están tabulados) se puede evaluar lo anterior y es igual a 0,021.

EJEMPLO 2.8. Generalicemos el problema anterior. Supóngase que tenemos N artículos. Si elegimos n de esos al azar, sin sustitución, hay $\binom{N}{n}$ muestras posibles diferentes, todas las cuales tienen la misma probabilidad de ser escogidas. Si los N artículos están formados de r_1 , A y r_2 B (con $r_1 + r_2 = N$), entonces la probabilidad de que los n artículos elegidos contengan exactamente s_1 A y $(n - s_1)$ B está dada por

$$\frac{\binom{r_1}{s_1} \binom{r_2}{n - s_1}}{\binom{N}{n}}$$

(La anterior se llama probabilidad hipergeométrica y se encontrará de nuevo.)

Observación: es muy importante especificar, cuando hablamos de escoger artículos al azar, si escogemos con o sin sustitución. En una descripción más realista propondremos esta última. Por ejemplo, cuando inspeccionamos un número de artículos manufacturados con el objeto de descubrir cuantos defectuosos podría haber, generalmente no pretendemos inspeccionar el mismo artículo dos veces. Previamente hemos observado que el número de maneras de escoger r objetos entre n , sin considerar el orden, está dado por $\binom{n}{r}$. El número de maneras de escoger r artículos entre n , con sustitución, está dado por n^r . Aquí estamos interesados en el orden en que se escogieron los artículos.

EJEMPLO 2.9. Supóngase que escogemos dos objetos al azar entre cuatro objetos clasificados a, b, c , y d .

(a) Si escogemos sin sustitución, el espacio muestral S se puede representar como sigue:

$$S = \{(a, b); (a, c); (b, c); (b, d); (c, d); (a, d)\}.$$

Hay $\binom{4}{2} = 6$ resultados posibles. Cada uno de los resultados individuales indica sólo cuáles fueron los dos objetos escogidos y no el orden en que se escogieron.

(b) Si escogemos con sustitución, el espacio muestral S' , se puede representar como sigue:

$$S' = \left\{ \begin{array}{l} (a, a); (a, b); (a, c); (a, d); (b, a); (b, b); (b, c); (b, d); \\ (c, a); (c, b); (c, c); (c, d); (d, a); (d, b); (d, c); (d, d) \end{array} \right\}.$$

Hay $4^2 = 16$ resultados posibles. Aquí cada uno de los resultados individuales indica cuáles objetos se escogieron y el orden de selección. Escoger al azar implica que si escogemos sin sustitución, todos los resultados en S son igualmente probables, mientras que si escogemos con sustitución, entonces todos los resultados en S' son igualmente probables. Así, si A es el suceso {el objeto c es elegido}, entonces tenemos de S , $P(A) = \frac{3}{6} = \frac{1}{2}$ si escogemos sin sustitución, y de S' , $P(A) = \frac{7}{16}$ si escogemos con sustitución.

E. Permutaciones cuando no todos los objetos son diferentes. En todos los métodos de enumeración presentados, hemos supuesto que todos los objetos considerados eran diferentes (esto es, distinguibles). Sin embargo, no siempre es este el caso.

Supongamos, entonces, que tenemos n objetos tales que hay n_1 de una clase, n_2 de una segunda clase, ..., n_k de una k -ésima clase, en donde $n_1 + n_2 + \dots + n_k = n$. Entonces el número de permutaciones de esos objetos está dada por

$$\frac{n!}{n_1! n_2! \cdots n_k!}.$$

La deducción de esta fórmula se deja al lector. Nótese que si todos los objetos son diferentes, tenemos $n_i = 1$, $i = 1, 2, \dots, k$, y, por tanto, la fórmula anterior se reduce a $n!$, el resultado obtenido previamente.

Observación: insistimos una vez más en que la asignación real de probabilidades a los resultados individuales de un espacio muestral (o a una colección de resultados, es decir, un suceso) es algo que no puede obtenerse matemáticamente; debe obtenerse de otras consideraciones. Por ejemplo, podemos utilizar ciertas características de simetría del experimento para observar que todos los resultados son igualmente probables. Nuevamente podemos hacer un método de muestreo (es decir, escoger uno o varios individuos de una población especificada) de tal manera que sea razonable suponer que todas las elecciones son igualmente probables. En muchos otros casos, cuando ninguna suposición básica es apropiada, debemos recurrir al enfoque de la frecuencia relativa. Repetimos n veces el experimento y anotamos la proporción de veces que ha ocurrido el resultado (o suceso) que se considera. Al usar ésta como una aproximación, sabemos que es altamente improbable que esta frecuencia relativa

se diferencie de la probabilidad «verdadera» (cuya existencia ha sido especificada por nuestro modelo teórico) en una cantidad apreciable si n es suficientemente grande. Cuando es imposible hacer suposiciones razonables acerca de la probabilidad de un resultado y es también imposible repetir el experimento un gran número de veces (debido al costo o a consideraciones de tiempo, por ejemplo) es realmente muy poco significativo proseguir con un estudio probabilístico del experimento excepto sobre una base completamente teórica. (Para una nota adicional sobre el mismo tema, ver sección 3.5.)

PROBLEMAS

2.1. En una habitación se encuentra el siguiente grupo de personas: 5 hombres mayores de 21, 4 hombres menores de 21, 6 mujeres mayores de 21 y 3 mujeres menores de 21. Se elige una persona al azar. Se definen los sucesos siguientes: $A = \{\text{la persona es mayor de 21}\}$; $B = \{\text{la persona es menor de 21}\}$; $C = \{\text{la persona es hombre}\}$; $D = \{\text{la persona es mujer}\}$. Evaluar las siguientes:

- (a) $P(B \cup D)$
- (b) $P(\bar{A} \cup \bar{C})$

2.2. En una habitación 10 personas tienen insignias numeradas del 1 al 10. Se eligen tres personas al azar y se les pide que dejen la habitación inmediatamente y se anotan los números de las insignias.

- (a) ¿Cuál es la probabilidad de que el número menor de las insignias sea 5?
- (b) ¿Cuál es la probabilidad de que el número mayor de las insignias sea 5?

2.3. (a) Supóngase que se escriben tres dígitos 1, 2 y 3 en un orden aleatorio. ¿Cuál es la probabilidad de que al menos un dígito ocupe su lugar propio?

- (b) Lo mismo que (a) con los dígitos 1, 2, 3, y 4.
- (c) Lo mismo que (a) con los dígitos 1, 2, 3, ..., n . [Indicación: usar (1.7).]
- (d) Discutir la respuesta de (c) si n es grande.

2.4. Un cargamento de 1500 lavadoras contiene 400 defectuosas y 1100 no defectuosas. Se eligen al azar doscientas lavadoras (sin sustitución) y se clasifican.

- (a) ¿Cuál es la probabilidad de que se encuentren exactamente 90 artículos defectuosos?
- (b) ¿Cuál es la probabilidad de que se encuentren al menos 2 artículos defectuosos?

2.5. Diez fichas numeradas del 1 al 10 se mezclan en una palangana. Se sacan de la palangana dos fichas numeradas (X, Y) una y otra vez sin sustitución. ¿Cuál es la probabilidad de que $X + Y = 10$?

2.6. Un lote consta de 10 artículos buenos, 4 con pequeños defectos, y 2 con defectos graves. Se elige un artículo al azar. Encontrar la probabilidad de que:

- (a) no tenga defectos,
- (b) tenga un defecto grave,
- (c) que sea bueno o que tenga un defecto grave.

2.7. Si del mismo lote de artículos descritos en el problema 2.6 se escogen dos artículos (sin sustitución) encuentre la probabilidad de que:

- | | |
|--------------------------------|------------------------------------|
| (a) ambos sean buenos, | (b) ambos tengan defectos graves, |
| (c) a lo menos uno sea bueno, | (d) a lo más uno sea bueno, |
| (e) exactamente uno sea bueno, | (f) ninguno tenga defectos graves, |
| | (g) ninguno sea bueno. |

2.8. Un producto se arma en tres etapas. En la primera etapa hay 5 líneas de armado, en la segunda etapa hay 4 líneas de armado, y en la tercera etapa hay 6 líneas de armado. ¿De cuántas maneras puede moverse el producto en el proceso de armado?

2.9. Un inspector visita 6 máquinas diferentes durante el día. A fin de impedir a los operadores que sepan cuándo inspeccionará, varía el orden de las visitas. ¿De cuántas maneras puede hacerlo?

2.10. Un mecanismo complejo puede fallar en 15 partes diferentes. Si falla en 3 partes, ¿de cuántas maneras puede suceder?

2.11. Hay 12 maneras en las cuales un artículo manufacturado puede tener un pequeño defecto y 10 maneras en las cuales puede tener un defecto mayor. ¿De cuántas maneras puede ocurrir un defecto menor y uno mayor? ¿2 defectos menores y 2 defectos mayores?

2.12. Un mecanismo puede ponerse en cuatro posiciones digamos, *a, b, c, y d*. Hay 8 de tales mecanismos en un sistema.

(a) ¿De cuántas maneras puede instalarse este sistema?

(b) Supóngase que dichos mecanismos estén instalados en algún orden (lineal) pre-asignado. ¿De cuántas maneras posibles se instalan los mecanismos si dos mecanismos adyacentes no están en la misma posición?

(c) ¿Cuántas maneras son posibles si sólo se usan las posiciones *a* y *b* con la misma frecuencia?

(d) ¿Cuántas maneras son posibles si sólo se usan dos posiciones diferentes y una de ellas aparece tres veces más a menudo que la otra?

2.13. Supongamos que de *N* objetos elegimos *n* al azar, con sustitución. ¿Cuál es la probabilidad de que ningún objeto sea elegido más de una vez? (Suponga que *n* < *N*.)

2.14. De las letras *a, b, c, d, e, y f*, ¿cuántas palabras de clave de 4 letras se pueden formar si,

(a) ninguna letra se puede repetir?

(b) cualquier letra se puede repetir cualquier número de veces?

2.15. Suponga que $\binom{99}{5} = a$ y $\binom{99}{4} = b$. Expresar $\binom{100}{95}$ en función de *a* y *b*. [Indicación: no calcule las expresiones anteriores para resolver este problema.]

2.16. Una caja contiene esferas numeradas 1, 2, ..., *n*. Se escogen dos esferas al azar. Encontrar la probabilidad de que los números sobre las esferas sean enteros consecutivos si

(a) las esferas se escogen sin sustitución,

(b) las esferas se escogen con sustitución.

2.17. ¿Cuántos subconjuntos que contengan al menos un elemento se pueden formar de un conjunto de 100 elementos?

2.18. Entre los números 1, 2, ..., 50 se escoge un número al azar. ¿Cuál es la probabilidad de que el número escogido sea divisible por 6 o por 8?

2.19. De 6 números positivos y 8 números negativos, se eligen 4 números al azar (sin sustitución) y se multiplican. ¿Cuál es la probabilidad de que el producto sea un número positivo?

2.20. Cierta sustancia química se forma mezclando 5 líquidos distintos. Se propone verter un líquido en un estanque y agregar sucesivamente los otros líquidos. Todas las combinaciones posibles se deben probar para establecer cuál da mejor resultado. ¿Cuántas pruebas deben hacerse?

2.21. Un lote contiene n artículos. Si se sabe que r artículos son defectuosos y se inspeccionan en un orden aleatorio, ¿cuál es la probabilidad de que el k -ésimo artículo ($k \geq r$) inspeccionado sea el último defectuoso en el lote?

2.22. r números ($0 < r < 10$) se escogen al azar (sin sustitución) entre los números 0, 1, 2, ..., 9. ¿Cuál es la probabilidad de que dos no sean iguales?

Probabilidad condicional e independencia

3.1 Probabilidad condicional

Consideremos de nuevo, la diferencia que existe entre elegir al azar en un lote, un artículo, con o sin sustitución. En el ejemplo 2.4, el lote que considerábamos tenía la siguiente composición: 80 artículos sin defectos y 20 defectuosos. Supóngase que elijamos dos artículos de este lote: (a) con sustitución; (b) sin sustitución.

Definamos los sucesos siguientes:

$$\begin{aligned} A &= \{\text{el primer artículo es defectuoso}\}, \\ B &= \{\text{el segundo artículo es defectuoso}\}. \end{aligned}$$

Si escogemos con sustitución, $P(A) = P(B) = \frac{20}{100} = \frac{1}{5}$. Cada vez que elegimos, en el lote hay 20 artículos defectuosos de un total de 100. Sin embargo, si elegimos sin sustitución, los resultados no son completamente inmediatos. Todavía es verdad, naturalmente, que $P(A) = \frac{1}{5}$. Pero, ¿cuál es el valor de $P(B)$? Es evidente que con el fin de calcular $P(B)$ deberíamos conocer la composición del lote, cuando se escoge el segundo artículo. En otras palabras, deberíamos saber si el suceso A ocurrió o no. Este ejemplo indica la necesidad de presentar el siguiente concepto importante.

Sean A y B dos sucesos asociados con un experimento ε . Indiquemos con $P(B | A)$ la probabilidad condicional del suceso B , dado que A ha ocurrido.

En el ejemplo anterior, $P(B | A) = \frac{19}{99}$. Porque si A ha ocurrido, entonces al sacar por segunda vez quedan sólo 99 artículos, de los cuales 19 son defectuosos.

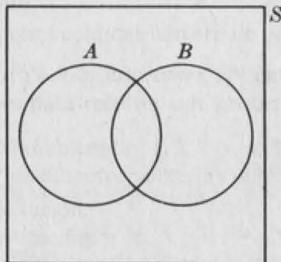


FIGURA 3.1

Cada vez que calculamos $P(B|A)$ estamos esencialmente calculando $P(B)$ con respecto al espacio muestral reducido de A en vez del espacio muestral original S . Consideremos el diagrama de Venn en la figura 3.1.

Cuando calculamos $P(B)$ nos preguntamos qué tan probable es que estemos en B , sabiendo que debemos estar en S , y cuando evaluamos $P(B|A)$ nos preguntamos qué tan probable es que estemos en B , sabiendo que debemos estar en A . (Esto es, el espacio muestral se ha *reducido* de S a A .)

Pronto daremos una definición formal de $P(B|A)$. Por el momento, sin embargo, seguiremos con nuestra noción intuitiva de probabilidad condicional y consideraremos un ejemplo:

EJEMPLO 3.1. Se lanzan dos dados normales y se anotan los resultados (x_1, x_2) en donde x_i es el resultado del i -ésimo dado $i = 1, 2$. Por tanto, el espacio muestral S se puede representar por el siguiente cuadro de 36 resultados igualmente posibles:

$$S = \left\{ \begin{array}{cccc} (1, 1) & (1, 2) & \cdots & (1, 6) \\ (2, 1) & (2, 2) & \cdots & (2, 6) \\ \vdots & & & \vdots \\ (6, 1) & (6, 2) & \cdots & (6, 6) \end{array} \right\}.$$

Consideremos los dos sucesos siguientes:

$$A = \{(x_1, x_2) \mid x_1 + x_2 = 10\}, \quad B = \{(x_1, x_2) \mid x_1 > x_2\}.$$

Así $A = \{(5, 5), (4, 6), (6, 4)\}$ y $B = \{(2, 1), (3, 1), (3, 2), \dots, (6, 5)\}$. Por tanto, $P(A) = \frac{3}{36}$ y $P(B) = \frac{15}{36}$. Además $P(B|A) = \frac{1}{3}$, ya que el espacio muestral es ahora A , (que son tres resultados) y sólo uno de ellos es consistente con el suceso B . De una manera semejante, nosotros podemos calcular $P(A|B) = \frac{1}{15}$.

Finalmente, calculemos $P(A \cap B)$. El suceso $(A \cap B)$ ocurre, si y sólo si la suma de los dos dados es 10 y el primer dado indica un valor mayor que el segundo dado. Hay solamente *un* resultado y, por tanto $P(A \cap B) = \frac{1}{36}$. Si observamos cuidadosamente los diversos números que hemos calculado anteriormente, deducimos que se cumple lo siguiente:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad \text{y} \quad P(B|A) = \frac{P(A \cap B)}{P(A)}.$$

Sucede que esas relaciones no solamente aparecen en los ejemplos particulares que hemos considerado, sino que son completamente generales y nos dan un medio de *definir formalmente* la probabilidad condicional.

Para justificar esta definición volvamos al concepto de frecuencia relativa. Supongamos que se ha repetido un experimento n veces. Sean n_A , n_B y $n_{A \cap B}$ el número respectivo de veces que los sucesos A , B , y $A \cap B$ han ocurrido en las n repeticiones. ¿Cuál es el significado de $n_{A \cap B}/n_A$? Representa la frecuencia relativa de B entre esos resultados en los que A ocurrió. Esto es, $n_{A \cap B}/n_A$ es la frecuencia relativa condicional de B , dado que A ocurrió.

Podemos escribir $n_{A \cap B}/n_A$ como sigue:

$$\frac{n_{A \cap B}}{n_A} = \frac{n_{A \cap B}/n}{n_A/n} = \frac{f_{A \cap B}}{f_A},$$

en donde $f_{A \cap B}$ y f_A son las frecuencias relativas de los sucesos $A \cap B$ y A , respectivamente. Como ya lo hemos indicado (y como lo demostraremos más adelante), si n , el número de repeticiones es grande, $f_{A \cap B}$ estará próximo a $P(A \cap B)$ y f_A estará próxima a $P(A)$. Por lo tanto, la relación anterior sugiere que $n_{A \cap B}/n_A$ estará próxima a $P(B|A)$. Así hacemos la siguiente definición formal:

Definición.

$$P(B|A) = \frac{P(A \cap B)}{P(A)}, \quad \text{dado que } P(A) > 0. \quad (3.1)$$

Observaciones: (a) Es importante darse cuenta de que lo anterior no es un teorema (no demostramos nada), ni tampoco un axioma. Simplemente presentamos la noción intuitiva de probabilidad condicional y luego hicimos la definición formal de lo que entendemos por esta noción. El hecho de que nuestra definición formal corresponde a nuestra noción intuitiva está comprobado por el párrafo que precede a la definición.

(b) Es muy sencillo comprobar que $P(B|A)$ para un valor de A fijo, satisface los diversos postulados de la probabilidad ecuación (1.3). (Ver problema 3.22.) Esto es, tenemos:

- (1') $0 \leq P(B|A) \leq 1$,
- (2') $P(S|A) = 1$,
- (3') $P(B_1 \cup B_2|A) = P(B_1|A) + P(B_2|A) \quad \text{si } B_1 \cap B_2 = \emptyset$,
- (4') $P(B_1 \cup B_2 \cup \dots|A) = P(B_1|A) + P(B_2|A) + \dots \quad \text{si } B_i \cap B_j = \emptyset$
para $i \neq j$.

(c) Si $A = S$, $P(B|S) = P(B \cap S)/P(S) = P(B)$.

(d) Con cada suceso $B \subset S$ podemos asociar dos números, $P(B)$, la probabilidad (no condicional) de B , y $P(B|A)$, la probabilidad condicional de B , dado que algún suceso A (para el cual $P(A) > 0$) ha ocurrido. En general, esas dos medidas de probabilidad asignarán probabilidades distintas al suceso B , como se indicó en los ejemplos precedentes. En breve estudiaremos un importante caso especial en el cual $P(B)$ y $P(B|A)$ son la misma.

(e) Nótese que la probabilidad condicional está definida en términos de la medida de probabilidad no condicional P . Esto es, si conocemos $P(B)$ para cada $B \subset S$ podemos calcular $P(B|A)$ para cada $B \subset S$.

Así tenemos dos maneras de calcular la probabilidad condicional $P(B|A)$:

(a) Directamente considerando la probabilidad de B con respecto al espacio muestral reducido A .

(b) Usando la definición anterior, donde $P(A \cap B)$ y $P(A)$ se calculan con respecto al espacio muestral original S .

Observación: si $A = S$, obtenemos $P(B|S) = P(B \cap S)/P(S) = P(B)$, puesto que $P(S) = 1$ y $B \cap S = B$. Así debe ser, porque decir que S ha ocurrido, sólo indica que el experimento ha sido realizado.

EJEMPLO 3.2. Supóngase que una oficina tiene 100 máquinas calculadoras. Algunas de esas máquinas son eléctricas (E), mientras que otras son manuales (M). Además, algunas son nuevas (N), mientras las otras son usadas (U). La tabla 3.1 da el número de máquinas de cada categoría. Una persona entra en la

TABLA 3.1

	E	M	
N	40	30	70
U	20	10	30
	60	40	100

oficina, escoge una máquina al azar, y descubre que es nueva. ¿Cuál es la probabilidad de que sea eléctrica? En términos de la notación introducida deseamos calcular $P(E | N)$.

Sólo considerando el espacio muestral reducido N (es decir, las 70 máquinas nuevas), tenemos que: $P(E | N) = \frac{40}{70} = \frac{4}{7}$. Usando la definición de probabilidad condicional, tenemos que:

$$P(E | N) = \frac{P(E \cap N)}{P(N)} = \frac{40/100}{70/100} = \frac{4}{7}.$$

La consecuencia más importante de la definición de probabilidad condicional se obtiene escribiéndola de la manera siguiente:

$$P(A \cap B) = P(B | A)P(A)$$

lo que equivale a,

(3.3a)

$$P(A \cap B) = P(A | B)P(B).$$

Esto a veces se conoce como el teorema de multiplicación de probabilidades.

Podemos aplicar este teorema al cálculo de la probabilidad de la ocurrencia simultánea de los dos sucesos A y B .

EJEMPLO 3.3. Consideremos otra vez el lote discutido al comienzo de la sección 3.1, el cual consta de 20 artículos defectuosos y 80 sin defectos. Si escogemos 2 artículos al azar, sin sustitución, ¿cuál es la probabilidad de que ambos artículos sean defectuosos?

Como antes, definimos los sucesos A y B como sigue:

$$A = \{\text{el primer artículo es defectuoso}\}, \quad B = \{\text{el segundo artículo es defectuoso}\}.$$

Por lo tanto, necesitamos $P(A \cap B)$, que puede calcularse de acuerdo con la fórmula anterior, como $P(B | A)P(A)$. Pero $P(B | A) = \frac{19}{99}$ mientras que $P(A) = \frac{1}{5}$. Por lo tanto

$$P(A \cap B) = \frac{19}{495}$$

Observación: se puede generalizar el anterior teorema de la multiplicación a más de dos sucesos de la siguiente manera:

$$P[A_1 \cap A_2 \cap \cdots \cap A_n] = P(A_1)P(A_2 | A_1)P(A_3 | A_1, A_2) \cdots P(A_n | A_1, \dots, A_{n-1}). \quad (3.3b)$$

Consideremos por un momento si podemos hacer una afirmación general acerca de la magnitud relativa de $P(A | B)$ y $P(A)$. Consideremos los 4 casos ilustrados por los diagramas de Venn en la figura 3.2. Tenemos:

- (a) $P(A | B) = 0 \leq P(A)$ puesto que A no puede ocurrir si B ha ocurrido.
- (b) $P(A | B) = P(A \cap B)/(P(B)) = [P(A)/P(B)] \geq P(A)$, puesto que $0 \leq P(B) \leq 1$.
- (c) $P(A | B) = P(A \cap B)/P(B) = P(B)/P(B) = 1 \geq P(A)$.
- (d) En este caso no podemos hacer ninguna afirmación acerca de la magnitud relativa de $P(A | B)$ y $P(A)$.

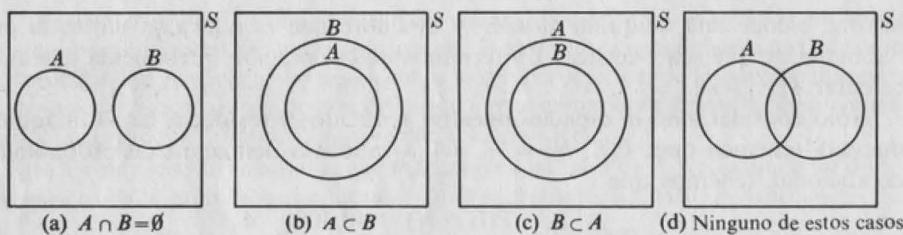


FIGURA 3.2

Nótese que en dos de los casos anteriores $P(A) \leq P(A | B)$, en un caso, $P(A) \geq P(A | B)$, y en el cuarto caso no podemos hacer ninguna clase de comparaciones.

Anteriormente, usamos el concepto de probabilidad condicional, con el fin de evaluar la probabilidad de la ocurrencia simultánea de los dos sucesos. Podemos aplicar este concepto de otra manera para calcular la probabilidad de un solo suceso A . Necesitamos la siguiente definición.

Definición. Decimos que los sucesos B_1, B_2, \dots, B_k representan una partición del espacio muestral S si:

- (a) $B_i \cap B_j = \emptyset$ para todo $i \neq j$.
- (b) $\bigcup_{i=1}^k B_i = S$.
- (c) $P(B_i) > 0$ para todo i .

En otras palabras: cuando se efectúa el experimento ε , ocurre *uno y sólo* uno de los sucesos B_i .

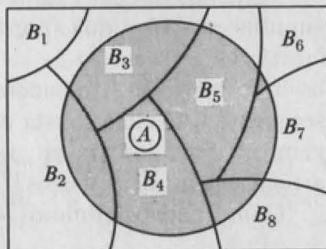


FIGURA 3.3

(Por ejemplo, en el lanzamiento de un dado $B_1 = \{1, 2\}$, $B_2 = \{3, 4, 5\}$, y $B_3 = \{6\}$ representarían una partición del espacio muestral, mientras que $C_1 = \{1, 2, 3, 4\}$ y $C_2 = \{4, 5, 6\}$ no).

Sea A algún suceso con respecto a S y sea B_1, B_2, \dots, B_k una partición de S . El diagrama de Venn de la figura 3.3 ilustra esto para $k = 8$. Por tanto, podemos escribir

$$A = A \cap B_1 \cup A \cap B_2 \cup \dots \cup A \cap B_k.$$

Por cierto que algunos de los conjuntos $A \cap B_j$ pueden ser vacíos, pero esto no invalida la anterior descomposición de A . Lo importante es que todos los sucesos $A \cap B_1, \dots, A \cap B_k$ son mutuamente excluyentes. Por lo tanto, podemos aplicar la propiedad aditiva para este tipo de sucesos (ecuación 1.3) y escribir:

$$P(A) = P(A \cap B_1) + P(A \cap B_2) + \dots + P(A \cap B_k).$$

Sin embargo, cada término $P(A \cap B_j)$ se puede expresar como $P(A | B_j)P(B_j)$ y, por lo tanto, obtenemos el llamado teorema de la probabilidad total:

$$P(A) = P(A | B_1)P(B_1) + P(A | B_2)P(B_2) + \dots + P(A | B_k)P(B_k). \quad (3.4)$$

Este resultado representa una relación muy útil, ya que frecuentemente cuando se busca $P(A)$ puede ser difícil calcularlo directamente. Sin embargo, con la información adicional de que B_j ha ocurrido, podemos calcular $P(A | B_j)$ y entonces usar la fórmula anterior.

EJEMPLO 3.4. Consideremos (por última vez) el lote de 20 artículos defectuosos y 80 sin defectos, de los cuales escogemos 2 artículos *sin sustitución*. Nuevamente definimos A y B :

$$\begin{aligned} A &= \{\text{el primer artículo elegido es defectuoso}\}, \\ B &= \{\text{el segundo artículo elegido es defectuoso}\}, \end{aligned}$$

podemos ahora calcular $P(B)$ como sigue:

$$P(B) = P(B | A)P(A) + P(B | \bar{A})P(\bar{A}).$$

Usando uno de los cálculos ya hechos en el ejemplo 3.3 encontramos que

$$P(B) = \frac{19}{99} \cdot \frac{1}{5} + \frac{20}{99} \cdot \frac{4}{5} = \frac{1}{5}.$$

Este resultado puede ser un poco sorprendente, particularmente si el lector recuerda que al comienzo de la sección 3.1 encontramos que $P(B) = \frac{1}{5}$ cuando escogemos los artículos *con sustitución*.

EJEMPLO 3.5. Cierto artículo es manufacturado por tres fábricas, sean 1, 2 y 3. Se sabe que la primera produce el doble de artículos que la segunda y que ésta y la tercera producen el mismo número de artículos (durante un período de producción especificado). Se sabe también que el 2 por ciento de los artículos producidos por las dos primeras es defectuoso mientras que el 4 por ciento de los manufac-

turados por la tercera es defectuoso. Se colocan juntos todos los artículos producidos en una fila y se escoge uno al azar. ¿Cuál es la probabilidad de que este artículo sea defectuoso?

Definamos los siguientes sucesos:

$$A = \{\text{el artículo es defectuoso}\}, \quad B_1 = \{\text{el artículo proviene de 1}\},$$

$$B_2 = \{\text{el artículo proviene de 2}\}, \quad B_3 = \{\text{el artículo proviene de 3}\}.$$

Nosotros necesitamos $P(A)$ y, usando el resultado anterior, podemos escribir:

$$P(A) = P(A | B_1)P(B_1) + P(A | B_2)P(B_2) + P(A | B_3)P(B_3).$$

Ahora $P(B_1) = \frac{1}{2}$, mientras que $P(B_2) = P(B_3) = \frac{1}{4}$. También $P(A | B_1) = P(A | B_2) = 0,02$, mientras que $P(A | B_3) = 0,04$. Poniendo sus valores en la expresión anterior obtenemos $P(A) = 0,025$.

Observación: en química se ha observado la siguiente analogía con el teorema de la probabilidad total. Supóngase que k matraces que contienen diferentes soluciones de la misma sal, hacen un litro. Sea $P(B_i)$ el volumen del i -ésimo matraz y $P(A | B_i)$ la concentración de la solución en el i -ésimo matraz. Si combinamos todas las soluciones en un matraz y suponemos que $P(A)$ indica la concentración de la solución resultante, obtenemos:

$$P(A) = P(A | B_1)P(B_1) + \cdots + P(A | B_k)P(B_k).$$

3.2 Teorema de Bayes

Podemos usar el ejemplo 3.5 para justificar otro resultado importante. Supongamos que del depósito se escoge un artículo y se encuentra que es defectuoso. ¿Cuál es la probabilidad de que se produjese en la primera fábrica?

Usando la notación, presentada previamente, necesitamos $P(B_1 | A)$. Podemos calcular esta probabilidad como una consecuencia de la siguiente. Sean B_1, \dots, B_k una partición del espacio muestral S . Sea A un suceso asociado con S . Aplicando la definición de probabilidad condicional, podemos escribir:

$$P(B_i | A) = \frac{P(A | B_i)P(B_i)}{\sum_{j=1}^k P(A | B_j)P(B_j)} \quad i = 1, 2, \dots, k. \quad (3.5)$$

Este resultado se conoce como *teorema de Bayes*. También se le llama fórmula para la probabilidad de las “causas”. Puesto que las B_i son una partición del espacio muestral, uno y sólo uno de los sucesos B_i ocurre. (Esto es, *uno* de los sucesos B_i debe ocurrir y solamente uno). Por lo tanto, la fórmula anterior nos da la probabilidad de un B_i particular (esto es, una “causa”), dado que el suceso A ha ocurrido. Para aplicar este teorema, debemos conocer los valores de las $P(B_i)$. Muy a menudo esos valores no son conocidos, y esto limita el uso del resultado. Ha habido considerable controversia acerca del teorema de Bayes. Matemática-

mente es perfectamente correcto; sólo la elección impropia para $P(B_i)$ hace el resultado objetable.

Volviendo a la pregunta propuesta anteriormente y aplicando ahora la ecuación (3.5), obtenemos:

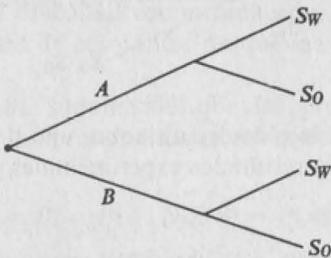
$$P(B_1 | A) = \frac{(0,02)(1/2)}{(0,02)(1/2) + (0,02)(1/4) + (0,04)(1/4)} = 0,40.$$

Observación: otra vez podemos encontrar una analogía con el teorema de Bayes, en química. En k matraces tenemos soluciones de la misma sal, pero de concentraciones diferentes. Supongamos que el volumen total de la solución es un litro. Indicando el volumen de la solución en el i -ésimo matraz por $P(B_i)$ e indicando la concentración de la sal en ese mismo matraz por $P(A | B_i)$, encontramos que la ecuación (3.5) da la proporción de la cantidad completa de sal encontrada en el i -ésimo matraz.

La siguiente ilustración del teorema de Bayes nos dará una oportunidad de presentar la idea de un diagrama de árbol, un método muy útil para analizar ciertos problemas.

Supóngase que muchas cajas están llenas de caramelos de dos tipos, digamos A y B . El tipo A contiene 70 por ciento dulce y 30 por ciento ácido, mientras que en el tipo B dichos porcentajes son al revés. Aún más, supóngase que el 60 por ciento de todas las cajas de caramelos son del tipo A mientras que el resto son del tipo B .

Ahora estamos ante el siguiente problema de decisión. Usted recibe una caja de dulces de tipo desconocido. Se le permite sacar una muestra de un caramelo (una situación ciertamente no real, pero que nos permite presentar las ideas importantes sin mucha complicación) y con esta información debe decir si cree que el tipo A o el tipo B le ha sido ofrecido. El siguiente "diagrama de árbol" (llamado así por las diversas trayectorias o ramas que aparecen) nos ayudará a analizar el problema. (S_W y S_O indican la elección de un caramelo dulce o ácido, respectivamente.)



Hagamos unos pocos cálculos:

$$P(A) = 0,6; P(B) = 0,4; P(S_W | A) = 0,7;$$

$$P(S_O | A) = 0,3; P(S_W | B) = 0,3; P(S_O | B) = 0,7.$$

Lo que realmente deseamos saber es $P(A | S_W)$, $P(A | S_O)$, $P(B | S_W)$ y $P(B | S_O)$. Esto es, suponiendo que realmente escogimos un caramelo dulce. ¿Qué decisión estariamos más inclinados a hacer? Comparemos $P(A | S_W)$ y $P(B | S_W)$. Utilizando la fórmula de Bayes tenemos

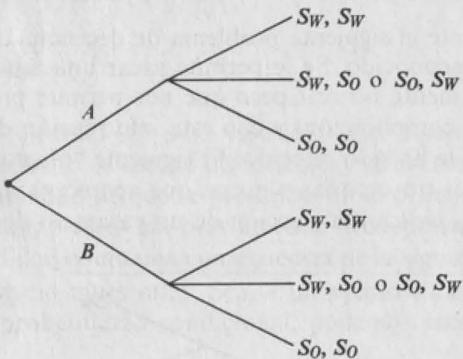
$$P(A | S_W) = \frac{P(S_W | A)P(A)}{P(S_W | A)P(A) + P(S_W | B)P(B)} = \frac{(0,7)(0,6)}{(0,7)(0,6) + (0,3)(0,4)} = \frac{7}{9}.$$

Un cálculo similar nos da $P(B | S_W) = 2/9$.

Así, con base en la evidencia que tenemos (es decir, la obtención de un caramelo dulce) es $2\frac{1}{2}$ veces más probable que estemos considerando un depósito del tipo A que del tipo B . Por lo tanto, decidiríamos, posiblemente, que el caramelo se obtuvo de una caja tipo A . (Podríamos estar equivocados, naturalmente. Lo interesante del análisis anterior es que elegimos la alternativa que parece más probable con base en los datos limitados que tenemos.)

En términos del diagrama de árbol, lo que realmente necesitábamos (e hicimos) en los cálculos precedentes fue un análisis “hacia atrás”. Esto es, dado lo que observamos, S_W en este caso, ¿qué tan probable era escoger el tipo A ?

Una situación un poco más interesante aparece si se nos permite elegir *dos* caramelos antes de decidir si el tipo A o el tipo B es escogido. En este caso el diagrama de árbol aparecerá como sigue.



En el problema 3.26 se le pide decidir sobre uno de los dos tipos, A o B , que usted muestra, según los tres resultados experimentales posibles que usted observa.

3.3 Sucesos independientes

Hemos considerado dos sucesos A y B que no pueden ocurrir simultáneamente, esto es $A \cap B = \emptyset$. Tales sucesos se designaron mutuamente excluyentes. Indicamos previamente que si A y B son mutuamente excluyentes, entonces $P(A|B) = 0$, porque la ocurrencia de B impide la ocurrencia de A . Por otra parte,

tenemos el caso, ya discutido anteriormente, en que $B \supset A$ y, por lo tanto, $P(B|A) = 1$.

En cada uno de los casos anteriores, sabiendo que B ocurrió, se nos dio una información precisa referente a la probabilidad de la ocurrencia de A . Sin embargo, hay muchos casos en los cuales se sabe que si un suceso B ocurre, no tiene influencia alguna en la ocurrencia o no ocurrencia de A .

EJEMPLO 3.6. Supongamos que se lanza un dado normal dos veces. Definiremos los sucesos A y B como sigue:

$$A = \{\text{el primer dado muestra un número par}\},$$

$$B = \{\text{el segundo dado muestra un 5 o un 6}\}.$$

Por intuición sabemos que los sucesos A y B no están relacionados. Saber que B ocurre, no proporciona información acerca de la ocurrencia de A . De hecho, el siguiente cálculo lo pone de manifiesto. Tomando como nuestro espacio muestral los 36 resultados igualmente posibles considerados en el ejemplo 3.1, encontraremos que $P(A) = \frac{18}{36} = \frac{1}{2}$, $P(B) = \frac{12}{36} = \frac{1}{3}$, mientras que $P(A \cap B) = \frac{6}{36} = \frac{1}{6}$. Por lo tanto

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\left(\frac{1}{6}\right)}{\left(\frac{1}{3}\right)} = \frac{1}{2}.$$

Así encontramos, como era de suponer, que la probabilidad no condicional es igual a la probabilidad condicional $P(A|B)$. De modo semejante

$$P(B|A) = \frac{P(B \cap A)}{P(A)} = \frac{\left(\frac{1}{6}\right)}{\left(\frac{1}{2}\right)} = \frac{1}{3} = P(B).$$

Por lo tanto, podríamos estar inclinados a decir que A y B son independientes si y sólo si $P(B|A) = P(A)$ y $P(B|A) = P(B)$. Aunque sería muy apropiado, hay otro método que evita la dificultad encontrada aquí, es decir que ambos $P(A)$ y $P(B)$, deben ser diferentes de cero antes de que las igualdades anteriores sean significativas.

Consideremos $P(A \cap B)$, suponiendo que las probabilidades condicionales anteriores sean iguales a las probabilidades no condicionales correspondientes. Tenemos

$$P(A \cap B) = P(A|B)P(B) = P(A)P(B),$$

$$P(A \cap B) = P(B|A)P(A) = P(B)P(A).$$

Así encontramos que, como ni $P(A)$ ni $P(B)$ son iguales a cero, las probabilidades no condicionales son iguales a las probabilidades condicionales si y sólo si $P(A \cap B) = P(A)P(B)$. Aquí hacemos la siguiente definición formal. [Si $P(A)$ o $P(B)$ es igual a cero, esta definición es aún válida.]

Definición. *A y B son sucesos independientes si y sólo si*

$$P(A \cap B) = P(A)P(B). \quad (3.6)$$

Observación: esta definición es equivalente a la sugerida anteriormente, es decir, que *A y B son independientes si $P(B|A) = P(B)$ y $P(A|B) = P(A)$.* Esta última forma es ligeramente más intuitiva, porque dice precisamente lo que hemos estado tratando de decir antes: *A y B son independientes si el conocimiento de la ocurrencia de A no influye de modo alguno en la probabilidad de la ocurrencia de B.*

Que la definición formal adoptada también tiene un cierto carácter intuitivo, puede verse al considerar el siguiente ejemplo.

EJEMPLO 3.7. Veamos otra vez el ejemplo 3.2. Consideraremos primero la tabla siguiente sólo con los valores marginales dados.

	E	M	
N			70
U			30
	60	40	100

Esto es, hay 60 máquinas eléctricas y 40 manuales. Del total, 70 son nuevas mientras que 30 son usadas. Hay muchas maneras de llenar los datos de la tabla, consistentemente con los totales marginales dados. Abajo anotamos algunas de esas posibilidades.

	E	M	
N	60	10	70
U	0	30	30
	60	40	100

(a)

	E	M	
N	30	40	70
U	30	0	30
	60	40	100

(b)

	E	M	
N	42	28	70
U	18	12	30
	60	40	100

(c)

Consideremos la tabla (a). Aquí *todas* las máquinas eléctricas son nuevas, y *todas* las usadas son manuales. Así hay una relación obvia (no necesariamente causal) entre las características de ser eléctricas y ser nuevas. Igualmente, en la tabla (b) *todas* las máquinas manuales son nuevas y *todas* las usadas son eléctricas. Otra vez parece existir una relación definida entre esas características. Sin embargo, cuando observamos la tabla (c) el panorama es muy diferente. Aquí no existe una relación aparente. Por ejemplo, 60 por ciento de todas las máquinas son eléctricas y exactamente el 60 por ciento de las máquinas usadas son eléctricas. De modo semejante, el 70 por ciento de todas las máquinas son nuevas, y exactamente el 70 por ciento de las máquinas manuales son nuevas, etc.... Así, ninguna indicación nos muestra que las características de «ser nuevas» y «ser eléctricas» tengan alguna relación entre sí. Naturalmente, esta tabla fue elaborada precisamente de modo que exhiba esta propiedad. ¿Cómo se obtuvieron los datos de esta tabla? Simplemente aplicando la ecuación (3.6); es decir, como $P(E) = \frac{60}{100}$ y $P(N) = \frac{70}{100}$ debemos tener,

por independencia, $P(E \cap N) = P(E)P(N) = \frac{42}{100}$. Por lo tanto, la entrada en la tabla que indica el número de máquinas eléctricas nuevas está dada por el número 42. Las otras entradas se obtuvieron de un modo semejante.

En la mayor parte de las aplicaciones *supondremos* la independencia de los dos sucesos A y B y luego usaremos esta hipótesis para calcular $P(A \cap B)$, como $P(A)P(B)$. Generalmente, las condiciones físicas bajo las cuales se realiza el experimento harán posible determinar si tal suposición es justificada o al menos aproximadamente justificada.

EJEMPLO 3.8. Consideremos un lote grande de artículos, digamos 10.000. Supongamos que el 10 por ciento de estos artículos es defectuoso y el 90 por ciento no. Se escogen dos artículos. ¿Cuál es la probabilidad de que ambos no sean defectuosos?

Definamos los sucesos A y B así:

$$A = \{\text{primer artículo no es defectuoso}\},$$

$$B = \{\text{segundo artículo no es defectuoso}\}.$$

Si suponemos que el primer artículo se sustituye antes de escoger el segundo, entonces se puede suponer que los sucesos A y B son independientes y, por tanto $P(A \cap B) = (0.9)(0.9) = 0.81$. Sin embargo, en forma más real, el segundo artículo se escoge sin sustituir el primero. En este caso,

$$P(A \cap B) = P(B | A)P(A) = \frac{8999}{9999}(0.9)$$

que es aproximadamente 0,81. Así, aunque A y B no son independientes en el segundo caso, la suposición de independencia que simplifica considerablemente los cálculos, causa sólo un error despreciable. (Recuerde el objetivo de un modelo matemático como el que se describió en la sección 1.1). Si hubiese habido sólo pocos artículos en el lote, digamos 30, la suposición de independencia habría producido un gran error. Así, se hace importante verificar cuidadosamente las condiciones bajo las cuales se realiza el experimento, a fin de establecer la validez de la suposición de independencia entre varios sucesos.

EJEMPLO 3.9. Supóngase que un mecanismo está formado por dos componentes acoplados en serie, como se indica en la figura 3.4. Cada componente tiene una probabilidad p de no funcionar. ¿Cuál es la probabilidad de que el mecanismo funcione?



FIGURA 3.4

Es evidente que el mecanismo funcionará si y sólo si *ambos* componentes funcionan. Por tanto

$$\text{Prob. (Mecanismo funcione)} = \text{Prob. (}C_1 \text{ funciona y } C_2 \text{ funciona}).$$

La información que hemos dado no nos permite seguir, a menos que sepamos (o supongamos) que los dos mecanismos trabajan independientemente. Esta puede o no ser una suposición realista, que depende de cómo se acoplan las dos partes. Si suponemos que los dos mecanismos trabajan independientemente, obtenemos para la probabilidad pedida $(1 - p)^2$.

Será muy importante para nosotros extender la noción de independencia a más de dos sucesos. Consideremos primero tres sucesos asociados con un experimento, digamos: A, B, C . Si A y B , A y C , B y C , son *mutuamente independientes* (en el sentido dado anteriormente), entonces no se deduce, en general, que no haya dependencia entre los tres sucesos A, B y C . El siguiente ejemplo (algo artificial) ilustra este punto.

EJEMPLO 3.10. Supongamos que lanzamos dos dados. Definamos los sucesos A, B y C , como sigue:

$$A = \{\text{el primer dado muestra un número par}\},$$

$$B = \{\text{el segundo dado muestra un número impar}\},$$

$$C = \{\text{ambos dados muestran números pares o números impares}\}.$$

Tenemos $P(A) = P(B) = P(C) = \frac{1}{2}$. Aún más, $P(A \cap B) = P(A \cap C) = P(B \cap C) = \frac{1}{4}$. Por lo tanto, los tres sucesos son mutuamente independientes. Sin embargo, $P(A \cap B \cap C) = 0 \neq P(A)P(B)P(C)$.

Este ejemplo motiva la siguiente definición.

Definición. Decimos que los tres sucesos A, B y C son *mutuamente independientes* si y sólo si todas las condiciones siguientes se mantienen:

$$\begin{aligned} P(A \cap B) &= P(A)P(B), & P(A \cap C) &= P(A)P(C), \\ P(B \cap C) &= P(B)P(C) & P(A \cap B \cap C) &= P(A)P(B)P(C). \end{aligned} \tag{3.7}$$

Finalmente generalizaremos esta noción a n sucesos, dando la siguiente definición.

Definición. Los n sucesos A_1, A_2, \dots, A_n son mutuamente independientes si y sólo si tenemos para $k = 2, 3, \dots, n$,

$$P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) = P(A_{i_1})P(A_{i_2}) \cdots P(A_{i_k}). \tag{3.8}$$

(Hay $2^n - n - 1$ condiciones juntas anotadas; ver el problema 3-18).

Observación: en la mayor parte de las aplicaciones, no necesitamos verificar todas estas condiciones, porque generalmente *suponemos* la independencia, (con base en lo que sabemos acerca del experimento). Nosotros, entonces, usamos esta suposición para calcular, digamos:

$$P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) \text{ como } P(A_{i_1})P(A_{i_2}) \cdots P(A_{i_k}).$$

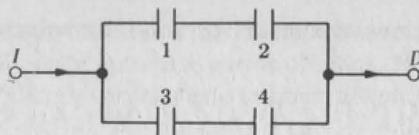


FIGURA 3.5

EJEMPLO 3.11. La probabilidad de cerrar cada uno de los relés del circuito que se indica en la figura 3.5 está dada por p . Si todos los relés funcionan independientemente, ¿cuál es la probabilidad de que exista una corriente en los terminales I y D ?

Sea A_i el suceso que representa $\{\text{relé } i \text{ está cerrado}\}$, $i = 1, 2, 3, 4$. Sea E el suceso que representa $\{\text{la corriente pasa de } I \text{ a } D\}$. Por tanto $E = (A_1 \cap A_2) \cup (A_3 \cap A_4)$. (Note que $A_1 \cap A_2$ y $A_3 \cap A_4$ no son mutuamente excluyentes.) Así

$$\begin{aligned} P(E) &= P(A_1 \cap A_2) + P(A_3 \cap A_4) - P(A_1 \cap A_2 \cap A_3 \cap A_4) \\ &= p^2 + p^2 - p^4 = 2p^2 - p^4. \end{aligned}$$

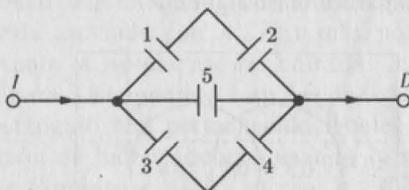


FIGURA 3.6

EJEMPLO 3.12. Supóngase otra vez que en el circuito de la figura 3.6, la probabilidad de que cada uno de los relés esté cerrado es p y que todos los relés funcionan independientemente. ¿Cuál es la probabilidad de que exista una corriente entre los terminales I y D ?

Usando la misma notación como en el ejemplo 3.11, tenemos que

$$\begin{aligned} P(E) &= P(A_1 \cap A_2) + P(A_5) + P(A_3 \cap A_4) - P(A_1 \cap A_2 \cap A_5) \\ &\quad - P(A_1 \cap A_2 \cap A_3 \cap A_4) - P(A_5 \cap A_3 \cap A_4) \\ &\quad + P(A_1 \cap A_2 \cap A_3 \cap A_4 \cap A_5) \\ &= p^2 + p + p^2 - p^3 - p^4 - p^3 + p^5 = p + 2p^2 - 2p^3 - p^4 + p^5. \end{aligned}$$

Para cerrar este capítulo indiquemos una solución muy común, pero errónea, del problema.

EJEMPLO 3.13. Suponga que entre seis pernos, dos son más cortos que una longitud específica. Si se escogen dos pernos al azar, ¿cuál es la probabilidad de que los dos más cortos sean los escogidos? Sea A_i el suceso $\{\text{el } i\text{-ésimo perno elegido es corto}\}$, $i = 1, 2$.

Por lo tanto queremos evaluar $P(A_1 \cap A_2)$. La solución correcta se obtiene, naturalmente, al escribir

$$P(A_1 \cap A_2) = P(A_2 | A_1)P(A_1) = \frac{1}{3} \cdot \frac{2}{6} = \frac{1}{15}.$$

La solución común pero *incorrecta* es escribir:

$$P(A_1 \cap A_2) = P(A_2)P(A_1) = \frac{1}{3} \cdot \frac{2}{6} = \frac{1}{15}.$$

Lo importante, en realidad, es que aunque la respuesta es correcta, la identificación de $\frac{1}{3}$ con $P(A_2)$ es incorrecta; $\frac{1}{3}$ representa $P(A_2 | A_1)$. Para evaluar $P(A_2)$ en forma correcta, escribimos

$$P(A_2) = P(A_2 | A_1)P(A_1) + P(A_2 | \bar{A}_1)P(\bar{A}_1) = \frac{1}{3} \cdot \frac{2}{6} + \frac{2}{3} \cdot \frac{4}{6} = \frac{1}{3}.$$

3.4 Consideraciones esquemáticas; probabilidad condicional e independencia

La solución esquemática siguiente puede ser útil para comprender el concepto de probabilidad condicional. Supóngase que A y B son dos sucesos asociados con un espacio muestral para el cual las distintas probabilidades se indican en el diagrama de Venn que aparece en la figura 3.7.

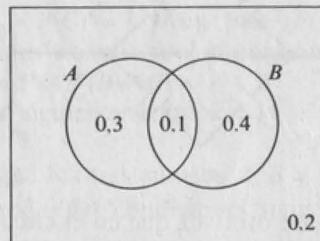


FIGURA 3.7

Por lo tanto $P(A \cap B) = 0.1$, $P(A) = 0.1 + 0.3 = 0.4$ y $P(B) = 0.1 + 0.4 = 0.5$.

En seguida, representemos las diversas probabilidades por las *áreas* de los rectángulos como en la figura 3.8. En cada caso, las regiones sombreadas indican el suceso B : en el rectángulo de la izquierda representamos $A \cap B$ y en el de la derecha $A' \cap B$.

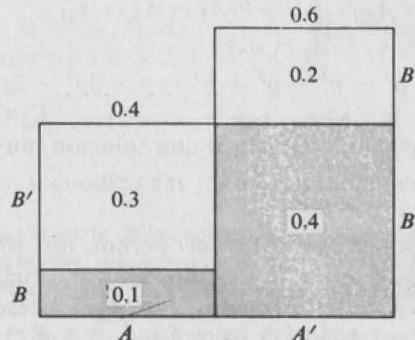


FIGURA 3.8

Supongamos ahora que deseamos calcular $P(B | A)$. Así sólo necesitamos considerar A ; esto es, A' puede ignorarse en los cálculos. Notemos que la proporción de B en A es $1/4$. (Podemos verificar esto también al aplicar Ec. (3.1): $P(B | A) = P(A \cap B)/P(A) = 0,1/0,4 = 1/4$). Por lo tanto $P(B' | A) = 3/4$, y el diagrama que representa esta probabilidad condicional estaría dado por la figura 3.9.

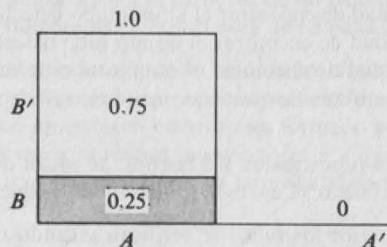


FIGURA 3.9

Nótese también que si se da A como ocurrido, todas las probabilidades (es decir 1) se deben asociar con el suceso A mientras que ninguna de las probabilidades (es decir 0) está asociada con A' . Aún más, nótese que en el rectángulo izquierdo, que representa A , solamente las entradas individuales han cambiado de la figura 3.8 a la figura 3.9 (sumando 1 en vez de 0,4). Sin embargo, las proporciones dentro del rectángulo han permanecido iguales (es decir, 3 : 1).

Ilustremos la noción de independencia, usando la solución esquemática presentada anteriormente. Supóngase que los sucesos A y B se dibujan en la figura 3.10. En ese caso las proporciones en los dos rectángulos, que representan A y A' , son las *mismas*: 3 : 1 en ambos casos. Así tenemos $P(B) = 0,1 + 0,15 = 0,25$, y $P(B \cap A) = 0,1/0,4 = 0,25$.

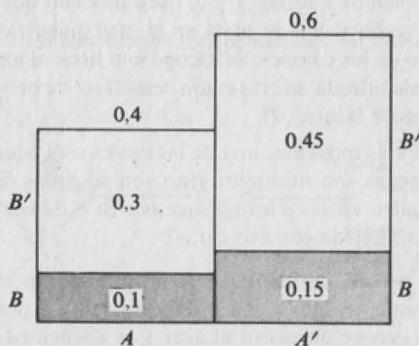


FIGURA 3.10

Finalmente, obsérvese que simplemente mirando la figura 3.8 podemos calcular las otras probabilidades condicionales: $P(A | B) = 1/5$ (puesto que $1/5$ del área total rectangular que representa B está ocupada por A), $P(A' | B) = 4/5$.

PROBLEMAS

3.1. La urna 1 contiene x esferas blancas e y rojas. La urna 2 contiene z esferas blancas y v rojas. Se escoge una esfera al azar de la urna 1 y se pone en la urna 2. Entonces se escoge una esfera al azar de la urna 2. ¿Cuál es la probabilidad de que esta esfera sea blanca?

3.2. Dos tubos defectuosos se confunden con dos buenos. Los tubos se prueban, uno por uno, hasta encontrar los defectuosos.

- (a) ¿Cuál es la probabilidad de encontrar el último tubo defectuoso en la segunda prueba?
- (b) ¿Cuál es la probabilidad de encontrar el último tubo defectuoso en la tercera prueba?
- (c) ¿Cuál es la probabilidad de encontrar el último tubo defectuoso en la cuarta prueba?
- (d) Agregue los números obtenidos anteriormente en (a), (b) y (c). ¿Es sorprendente el resultado?

3.3. Una caja contiene 4 tubos malos y 6 buenos. Se sacan dos a la vez. Se prueba uno de ellos y se encuentra que es bueno. ¿Cuál es la probabilidad de que el otro también sea bueno?

3.4. En el problema anterior los tubos se verifican sacando uno al azar, se prueba y se repite el proceso hasta que se encuentran los cuatro tubos malos. ¿Cuál es la probabilidad de encontrar el cuarto tubo malo,

- (a) en la quinta prueba?
- (b) en la décima prueba?

3.5. Supóngase que A y B son dos sucesos independientes asociados con un experimento. Si la probabilidad de que A o B ocurra es igual a 0,6, mientras que la probabilidad de que A ocurra es igual a 0,4, determinar la probabilidad de que B ocurra.

3.6. Veinte artículos, 12 de los cuales son defectuosos y 8 no defectuosos, se inspeccionan uno después de otro. Si esos artículos se escogen al azar, ¿cuál es la probabilidad de que:

- (a) los dos primeros artículos inspeccionados sean defectuosos?
- (b) los dos primeros artículos inspeccionados sean no defectuosos?
- (c) entre los dos primeros artículos inspeccionados haya uno defectuoso y uno no defectuoso?

3.7. Supóngase que tenemos 2 urnas, 1 y 2, cada una con dos cajones. La urna 1 tiene una moneda de oro en un cajón y una de plata en el otro, mientras que la urna 2 tiene una moneda de oro en cada uno de los cajones. Se escoge una urna al azar; y de ésta se escoge un cajón al azar. La moneda encontrada en este cajón resulta ser de oro. ¿Cuál es la probabilidad de que la moneda provenga de la urna 2?

3.8. Un bolso contiene tres monedas, una de las cuales está acuñada con dos caras mientras que las otras dos monedas son normales y no son sesgadas. Se escoge una moneda al azar del bolso y se lanza cuatro veces en forma sucesiva. Si cada vez sale cara, ¿cuál es la probabilidad de que ésta sea la moneda con dos caras?

3.9. En una fábrica de pernos, las máquinas A , B y C fabrican 25, 35 y 40 por ciento de la producción total, respectivamente. De lo que producen, 5, 4 y 2 por ciento, respectivamente, son pernos defectuosos. Se escoge un perno al azar y se encuentra que es defectuoso. ¿Cuál es la probabilidad que el perno provenga de la máquina A ? B ? C ?

3.10. Sean A y B dos sucesos asociados con un experimento. Supóngase que $P(A) = 0,4$ mientras que $P(A \cup B) = 0,7$. Sea $P(B) = p$.

- (a) ¿Para qué elección de p son A y B mutuamente excluyentes?
- (b) ¿Para qué elección de p son A y B independientes?

3.11. Tres componentes de un mecanismo, digamos C_1 , C_2 y C_3 están colocados en serie (en una línea recta). Supóngase que esos mecanismos están agrupados en un orden aleatorio. Sea R el suceso $\{C_2$ está a la derecha de $C_1\}$, y sea S el suceso $\{C_3$ está a la derecha de $C_1\}$. ¿Son independientes los sucesos R y S ? ¿Por qué?

3.12. Se lanza un dado e, independientemente, se escoge al azar una carta de una baraja normal. ¿Cuál es la probabilidad de que:

- el dado muestre un número par y la carta sea de un palo rojo?
- el dado muestre un número par o la carta sea de un palo rojo?

3.13. Un número binario está compuesto sólo de los dígitos cero y uno. (Por ejemplo, 1011, 1100, etc.) Esos números juegan un papel importante en el uso de los computadores electrónicos. Supóngase que un número binario está formado por n dígitos. Supóngase que la probabilidad de que aparezca un dígito incorrecto es p y que los errores en dígitos diferentes son independientes uno de otro. ¿Cuál es la probabilidad de formar un número incorrecto?

3.14. Se lanza un dado n veces. ¿Cuál es la probabilidad de que «6» salga al menos una vez en los n lanzamientos?

3.15. Dos personas lanzan tres monedas regulares cada una. ¿Cuál es la probabilidad de que obtengan el mismo número de caras?

3.16. Se lanzan dos dados. Puesto que las caras muestran números diferentes, ¿cuál es la probabilidad de que una cara sea 4?

3.17. En la fabricación de cierto artículo se encuentra que se presenta un tipo de defectos con una probabilidad de 0,1 y defectos de un segundo tipo con probabilidad 0,05. (Se supone la independencia entre los tipos de defectos). ¿Cuál es la probabilidad de que:

- un artículo no tenga ambas clases de defectos?
- un artículo sea defectuoso?
- suponiendo que un artículo sea defectuoso, tenga sólo un tipo de defecto?

3.18. Verifique que el número de condiciones indicadas en la ecuación (3.8) esté dado por $2^n - n - 1$.

3.19. Probar que si A y B son sucesos independientes, también lo son A y \bar{B} , \bar{A} y B , \bar{A} y \bar{B} .

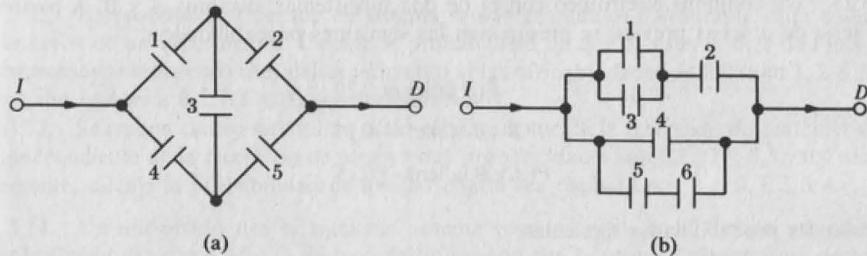


FIGURA 3.11

3.20. En la figura 3.11 (a) y (b) se supone que la probabilidad de que cada relé esté cerrado es p y que cada relé se abre o se cierra independientemente de cualquier otro. Encontrar en cada caso la probabilidad de que la corriente pase de I a D .

TABLA 3.2

Número de fallas	0	1	2	3	4	5	6
<i>A</i>	0,1	0,2	0,3	0,2	0,09	0,07	0,04
<i>B</i>	0,3	0,1	0,1	0,1	0,1	0,15	0,15

3.21. Dos máquinas, *A*, *B*, que se accionan independientemente, pueden tener un cierto número de fallas cada día. La tabla 3.2 da la distribución de probabilidades de las fallas de cada una. Calcular las siguientes probabilidades:

- (a) *A* y *B* tienen el mismo número de fallas.
- (b) El número de fallas es menor que 4; menor que 5.
- (c) *A* tiene más fallas que *B*.
- (d) *B* tiene el doble de fallas que *A*.
- (e) *B* tiene 4 fallas, cuando se sabe que *B* tiene por lo menos 2 fallas.
- (f) El número mínimo de fallas de las dos máquinas es 3; es menor que 3.
- (g) El número máximo de fallas de las máquinas es 3; es más que 3.

3.22. Al verificar la ecuación (3.2), se observa que para *A* fijo, $P(B | A)$ satisface los diversos postulados de la probabilidad.

3.23. Si cada uno de los elementos de un determinante de segundo orden es cero o uno, ¿cuál es la probabilidad de que el valor del determinante sea positivo? (Supóngase que las entradas individuales del determinante se escogen independientemente, se supone que cada uno de los valores tiene probabilidad $\frac{1}{2}$.)

3.24. Verifique que el teorema de la multiplicación $P(A \cap B) = P(A | B)P(B)$, establecido para dos sucesos, se puede generalizar para tres sucesos como sigue:

$$P(A \cap B \cap C) = P(A | B \cap C)P(B | C)P(C).$$

3.25. Un conjunto electrónico consta de dos subsistemas, digamos *A* y *B*. A partir de una serie de pruebas previas, se presuponen las siguientes probabilidades:

$$P(A \text{ falle}) = 0,20$$

$$P(B \text{ sólo falle}) = 0,15$$

$$P(A \text{ y } B \text{ fallen}) = 0,15$$

Calcular las probabilidades siguientes.

- (a) $P(A \text{ falle} | B \text{ haya fallado}),$
- (b) $P(A \text{ falle solamente}).$

3.26. Finalice el análisis del ejemplo dado en la sección 3.2 decidiendo cuál de los tipos de tarros de caramelos, *A* o *B*, es el escogido con base en el conocimiento de los dos caramelos que fueron muestrados.

3.27. Cada vez que se hace un experimento, la ocurrencia de un suceso particular A es igual a 0,2. El experimento se repite, independientemente, hasta que A ocurre. Calcule la probabilidad de que sea necesario realizar un cuarto experimento.

3.28. Supóngase que un mecanismo tiene N tubos, todos los cuales son necesarios para su funcionamiento. Para localizar el tubo que funciona mal, se reemplaza sucesivamente cada uno de ellos por uno nuevo. Calcular la probabilidad de que sea necesario verificar N tubos si la probabilidad (constante) es p de que un tubo esté dañado.

3.29. Probar: Si $P(A|B) > P(A)$ entonces $P(B|A) > P(B)$.

3.30. Un tubo al vacío puede provenir de uno cualquiera de tres fabricantes con probabilidades $p_1 = 0,25$, $p_2 = 0,50$ y $p_3 = 0,25$. Las probabilidades de que el tubo funcione correctamente durante un período de tiempo especificado son iguales a 0,1, 0,2 y 0,4, respectivamente, para los tres fabricantes. Calcular la probabilidad de que un tubo elegido aleatoriamente funcione durante el período de tiempo especificado.

3.31. Un sistema eléctrico consta de dos interruptores del tipo A , uno del tipo B , y cuatro del tipo C , conectados como aparece en la figura 3.12. Calcule la probabilidad de que no se pueda eliminar una falla en el circuito, con la llave K si los interruptores A , B y C están abiertos (es decir, fuera de servicio) con probabilidades 0,3, 0,4, 0,2, respectivamente y si ellos funcionan independientemente.

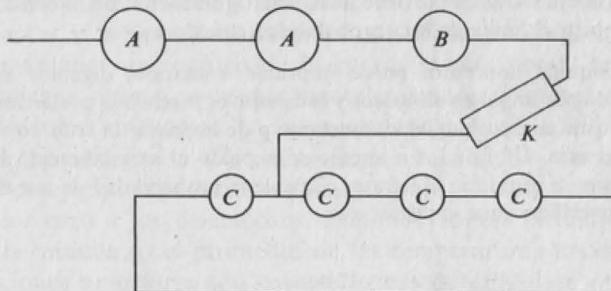


FIGURA 3.12

3.32. La probabilidad de que un sistema se sobrecargue es 0,4 durante cada conjunto de ensayos de un experimento. Calcule la probabilidad de que el sistema deje de funcionar en tres ensayos independientes del experimento si las probabilidades de fallar en 1, 2 ó 3 ensayos son iguales a 0,2, 0,5 y 0,8, respectivamente.

3.33. Se emiten cuatro señales de radio sucesivamente. Si la recepción de cualquier señal es independiente de la recepción de otra y estas probabilidades son 0,1; 0,2; 0,3 y 0,4 respectivamente, calcule la probabilidad de que la señal k sea recibida por $k = 0, 1, 2, 3, 4$.

3.34. Un aficionado usa el siguiente sistema para pronosticar el tiempo atmosférico. Clasifica cada día como «seco» o «mojado» y supone que la probabilidad de que cualquier día dado sea igual al precedente está dada por una constante p ($0 < p < 1$). Con base en anotaciones anteriores, se supone que el 1º de enero tiene una probabilidad β de ser «seco». Suponiendo que β_n = probabilidad (el n -ésimo día del año es «seco»), obtener una expresión para β_n en función de β y p . Evaluar también $\lim_{n \rightarrow \infty} \beta_n$ e interpretar su resultado. (Indicación: expresar β_n en función de β_{n-1}).

3.35. En una ciudad se publican los periódicos A , B y C . Una encuesta reciente de lectores indica lo siguiente: 20 por ciento lee A , 16 por ciento lee B , 14 por ciento lee C , 8 por ciento lee A y B , 5 por ciento lee A y C , y 2 por ciento lee A , B y C . Para un adulto escogido al azar, calcular la probabilidad de que (a) no lea ninguno de los periódicos (b) lea exactamente uno de los periódicos (c) lea al menos A y B si se sabe que lee al menos uno de los periódicos publicados.

3.36. Una moneda normal se lanza 2_n veces.

- (a) Obtener la probabilidad de que haya un número igual de caras y sellos.
- (b) Verificar que la probabilidad calculada en (a) es una función decreciente de n .

3.37. Cada una de las urna 1, urna 2, ..., urna n , contiene α esferas blancas y β esferas negras. Se pasa una esfera de la urna 1 a la urna 2 y *luego* se pasa una de la urna 2 a la urna 3, etc. Finalmente, se escoge una esfera de la urna n . Si la primera esfera que se pasó era blanca, ¿cuál es la probabilidad de que la última esfera elegida sea blanca? ¿Qué sucede cuando $n \rightarrow \infty$? [Indicación: sea $p_n = \text{Prob}(n\text{-ésima esfera pasada sea blanca})$ y exprese p_n en función de p_{n-1} .]

3.38. La urna 1 contiene α esferas blancas y β esferas negras mientras que una urna 2 contiene α esferas blancas y β negras. Se escoge una esfera (de una de las urnas) y luego se devuelve a esa urna. Si la esfera escogida es blanca, se escoge la esfera siguiente de la urna 1; si la esfera escogida es negra, se escoge la siguiente de la urna 2. Continúe de esta manera. Como la primera esfera escogida proviene de la urna 1, obtener $\text{Prob}(n\text{-ésima esfera escogida sea blanca})$ y también el límite de esta probabilidad cuando $n \rightarrow \infty$.

3.39. Una máquina impresora puede imprimir n «letras», digamos $\alpha_1, \dots, \alpha_n$. Esta máquina es operada por impulsos eléctricos y cada letra es producida por un impulso *diferente*. Supóngase que existe una probabilidad constante p de imprimir la letra correcta y también suponga independencia. Uno de los n impulsos, escogido al azar, alimentó la máquina dos veces y las dos veces se imprimió la letra α_1 . Calcule la probabilidad de que el impulso escogido estuviese proyectado para imprimir α_1 .

Variables aleatorias unidimensionales

4.1 Noción general de una variable aleatoria

Al describir el espacio muestral de un experimento, no especificamos que un resultado individual necesariamente tiene que ser un número. De hecho, hemos citado varios ejemplos en los cuales el resultado del experimento no fue una cantidad numérica. Por ejemplo, al clasificar un artículo manufacturado simplemente podríamos usar las categorías «defectuoso» y «no defectuoso». En otro caso, para observar la temperatura durante un período de 24 horas sencillamente podríamos mantener un registro de la curva trazada por el termógrafo. Sin embargo, en muchas situaciones experimentales vamos a interesarnos en medir algo y anotarlo como un *número*. Aún en los casos citados anteriormente podremos asignar un número a cada uno de los resultados (no numéricos) del experimento. Por ejemplo, pudimos asignar el valor uno a artículos no defectuosos y el valor cero a los defectuosos. Pudimos anotar la temperatura máxima del día, o la mínima, o el promedio de las temperaturas máxima y mínima.

Las ilustraciones anteriores son características de una clase muy general de problemas. En muchas situaciones experimentales deseamos asignar un número real x a cada uno de los elementos s del espacio muestral S . Esto es, $x = X(s)$ es el valor de una función X del espacio muestral a los números reales. Teniendo esto presente, hagamos la siguiente definición formal.

Definición. Sea un experimento ε y S el espacio muestral asociado con el experimento. Una *función* X que asigna a cada uno de los elementos $s \in S$, un número real $X(s)$, se llama variable aleatoria.

Observaciones: (a) La terminología anterior es algo desafortunada, pero como es tan universalmente aceptada, nosotros no nos apartaremos de ella. Hemos hecho lo más claro posible que X es una *función* y todavía la llamamos variable (aleatoria).

(b) Resulta que *no toda* función que se conciba puede ser considerada como una variable aleatoria. Una exigencia (aunque no la más general) es que para todo número real x el suceso $\{X(s) = x\}$ y para todo intervalo I , el suceso $\{X(s) \in I\}$ tiene probabilidades bien definidas y consecuentes con los axiomas básicos. En la mayoría de las aplicaciones no aparece esta dificultad y no haremos referencia posterior.

(c) En algunos casos el resultado s del espacio muestral es ya la característica numérica que queremos anotar. Sencillamente tomemos $X(s) = s$, la función identidad.

(d) En la mayoría de las discusiones de variables aleatorias que siguen, no necesitamos indicar la naturaleza funcional de X . Generalmente nos interesamos en los valores *posibles* de X , en vez de indagar de dónde provienen esos valores. Por ejemplo, supongamos que lanzamos dos monedas y consideremos el espacio muestral asociado con este experimento. Esto es,

$$S = \{CC, CS, SC, SS\}.$$

Definamos la variable aleatoria X como sigue: X es el número de caras obtenidas en los dos lanzamientos. Por lo tanto $X(CC) = 2$, $X(CS) = X(SC) = 1$, y $X(SS) = 0$.

$$S = \text{espacio muestral de } \varepsilon \quad R_X = \text{valores posibles de } X$$

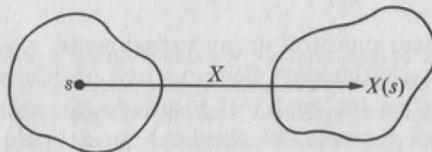


FIGURA 4.1

(e) Es muy importante comprender que una exigencia básica de una función (uni-variada): a cada $s \in S$ le corresponde *exactamente* un valor $X(s)$. Esto se demuestra esquemáticamente en la figura 4.1. Valores diferentes de s pueden dar el mismo valor de X . Por ejemplo en la ilustración anterior encontramos que $X(CS) = X(SC) = 1$.

El espacio R_X , el conjunto de todos los valores posibles de X , se llama algunas veces el *recorrido*. En ciertos sentidos podemos considerar a R_X como otro espacio muestral. El espacio muestral (original) S corresponde a los resultados no numéricos (posiblemente) del experimento, mientras que R_X es el espacio muestral asociado con la variable aleatoria X , que representa la característica numérica que puede ser de interés. Si $X(s) = s$, tenemos $S = R_X$.

Aunque estamos conscientes del peligro pedagógico que existe al dar muchas explicaciones de la misma cosa, señalemos sin embargo que podemos concebir una variable aleatoria X de dos maneras:

(a) Realizamos el experimento ε que tiene como resultado $s \in S$. Luego evaluamos el número $X(s)$.

(b) Efectuamos ε , y obtenemos el resultado s , e (inmediatamente) calculamos $X(s)$. El número $X(s)$ se considera entonces como el resultado obtenido en el experimento y R_X llega a ser el espacio muestral del experimento.

La diferencia entre las interpretaciones (a) y (b) es muy difícil de captar. Relativamente es muy pequeña, pero es digna de atención. En (a) el experimento termina, de hecho, con la observación de s . La evaluación de $X(s)$ se considera como algo que se hace posteriormente y que no está afectada por la aleatoriedad de ε . En (b) se considera que el experimento no está terminado hasta que

el número $X(s)$ ya se ha calculado y ha resultado así en el espacio muestral R_X . Aunque la primera interpretación, (a), es la que se pretende corrientemente, el segundo punto de vista, (b), puede ser muy útil y el lector deberá recordarlo. Lo que queremos decir, y esto será más evidente en las últimas secciones, es que al estudiar variables aleatorias estamos más interesados respecto a los valores que toma X que de su forma funcional. Por lo tanto, en muchos casos ignoraremos completamente el espacio muestral sobre el cual se puede definir X .

EJEMPLO 4.1. Supongamos que se pone una bombilla en un portalámparas. Se considera que el experimento termina cuando la bombilla se apaga. ¿Cuál es un resultado posible, digamos s ? Una manera de describir s sería anotando simplemente la fecha y la hora del día en la cual la bombilla se quema, por ejemplo 19 de mayo, 4:32 p.m. Por lo tanto el espacio muestral se puede representar como $S = \{(d, t) | d = \text{fecha}, t = \text{hora del día}\}$. Posiblemente la variable aleatoria de interés es X , la duración del encendido. Nótese que una vez que se ha observado $s = (d, t)$, la evaluación de $X(s)$ no indica ninguna aleatoriedad. Cuando está especificado s , $X(s)$ está completamente determinado.

Los dos puntos de vista expresados anteriormente pueden aplicarse a este ejemplo como sigue. En (a) consideramos que el experimento termina con la observación $s = (d, t)$, la fecha y la hora del día. El cálculo de $X(s)$ se efectúa, entonces, haciendo una sencilla operación aritmética. En (b) consideramos que el experimento está terminado sólo después de que se haya calculado $X(s)$ y el número $X(s) = 107$ horas, por ejemplo, se considera como el resultado del experimento.

Podría señalarse que un análisis similar se podría aplicar a otra variable aleatoria de interés, por ejemplo, $Y(s)$ la temperatura en la habitación en el instante en que la bombilla se quemó.

EJEMPLO 4.2. En una mesa se lanzan tres monedas. Tan pronto como las monedas caen en la mesa, la fase "aleatoria" del experimento ha terminado. Un solo resultado s podría consistir en una descripción detallada de cómo y dónde cayeron las monedas. Posiblemente estamos interesados sólo en ciertas características numéricas asociadas con este experimento. Por ejemplo, podríamos calcular

$X(s)$ = número de caras que aparecen,

$Y(s)$ = distancia máxima entre dos monedas cualesquiera,

$Z(s)$ = distancia mínima de las monedas de cualquier arista de la mesa.

Si la variable aleatoria X es de interés, como indicamos en el ejemplo anterior podríamos incluir la evaluación de $X(s)$ en la descripción del experimento y por lo tanto indicar simplemente que el espacio muestral asociado con el experimento es $\{0, 1, 2, 3\}$, que corresponden a los valores de X . Aunque muy a menudo

adoptaremos precisamente este punto de vista, es importante establecer que la cuenta del número de caras se hace *después* de que ha terminado la parte aleatoria del experimento.

Observación: al referirnos a variables aleatorias usaremos, casi sin excepción, letras mayúsculas tales como X, Y, Z , etc. Sin embargo, cuando hablamos del *valor* de esas variables aleatorias suponemos que en general usaremos letras minúsculas tales como x, y, z , etc. Esta es una *distinción muy importante* que debemos hacer y el estudiante debe considerarla detenidamente. Por ejemplo, cuando hablamos de escoger al azar una persona de alguna población señalada y medir su altura (en pulgadas, por ejemplo), podríamos referirnos a los resultados *posibles* como una variable aleatoria X . Podríamos luego hacer varias preguntas acerca de X , tales como $P(X) \geq 60$. Sin embargo, una vez que escogemos una persona y medimos su altura, obtenemos un valor específico de X , por ejemplo x . Así no sería importante preguntar por $P(x \geq 60)$ puesto que x es o no ≥ 60 . Esta distinción entre una variable aleatoria y su valor es importante, y haremos referencias posteriores a ella.

Así como nos interesamos por los sucesos asociados con el espacio muestral S , también será necesario tratar sucesos con respecto a la variable aleatoria X , esto es, subconjunto del recorrido R_X . Muy a menudo ciertos sucesos asociados con S están «relacionados» (en un sentido que luego describiremos) con sucesos asociados con R_X de la manera siguiente.

Definición. Sea ε un experimento y S su espacio muestral. Sea X una variable aleatoria definida en S y sea R_X su recorrido. Sea B un suceso respecto a R_X ; esto es, $B \subset R_X$. Supongamos que A se define

$$A = \{s \in S \mid X(s) \in B\}. \quad (4.1)$$

En palabras: A consta de todos los resultados en S para los cuales $X(s) \in B$ (figura 4.2). En este caso decimos que A y B son sucesos equivalentes.

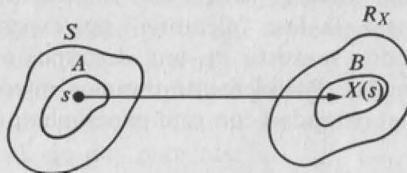


FIGURA 4.2

Observaciones: (a) Expresando más informalmente lo anterior, A y B son sucesos equivalentes siempre que ocurran juntos. Esto es, siempre que A ocurre, ocurre B y viceversa. Si ocurrió A , entonces se obtuvo un resultado s para el cual $X(s) \in B$ y por lo tanto ocurrió B . Recíprocamente, si B ocurrió, se observó un valor $X(s)$ para el cual $s \in A$ y por lo tanto ocurrió A .

(b) Es importante destacar que en nuestra definición de sucesos equivalentes, A y B están asociados con espacios muestrales *diferentes*.

EJEMPLO 4.3. Consideremos el lanzamiento de dos monedas. En este caso $S = \{CC, CS, SC, SS\}$. Sea X el número de caras obtenidas. En este caso $R_X = \{0, 1, 2\}$. Sea $B = \{1\}$. Puesto que $X(CS) = X(SC) = 1$ si y sólo si $X(s) = 1$, tenemos que $A = \{CS, SC\}$ es equivalente a B .

Ahora damos la siguiente definición importante.

Definición. Sea B un suceso en el recorrido R_X . Definimos entonces $P(B)$ como sigue:

$$P(B) = P(A), \quad \text{en donde} \quad A = \{s \in S \mid X(s) \in B\}. \quad (4.2)$$

En palabras: Definimos $P(B)$ igual a la probabilidad del suceso $A \subset S$, que es equivalente a B , en el sentido de la ecuación (4.1).

Observaciones: (a) Suponemos que las probabilidades se pueden asociar con sucesos en S . Por tanto la definición anterior hace posible asignar probabilidades a sucesos asociados con R_X en función de probabilidades definidas en S .

(b) Realmente es posible *probar* que $P(B)$ debe ser como se definió. Sin embargo, esto implicaría alguna dificultad teórica que queremos evitar y, por tanto, proseguiremos como antes.

(c) Puesto que en la formulación de la ecuación (4.2) los sucesos A y B se refieren a espacios muestrales diferentes, realmente deberíamos usar una notación diferente cuando nos referimos a las probabilidades definidas en S y para las definidas en R_X , por ejemplo algo como $P(A)$ y $P_X(B)$. Sin embargo, no haremos esto sino que continuaremos escribiendo simplemente $P(A)$ y $P(B)$. El contexto dentro del cual aparecerán estas expresiones harán evidente la interpretación.

(d) Las probabilidades asociadas con sucesos en el espacio muestral S (original) están, en un sentido, determinadas por «fuerzas que escapan a nuestro control» o, como se dice algunas veces «por naturaleza». La constitución de una fuente radioactiva que emite partículas, la distribución de un gran número de personas que podría hacer una llamada telefónica durante cierta hora y la agitación térmica que resulta de una corriente o las condiciones atmosféricas que dan origen a una fuente de tormenta, ilustran este punto. Cuando introducimos una variable aleatoria X y su recorrido asociado R_X , *inducimos* probabilidades sobre los sucesos asociados con R_X que se determinan estrictamente si las probabilidades asociadas con los sucesos en S están especificadas.

EJEMPLO 4.4. Si las monedas consideradas en el ejemplo 4.3 son «normales», tenemos $P(CS) = P(SC) = \frac{1}{4}$. Por tanto $P(CS, SC) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$. (Los cálculos anteriores son una consecuencia directa de nuestra suposición básica que se refiere a la regularidad de las monedas.) Puesto que el suceso $\{X = 1\}$ es equivalente al suceso $\{CS, SC\}$, usando la ecuación (4.1), tenemos que $P(X = 1) = P(CS, SC) = \frac{1}{2}$. [Realmente no había elección acerca del valor de $P(X = 1)$, consistente con la ecuación (4.2), una vez que se determinó $P(CS, SC)$. En este sentido se *inducen* las probabilidades asociadas con sucesos R_X .]

Observación: ahora que hemos establecido la existencia de una función de probabilidades inducidas sobre el recorrido de X (ecuaciones 4.1 y 4.2) encontraremos conveniente suprimir la naturaleza funcional de X . Por lo tanto escribiremos (como lo hicimos en el ejem-

plo anterior), $P(X = 1) = \frac{1}{2}$. Lo que se quiere decir es que un suceso en el espacio muestral S , llamado $\{CS, SC\} = \{s | X(s) = 1\}$ ocurre con probabilidad $\frac{1}{2}$. Por lo tanto asignamos esa misma probabilidad al suceso $\{X = 1\}$ en el recorrido. Continuaremos escribiendo expresiones como $P(X = 1)$, $P(X \leq 5)$, etc. Es muy importante que el lector se dé cuenta de lo que esas expresiones representan realmente.

Una vez que se han determinado (o más exactamente inducido) las probabilidades asociadas con varios resultados (o sucesos) en el recorrido R_X ignoraremos a menudo el espacio muestral original S que dio lugar a esas probabilidades. Así, en el ejemplo anterior, estamos simplemente interesados en $R_X = \{0, 1, 2\}$ y las probabilidades asociadas ($\frac{1}{4}, \frac{1}{2}, \frac{1}{4}$). El hecho de que estas probabilidades estén determinadas por una función de probabilidad definida en el espacio muestral original S no nos preocupa si estamos interesados sólo en estudiar los *valores* de la variable aleatoria X .

Al discutir, en detalle, muchos de los conceptos importantes asociados con variables aleatorias, encontraremos conveniente distinguir dos casos importantes: las variables aleatorias discretas y las continuas.

4.2 Variables aleatorias discretas

Definición. Sea X una variable aleatoria. Si el número de valores posibles de X (esto es, R_X , el recorrido) es finito o infinito numerable, llamamos a X una *variable aleatoria discreta*. Esto es, se pueden anotar los valores posibles de X como $x_1, x_2, \dots, x_n, \dots$. En el caso finito la lista termina y en el caso infinito numerable la lista continúa indefinidamente.

EJEMPLO 4.5. Una fuente radioactiva emite partículas alfa. Un contador observa la emisión de esas partículas durante un período de tiempo especificado. La siguiente variable aleatoria es de interés: X = número de partículas observadas. ¿Cuáles son los valores posibles de X ? Supondremos que estos valores constan de todos los enteros no negativos. Esto es, $R_X = \{0, 1, 2, \dots, n, \dots\}$. Una objeción que vimos anteriormente puede aparecer de nuevo en este punto. Podría argüirse que durante un intervalo de tiempo especificado (finito) es imposible observar más de N partículas, por ejemplo, en donde N puede ser un entero positivo muy grande. Por tanto los valores posibles de X realmente serían: $0, 1, 2, \dots, N$. Sin embargo resulta que matemáticamente es más sencillo considerar la descripción idealizada dada anteriormente. De hecho, cada vez que suponemos que los valores posibles de una variable aleatoria X son infinitos numerables estamos considerando en ese momento una representación idealizada de X .

En vista de nuestros comentarios previos sobre la descripción probabilística de sucesos con número finito de elementos o infinitos numerables, la descripción probabilística de una variable aleatoria discreta no nos causará ninguna dificultad. Procederemos como se indica a continuación.

Definición. Sea X una variable aleatoria discreta. Por tanto R_X , el recorrido de X , consta, a lo más, de un número de valores, x_1, x_2, \dots infinito numerable. Con cada resultado posible x_i asociamos un número $p(x_i) = P(X = x_i)$, llamado la probabilidad de x_i . Los números $p(x_i), i = 1, 2, \dots$ deben satisfacer las condiciones siguientes:

$$(a) \quad p(x_i) \geq 0 \quad \text{para todo } i, \quad (4.3)$$

$$(b) \quad \sum_{i=1}^{\infty} p(x_i) = 1.$$

La función p definida anteriormente se llama función de probabilidad (o función de probabilidad puntual) de la variable aleatoria X . La colección de pares $(x_i, p(x_i))$, $i = 1, 2, \dots$, se llama algunas veces distribución de probabilidad de X .

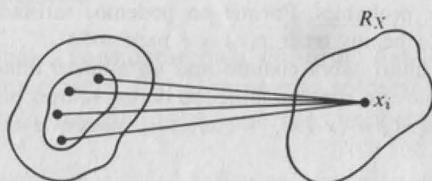


FIGURA 4.3

Observaciones: (a) La elección particular de los números $p(x_i)$ posiblemente está determinada por la función de probabilidad asociada con sucesos en el espacio muestral S sobre el cual se define X . Esto es, $p(x_i) = P[s | X(s) = x_i]$. (Ver ecuaciones 4.1 y 4.2.) Sin embargo, puesto que sólo estamos interesados en los valores de X , esto es R_X , y en las probabilidades asociadas con esos valores, omitimos nuevamente la naturaleza funcional de X . (Ver figura 4.3). Aunque en la mayoría de los casos, de hecho, los números se determinarán de la distribución de probabilidades en algún espacio muestral fundamental S , cualquier conjunto de números $p(x_i)$ que satisfaga la ecuación (4.3) puede servir como una descripción probabilística propia de una variable aleatoria discreta.

(b) Si X toma sólo un número finito de valores, por ejemplo x_1, \dots, x_N , entonces $p(x_i) = 0$ para $i > N$, y por lo tanto la serie infinita en la ecuación (4.3) llega a ser una suma finita.

(c) Nuevamente podemos observar una analogía con la mecánica al considerar una masa total unitaria distribuida sobre la recta real con la masa completa ubicada en los puntos x_1, x_2, \dots . Los números $p(x_i)$ representan la cantidad de masa ubicada en x_i .

(d) La interpretación geométrica (figura 4.4) de una distribución de probabilidades es siempre útil.

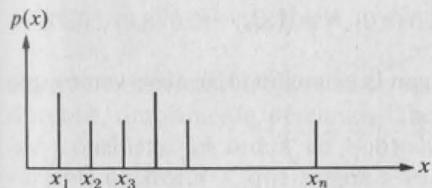


FIGURA 4.4

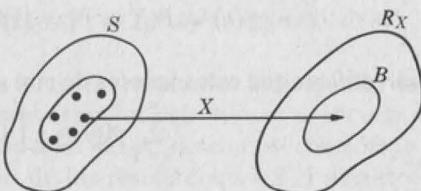


FIGURA 4.5

Sea B un suceso asociado con la variable aleatoria X . Esto es $B \subset R_X$ (figura 4.5). Específicamente, supongamos que $B = \{x_{i_1}, x_{i_2}, \dots\}$. Por tanto

$$P(B) = P[s | X(s) \in B] \quad (\text{puesto que estos sucesos son equivalentes})$$

$$= P[s | X(s) = x_{i_j}, j = 1, 2, \dots] = \sum_{j=1}^{\infty} p(x_{i_j}). \quad (4.4)$$

En palabras: la probabilidad de un suceso B es igual a la suma de las probabilidades de los resultados individuales asociados con B .

Observaciones: (a) Supóngase que la variable aleatoria discreta X puede tomar sólo un número finito de valores, por ejemplo x_1, \dots, x_N . Si cada resultado es igualmente probable, obviamente tenemos $p(x_1) = \dots = p(x_N) = 1/N$.

(b) Si X toma un número infinito numerable de valores, entonces es *imposible* tener todos los resultados igualmente probables. Porque no podemos satisfacer posiblemente la condición $\sum_{i=1}^{\infty} p(x_i) = 1$ si debemos tener $p(x_i) = c$ para todo i .

(c) En cada intervalo finito habrá cuando más un número finito de valores posibles de X . Si uno de tales intervalos no contiene ninguno de los valores posibles le asignamos probabilidad cero. Esto es, si $R_X = \{x_1, x_2, \dots, x_n\}$ y si ningún $x_i \in [a, b]$, entonces $P[a \leq X \leq b] = 0$.

EJEMPLO 4.6. Supóngase que se pone un tubo de radio en un soporte y se prueba. Supóngase que la probabilidad de que el control sea positivo es igual a $\frac{3}{4}$; por tanto la probabilidad de que el control sea negativo es $\frac{1}{4}$. Supóngase además, que probamos una gran cantidad de tales tubos. La prueba continúa hasta que aparece el primer tubo positivo. Definase la variable aleatoria X como sigue: X es el número de pruebas necesarias para finalizar el experimento. El espacio muestral asociado con este experimento es

$$S = \{+, -, +, -, +, -, - +, \dots\}.$$

Para determinar la distribución de probabilidades de X razonamos como sigue. Los valores posibles de X son $1, 2, \dots, n, \dots$ (obviamente estamos considerando el espacio muestral idealizado). Y $X = n$ si y sólo si los primeros $(n-1)$ tubos son negativos y el n -ésimo tubo es positivo. Si suponemos que la condición de un tubo no afecta la condición de otro podemos escribir

$$p(n) = P(X = n) = \left(\frac{1}{4}\right)^{n-1} \left(\frac{3}{4}\right), \quad n = 1, 2, \dots$$

Para verificar que estos valores de $p(n)$ satisfagan la ecuación (4.3), observemos que

$$\sum_{n=1}^{\infty} p(n) = \frac{3}{4} \left(1 + \frac{1}{4} + \frac{1}{16} + \dots\right)$$

$$= \frac{3}{4} \frac{1}{1 - \frac{1}{4}} = 1.$$

Observaciones: aquí empleamos el resultado de que la serie geométrica $1 + r + r^2 + \dots$ converge a $1/(1 - r)$ siempre que $|r| < 1$. A este resultado nos referiremos varias veces. Supóngase que queremos calcular $P(A)$, en donde A se define como {El experimento termina después de un número par de repeticiones}. Usando la ecuación (4.4), tenemos

$$\begin{aligned} P(A) &= \sum_{n=1}^{\infty} p(2n) = \frac{3}{16} + \frac{3}{256} + \dots \\ &= \frac{3}{16}(1 + \frac{1}{16} + \dots) \\ &= \frac{3}{16} \cdot \frac{1}{1 - \frac{1}{16}} = \frac{1}{5}. \end{aligned}$$

4.3 La distribución binomial

En los capítulos finales consideraremos, en forma detallada diversas variables aleatorias discretas importantes. Por el momento sólo estudiaremos una de estas y luego la usaremos para ilustrar varios conceptos importantes.

EJEMPLO 4.7. Supóngase que los artículos que salen de una línea de producción se clasifican como defectuosos (D) o no defectuosos (N). Supongamos que se eligen al azar tres artículos de la producción de un día y se clasifican de acuerdo con este esquema. El espacio muestral para este experimento, digamos S , puede describirse así:

$$S = \{DDD, DDN, DND, NDD, NND, NDN, DNN, NNN\}.$$

(Otra manera de describir S es como $S = S_1 \times S_2 \times S_3$, el producto cartesiano de S_1, S_2 , y S_3 , en donde cada $S_i = \{D, N\}$.)

Supongamos que con probabilidad 0,2 un artículo es defectuoso y por lo tanto con probabilidad 0,8 un artículo no es defectuoso. Supongamos que esas probabilidades son *iguales* para cada artículo al menos durante nuestro estudio. Finalmente, supongamos que la clasificación de cualquier artículo particular es independiente de la clasificación de cualquier otro artículo. Usando estas suposiciones, se deduce que las probabilidades asociadas con los diversos resultados del espacio muestral S como se describió anteriormente son

$$(0,2)^3, (0,8)(0,2)^2, (0,8)(0,2)^2, (0,8)(0,2)^2, (0,2)(0,8)^2, (0,2)(0,8)^2, (0,2)(0,8)^2, (0,8)^3.$$

Corrientemente nuestro interés no se enfoca hacia los resultados individuales de S ; sino que, simplemente, deseamos saber *cuántos* artículos defectuosos se encuentran (sin considerar el orden en que ocurrieron). Es decir, deseamos considerar la variable aleatoria X que asigna a cada uno de los resultados $s \in S$ el número de artículos defectuosos encontrados en s . Por tanto el conjunto de valores posibles de X es $\{0, 1, 2, 3\}$.

Podemos obtener la distribución de probabilidades para X , $p(x_i) = P(X = x_i)$ como sigue:

$X = 0$	si y sólo si ocurre	NNN ;
$X = 1$	si y sólo si ocurre	DNN, NDN , o NND ;
$X = 2$	si y sólo si ocurre	DDN, DND , o NDD ;
$X = 3$	si y sólo si ocurre	DDD .

(Nótese que $\{NNN\}$ es equivalente a $\{X = 0\}$, etc.) Por lo tanto

$$\begin{aligned} p(0) &= P(X = 0) = (0.8)^3, & p(1) &= P(X = 1) = 3(0.2)(0.8)^2, \\ p(2) &= P(X = 2) = 3(0.2)^2(0.8), & p(3) &= P(X = 3) = (0.2)^3. \end{aligned}$$

Nótese que la suma de estas probabilidades es igual a 1, porque la suma puede escribirse $(0.8 + 0.2)^3$.

Observación: la presentación anterior ilustra cómo las probabilidades en el recorrido R_X (en este caso $\{0, 1, 2, 3\}$) son *inducidas* por las probabilidades definidas en el espacio muestral S , porque la suposición de que los ocho resultados de

$$S = \{DDD, DDN, DND, NDD, NND, NDN, DNN, NNN\}$$

tienen las probabilidades dadas por el ejemplo 4.7, que *determina* el valor de $p(x)$ para todo $x \in R_X$.

Generalicemos ahora las nociones presentadas en el ejemplo anterior.

Definición. Consideremos un experimento ε y sea A un suceso asociado con ε . Supongamos que $P(A) = p$ y por lo tanto $P(\bar{A}) = 1 - p$. Consideremos n repeticiones independientes de ε . Por lo tanto el espacio muestral consiste en todas las sucesiones posibles $\{a_1, a_2, \dots, a_n\}$, en donde cada a_i es A o \bar{A} , según que A o \bar{A} ocurra en la i -ésima repetición de ε . (Hay 2^n de tales sucesiones.) Aún más, supongamos que $P(A) = p$ es el mismo para todas las repeticiones. Definamos la variable aleatoria X como sigue: X = número de veces que ocurrió el suceso A . Llamamos a X una variable aleatoria *binomial* con los parámetros n y p . Sus valores posibles obviamente son $0, 1, 2, \dots, n$. (Decimos en forma equivalente que X tiene una *distribución binomial*.) Las repeticiones individuales de ε se llamarán *ensayos de Bernoulli*.

Teorema 4.1. Sea X una variable binomial basada en n repeticiones. Entonces

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}, \quad k = 0, 1, \dots, n. \quad (4.5)$$

Demostración: consideremos un elemento particular del espacio muestral ε que satisfaga la condición de que $X = k$. Tal resultado aparecería, por ejemplo, si las primeras k repeticiones de ε resultasen en la ocurrencia de A mientras que las últimas $n - k$ repeticiones resultasen en \bar{A} , es decir:

$$\underbrace{AAA \cdots A}_k \underbrace{\bar{A} \bar{A} \bar{A} \cdots \bar{A}}_{n-k}.$$

Puesto que todas las repeticiones son independientes, la probabilidad de esta sucesión particular sería $p^k(1-p)^{n-k}$. Pero exactamente la misma probabilidad estaría asociada con cualquier otro resultado para el cual $X = k$. El número total de tales resultados es igual a $\binom{n}{k}$, por lo que debemos elegir exactamente k posiciones (entre n) para las A . Pero esto produce el resultado anterior ya que esos $\binom{n}{k}$ resultados son mutuamente excluyentes.

Observaciones: (a) Para verificar nuestros cálculos observemos que, usando el teorema del binomio, tenemos $\sum_{k=0}^n P(X = k) = \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} = [p + (1-p)]^n = 1^n = 1$, como debería ser. Puesto que las probabilidades $\binom{n}{k} p^k (1-p)^{n-k}$ se obtienen al desarrollar la expresión binomial $[p + (1-p)]^n$, a ésta llamamos distribución binomial.

(b) Cada vez que realizamos repeticiones independientes de un experimento y nos interesamos sólo en una dicotomía —defectuosos o no defectuosos, dureza sobre o bajo un cierto valor estándar, nivel de ruido en un sistema de comunicaciones sobre o bajo un margen preasignado— consideraremos potencialmente un espacio muestral en el cual podemos definir una variable aleatoria binomial. Si las condiciones del experimento permanecen suficientemente uniformes de modo que la probabilidad de algún atributo, por ejemplo A , permanezca constante, podemos usar el modelo anterior.

(c) Si n es pequeño, los términos individuales de la distribución binomial son relativamente sencillos de calcular. Sin embargo, si n es razonablemente grande estos cálculos llegan a ser un poco complicados. Afortunadamente, las probabilidades binomiales han sido calculadas. Existen muchas de tales tabulaciones. (Ver Apéndice.)

EJEMPLO 4.8. Supongamos que un tubo de radio puesto en cierto tipo de equipo tiene una probabilidad de 0,2 de funcionar más de 500 horas. Si probamos 20 tubos, ¿cuál es la probabilidad de que exactamente k de ellos funcionen más de 500 horas, $k = 0, 1, 2, \dots, 20$?

Si X es el número de tubos que funcionan más de 500 horas, supondremos que X tiene una distribución binomial. Así $P(X = k) = \binom{20}{k} (0,2)^k (0,8)^{20-k}$.

Los siguientes valores pueden leerse en la tabla 4.1.

TABLA 4.1

$P(X = 0) = 0,012$	$P(X = 4) = 0,218$	$P(X = 8) = 0,022$
$P(X = 1) = 0,058$	$P(X = 5) = 0,175$	$P(X = 9) = 0,007$
$P(X = 2) = 0,137$	$P(X = 6) = 0,109$	$P(X = 10) = 0,002$
$P(X = 3) = 0,205$	$P(X = 7) = 0,055$	$P(X = k) = 0^+ \text{ para } k \geq 11$
(Las probabilidades restantes son menores que 0,001).		

Si dibujamos esta distribución de probabilidades obtenemos el gráfico que se muestra en la figura 4.6. El modelo que observamos aquí es muy general: las probabilidades binomiales aumentan monótonamente hasta que alcanzan un valor máximo y luego disminuyen monótonamente. (Ver problema 4.8.)

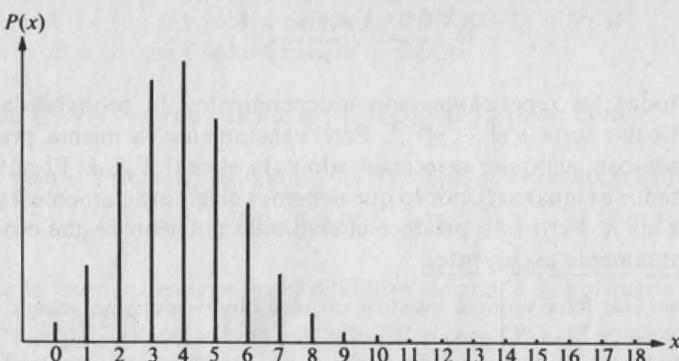


FIGURA 4.6

EJEMPLO 4.9. Al operar una máquina, existe cierta probabilidad de que el operario cometa un error. Puede suponerse de manera realista que el operario aprende en cuanto sabe que la probabilidad de cometer errores disminuye cuando él use repetidamente la máquina. Supongamos que el operario hace n intentos y que los n ensayos son estadísticamente independientes. Supongamos específicamente que $P(\text{un error cometido en la } i\text{-ésima repetición}) = 1/(i + 1)$, $i = 1, 2, \dots, n$. Supongamos que se consideran 4 intentos (esto es, $n = 4$) y que se define la variable aleatoria X como el número de operaciones hechas sin error en la máquina. Observemos que X no está distribuida binomialmente porque la probabilidad de «éxito» no es constante.

Para calcular la probabilidad de $X = 3$, por ejemplo, procedemos como sigue: $X = 3$ si y sólo si hay exactamente un intento no exitoso. Esto puede suceder en el primero, segundo, tercero, o cuarto ensayo. Por lo tanto

$$P(X = 3) = \frac{1}{2} \cdot \frac{2}{3} \cdot \frac{3}{4} \cdot \frac{4}{5} + \frac{1}{2} \cdot \frac{1}{3} \cdot \frac{3}{4} \cdot \frac{4}{5} + \frac{1}{2} \cdot \frac{2}{3} \cdot \frac{1}{4} \cdot \frac{4}{5} + \frac{1}{2} \cdot \frac{2}{3} \cdot \frac{3}{4} \cdot \frac{1}{5} = \frac{5}{12}.$$

EJEMPLO 4.10. Consideremos una situación semejante a la descrita en el ejemplo 4.9. Esta vez supondremos que hay una probabilidad constante p_1 de no cometer error en la máquina durante cada uno de los primeros n_1 intentos y una probabilidad constante $p_2 \leq p_1$ de no cometer error en cada una de las siguientes n_2 repeticiones. Sea X el número de operaciones exitosas de la máquina durante los $n = n_1 + n_2$ intentos independientes. Encontramos una expresión general para $P(X = k)$. Por la misma razón dada en el ejemplo precedente, X no está distribuida binomialmente. Para obtener $P(X = k)$ procedemos como sigue.

Sea Y_1 el número de operaciones correctas durante los primeros n_1 intentos y sea Y_2 el número de operaciones correctas durante los segundos n_2 intentos.

Por lo tanto, Y_1 e Y_2 son variables aleatorias independientes y $X = Y_1 + Y_2$. Así $X = k$ si y sólo si $Y_1 = r$ e $Y_2 = k - r$, para cualquier entero r que satisfaga $0 \leq r \leq n_1$ y $0 \leq k - r \leq n_2$.

Las restricciones anteriores sobre r son equivalentes a $0 \leq r \leq n_1$ y $k - n_2 \leq r \leq k$. Combinándolas podemos escribir

$$\max(0, k - n_2) \leq r \leq \min(k, n_1).$$

Por tanto tenemos

$$P(X = k) = \sum_{r=\max(0, k-n_2)}^{\min(k, n_1)} \binom{n_1}{r} p_1^r (1-p_1)^{n_1-r} \binom{n_2}{k-r} p_2^{k-r} (1-p_2)^{n_2-(k-r)}.$$

Con la convención corriente de que $\binom{a}{b} = 0$ cada vez que $b > a$ o $b < 0$, podemos escribir la probabilidad anterior como

$$P(X = k) = \sum_{r=0}^{n_1} \binom{n_1}{r} p_1^r (1-p_1)^{n_1-r} \binom{n_2}{k-r} p_2^{k-r} (1-p_2)^{n_2-k+r}. \quad (4.6)$$

Por ejemplo, si $p_1 = 0,2$, $p_2 = 0,1$, $n_1 = n_2 = 10$, y $k = 2$, la probabilidad anterior llega a ser

$$P(X = 2) = \sum_{r=0}^2 \binom{10}{r} (0,2)^r (0,8)^{10-r} \binom{10}{2-r} (0,1)^{2-r} (0,9)^{8+r} = 0,27,$$

después de un cálculo elemental.

Observación: supongamos que $p_1 = p_2$. En este caso, la ecuación (4.6) se reduciría a $\binom{n}{k} p_1^k (1-p_1)^{n-k}$, puesto que ahora la variable aleatoria X tiene una distribución binomial. Para ver que esto es así, obsérvese que podemos escribir (puesto que $n_1 + n_2 = n$)

$$P(X = k) = p_1^k (1-p_1)^{n-k} \sum_{r=0}^{n_1} \binom{n_1}{r} \binom{n_2}{k-r}.$$

Para mostrar que la suma anterior iguala $\binom{n}{k}$ compárense simplemente los coeficientes de las potencias de x^k en ambos lados de la identidad $(1+x)^{n_1}(1+x)^{n_2} = (1+x)^{n_1+n_2}$.

4.4 Variables aleatorias continuas

Supongamos que el recorrido de X está formado por un gran número de valores, por ejemplo todos los valores x en el intervalo $0 \leq x \leq 1$ de la forma $0; 0,01; 0,02; \dots, 0,98; 0,99; 1,00$. Con cada uno de esos valores está asociado un número no negativo $p(x_i) = P(X = x_i)$, $i = 1, 2, \dots$, cuya suma es igual a 1. Esta situación se representa geométricamente en la figura 4.7.

Hemos señalado antes que podría ser matemáticamente más fácil idealizar la

anterior descripción probabilística de X al suponer que X puede tomar *todos* los valores posibles, $0 \leq x \leq 1$. Si hacemos esto, ¿qué le sucede a las probabilidades puntuales $p(x_i)$? Puesto que los valores posibles de X no son contables, no podemos hablar realmente del i -ésimo valor de X y por lo tanto $p(x_i)$ pierde significado. Lo que haremos es sustituir la función p , definida sólo para x_1, x_2, \dots , por una función f definida (en el contexto presente) para *todos* los valores de x , $0 \leq x \leq 1$. Las propiedades de la ecuación (4.3) se sustituirán por $f(x) \geq 0$ y $\int_0^1 f(x)dx = 1$. Procederemos formalmente como sigue.

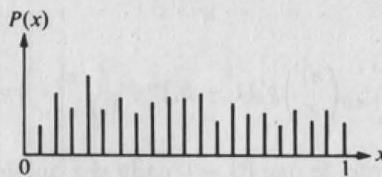


FIGURA 4.7

Definición. Se dice que X es una *variable aleatoria continua* si existe una función f , llamada función de densidad de probabilidad (fdp) de X , que satisface las siguientes condiciones:

$$(a) \quad f(x) \geq 0 \quad \text{para todo } x,$$

$$(b) \quad \int_{-\infty}^{+\infty} f(x)dx = 1. \quad (4.7)$$

$$(c) \quad \text{Para cualquier } a, b, \text{ tal que } -\infty < a < b < +\infty, \\ \text{tenemos } P(a \leq X \leq b) = \int_a^b f(x)dx. \quad (4.8)$$

Observaciones: (a) Fundamentalmente queremos decir que X es una variable aleatoria continua si X puede tomar todos los valores en algún intervalo (c, d) en donde c y d pueden ser $-\infty$ y $+\infty$, respectivamente. La existencia estipulada de una fdp es un método matemático que tiene una base intuitiva considerable y hacen más sencillos nuestros cálculos. En relación con esto, de nuevo se debe señalar que cuando suponemos que X es una variable aleatoria continua, estamos considerando la descripción *idealizada* de X .

(b) $P(c < X < d)$ representa el área bajo el gráfico en la figura 4.8 de la fdp f entre $x = c$ y $x = d$.

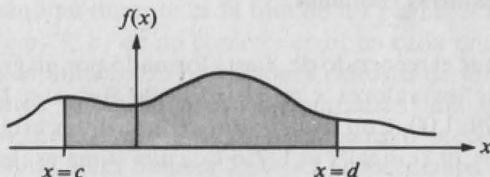


FIGURA 4.8

(c) Una consecuencia de la descripción probabilística de X para cualquier valor específico de X , por ejemplo x_0 es que tenemos $P(X = x_0) = 0$, puesto que $P(X = x_0) = \int_{x_0}^{x_0} f(x)dx = 0$. Este resultado puede parecer muy contrario a nuestra intuición. Debemos establecer, sin embargo, que si permitimos que X tome todos los valores en un intervalo, entonces la probabilidad cero no es equivalente con la imposibilidad. Por tanto, en el caso continuo, $P(A) = 0$ no implica $A = \emptyset$, el conjunto vacío. (Ver teorema 1.1.) Para decirlo de un modo más informal, consideremos que elegimos un punto al azar en el segmento $\{x | 0 \leq x \leq 2\}$. Aunque quisieramos estar de acuerdo (como un objetivo matemático) con que cada punto conceivable del segmento pudiese ser el resultado de nuestro experimento, nos sorprenderíamos mucho si en realidad escogieramos precisamente el punto medio del segmento, o cualquier otro punto específico de ese elemento. Cuando indicamos esto en un lenguaje matemático preciso, decimos que el suceso tiene «probabilidad 0». En vista de estas consideraciones, las siguientes probabilidades son todas iguales si X es una variable aleatoria continua:

$$P(c \leq X \leq d), \quad P(c \leq X < d), \quad P(c < X \leq d), \quad \text{y} \quad P(c < X < d).$$

(d) Aunque no verificaremos los detalles aquí, puede demostrarse que la anterior asignación de probabilidades a los sucesos en R_X satisface los axiomas básicos de probabilidades (ecuación 1.3), en donde podemos tomar $\{x | -\infty < x < +\infty\}$ como el espacio muestral.

(e) Si una función f^* satisface las condiciones, $f^*(x) \geq 0$, para todo x , y $\int_{-\infty}^{+\infty} f^*(x)dx = K$, en donde K es un número positivo real (no necesariamente igual a 1), entonces f^* no satisface todas las condiciones para ser una fdp. Sin embargo, podemos definir fácilmente una nueva función, por ejemplo f , en función de f^* como sigue:

$$f(x) = \frac{f^*(x)}{K} \quad \text{para todo } x.$$

Por tanto f satisface todas las condiciones de una fdp.

(f) Si X sólo toma valores en un intervalo finito $[a, b]$, simplemente podemos establecer $f(x) = 0$ para todo $x \notin [a, b]$. Por tanto la fdp está definida para todos los valores reales de x , y debemos exigir que $\int_{-\infty}^{+\infty} f(x)dx = 1$. Cuando quiera que la fdp se especifique sólo para ciertos valores de x , supondremos que es cero para cualquier otro.

(g) ¡ $f(x)$ no representa la probabilidad de nada! Hemos observado antes que $P(X = 2) = 0$, por ejemplo, y por tanto $f(2)$ ciertamente no representa esta probabilidad. Sólo cuando la función se integra entre dos límites produce una probabilidad. Sin embargo, podemos dar una interpretación de $f(x)\Delta x$ como sigue. Del teorema del valor medio del cálculo se deduce que

$$P(x \leq X \leq x + \Delta x) = \int_x^{x + \Delta x} f(s)ds = \Delta x f(\xi), \quad x \leq \xi \leq x + \Delta x.$$

Si Δx es pequeño, $f(x)\Delta x$ es aproximadamente igual a $P(x \leq X \leq x + \Delta x)$. Si f es continuo por la derecha, esta aproximación llega a ser más segura cuando $\Delta x \rightarrow 0$.

(h) Deberíamos señalar nuevamente que la distribución de probabilidades (en este caso la fdp) es inducida en R_X por las probabilidades asociadas con sucesos en S . Así cuando escribimos $P(c < X < d)$, queremos decir, como siempre $P[c < X(s) < d]$, que a su vez es igual a $P[s | c < X(s) < d]$, puesto que estos sucesos son equivalentes. La definición anterior, ecuación (4.8), estipula esencialmente la existencia de una fdp f definida en una R_X tal que

$$P[s | c < X(s) < d] = \int_c^d f(x) dx.$$

Nuevamente eliminaremos la naturaleza funcional de X y por tanto estaremos interesados sólo en R_X y la fdp f .

(i) En el caso continuo, nuevamente podemos considerar la siguiente *analogía con la mecánica*: supóngase que tenemos una masa total de una unidad, distribuida continuamente en el intervalo $a \leq x \leq b$. Entonces $f(x)$ representa la densidad de la masa en el punto x y $\int_a^x f(x)dx$ representa la masa total contenida en el intervalo $c \leq x \leq d$.

EJEMPLO 4.11. En la discusión precedente sobre variable aleatoria, se supuso la existencia de una fdp. Consideremos un ejemplo simple en el cual podamos determinar fácilmente la fdp, haciendo una suposición apropiada acerca de la conducta probabilística de la variable aleatoria. Supóngase que escogemos un punto en el intervalo $(0, 1)$. Representemos la variable aleatoria X cuyo valor es la coordenada x del punto escogido.

Suposición: si I es cualquier intervalo entre $(0, 1)$, entonces la $\text{Prob}[X \in I]$ es directamente proporcional a la longitud de I , por ejemplo $L(I)$. Esto es, $\text{Prob}[X \in I] = kL(I)$, en donde k es la constante de proporcionalidad. (Es fácil ver, tomando $I = (0, 1)$ y observando que $L((0, 1)) = 1$ y $\text{Prob}[X \in (0, 1)] = 1$, que $k = 1$.)

Obviamente X toma todos los valores en $(0, 1)$. ¿Cuál es su fdp? Esto es, ¿podemos encontrar una función f tal que

$$P(a < X < b) = \int_a^b f(x)dx?$$

Nótese que si $a < b < 0$ o $1 < a < b$, $P(a < X < b) = 0$ y por tanto $f(x) = 0$. Si $0 < a < b < 1$, $P(a < X < b) = b - a$ y por tanto $f(x) = 1$. Así encontramos,

$$f(x) = \begin{cases} 1, & 0 < x < 1, \\ 0, & \text{para cualquier otro valor.} \end{cases}$$

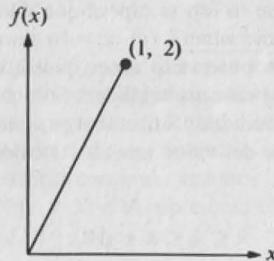


FIGURA 4.9

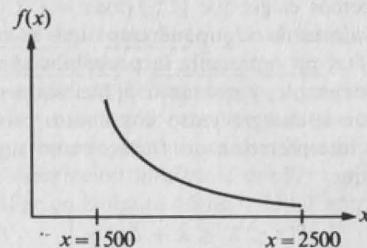


FIGURA 4.10

EJEMPLO 4.12. Supóngase que la variable aleatoria X es continua. (Ver figura 4.9.) Sea la fdp f dada por

$$\begin{aligned} f(x) &= 2x, & 0 < x < 1, \\ &= 0, & \text{para cualquier otro valor.} \end{aligned}$$

Claramente, $f(x) \geq 0$ y $\int_{-\infty}^{+\infty} f(x)dx = \int_0^1 2xdx = 1$. Para calcular $P(X \leq \frac{1}{2})$, debemos evaluar simplemente la integral $\int_0^{1/2} (2x)dx = \frac{1}{4}$.

El concepto de probabilidad condicional tratado en el capítulo 3 se puede aplicar debidamente a las variables aleatorias. Por ejemplo, en el ejemplo anterior debimos evaluar $P(X \leq \frac{1}{2} | \frac{1}{3} \leq X \leq \frac{2}{3})$. Aplicando directamente la definición de probabilidad individual, tenemos

$$\begin{aligned} P(X \leq \frac{1}{2} | \frac{1}{3} \leq X \leq \frac{2}{3}) &= \frac{P(\frac{1}{3} \leq X \leq \frac{1}{2})}{P(\frac{1}{3} \leq X \leq \frac{2}{3})} \\ &= \frac{\int_{1/3}^{1/2} 2xdx}{\int_{1/3}^{2/3} 2xdx} = \frac{5/36}{1/3} = \frac{5}{12}. \end{aligned}$$

EJEMPLO 4.13. Sea X la duración (en horas) de cierto tipo de bombilla eléctrica. Supóngase que X es una variable aleatoria continua y supóngase que la fdp f de X está dada por

$$\begin{aligned} f(x) &= a/x^3, & 1500 \leq x \leq 2500, \\ &= 0, & \text{para cualquier otro valor.} \end{aligned}$$

(Es decir, asignamos probabilidad cero al suceso $\{X \leq 1500\}$ y $\{X > 2500\}$.) Para evaluar la constante a , debemos acudir a la condición $\int_{-\infty}^{+\infty} f(x)dx = 1$ que en este caso llega a ser $\int_{1500}^{2500} a/x^3 dx = 1$. De esto obtenemos $a = 7,031,250$. El gráfico de f aparece en la figura 4.10.

En un capítulo posterior estudiaremos, con mucho detalle, diversas variables aleatorias importantes, discretas y continuas. Mediante el uso de los determinísticos sabemos que ciertas funciones desempeñan un papel mucho más importante que otras. Por ejemplo, las funciones lineales, cuadráticas, exponenciales y trigonométricas juegan un papel primordial al describir modelos determinísticos. Al desarrollar los modelos no determinísticos (es decir, probabilísticos) encontraremos que ciertas variables aleatorias son particularmente importantes.

4.5 Función de distribución acumulativa

Queremos presentar otro importante concepto general en este capítulo.

Definición. Sea X una variable aleatoria, discreta o continua. Definimos que F es la función de distribución acumulativa de la variable aleatoria X (abreviada fda) como $F(x) = P(X \leq x)$.

Teorema 4.2. (a) Si X es una variable aleatoria discreta,

$$F(x) = \sum_j p(x_j), \tag{4.9}$$

en donde la suma se toma sobre todos los índices j que satisfacen $x_j \leq x$.

(b) Si X es una variable aleatoria continua con fdp f ,

$$F(x) = \int_{-\infty}^x f(s) ds. \quad (4.10)$$

Demostración: ambos resultados se deducen inmediatamente de la definición.

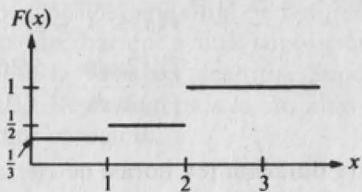


FIGURA 4.11

EJEMPLO 4.14. Supongamos que la variable aleatoria X toma los tres valores 0, 1, y 2 con probabilidades $\frac{1}{3}$, $\frac{1}{6}$, y $\frac{1}{2}$, respectivamente. Entonces

$$\begin{aligned} F(x) &= 0 && \text{si } x < 0, \\ &= \frac{1}{3} && \text{si } 0 \leq x < 1, \\ &= \frac{1}{2} && \text{si } 1 \leq x < 2, \\ &= 1 && \text{si } x \geq 2. \end{aligned}$$

(Nótese que es muy importante indicar la inclusión o exclusión de los puntos extremos al describir los diversos intervalos.) El gráfico de F se da en la figura 4.11.

EJEMPLO 4.15. Supongamos que X es una variable aleatoria continua con fdp

$$\begin{aligned} f(x) &= 2x, && 0 < x < 1, \\ &= 0, && \text{para cualquier otro valor.} \end{aligned}$$

Por lo tanto la fda F está dada por

$$\begin{aligned} F(x) &= 0 && \text{si } x \leq 0, \\ &= \int_0^x 2s ds = x^2 && \text{si } 0 < x \leq 1, \\ &= 1 && \text{si } x > 1. \end{aligned}$$

El gráfico aparece en la figura 4.12.

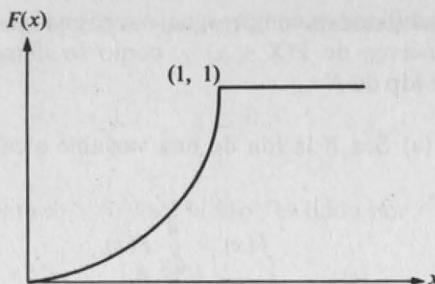


FIGURA 4.12

Los gráficos obtenidos en las figuras 4.11 y 4.12 para las fda (en cada uno de los casos) son muy típicos en el siguiente sentido.

(a) Si X es una variable aleatoria discreta con un número finito de valores posibles, el gráfico de la fda F se formará de trazos horizontales (se llama función escalonada). La función F es continua excepto para los valores posibles de X , llamados, x_1, \dots, x_n . En el valor x_j el gráfico tendrá un «salto» de magnitud $p(x_j) = P(X = x_j)$.

(b) Si X es una función aleatoria continua, F será una función continua para todo x .

(c) La fda F está definida para todos los valores de x , lo cual es una razón importante para considerarla.

Hay dos propiedades importantes de la fda que resumiremos en el teorema siguiente.

Teorema 4.3. (a) La función F es no decreciente. Esto es, si $x_1 \leq x_2$, tenemos $F(x_1) \leq F(x_2)$.

(b) $\lim_{x \rightarrow -\infty} F(x) = 0$ y $\lim_{x \rightarrow \infty} F(x) = 1$. [A menudo esto lo escribimos como $F(-\infty) = 0$, $F(\infty) = 1$.]

*Demuestra*ción: (a) Definamos los sucesos A y B como sigue: $A = \{X \leq x_1\}$, $B = \{X \leq x_2\}$. Entonces, puesto que $x_1 \leq x_2$, tenemos $A \subset B$ y por el teorema 1.5, $P(A) \leq P(B)$, que es el resultado pedido.

(b) En el caso continuo tenemos

$$F(-\infty) = \lim_{x \rightarrow -\infty} \int_{-\infty}^x f(s) ds = 0,$$

$$F(\infty) = \lim_{x \rightarrow \infty} \int_{-\infty}^x f(s) ds = 1$$

En el caso discreto el argumento es análogo.

La función de distribución acumulativa es importante por varias razones. Esto es particularmente cierto cuando tratamos con una variable aleatoria continua, porque en este caso no podemos estudiar la conducta probabilística de X al calcular

$P(X = x)$. Esta probabilidad es siempre igual a cero en el caso continuo. Sin embargo, podemos inquirir acerca de $P(X \leq x)$ y, como lo demostramos en el próximo teorema, obtener la fdp de X .

Teorema 4.4. (a) Sea F la fda de una variable aleatoria continua con fdp f . Luego

$$f(x) = \frac{d}{dx} F(x),$$

para toda x en la cual F es diferenciable.

(b) Sea X una variable aleatoria discreta con valores posibles x_1, x_2, \dots , y supongamos que es posible rotular esos valores de modo que $x_1 < x_2 < \dots$ Sea F la fda de X . Entonces

$$p(x_j) = P(X = x_j) = F(x_j) - F(x_{j-1}). \quad (4.12)$$

Demostración: (a) $F(x) = P(X \leq x) = \int_{-\infty}^x f(s) ds$. Así, aplicando al teorema fundamental del cálculo obtenemos, $F'(x) = f(x)$.

(b) Puesto que supusimos $x_1 < x_2 < \dots$, tenemos

$$\begin{aligned} F(x_j) &= P(X = x_j \cup X = x_{j-1} \cup \dots \cup X = x_1) \\ &= p(x_j) + p(x_{j-1}) + \dots + p(x_1). \end{aligned}$$

y

$$\begin{aligned} F(x_{j-1}) &= P(X = x_{j-1} \cup X = x_{j-2} \cup \dots \cup X = x_1) \\ &= p(x_{j-1}) + p(x_{j-2}) + \dots + p(x_1). \end{aligned}$$

Luego $F(x_j) - F(x_{j-1}) = p(x_j) = p(x_j)$.

Observación: reconsideraremos brevemente (a) del teorema anterior. Recordemos la definición de derivada de la función F :

$$\begin{aligned} F'(x) &= \lim_{h \rightarrow 0} \frac{F(x+h) - F(x)}{h} \\ &= \lim_{h \rightarrow 0^+} \frac{P(X \leq x+h) - P(X \leq x)}{h} \\ &= \lim_{h \rightarrow 0^+} \frac{1}{h} [P(x < X \leq x+h)]. \end{aligned}$$

Así si h es pequeño y positivo

$$F'(x) = f(x) \cong \frac{P(x < X \leq x+h)}{h}.$$

Es decir, $f(x)$ es aproximadamente igual a la «cantidad de probabilidad en el intervalo $(x, x+h)$ por longitud h ». De aquí el nombre de *función densidad de probabilidad*.

EJEMPLO 4.16. Supongamos que una variable aleatoria continua tiene fda F dada por

$$\begin{aligned} F(x) &= 0, & x \leq 0, \\ &= 1 - e^{-x}, & x > 0. \end{aligned}$$

Entonces $F'(x) = e^{-x}$ para $x > 0$, y así la fdp f es dada por

$$\begin{aligned} f(x) &= e^{-x}, & x \geq 0, \\ &= 0, & \text{para cualquier otro valor} \end{aligned}$$

Observación: una advertencia sobre la terminología puede ser de utilidad. Esta terminología, aunque no es muy uniforme, ha llegado a estandarizarse. Cuando hablamos de la *distribución de probabilidades* de una variable aleatoria X indicamos su fdp f si X es continua, o de su función de probabilidad puntual p definida para x_1, x_2, \dots si X es discreta. Cuando hablamos de la función de distribución acumulativa, o algunas veces sólo de la *función de distribución*, siempre nos referimos a F , en donde $F(x) = P(X \leq x)$.

4.6 Distribuciones mixtas

Hemos restringido nuestra exposición sólo a variables aleatorias que son discretas o continuas. Tales variables aleatorias ciertamente son las más importantes en las aplicaciones. Sin embargo, hay situaciones en las cuales podemos encontrar variables del tipo *mixto*: la variable aleatoria X puede tomar ciertos valores distintos, por ejemplo x_1, \dots, x_n , con probabilidad positiva y tomar también todos los valores en algún intervalo, por ejemplo $a \leq x \leq b$. La distribución de probabilidades de tales variables aleatorias se obtendría al combinar las ideas consideradas anteriormente para la descripción de las variables aleatorias discretas y continuas como sigue. A cada uno de los valores x_i se le asigna un número $p(x_i)$ tal que $p(x_i) \geq 0$, para todo i , y tal que $\sum_{i=1}^n p(x_i) = p < 1$. Entonces definimos una función f que satisface $f(x) \geq 0$, $\int_a^b f(x) dx = 1 - p$. Para todo a, b , con $-\infty < a < b < +\infty$,

$$P(a \leq X \leq b) = \int_a^b f(x) dx + \sum_{\{(i: a \leq x_i \leq b)\}} p(x_i).$$

De esta manera satisfacemos la condición

$$P(S) = P(-\infty < X < \infty) = 1.$$

Una variable aleatoria del tipo mixto puede aparecer como sigue. Supóngase que estamos probando un equipo y sea X el tiempo de funcionamiento. En la mayoría de los problemas describiríamos X como una variable aleatoria continua con valores posibles $x \geq 0$. Sin embargo, hay algunos casos en los cuales hay una probabilidad positiva, de que el artículo no funcione del todo, es decir, falla al tiempo $X = 0$. En tal caso desearíamos modificar nuestro modelo y asignar una

probabilidad positiva, por ejemplo p , al resultado $X = 0$. Por tanto, tendríamos $P(X = 0) = p$ y $P(X > 0) = 1 - p$. Así el número p describiría la distribución de X en 0, mientras que la fdp f describiría la distribución de valores de $X > 0$ (figura 4.13).

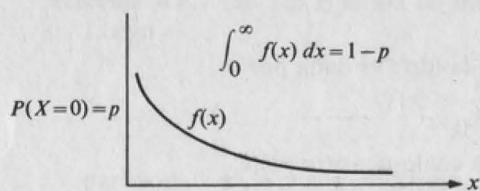


FIGURA 4.13

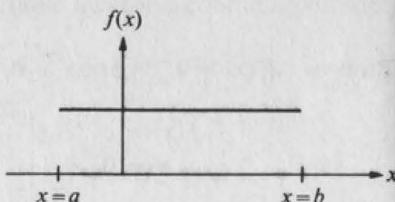


FIGURA 4.14

4.7 Variables aleatorias distribuidas uniformemente

En los capítulos 8 y 9 estudiaremos detalladamente diversas variables aleatorias importantes, discretas y continuas. Ya hemos presentado la importante variable aleatoria binomial. Consideremos brevemente ahora una variable continua de importancia.

Definición. Supongamos que X es una variable aleatoria continua que toma todos los valores en el intervalo $[a, b]$, en donde ambos a y b son finitos. Si la fdp de X está dada por

$$f(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b, \\ 0, & \text{para cualquier otro valor} \end{cases} \quad (4.13)$$

decimos que X está *distribuida uniformemente* en el intervalo $[a, b]$. (Ver figura 4.14.)

Observaciones: (a) Una variable aleatoria uniformemente distribuida tiene una fdp que es una constante en el intervalo de definición. A fin de satisfacer la condición $\int_{-\infty}^{+\infty} f(x)dx = 1$, esta constante debe ser igual al recíproco de la longitud del intervalo.

(b) Una variable aleatoria distribuida uniformemente representa la analogía continua a los resultados igualmente posibles en el sentido siguiente. Para cualquier sub-intervalo $[c, d]$, en donde $a \leq c < d \leq b$, $P(c \leq X \leq d)$ es la misma para todos los sub-intervalos que tienen la misma longitud. Esto es,

$$P(c \leq X \leq d) = \int_c^d f(x)dx = \frac{d-c}{b-a}$$

y así sólo depende de la longitud del intervalo y no de la ubicación del intervalo.

(c) Ahora podemos hacer precisa la noción intuitiva de *elegir un punto P al azar* en un intervalo, por ejemplo $[a, b]$. Con esto sencillamente indicaremos que la coordenada x del punto elegido, por ejemplo X , está distribuida uniformemente en $[a, b]$.

EJEMPLO 4.17. Un punto se elige al azar sobre el segmento de línea $[0, 2]$. ¿Cuál es la probabilidad de que el punto escogido quede entre 1 y $\frac{3}{2}$?

Representando la coordenada del punto elegido por X , tenemos que la fdp de X está dada por $f(x) = \frac{1}{2}$, $0 < x < 2$, y por tanto $P(1 \leq X \leq \frac{3}{2}) = \frac{1}{4}$.

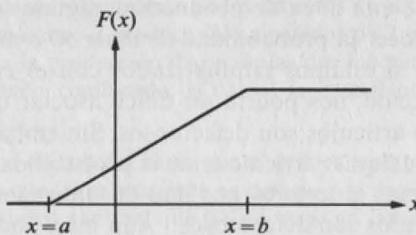


FIGURA 4.15

EJEMPLO 4.18. Se puede suponer que la dureza, H , de una muestra de acero (medida en la escala Rockwell) es una variable aleatoria continua distribuida uniformemente sobre $[50, 70]$ en la escala B. Por tanto

$$f(h) = \frac{1}{20}, \quad 50 < h < 70,$$

$$= 0, \quad \text{para cualquier otro valor.}$$

EJEMPLO 4.19. Obtengamos una expresión para la fda de una variable aleatoria distribuida uniformemente

$$\begin{aligned} F(x) &= P(X \leq x) = \int_{-\infty}^x f(s) ds \\ &= 0 \quad \text{si} \quad x < a, \\ &= \frac{x - a}{b - a} \quad \text{si} \quad a \leq x < b, \\ &= 1 \quad \text{si} \quad x \geq b. \end{aligned}$$

El gráfico aparece en la figura 4.15.

4.8 Una observación

Hemos señalado repetidamente que en alguna etapa del desarrollo de nuestro modelo probabilístico, deben asignarse algunas probabilidades a los resultados sobre la base de alguna evidencia experimental (tal como la frecuencia relativa, por ejemplo) o algunas otras consideraciones, tales como experiencias pasadas

con los fenómenos estudiados. El interrogante siguiente se le podría presentar al estudiante: ¿por qué no podríamos obtener todas las probabilidades en las cuales estamos interesados por un medio no deductivo? La respuesta es que muchos sucesos cuyas probabilidades deseamos conocer son tan complicados que nuestro conocimiento intuitivo es insuficiente. Por ejemplo, supongamos que diariamente salen 1000 artículos de una línea de producción, algunos de los cuales son defectuosos. Deseamos conocer la probabilidad de tener 50 o más artículos defectuosos en un día dado. Aun si estamos familiarizados con el comportamiento general del proceso de producción, nos podría ser difícil asociar una medida cuantitativa al suceso: 50 o menos artículos son defectuosos. Sin embargo, podríamos afirmar que individualmente cualquier artículo tenía la probabilidad 0,10 de ser defectuoso. (Esto es, las experiencias anteriores nos dan la información de que cerca del 10 por ciento de los artículos son defectuosos.) Aún más, podríamos suponer que los artículos individuales son defectuosos o no defectuosos independientemente, uno de otro. *Ahora podemos proceder deductivamente y derivar la probabilidad del suceso que se considera.* Esto es, si X = número de defectuosos,

$$P(X \leq 50) = \sum_{k=0}^{50} \binom{1000}{k} (0,10)^k (0,90)^{1000-k}.$$

Lo que se destaca aquí es que los diversos métodos para calcular probabilidades que hemos derivado (y los que estudiaremos posteriormente), son de mucha importancia puesto que con ellos podemos calcular las probabilidades asociadas con sucesos más complicados lo cual sería difícil de obtener con medios intuitivos o empíricos.

PROBLEMAS

4.1. Se sabe que en una moneda sale cara tres veces más a menudo que sello. Esta moneda se lanza tres veces. Sea X el número de caras que aparecen. Establecer la distribución de probabilidades de X y también la fda. Hacer un gráfico de ambas.

4.2. De un lote que contiene 25 artículos, 5 de los cuales son defectuosos, se eligen 4 al azar. Sea X el número de artículos defectuosos encontrados. Obtener la distribución de probabilidades de X si

- (a) los artículos se escogen con sustitución,
- (b) los artículos se escogen sin sustitución.

4.3. Supóngase que la variable aleatoria X tiene valores posibles $1, 2, 3, \dots$, y $P(X = j) = 1/2^j$, $j = 1, 2, \dots$

- (a) Calcular $P(X \text{ es par})$.
- (b) Calcular $P(X \geq 5)$.
- (c) Calcular $P(X \text{ es divisible por } 3)$.

4.4. Considérese una variable aleatoria X con resultados posibles: $0, 1, 2, \dots$ Supongamos que $P(X = j) = (1 - a)a^j$, $j = 0, 1, 2, \dots$

- (a) ¿Para qué valores de a es significativo el modelo anterior?
 (b) Verificar que lo anterior representa una distribución de probabilidades legítima.
 (c) Demotrar que para dos enteros positivos cualesquiera s y t ,

$$P(X > s + t \mid X > s) = P(X \geq t).$$

4.5. Supóngase que la máquina 1 produce (diariamente) el doble de artículos que la máquina 2. Sin embargo, cerca del 4 por ciento de los artículos de la máquina 1 tiende a ser defectuoso mientras que la máquina 2 produce sólo alrededor de 2 por ciento defectuosos. Supongamos que se combina la producción diaria de las dos máquinas. Se toma una muestra aleatoria de 10 del resultado combinado. ¿Cuál es la probabilidad de que esta muestra contenga 2 defectuosos?

4.6. Se lanza una serie de cohetes hasta que el primer lanzamiento exitoso tiene lugar. Si esto no ocurre en 5 ensayos, el experimento se detiene y se inspecciona el equipo. Supóngase que hay una probabilidad constante de 0,8 de tener un lanzamiento exitoso y que los ensayos sucesivos son independientes. Además, el costo del primer lanzamiento es K dólares mientras que los lanzamientos que siguen cuestan $K/3$ dólares. Cada vez que hay un lanzamiento exitoso, se obtiene cierta cantidad de información que puede expresarse como una ganancia financiera de C dólares. Si T es el costo neto del experimento, encontrar la distribución de probabilidades de T .

4.7. Calcular $P(X = 5)$, en donde X es la variable aleatoria definida en el ejemplo 4.10. Supongamos que $n_1 = 10$, $n_2 = 15$, $p_1 = 0,3$ y $p_2 = 0,2$.

4.8. (*Propiedades de las probabilidades binomiales.*) En la discusión del ejemplo 4.8 se sugirió un modelo general para las probabilidades binomiales $\binom{n}{k} p^k (1-p)^{n-k}$. Indiquemos estas probabilidades por $p_n(k)$.

- (a) Demostrar que para $0 \leq k < n$ tenemos

$$p_n(k+1)/p_n(k) = [(n-k)/(k+1)][p/(1-p)].$$

- (b) Usando (a), demostrar que

- (i) $p_n(k+1) > p_n(k)$ si $k < np - (1-p)$,
- (ii) $p_n(k+1) = p_n(k)$ si $k = np - (1-p)$,
- (iii) $p_n(k+1) < p_n(k)$ si $k > np - (1-p)$.

(c) Demostrar que si $np - (1-p)$ es un entero, $p_n(k)$ toma su valor máximo para dos valores de k , llamados $k_0 = np - (1-p)$ y $k'_0 = np - (1-p) + 1$.

(d) Demostrar que si $np - (1-p)$ no es un entero entonces $p_n(k)$ toma su valor máximo cuando k es igual al entero más pequeño mayor que k_0 .

(e) Demostrar que si $np - (1-p) < 0$, $p_n(0) > p_n(1) > \dots > p_n(n)$ mientras que si $np - (1-p) = 0$, $p_n(0) = p_n(1) > p_n(2) > \dots > p_n(n)$.

4.9. La variable aleatoria continua X tiene fdp $f(x) = x/2$, $0 \leq x \leq 2$. Se hacen dos determinaciones independientes de X . ¿Cuál es la probabilidad de que ambas determinaciones sean mayores que uno? Si se han hecho tres determinaciones independientes, ¿cuál es la probabilidad de que exactamente dos sean mayores que uno?

4.10. Sea X la duración de un tubo electrónico y supongamos que X se puede representar como una variable aleatoria continua con fdp $f(x) = be^{-bx}$, $x \geq 0$. Sea $p_j = P(j \leq X < j+1)$. Demostrar que p_j es de la forma $(1-a)a^j$ y determine a .

4.11. La variable aleatoria continua X tiene la fdp $f(x) = 3x^2$, $-1 \leq x \leq 0$. Si b es un número que satisface $-1 < b < 0$, calcular $P(X > b \mid X < b/2)$.

4.12. Supongamos que f y g son fdp en el mismo intervalo, $a \leq x \leq b$.

(a) Demostrar que $f + g$ no es una fdp en ese intervalo.

(b) Demostrar que para todo número β , $0 < \beta < 1$, $\beta f(x) + (1 - \beta)g(x)$ es una fdp en ese intervalo.

4.13. Supongamos que el gráfico de la figura 4.16 representa la fdp de una variable aleatoria X .

(a) ¿Cuál es la relación entre a y b ?

(b) Si $a > 0$ y $b > 0$, ¿qué puede Ud. decir acerca del mayor valor que puede tomar b ? (Ver figura 4.16.)

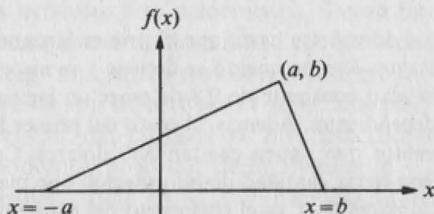


FIGURA 4.16

4.14. El porcentaje de alcohol ($100X$) en cierto compuesto se puede considerar como una variable aleatoria, en donde X , $0 < X < 1$, tiene la siguiente fdp:

$$f(x) = 20x^3(1 - x), \quad 0 < x < 1.$$

(a) Obtener una expresión para fda F y dibujar su gráfico.

(b) Calcular $P(X \leq \frac{2}{3})$.

(c) Supóngase que el precio de venta del compuesto anterior depende del contenido alcohol. Específicamente, si $\frac{1}{3} < X < \frac{2}{3}$, el compuesto se vende por C_1 dólares/galón; de otro modo se vende por C_2 dólares/galón. Si el costo es C_3 dólares/galón, encontrar la distribución de probabilidades de la utilidad neta por galón.

4.15. Sea X una variable aleatoria continua con fdp f dada por:

$$\begin{aligned} f(x) &= ax, & 0 \leq x \leq 1, \\ &= a, & 1 \leq x \leq 2, \\ &= -ax + 3a, & 2 \leq x \leq 3, \\ &= 0, & \text{para cualquier otro valor.} \end{aligned}$$

(a) Determinar la constante a . (b) Determinar F , la fda y dibujar el gráfico.

(c) Si X_1 , X_2 , y X_3 son tres observaciones independientes de X ¿cuál es la probabilidad de que exactamente uno de esos tres números sea mayor que 1,5?

4.16. Se supone que el diámetro de un cable eléctrico, digamos X es una variable aleatoria continua con fdp $f(x) = 6x(1 - x)$, $0 \leq x \leq 1$.

(a) Verificar que la anterior es una fdp y dibujarla.

(b) Obtener una expresión para la fda de X y dibujarla.

(c) Determinar un número b tal que $P(X < b) = 2P(X > b)$.

(d) Calcular $P(X \leq \frac{1}{2} | \frac{1}{3} < X < \frac{2}{3})$.

4.17. Cada una de las siguientes funciones representa la fda de una variable aleatoria continua. En cada caso $F(x) = 0$ para $x < a$ y $F(x) = 1$ para $x > b$, en donde $[a, b]$ es el intervalo indicado. En cada uno de los casos, dibujar la función F , determinar la fdp f y dibujarla. También verificar que f es una fdp.

- (a) $F(x) = x/5, 0 \leq x \leq 5$ (b) $F(x) = (2/\pi) \operatorname{sen}^{-1}(\sqrt{x}), 0 \leq x \leq 1$
 (c) $F(x) = e^{3x}, -\infty < x \leq 0$ (d) $F(x) = x^3/2 + \frac{1}{2}, -1 \leq x \leq 1$.

4.18. Sea X la duración de un instrumento electrónico (medido en horas). Supóngase que X es una variable aleatoria continua con fdp $f(x) = k/x^n, 2000 \leq x \leq 10.000$.

- (a) Para $n = 2$, determinar k .
- (b) Para $n = 3$, determinar k .
- (c) Para n en general, determinar k .
- (d) ¿Cuál es la probabilidad de que el instrumento falle antes de 5000 horas de funcionamiento?
- (e) Dibujar la fda $F(t)$ para (c) y determinar su forma algebraica.

4.19. Sea X una variable aleatoria distribuida binomialmente con base en 10 repeticiones de un experimento. Si $p = 0,3$, calcular las siguientes probabilidades, usando la tabla de la distribución binomial del Apéndice.

- (a) $P(X \leq 8)$
- (b) $P(X = 7)$
- (c) $P(X > 6)$.

4.20. Supóngase que X está distribuida uniformemente en $[-\alpha, +\alpha]$, en donde $\alpha > 0$. Cada vez que sea posible, determinar α de modo que se satisfaga lo siguiente.

- (a) $P(X > 1) = \frac{1}{3}$
- (b) $P(X > 1) = \frac{1}{2}$
- (c) $P(X < \frac{1}{2}) = 0,7$
- (d) $P(X < \frac{1}{2}) = 0,3$
- (e) $P(|X| < 1) = P(|X| > 1)$.

4.21. Suponiendo que X está distribuida uniformemente en $[0, \alpha], \alpha > 0$ responder las preguntas del problema 4.20.

4.22. Se escoge un punto al azar de un segmento de longitud L . ¿Cuál es la probabilidad de que la razón del segmento más corto con relación al más largo sea menor que $\frac{1}{4}$?

4.23. Una fábrica produce diariamente 10 recipientes de vidrio. Se puede suponer que hay una probabilidad constante $p = 0,1$ de producir uno defectuoso. Antes de que estos depósitos se almacenen son inspeccionados y los defectuosos puestos aparte. Supongamos que hay una probabilidad constante $r = 0,1$ de que un recipiente defectuoso sea mal clasificado. Sea X igual al número de recipientes clasificados como defectuosos al término de un día de producción. (Suponemos que todos los recipientes que se fabrican en un día se inspeccionan ese mismo día.)

- (a) Calcular $P(X = 3)$ y $P(X > 3)$.
- (b) Obtener una expresión para $P(X = k)$.

4.24. Supóngase que el 5 por ciento de los artículos que salen de una línea de producción son defectuosos. Se escogen 10 de tales artículos y se inspeccionan. ¿Cuál es la probabilidad de que se encuentren a lo más dos defectuosos?

4.25. Suponiendo que la duración (en horas) de cierto tubo de radio es una variable aleatoria continua X con fdp $f(x) = 100/x^2, x > 100$ y 0 para cualquier otro valor.

(a) ¿Cuál es la probabilidad de que un tubo dure menos de 200 horas si se sabe que el tubo todavía funciona después de 150 horas de servicio?

(b) ¿Cuál es la probabilidad de que si se instalan 3 de tales tubos en un conjunto, exactamente uno tenga que ser sustituido después de 150 horas de servicio?

(c) ¿Cuál es el número máximo de tubos que se pueden poner en un conjunto de modo que haya una probabilidad 0,5 de que después de 150 horas de servicio todos ellos todavía funcionen?

4.26. Un experimento consta de n ensayos independientes. Se puede suponer que debido al «aprendizaje», la probabilidad de obtener un resultado exitoso aumenta con el núme-

ro de ensayos realizados. Específicamente supongamos que $P(\text{éxito en la } i\text{-ésima repetición}) = (i+1)/(i+2)$, $i = 1, 2, \dots, n$.

(a) ¿Cuál es la probabilidad de tener tres resultados exitosos a lo menos en ocho repeticiones?

(b) ¿Cuál es la probabilidad de que el primer resultado exitoso ocurra en la octava repetición?

4.27. Refiéndonos al ejemplo 4.10,

(a) calcular $P(X = 2)$ si $n = 4$,

(b) para un n arbitrario, demostrar que $P(X = n - 1) = P$ (exactamente un intento no exitoso) es igual a $[1/(n+1)] \sum_{i=1}^n (1/i)$.

4.28. Si la variable aleatoria K está distribuida uniformemente en $(0, 5)$ ¿cuál es la probabilidad de que las raíces de la ecuación $4x^2 + 4xK + K + 2 = 0$ sean reales?

4.29. Supóngase que la variable aleatoria X tiene valores posibles $1, 2, 3, \dots$ y que $P(X = r) = k(1 - \beta)^{r-1}$, $0 < \beta < 1$.

(a) Determinar la constante k .

(b) Encontrar la *moda* de esta distribución (es decir, el valor de r que da el mayor valor de $P(X = r)$).

4.30. Una variable aleatoria X puede tomar cuatro valores con probabilidades $(1 + 3x)/4$, $(1 - x)/4$, $(1 + 2x)/4$, y $(1 - 4x)/4$. ¿Para qué valores de X es ésta una distribución de probabilidades?

Funciones de variables aleatorias

5.1 Un ejemplo

Supongamos que el radio X de la entrada de un tubo muy bien calibrado se considera como una variable aleatoria continua con fdp f . Sea $A = \pi X^2$ el área de la sección de la entrada. Es evidente que, puesto que el valor de X es el resultado de un experimento aleatorio, también lo es el valor de A . Es decir, A es una variable aleatoria (continua), y podríamos obtener su fdp, digamos g . Esperaríamos que puesto que A es una función de X , la fdp g es obtenible de alguna manera, conociendo la fdp f . En este capítulo trataremos problemas de esta naturaleza. Antes de familiarizarnos con algunas técnicas específicas necesarias, formularemos más precisamente los conceptos anteriores.

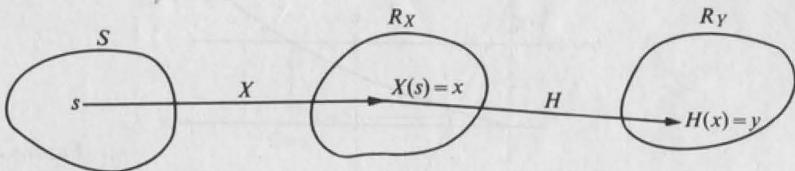


FIGURA 5.1

5.2 Sucesos equivalentes

Sea ε un experimento y sea S un espacio muestral asociado con ε . Sea X una *variable aleatoria* definida en S . Supongamos que $y = H(x)$ es una función real de x . Entonces $Y = H(X)$ es una variable aleatoria puesto que para cada $s \in S$, se determina un valor de Y , sea $y = H[X(s)]$. Esquemáticamente tenemos la figura 5.1. Como antes, llamamos R_X al *recorrido de X* , el conjunto de valores posibles de la función X . De igual modo, definimos como R_Y al *recorrido de la variable aleatoria Y* , el conjunto de valores posibles de Y . Previamente (ecuación 4.1) definimos la noción de sucesos equivalentes en S y en R_X . Extendamos ahora este concepto de la siguiente manera natural.

Definición. Sea C un suceso (subconjunto) asociado con el recorrido de Y , R_Y , como se describió anteriormente. Definase $B \subset R_X$ como sigue:

$$B = \{x \in R_X : H(x) \in C\}. \quad (5.1)$$

En palabras: B es el conjunto de los valores de X tales que $H(x) \in C$. Si B y C están relacionados de esta manera los llamamos sucesos equivalentes.

Observaciones: (a) Como dijimos antes, la interpretación informal de lo anterior es que B y C son sucesos equivalentes si y sólo si B y C ocurren juntos. Esto es, cuando ocurre B , ocurre C y reciprocamente.

(b) Supóngase que A es un suceso asociado con S lo que equivale a un suceso B asociado con R_X . Luego, si C es un suceso asociado con R_Y que es equivalente a B , tenemos que A es equivalente a C .

(c) Nuevamente, es importante reafirmar que cuando hablamos de sucesos equivalentes (en el sentido anterior), esos sucesos están asociados con espacios muestrales diferentes.

EJEMPLO 5.1. Supongamos que $H(x) = \pi x^2$ como en la sección 5.1. Luego los sucesos $B: \{X > 2\}$ y $C: \{Y > 4\pi\}$ son equivalentes. Porque si $Y = \pi X^2$, entonces ocurre $\{X > 2\}$ si y sólo si ocurre $\{Y > 4\pi\}$, puesto que en el presente contexto X no puede tomar valores negativos. (Ver figura 5.2).

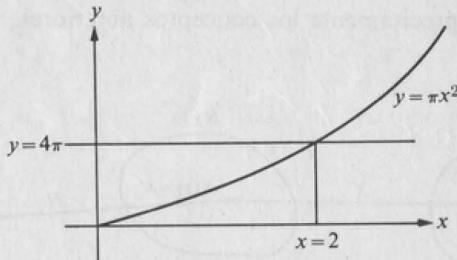


FIGURA 5.2

Observación: otra vez es importante señalar que se usa una notación taquigráfica cuando escribimos expresiones tales como $\{X > 2\}$ y $\{Y > 4\pi\}$. Cuando nos referimos a los valores de X y de Y , o sea $\{s \mid X(s) > 2\}$ y $\{x \mid Y(x) > 4\pi\}$.

Como lo hicimos en el capítulo 4, (ecuación 4.2), nuevamente haremos la definición siguiente.

Definición. Sea X una variable aleatoria en el espacio muestral S . Sea R_X el recorrido de X . Sea H una función real y consideremos la variable aleatoria $Y = H(X)$ con recorrido R_Y . Para cualquier suceso $C \subset R_Y$, definimos $P(C)$ como sigue:

$$P(C) = P[\{x \in R_X : H(x) \in C\}]. \quad (5.2)$$

En palabras: la probabilidad de un suceso asociado con el recorrido de Y está definida como la probabilidad del suceso equivalente (en función de X), como aparece en la ecuación (5.2).

Observaciones: (a) La definición anterior hará posible calcular las probabilidades relacionadas con sucesos con Y si conocemos la distribución de probabilidades de X y podemos determinar el suceso equivalente en cuestión.

(b) Puesto que lo discutido previamente (ecuaciones 4.1 y 4.2) relaciona probabilidades asociadas con R_X con probabilidades asociadas con S , podemos escribir la ecuación (5.2) como sigue:

$$P(C) = P[\{x \in R_X : H(x) \in C\}] = P[\{s \in S : H(X(s)) \in C\}].$$

EJEMPLO 5.2. Sea X una variable aleatoria con fdp

$$f(x) = e^{-x}, \quad x > 0.$$

(Una integración sencilla indica que $\int_0^\infty e^{-x} dx = 1$.)

Supóngase que $H(x) = 2x + 1$. Por tanto $R_X = \{x | x > 0\}$, mientras que $R_Y = \{y | y > 1\}$. Supóngase que el suceso C está definido como sigue: $C = \{Y \geq 5\}$. Ahora $y \geq 5$ si y sólo si $2x + 1 \geq 5$ lo que a su vez da $x \geq 2$. Por lo tanto C es equivalente a $B = \{X \geq 2\}$. (Ver figura 5.3). Ahora $P(X \geq 2) = \int_2^\infty e^{-x} dx = 1/e^2$. Por lo tanto aplicando la ecuación (5.2) encontramos que

$$P(Y \geq 5) = 1/e^2.$$

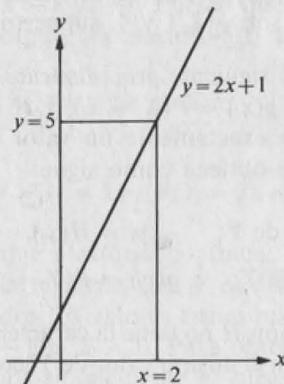


FIGURA 5.3

Observaciones: (a) Nuevamente es conveniente señalar que debemos considerar incorporadas la evaluación de $x = X(s)$ y la evaluación de $y = H(x)$ en nuestro experimento y, por tanto, considerar simplemente R_Y , el recorrido de Y , como el espacio muestral de nuestro experimento.

Estrictamente hablando, el espacio muestral del experimento es S y el resultado del experimento es s . Todo lo que hagamos a continuación no está influenciado por la naturaleza aleatoria del experimento. La determinación de $x = X(s)$ y la evaluación de $y = H(x)$ son

estrictamente procesos determinísticos una vez que se ha observado s . Sin embargo, como discutimos previamente, podemos incorporar estos cálculos en la descripción de nuestro experimento y así relacionarlos directamente con el recorrido R_Y .

(b) Así como la distribución de probabilidades fue inducida en R_X por la distribución de probabilidades sobre el espacio muestral original S , así la distribución de probabilidades de Y , se determina si se conoce la distribución de probabilidades de X . Por ejemplo, en el anterior ejemplo 5.2, la distribución específica de X determinó completamente el valor de $P(Y \geq 5)$.

(c) Al considerar una función de una variable aleatoria X , digamos $Y = H(X)$, debemos comentar que no se puede permitir toda función H concebible. Sin embargo, las funciones que aparecen en las aplicaciones están inmediatamente entre aquellas que podemos considerar y, por tanto, no nos referiremos posteriormente a esta pequeña dificultad.

5.3 Variables aleatorias discretas

Caso 1. X es una variable aleatoria discreta. Si X es una variable aleatoria discreta e $Y = H(X)$, entonces se deduce de inmediato que Y es también una variable aleatoria discreta.

Porque supóngase que se pueden enumerar los valores posibles de X como $x_1, x_2, \dots, x_n, \dots$ Con seguridad los valores posibles de Y se pueden enumerar como $y_1 = H(x_1), y_2 = H(x_2), \dots$ (Algunos de los valores anteriores de Y pueden ser iguales, pero esto ciertamente no impide el hecho de que esos valores puedan enumerarse).

EJEMPLO 5.3. Supóngase que la variable aleatoria X tome los tres valores $-1, 0$ y 1 con probabilidades $\frac{1}{3}, \frac{1}{2}$, y $\frac{1}{6}$, respectivamente. Sea $Y = 3X + 1$. Entonces los valores posibles de Y son $-2, 1$ y 4 , supuestos con probabilidades $\frac{1}{3}, \frac{1}{2}$ y $\frac{1}{6}$.

Este ejemplo sugiere el siguiente *procedimiento general*: si x_1, \dots, x_n, \dots son los valores posibles de X , $p(x_i) = P(X = x_i)$, y H es una función tal que a cada valor de y le corresponda exactamente un valor de x , entonces la distribución de probabilidades de Y se obtiene como sigue:

$$\text{Valores posibles de } Y: \quad y_i = H(x_i), \quad i = 1, 2, \dots, n, \dots;$$

$$\text{Probabilidades de } Y: \quad q(y_i) = P(Y = y_i) = p(x_i).$$

Muy a menudo la función H no tiene la característica anterior, y puede suceder que varios valores de X den el mismo valor de Y , como ilustra el siguiente ejemplo.

EJEMPLO 5.4. Supóngase que consideramos la misma variable aleatoria X que en el anterior ejemplo 5.3. Sin embargo, introducimos $Y = X^2$. En este caso los valores posibles de Y son cero y uno, supuestos con probabilidades $\frac{1}{2}, \frac{1}{2}$. Porque $Y = 1$ si y sólo si $X = -1$ o $X = 1$ y la probabilidad de este último suceso es $\frac{1}{3} + \frac{1}{6} = \frac{1}{2}$. En razón de nuestra terminología anterior los sucesos $B: \{X = \pm 1\}$ y $C: \{Y = 1\}$ son sucesos equivalentes y por lo tanto según la ecuación (5.2) tienen probabilidades iguales.

El procedimiento general para situaciones como la descrita en el ejemplo anterior es como sigue: representemos como $x_{i_1}, x_{i_2}, \dots, x_{i_k}, \dots$, los valores de X que tienen la propiedad $H(x_{i_j}) = y_i$ para todo j . Luego

$$q(y_i) = P(Y = y_i) = p(x_{i_1}) + p(x_{i_2}) + \dots$$

En palabras: para calcular la probabilidad del suceso $\{Y = y_i\}$, encuentre el suceso equivalente en función de X (en el recorrido R_X) y luego agregue todas las probabilidades correspondientes. (Ver figura 5.4.)

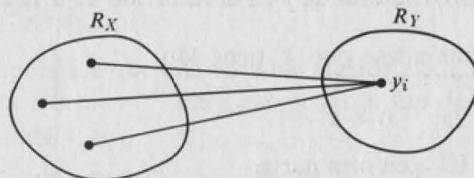


FIGURA 5.4

EJEMPLO 5.5. Tenga X los valores posibles $1, 2, \dots, n, \dots$ y supóngase que $P(X = n) = \frac{1}{2^n}$. Sea

$$\begin{aligned} Y &= 1 && \text{si } X \text{ es par,} \\ &= -1 && \text{si } X \text{ es impar.} \end{aligned}$$

Luego Y toma los dos valores -1 y $+1$. Puesto que $Y = 1$ si y sólo si $X = 2$, o $X = 4$, o $X = 6$, o \dots , aplicando la ecuación (5.2) se tiene

$$P(Y = 1) = \frac{1}{4} + \frac{1}{16} + \frac{1}{64} + \dots = \frac{1}{3}.$$

Luego

$$P(Y = -1) = 1 - P(Y = 1) = \frac{2}{3}.$$

Caso 2. X es una variable aleatoria continua. Puede suceder que X sea una variable aleatoria continua mientras que Y es discreta. Por ejemplo, supongamos que X puede tomar todos los valores reales mientras que se define que Y sea $+1$ si $X \geq 0$ y que $Y = -1$ si $X < 0$. Para obtener la distribución de probabilidades de Y , determinemos simplemente el suceso equivalente (en el recorrido R_X) que corresponde a los diferentes valores de Y . En el caso anterior, $Y = 1$ si y sólo si $X \geq 0$, mientras que $Y = -1$ si y sólo si $X < 0$. Por lo tanto $P(Y = 1) = P(X \geq 0)$ mientras que $P(Y = -1) = P(X < 0)$. Si se conoce la fdp de X , pueden calcularse estas probabilidades. En el caso general, si $\{Y = y_i\}$ es equivalente a un suceso, sea A , en el recorrido de X , entonces

$$q(y_i) = P(Y = y_i) = \int_A f(x) dx.$$

5.4 Variables aleatorias continuas

El caso más importante (y el que se encuentra con más frecuencia) aparece cuando X es una variable aleatoria continua con fdp f y H es una función continua. Por tanto $Y = H(X)$ es una variable aleatoria continua, y será nuestro propósito obtener su fdp, digamos g .

El procedimiento general será así:

- Obtener G , la fda de Y , en donde $G(y) = P(Y \leq y)$, al encontrar el suceso A (en el recorrido de X) que es equivalente al suceso $\{Y \leq y\}$.
- Diferenciar $G(y)$ con respecto a y para obtener $g(y)$.
- Determinar estos valores de y en el recorrido de Y para los cuales $g(y) > 0$.

EJEMPLO 5.6. Supóngase que X tiene fdp

$$\begin{aligned}f(x) &= 2x, \quad 0 < x < 1, \\&= 0, \quad \text{en otra parte.}\end{aligned}$$

Sea $H(x) = 3x + 1$. Por tanto para encontrar la fdp de $Y = H(X)$ tenemos (ver figura 5.5)

$$\begin{aligned}G(y) &= P(Y \leq y) = P(3X + 1 \leq y) \\&= P(X \leq (y - 1)/3) \\&= \int_0^{(y-1)/3} 2x \, dx = [(y - 1)/3]^2.\end{aligned}$$

Así

$$g(y) = G'(y) = \frac{2}{3}(y - 1).$$

Puesto que $f(x) > 0$ para $0 < x < 1$, encontramos que $g(y) > 0$ para $1 < y < 4$.

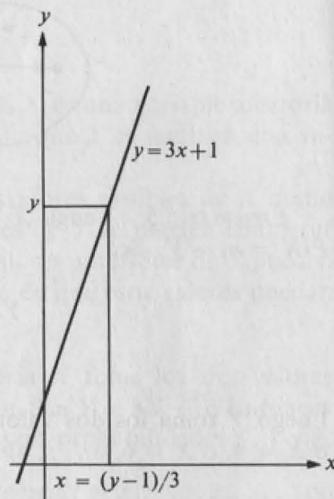


FIGURA 5.5

Observación: el suceso A , mencionado atrás, que es equivalente al suceso $\{Y \leq y\}$, es simplemente $\{X \leq (y - 1)/3\}$.

Hay otro método ligeramente diferente, para obtener el mismo resultado y que posteriormente será útil. Consideremos nuevamente

$$G(y) = P(Y \leq y) = P\left(X \leq \frac{y-1}{3}\right) = F\left(\frac{y-1}{3}\right),$$

en donde F es la fda de X , esto es,

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

A fin de evaluar la derivada de G , $G'(y)$, usemos la regla de la cadena para la derivación como sigue:

$$\frac{dG(y)}{dy} = \frac{dG(y)}{du} \cdot \frac{du}{dy}, \quad \text{en donde } u = \frac{y-1}{3}$$

por tanto

$$G'(y) = F'(u) \cdot \frac{1}{3} = f(u) \cdot \frac{1}{3} = 2\left(\frac{y-1}{3}\right) \cdot \frac{1}{3},$$

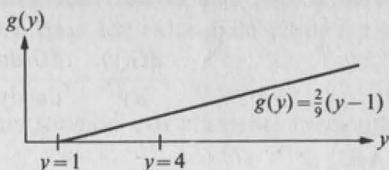


FIGURA 5.6

como antes. En la figura 5.6, aparece el gráfico de la fdp de Y . (Para verificar los cálculos nótese que $\int_1^4 g(y) dy = 1$.)

EJEMPLO 5.7. Supóngase que una variable aleatoria continua tiene fdp como en el ejemplo 5.6. Sea $H(x) = e^{-x}$. Para encontrar la fdp de $Y = H(X)$ procedemos como sigue (ver figura 5.7):

$$\begin{aligned} G(y) &= P(Y \leq y) = P(e^{-X} \leq y) \\ &= P(X \geq -\ln y) = \int_{-\ln y}^1 2x dx \\ &= 1 - (-\ln y)^2. \end{aligned}$$

Por tanto $g(y) = G'(y) = -2 \ln y / y$. Puesto que $f(x) > 0$ para $0 < x < 1$, encontramos que $g(y) > 0$ para $1/e < y < 1$. (Nótese que el signo algebraico para $g(y)$ es correcto puesto que $\ln y < 0$ para $1/e < y < 1$.) El gráfico de $g(y)$ está dibujado en la figura 5.8.

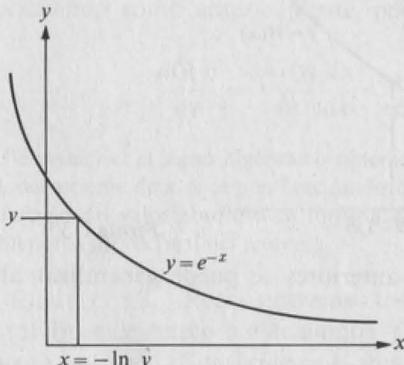


FIGURA 5.7

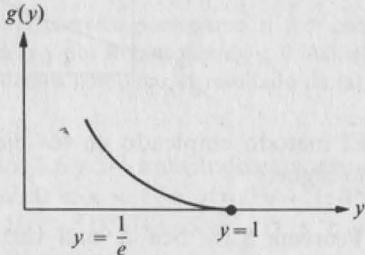


FIGURA 5.8

Otra vez podemos obtener el resultado anterior con un enfoque ligeramente diferente que esbozaremos brevemente. Como antes,

$$\begin{aligned} G(y) &= P(Y \leq y) = P(X \geq -\ln y) \\ &= 1 - P(X \leq -\ln y) = 1 - F(-\ln y), \end{aligned}$$

en donde como antes F es la fda de X . A fin de obtener la derivada de G usamos otra vez la regla de derivación en cadena como sigue:

$$\frac{dG(y)}{dy} = \frac{dG}{du} \frac{du}{dy}, \quad \text{en donde } u = -\ln y.$$

Así

$$G'(y) = -F'(u) \left(-\frac{1}{y} \right) = +2 \ln y \cdot \left(-\frac{1}{y} \right),$$

como antes.

Generalicemos ahora el enfoque sugerido por los ejemplos anteriores. El paso crucial en cada uno de los ejemplos se efectuó al sustituir el suceso $\{Y \leq y\}$ por el suceso equivalente en función de la variable aleatoria X . En los problemas anteriores esto fue relativamente sencillo puesto que en cada uno de los casos la función fue una función de X estrictamente creciente o estrictamente decreciente.

En la figura 5.9 y es una función estrictamente creciente de x . Por lo tanto podemos resolver $y = H(x)$ para x en función de y , digamos $x = H^{-1}(y)$, en donde H^{-1} se llama función inversa de H . Así, si H es estrictamente creciente, $\{H(X) \leq y\}$ es equivalente a $\{X \leq H^{-1}(y)\}$ mientras que si H es estrictamente decreciente, $\{H(X) \leq y\}$ es equivalente a $\{X \geq H^{-1}(y)\}$.

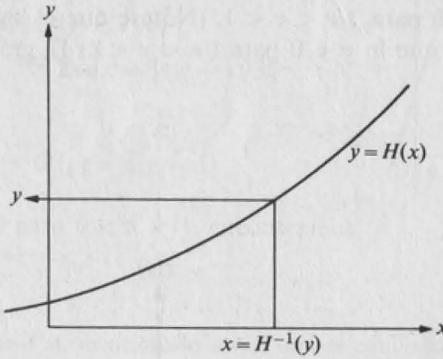


FIGURA 5.9

El método empleado en los ejemplos anteriores se puede generalizar ahora como sigue.

Teorema 5.1. Sea X una variable aleatoria continua con fdp f , en donde $f(x) > 0$ para $a < x < b$. Supóngase que $y = H(x)$ sea una función de x estrictamente monótona (creciente o decreciente). Supóngase que esta función es derivable (y por tanto continua) para toda x . Luego la variable aleatoria Y definida como $Y = H(X)$ tiene una fdp g dada por,

$$g(y) = f(x) \left| \frac{dx}{dy} \right|, \quad (5.3)$$

en donde x se expresa en función de y . Si H es creciente, entonces g es distinta de cero para los valores de y que satisfacen $H(a) < y < H(b)$. Si H es decreciente, entonces g es distinta de cero para los valores de y que satisfacen $H(b) < y < H(a)$.

Demostración: (a) Supongamos que H es una función estrictamente creciente. Por tanto

$$\begin{aligned} G(y) &= P(Y \leq y) = P(H(X) \leq y) \\ &= P(X \leq H^{-1}(y)) = F(H^{-1}(y)). \end{aligned}$$

Diferenciando $G(y)$ con respecto a y , obtenemos, al usar la regla de la cadena para las derivadas,

$$\frac{dG(y)}{dy} = \frac{dG(y)}{dx} \frac{dx}{dy}, \quad \text{en donde } x = H^{-1}(y).$$

Así

$$G'(y) = \frac{dF(x)}{dx} \frac{dx}{dy} = f(x) \frac{dx}{dy}.$$

(b) Supongamos que H es una función decreciente. Por lo tanto

$$\begin{aligned} G(y) &= P(Y \leq y) = P(H(X) \leq y) = P(X \leq H^{-1}(y)) \\ &= 1 - P(X \leq H^{-1}(y)) = 1 - F(H^{-1}(y)). \end{aligned}$$

Procediendo como anteriormente, podemos escribir

$$\frac{dG(y)}{dy} = \frac{dG(y)}{dx} \frac{dx}{dy} = \frac{d}{dx} [1 - F(x)] \frac{dx}{dy} = -f(x) \frac{dx}{dy}.$$

Observación: el signo algebraico obtenido en (b) es correcto puesto que, si y es una función decreciente de x , x es una función decreciente de y y por lo tanto $dx/dy < 0$. Así, al usar el símbolo del valor absoluto en torno a dx/dy , podemos combinar el resultado de (a) y (b) y obtener la forma final del teorema.

EJEMPLO 5.8. Reconsideremos los ejemplos 5.6 y 5.7, aplicando el teorema 5.1.

(a) En el ejemplo 5.6 teníamos $f(x) = 2x$, $0 < x < 1$, e $y = 3x + 1$. Por lo tanto $x = (y - 1)/3$ y $dx/dy = \frac{1}{3}$. Así $g(y) = 2[(y - 1)/3]\frac{1}{3} = \frac{2}{9}(y - 1)$, $1 < y < 4$, lo que concuerda con el resultado obtenido previamente.

(b) En el ejemplo 5.7, teníamos $f(x) = 2x$, $0 < x < 1$, e $y = e^{-x}$. Por tanto $x = -\ln y$ y $dx/dy = -1/y$. Así $g(y) = -2(\ln y)/y$, $1/e < y < 1$ lo que de nuevo está de acuerdo con el resultado anterior.

Si $y = H(x)$ no es una función monótona de x , no podemos aplicar directamente el método anterior. En su lugar, volveremos al método general bosquejado anteriormente. El ejemplo siguiente ilustra este procedimiento.

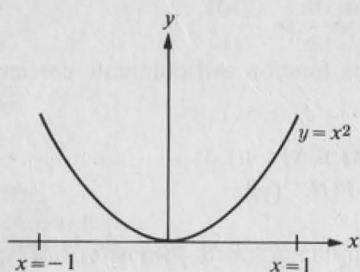


FIGURA 5.10

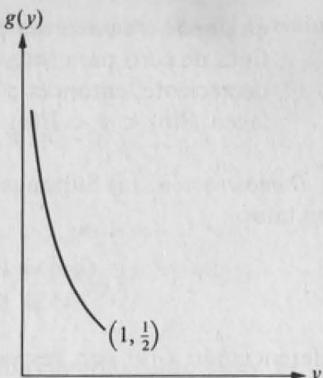


FIGURA 5.11

EJEMPLO 5.9. Supongamos que

$$\begin{aligned} f(x) &= \frac{1}{2}, & -1 < x < 1, \\ &= 0, & \text{para cualquier otro valor.} \end{aligned}$$

Sea $H(x) = x^2$. Esta obviamente *no* es una función monótona en el intervalo $[-1, 1]$ (figura 5.10). Por lo tanto obtenemos la fdp de $Y = X^2$ como sigue:

$$\begin{aligned} G(y) &= P(Y \leq y) = P(X^2 \leq y) \\ &= P(-\sqrt{y} \leq X \leq \sqrt{y}) \\ &= F(\sqrt{y}) - F(-\sqrt{y}), \end{aligned}$$

en donde F es la fda de la variable aleatoria X . Por tanto

$$\begin{aligned} g(y) &= G'(y) = \frac{f(\sqrt{y})}{2\sqrt{y}} - \frac{f(-\sqrt{y})}{-2\sqrt{y}} \\ &= \frac{1}{2\sqrt{y}} [f(\sqrt{y}) + f(-\sqrt{y})]. \end{aligned}$$

Así $g(y) = (1/2\sqrt{y})(\frac{1}{2} + \frac{1}{2}) = 1/2\sqrt{y}$, $0 < y < 1$. (Ver figura 5.11.)

El método usado en el ejemplo anterior da el siguiente resultado general.

Teorema 5.2. Sea X una variable aleatoria continua con fdp f . Sea $Y = X^2$. Entonces la variable aleatoria Y tiene fdp dada por

$$g(y) = \frac{1}{2\sqrt{y}} [f(\sqrt{y}) - f(-\sqrt{y})].$$

Demostración: ver ejemplo 5.9.

PROBLEMAS

5.1. Supóngase que X está distribuida uniformemente en $(-1, 1)$. Sea $Y = 4 - X^2$. Encontrar la fdp de Y , sea $g(y)$, y dibujarla. También verificar que $g(y)$ es una fdp.

5.2. Supóngase que X está distribuida uniformemente en $(1, 3)$. Obtener las fdp de las siguientes variables aleatorias:

$$(a) Y = 3X + 4 \quad (b) Z = e^X.$$

Verificar en cada uno de los casos que la función obtenida es una fdp.

5.3. Supóngase que la variable aleatoria X tiene fdp $f(x) = e^{-x}$, $x > 0$. Encontrar las fdp de las siguientes variables aleatorias:

$$(a) Y = X^3 \quad (b) Z = 3/(X + 1)^2.$$

5.4. Supóngase que la variable aleatoria discreta X tome los valores 1, 2, y 3 con igual probabilidad. Encontrar la distribución de probabilidades de $Y = 2X + 3$.

5.5. Supóngase que X está distribuida uniformemente en el intervalo $(0, 1)$. Encontrar la fdp de las siguientes variables aleatorias:

$$(a) Y = X^2 + 1 \quad (b) Z = 1/(X + 1).$$

5.6. Supóngase que X está distribuida uniformemente en $(-1, 1)$. Encontrar la fdp de las siguientes variables aleatorias:

$$(a) Y = \operatorname{sen}(\pi/2)X \quad (b) Z = \cos(\pi/2)X \quad (c) W = |X|.$$

5.7. Supóngase que el radio de una esfera es una variable aleatoria continua. (Debido a la imprecisión del proceso de fabricación, los radios de esferas diferentes pueden ser diferentes.) Supóngase que el radio R tenga fdp $f(r) = 6r(1-r)$, $0 < r < 1$. Encontrar la fdp del volumen V y del área superficial S de la esfera.

5.8. Una corriente eléctrica I que fluctúa se puede considerar como una variable aleatoria distribuida uniformemente en el intervalo $(9, 11)$. Si esta corriente pasa por una resistencia de 2-ohm, encontrar la fdp de la potencia $P = 2I^2$.

5.9. La velocidad de una molécula en un gas uniforme en equilibrio es una variable aleatoria V cuya fdp está dada por

$$f(v) = av^2e^{-bv^2}, \quad v > 0,$$

en donde $b = m/2kT$ y k , T , y m denotan la constante de Boltzman, la temperatura absoluta, y la masa de la molécula, respectivamente.

(a) Calcular la constante a (en función de b). [Indicación: use el hecho de que $\int_0^\infty e^{-x^2} dx = \sqrt{\pi}/2$ e integre por partes.]

(b) Derivar la distribución de la variable aleatoria $W = mV^2/2$, que representa la energía cinética de la molécula.

5.10. Un voltaje aleatorio X está distribuido uniformemente en el intervalo $(-k, k)$. Si X es la energía recibida de un artefacto no lineal, con las características que se indican en la figura 5.12, encontrar la distribución de probabilidades de Y en los tres casos siguientes

$$(a) k < a \quad (b) a < k < x_0 \quad (c) k > x_0.$$

Observación: la distribución de probabilidades de Y es un ejemplo de una distribución mixta. Y toma el valor cero con una probabilidad positiva y también toma todos los valores en ciertos intervalos. (Ver sección 4.6.)

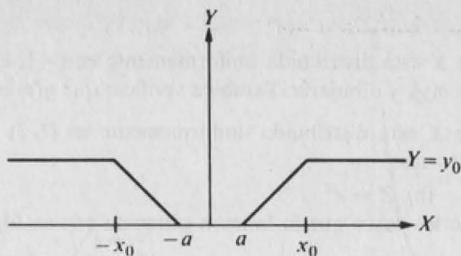


FIGURA 5.12

5.11. La energía radiante (en Btu/hr/ft²) está dada como la función siguiente de la temperatura T (en grados fahrenheit): $E = 0,173(T/100)^4$. Supongamos que la temperatura T está considerada que es una variable aleatoria continua con fdp

$$\begin{aligned}f(t) &= 200t^{-2}, \quad 40 \leq t \leq 50, \\&= 0, \quad \text{para cualquier otro valor.}\end{aligned}$$

Encontrar la fdp de la energía radiante E .

5.12. Para medir las velocidades del aire, se usa un tubo (conocido como el tubo estático de Pitot) que nos permite medir la diferencia de presión. Esta diferencia de presión está dada por $P = (1/2)dV^2$, donde d es la densidad del aire y V es la velocidad del viento (kph). Si V es una variable aleatoria distribuida uniformemente en (10, 20), encontrar la fdp de P .

5.13. Supóngase que $P(X \leq 0,29) = 0,75$, en donde X es una variable aleatoria continua con alguna distribución definida en (0, 1). Si $Y = 1 - X$, determinar k de modo que $P(Y \leq k) = 0,25$.

Variables aleatorias bidimensionales y de mayor dimensión

6.1 Variables aleatorias bidimensionales

En nuestro estudio de las variables aleatorias hemos considerado, hasta aquí, sólo el caso unidimensional. Es decir, el resultado del experimento se podía registrar como un solo número x .

En muchos casos, sin embargo, nos interesa observar dos o más características numéricas simultáneamente. Por ejemplo, la dureza H y la resistencia a la tensión T de una pieza manufacturada de acero pueden ser de interés y consideraríamos (h, t) como un solo resultado experimental. Podríamos estudiar la altura A y el peso P de alguna persona determinada, que daría lugar al resultado (p, a) . Finalmente, podríamos observar la cantidad de lluvia total, LL y el promedio de temperatura T en cierta región durante un mes específico, que daría lugar al resultado (ll, t) .

Haremos la siguiente definición formal.

Definición. Sea ε un experimento y S un espacio muestral asociado con ε .

Sean $X = X(s)$ e $Y = Y(s)$ dos funciones que asignan un número real a cada uno de los resultados $s \in S$ (figura 6.1). Llamamos a (X, Y) variable aleatoria bidimensional (algunas veces llamada vector aleatorio.)

Si $X_1 = X_1(s)$, $X_2 = X_2(s)$, ..., $X_n = X_n(s)$ son n funciones cada una de las cuales asigna un número real a cada resultado $s \in S$, llamamos (X_1, \dots, X_n) variable aleatoria n -dimensional (o un vector aleatorio n -dimensional).

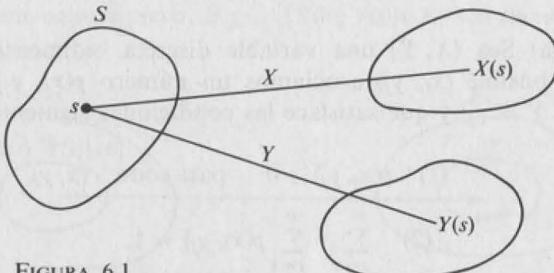


FIGURA 6.1

Observación: como en el caso unidimensional, no nos interesaría la naturaleza funcional de $X(s)$ e $Y(s)$, sino los valores que toman X y Y . Hablaremos nuevamente del *recorrido* de (X, Y) , digamos $R_{X,Y}$, como el conjunto de todos los valores posibles de (X, Y) . En el caso bidimensional, por ejemplo, el recorrido de (X, Y) será un subconjunto del plano euclíadiano. Cada uno de los resultados $X(s), Y(s)$ se puede representar como un punto (x, y) en el plano. Suprimiremos nuevamente la naturaleza funcional de X y Y al escribir, por ejemplo, $P[X \leq a, Y \leq b]$ en vez de $P[X(s) \leq a, Y(s) \leq b]$.

Como en el caso unidimensional, distinguiremos entre dos tipos básicos de variables aleatorias: las variables aleatorias discretas y las continuas.

Definición. (X, Y) es una variable aleatoria *bidimensional discreta* si los valores posibles de (X, Y) son finitos o infinitos numerables. Es decir, los valores posibles de (X, Y) se pueden representar como $(x_i, y_j), i = 1, 2, \dots, n, \dots; j = 1, 2, \dots, m, \dots$

(X, Y) es una variable aleatoria *bidimensional continua* si (X, Y) puede tomar todos los valores en un conjunto no numerable del plano euclíadiano. [Por ejemplo, si (X, Y) toma todos los valores en el rectángulo $\{(x, y) | a \leq x \leq b, c \leq y \leq d\}$ o todos los valores en el círculo $\{(x, y) | x^2 + y^2 \leq 1\}$, diríamos que (X, Y) es una variable aleatoria continua.]

Observaciones: (a) En otras palabras, (X, Y) es una variable aleatoria bidimensional si representa el resultado de un experimento aleatorio en el cual hemos medido las *dos* características numéricas X y Y .

(b) Puede suceder que una de las componentes de (X, Y) , digamos X , sea discreta, mientras que la otra es continua. Sin embargo, en la mayor parte de las aplicaciones nos interesamos sólo por los casos discutidos anteriormente, en los cuales ambas componentes son discretas, o ambas continuas.

(c) En muchas situaciones las dos variables aleatorias X y Y , consideradas en conjunto, son el resultado de un solo experimento, como se ilustró en los ejemplos anteriores. Es así como X y Y pueden representar la altura y el peso de un mismo individuo, etc. Sin embargo, no es necesario que exista esta clase de conexión. Por ejemplo, X podría ser la corriente que circula por un circuito en un momento específico, mientras que Y podría ser la temperatura en la habitación en ese instante, entonces podríamos considerar la variable aleatoria bidimensional (X, Y) . En la mayor parte de las aplicaciones hay una razón importante para considerar X y Y en conjunto.

Para describir la distribución de probabilidades de (X, Y) procedemos de un modo análogo al caso unidimensional.

Definición. (a) Sea (X, Y) una variable discreta bidimensional. Con cada resultado posible (x_i, y_j) asociamos un número $p(x_i, y_j)$ que representa $P(X = x_i, Y = y_j)$ y que satisface las condiciones siguientes:

$$(1) \quad p(x_i, y_j) \geq 0 \quad \text{para todo } (x, y),$$

$$(2) \quad \sum_{j=1}^{\infty} \sum_{i=1}^{\infty} p(x_i, y_j) = 1. \tag{6.1}$$

La función p definida para todo (x_i, y_j) en el recorrido de (X, Y) se llama *función de probabilidad* de (X, Y) . El conjunto de ternas $(x_i, y_j, p(x_i, y_j))$, $i, j = 1, 2, \dots$, es llamada en algunos casos la *distribución de probabilidades* de (X, Y) .

(b) Sea (X, Y) una variable aleatoria continua que toma todos los valores en una región R del plano euclíadiano. La *función de densidad de probabilidades conjuntas* f es una función que satisface las siguientes condiciones:

$$(3) \quad f(x, y) \geq 0 \quad \text{para todo } (x, y) \in R,$$

$$(4) \quad \iint_R f(x, y) dx dy = 1. \quad (6.2)$$

Observaciones: (a) La analogía a una distribución de masas es otra vez indudable. Tenemos una masa unitaria distribuida en una región en el plano. En el caso discreto, toda la masa está concentrada en un número finito o infinito numerable, de puntos con masa $p(x_i, y_j)$ ubicados en (x_i, y_j) . En el caso continuo, la masa se encuentra en todos los puntos de un conjunto no numerable en el plano.

(b) La condición 4 indica que el *volumen* total bajo la superficie dada por la ecuación $z = f(x, y)$ es igual a 1.

(c) Como en el caso unidimensional, $f(x, y)$ no representa la probabilidad de algo. Sin embargo, para Δx y Δy suficientemente pequeños y positivos, $f(x, y) \Delta x \Delta y$ es aproximadamente igual a $P(x \leq X \leq x + \Delta x, y \leq Y \leq y + \Delta y)$.

(d) Como en el caso unidimensional adoptaremos la convención de que $f(x, y) = 0$ si $(x, y) \notin R$. Por tanto, podemos considerar f definida para todo (x, y) en el plano y la condición 4 anterior llega a ser $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) dx dy = 1$.

(e) Nuevamente *suprimiremos* la naturaleza funcional de la variable aleatoria bidimensional (X, Y) . Deberíamos escribir proposiciones de la forma $P[X(s) = x_i, Y(s) = y_j]$, etc. Sin embargo, si se entiende nuestra notación taquigráfica no debe surgir, ninguna dificultad.

(f) De nuevo, como en el caso unidimensional, la distribución de probabilidades de (X, Y) es realmente inducida por la probabilidad de sucesos asociados con el espacio muestral original S . Sin embargo, principalmente nos interesaremos por los valores de (X, Y) y por tanto relacionados directamente con el recorrido de (X, Y) . No obstante, el lector no debe perder de vista el hecho de que si $P(A)$ está especificada para todos los sucesos $A \subset S$, entonces está determinada la probabilidad asociada con sucesos en el recorrido de (X, Y) . Es decir, si B está en el recorrido de (X, Y) , tenemos

$$P(B) = P[(X(s), Y(s)) \in B] = P[s | (X(s), Y(s)) \in B].$$

Esta última probabilidad se refiere a un suceso en S y por tanto *determina* la probabilidad de B . En nuestra terminología previa, B y $\{s | (X(s), Y(s)) \in B\}$ son sucesos *equivalentes* (figura 6.2).

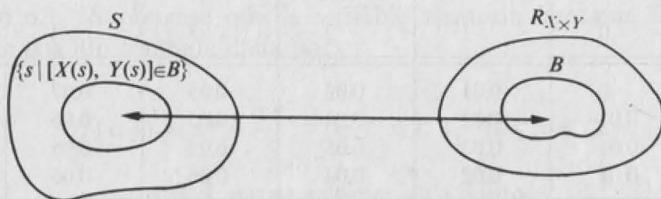


FIGURA 6.2

Si B está en el recorrido de (X, Y) tenemos

$$P(B) = \sum_B \sum p(X_i, y_j), \quad (6.3)$$

si (X, Y) es discreta, la suma se toma con todos los índices (i, j) para los cuales $(x_i, y_j) \in B$.

$$P(B) = \iint_B f(x, y) dx dy, \quad (6.4)$$

si (X, Y) es continua.

EJEMPLO 6.1. Dos líneas de producción manufacturan cierto tipo de artículo. Supóngase que la capacidad (en cualquier día dado) es 5 artículos para la línea I y 3 artículos para la línea II. Supóngase que el número verdadero de artículos producidos por cada una de las líneas de producción es una variable aleatoria. Sea (X, Y) la representación de la variable aleatoria bidimensional que da el número de artículos producidos por la línea I y por la línea II respectivamente. La tabla 6.1 da la distribución de probabilidades conjunta de (X, Y) . Cada entrada representa

$$p(x_i, y_j) = P(X = x_i, Y = y_j).$$

Así $p(2, 3) = P(X = 2, Y = 3) = 0,04$, etc. Por lo tanto si B está definida como

$$B = \{\text{Más artículos producidos por la línea I que por la línea II}\}$$

encontramos que

$$\begin{aligned} P(B) &= 0,01 + 0,03 + 0,05 + 0,07 + 0,09 + 0,04 + 0,05 + 0,06 \\ &\quad + 0,08 + 0,05 + 0,05 + 0,06 + 0,06 + 0,05 \\ &= 0,75. \end{aligned}$$

EJEMPLO 6.2. Supóngase que un fabricante de bombillas está interesado en el número de éstas que le han sido pedidas durante los meses de enero y febrero. X y Y indican el número de bombillas ordenadas durante esos dos meses respectivamente. Supondremos que (X, Y) es una variable aleatoria bidimensional

TABLA 6.1

$X \backslash Y$	0	1	2	3	4	5
0	0	0,01	0,03	0,05	0,07	0,09
1	0,01	0,02	0,04	0,05	0,06	0,08
2	0,01	0,03	0,05	0,05	0,05	0,06
3	0,01	0,02	0,04	0,06	0,06	0,05

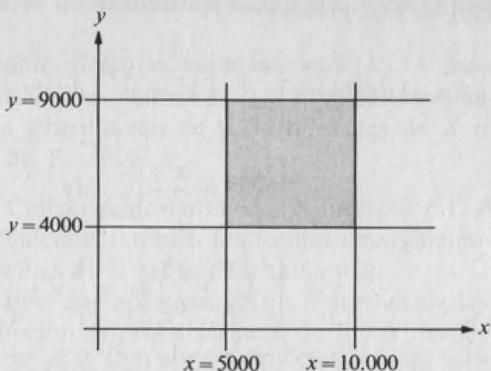


FIGURA 6.3

con la siguiente fdp conjunta (ver figura 6.3):

$$\begin{aligned}f(x, y) &= c && \text{si } 5000 \leq x \leq 10,000 \quad \text{y} \quad 4000 \leq y \leq 9000, \\&= 0 && \text{en otro caso.}\end{aligned}$$

Para determinar c usamos el hecho de que $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) dx dy = 1$. Por tanto

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) dx dy = \int_{4000}^{9000} \int_{5000}^{10,000} f(x, y) dx dy = c[5000]^2.$$

Así $c = (5000)^{-2}$. Por tanto si $B = \{X \geq Y\}$, tenemos

$$\begin{aligned}P(B) &= 1 - \frac{1}{(5000)^2} \int_{5000}^{9000} \int_{5000}^y dx dy \\&= 1 - \frac{1}{(5000)^2} \int_{5000}^{9000} [y - 5000] dy = \frac{17}{25}.\end{aligned}$$

Observación: en el ejemplo anterior X y Y obviamente deben ser números enteros puesto que no podemos ordenar un número fraccionario de bombillas. Sin embargo, nuevamente tratamos con una situación idealizada en la cual permitimos que X tome *todos* los valores entre 5000 y 10,000 (inclusive).

EJEMPLO 6.3. Supóngase que la variable aleatoria continua bidimensional (X, Y) tiene una fdp conjunta dada por

$$\begin{aligned}f(x, y) &= x^2 + \frac{xy}{3}, && 0 \leq x \leq 1, \quad 0 \leq y \leq 2, \\&= 0, && \text{para cualquier otro punto.}\end{aligned}$$

Para verificar que $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) dx dy = 1$:

$$\begin{aligned}\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) dx dy &= \int_0^2 \int_0^1 \left(x^2 + \frac{xy}{3} \right) dx dy \\&= \int_0^2 \frac{x^3}{3} + \frac{x^2 y}{6} \Big|_{x=0}^{x=1} dy \\&= \int_0^2 \left(\frac{1}{3} + \frac{y}{6} \right) dy = \frac{1}{3} y + \frac{y^2}{12} \Big|_0^2 \\&= \frac{2}{3} + \frac{4}{12} = 1.\end{aligned}$$

Sea $B = \{X + Y \geq 1\}$. (Ver figura 6.4) Calcularemos $P(B)$ al evaluar $1 - P(\bar{B})$, en donde $\bar{B} = \{X + Y < 1\}$. Por lo tanto

$$\begin{aligned}P(B) &= 1 - \int_0^1 \int_0^{1-x} \left(x^2 + \frac{xy}{3} \right) dy dx \\&= 1 - \int_0^1 \left[x^2(1-x) + \frac{x(1-x)^2}{6} \right] dx \\&= 1 - \frac{7}{2} = \frac{9}{2}.\end{aligned}$$

Al estudiar variables aleatorias unidimensionales, encontramos que F , la función de distribución acumulativa, jugaba un papel importante. En el caso bidimensional podemos definir otra vez una función acumulativa como se indica a continuación.

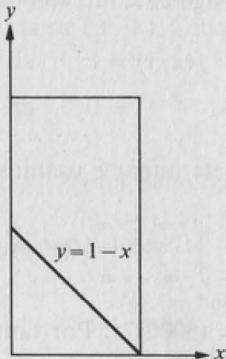


FIGURA 6.4

Definición. Sea (X, Y) una variable aleatoria bidimensional. La *función de distribución acumulativa* (fda) F de la variable aleatoria bidimensional (X, Y) está definida por

$$F(x, y) = P(X \leq x, Y \leq y).$$

Observación: F es una función de *dos* variables y tiene un número de propiedades análogas a las discutidas para la fda unidimensional. (Ver sección 4.5). Mencionaremos sólo la siguiente propiedad importante.

Si F es fda de una variable bidimensional con fdp conjunta f , entonces

$$\partial^2 F(x, y)/\partial x \partial y = f(x, y)$$

dondequiero que F sea diferenciable. Este resultado es análogo al del teorema 4.4 en el cual probamos que $(d/dx)F(x) = f(x)$, en donde f es la fdp de una variable aleatoria unidimensional X .

6.2 Distribuciones de probabilidades marginales y condicionales

Con cada variable aleatoria bidimensional (X, Y) asociamos dos variables aleatorias unidimensionales, llamadas X y Y respectivamente. Es decir, podemos interesarnos por la distribución de probabilidades de X o por la distribución de probabilidades de Y .

EJEMPLO 6.4. Consideremos otra vez el ejemplo 6.1. Además de los datos de la tabla 6.1, calculemos también los totales «marginales», esto es, la suma de las 6 columnas y 4 filas de la tabla. (Ver tabla 6.2).

Las probabilidades que aparecen en los márgenes de las filas y columnas representan la distribución de probabilidades de Y y X , respectivamente. Por ejemplo, $P(Y = 1) = 0,26$, $P(X = 3) = 0,21$, etc. Debido a la forma de la tabla 6.2 nos referimos, generalmente, a la distribución *marginal* de X o a la distribución *marginal* de Y , cada vez que tenemos una variable aleatoria bidimensional (X, Y) sea discreta o continua.

TABLA 6.2

$Y \backslash X$	0	1	2	3	4	5	Suma
0	0	0,01	0,03	0,05	0,07	0,09	0,25
1	0,01	0,02	0,04	0,05	0,06	0,08	0,26
2	0,01	0,03	0,05	0,05	0,05	0,06	0,25
3	0,01	0,02	0,04	0,06	0,06	0,05	0,24
Suma	0,03	0,08	0,16	0,21	0,24	0,28	1,00

En el caso *discreto* procedemos así: puesto que $X = x_i$ debe ocurrir con $Y = y_j$ para una j , y puede ocurrir con $Y = y_j$ para una j , tenemos:

$$\begin{aligned} p(x_i) &= P(X = x_i) = P(X = x_i, Y = y_1 \text{ o } X = x_i, Y = y_2 \text{ o } \dots) \\ &= \sum_{j=1}^{\infty} p(x_i, y_j). \end{aligned}$$

Definida la función p para x_1, x_2, \dots , representa la *distribución marginal de probabilidades* de X . Análogamente definimos $q(y_j) = P(Y = Y_j) = \sum_{i=1}^{\infty} p(x_i, y_j)$ como la *distribución marginal de probabilidades* de Y .

En el caso *continuo* procedemos como sigue: sea f la función densidad de probabilidades conjunta de la variable aleatoria continua bidimensional (X, Y) . Definimos g y h las *funciones densidad de probabilidades marginales* de X y Y , respectivamente, como sigue:

$$g(x) = \int_{-\infty}^{+\infty} f(x, y) dy; \quad h(y) = \int_{-\infty}^{+\infty} f(x, y) dx.$$

Estas fdp corresponden a las fdp básicas de las variables aleatorias unidimensionales X y Y , respectivamente. Por ejemplo

$$\begin{aligned} P(c \leq X \leq d) &= P[c \leq X \leq d, -\infty < Y < \infty] \\ &= \int_c^d \int_{-\infty}^{+\infty} f(x, y) dy dx \\ &= \int_c^d g(x) dx. \end{aligned}$$

EJEMPLO 6.5. El empuje X y la razón de la mezcla Y son dos características del funcionamiento de un motor a reacción. Supóngase que (X, Y) es una variable aleatoria bidimensional continua con fdp conjunta:

$$\begin{aligned} f(x, y) &= 2(x + y - 2xy), \quad 0 \leq x \leq 1, \quad 0 \leq y \leq 1, \\ &= 0, \quad \text{en otro caso.} \end{aligned}$$

(Las unidades se han ajustado para usar valores entre 0 y 1.) La fdp marginal de X está dada por

$$\begin{aligned} g(x) &= \int_0^1 2(x + y - 2xy) dy = 2(xy + y^2/2 - xy^2)|_0^1 \\ &= 1, \quad 0 \leq x \leq 1. \end{aligned}$$

Es decir, X está distribuida uniformemente en $[0, 1]$.

La fdp marginal de Y está dada por

$$\begin{aligned} h(y) &= \int_0^1 2(x + y - 2xy) dx = 2(x^2/2 + xy - x^2y)|_0^1 \\ &= 1, \quad 0 \leq y \leq 1. \end{aligned}$$

Por lo tanto Y también está distribuida uniformemente sobre $[0, 1]$.

Definición. Decimos que una variable aleatoria continua bidimensional está *distribuida uniformemente* en una región R en un plano eucliano, si

$$\begin{aligned} f(x, y) &= \text{const.} \quad \text{para } (x, y) \in R, \\ &= 0; \quad \text{para todo otro punto.} \end{aligned}$$

Debido a la condición $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) dx dy = 1$, lo anterior implica que la constante es igual a $1/\text{área}(R)$. Suponemos que R es una región con área finita distinta de cero.

Observación: esta definición presenta la analogía bidimensional con la variable aleatoria unidimensional distribuida uniformemente.

EJEMPLO 6.6. Supóngase que la variable aleatoria bidimensional (X, Y) está distribuida uniformemente en la región sombreada R , indicada en la figura 6.5. Por tanto

$$f(x, y) = \frac{1}{\text{área}(R)}, \quad (x, y) \in R.$$

Encontramos que

$$\text{área}(R) = \int_0^1 (x - x^2) dx = \frac{1}{6}.$$

Luego la fdp está dada por

$$\begin{aligned} f(x, y) &= 6, \quad (x, y) \in R \\ &= 0, \quad (x, y) \notin R. \end{aligned}$$

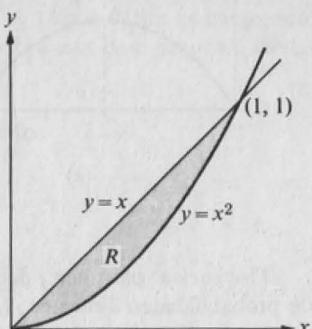


FIGURA 6.5

En las ecuaciones siguientes encontramos las fdp marginales de X y Y .

$$\begin{aligned} g(x) &= \int_{-\infty}^{+\infty} f(x, y) dy = \int_{x^2}^x 6 dy \\ &= 6(x - x^2), \quad 0 \leq x \leq 1; \\ h(y) &= \int_{-\infty}^{+\infty} f(x, y) dx = \int_y^{\sqrt{y}} 6 dx \\ &= 6(\sqrt{y} - y), \quad 0 \leq y \leq 1. \end{aligned}$$

Los gráficos de estas fdp aparecen en la figura 6.6.

El concepto de probabilidad condicional se puede presentar de una manera muy natural.

EJEMPLO 6.7. Consideremos otra vez los ejemplos 6.1 y 6.4. Supóngase que queremos evaluar la probabilidad condicional $P(X = 2 | Y = 2)$. De acuerdo con la definición de probabilidad condicional tenemos

$$P(X = 2 | Y = 2) = \frac{P(X = 2, Y = 2)}{P(Y = 2)} = \frac{0,05}{0,25} = 0,20.$$

Podemos efectuar tal cálculo muy general para el caso discreto. Tenemos

$$\begin{aligned} p(x_i | y_j) &= P(X = x_i | Y = y_j) \\ &= \frac{p(x_i, y_j)}{q(y_j)} \quad \text{si } q(y_j) > 0, \end{aligned} \tag{6.5}$$

$$\begin{aligned} q(y_j | x_i) &= P(Y = y_j | X = x_i) \\ &= \frac{p(x_i, y_j)}{p(x_i)} \quad \text{si } p(x_i) > 0. \end{aligned} \tag{6.6}$$

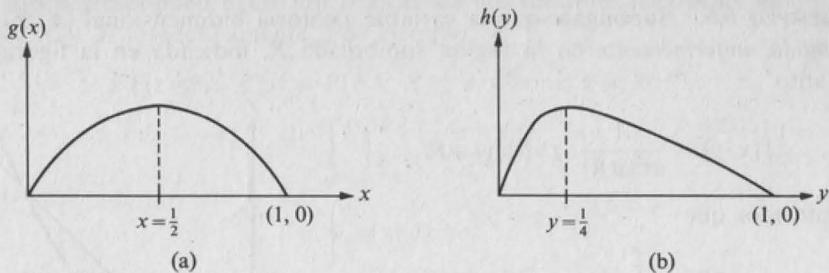


FIGURA 6.6

Observación: para una j dada, $p(x_i | y_j)$ satisface todas las condiciones de una distribución de probabilidades. Tenemos $p(x_i | y_j) \geq 0$ y también

$$\sum_{i=1}^{\infty} p(x_i | y_j) = \sum_{i=1}^{\infty} \frac{p(x_i, y_j)}{q(y_j)} = \frac{q(y_j)}{q(y_j)} = 1.$$

En el *caso continuo* la formulación de la probabilidad condicional presenta alguna dificultad puesto que para cualquier x_0, y_0 , dado, tenemos $P(X = x_0) = P(Y = y_0) = 0$. Enunciemos las definiciones formales siguientes.

Definición. Sea (X, Y) una variable bidimensional continua con fdp conjunta f . Sean g y h las fdp marginales de X y Y respectivamente.

La fdp *condicional* de X para $Y = y$ dada, está definida por

$$g(x | y) = \frac{f(x, y)}{h(y)}, \quad h(y) > 0. \quad (6.7)$$

La fdp *condicional* de Y para una $X = x$ dada se define

$$h(y | x) = \frac{f(x, y)}{g(x)}, \quad g(x) > 0. \quad (6.8)$$

Observaciones: (a) Las fdp condicionales anteriores satisfacen todas las exigencias de una fdp unidimensional. Así, para y fijo, tenemos $g(x | y) \geq 0$ y

$$\int_{-\infty}^{+\infty} g(x | y) dx = \int_{-\infty}^{+\infty} \frac{f(x, y)}{h(y)} dx = \frac{1}{h(y)} \int_{-\infty}^{+\infty} f(x, y) dx = \frac{h(y)}{h(y)} = 1.$$

Un cálculo análogo se puede hacer para $h(y | x)$. Por tanto, las ecuaciones (6.7) y (6.8) definen las fdp en R_X y R_Y , respectivamente.

(b) Una interpretación intuitiva de $g(x | y)$ se obtiene si consideramos que la superficie representada por la fdp conjunta f es cortada digamos por el plano $y = c$. La intersección del plano con la superficie $z = f(x, y)$ resultará en una fdp unidimensional, llamada la fdp de X para $Y = c$. Esta será precisamente $g(x | c)$.

(c) Supongamos que (X, Y) representen respectivamente la altura y el peso de una persona. Sea f la fdp conjunta de (X, Y) y sea g la fdp marginal de X (prescindiendo de Y).

Por tanto $\int_{5.8}^6 g(x)dx$ representaría la probabilidad del suceso $\{5.8 \leq X \leq 6\}$ prescindiendo del peso Y . $\int_{5.8}^6 g(x|150)dx$ se interpretaría como $P(5.8 \leq X \leq 6 | Y = 150)$. Estrictamente hablando, esta probabilidad condicional no está definida en los términos de nuestra convención previa con la probabilidad condicional, puesto que $P(Y = 150) = 0$. Sin embargo, sólo usamos la integral anterior para *definir* esta probabilidad. Sobre una base intuitiva, ciertamente éste debe ser el significado de este número.

EJEMPLO 6.8. Refiriéndonos al ejemplo 6.3, tenemos

$$g(x) = \int_0^2 \left(x^2 + \frac{xy}{3} \right) dy = 2x^2 + \frac{2}{3}x,$$

$$h(y) = \int_0^1 \left(x^2 + \frac{xy}{3} \right) dx = \frac{y}{6} + \frac{1}{3}.$$

Por tanto,

$$g(x|y) = \frac{x^2 + xy/3}{1/3 + y/6} = \frac{6x^2 + 2xy}{2 + y}, \quad 0 \leq x \leq 1, \quad 0 \leq y \leq 2;$$

$$h(y|x) = \frac{x^2 + xy/3}{2x^2 + 2/3(x)} = \frac{3x^2 + xy}{6x^2 + 2x} = \frac{3x + y}{6x + 2},$$

$$0 \leq y \leq 2, \quad 0 \leq x \leq 1.$$

Para verificar que $g(x|y)$ es una fdp, tenemos

$$\int_0^1 \frac{6x^2 + 2xy}{2 + y} dx = \frac{2 + y}{2 + y} = 1 \quad \text{para todo } y.$$

Un cálculo semejante se puede hacer para $h(y|x)$.

6.3 Variables aleatorias independientes

Tal como definimos el concepto de independencia entre dos sucesos A y B , definiremos ahora las *variables aleatorias independientes*. Lo que queremos decir intuitivamente es que X y Y son variables aleatorias independientes si el resultado de X , digamos, de ninguna manera influye en el resultado de Y . Esta es una noción extremadamente importante y hay muchas situaciones en que tal suposición se justifica.

EJEMPLO 6.9. Consideremos dos fuentes de material radiactivo situadas a cierta distancia una de otra, que emiten partículas α . Supongamos que estas dos fuentes se observan durante un período de dos horas y se anota el número de partículas emitidas. Supóngase que las siguientes variables aleatorias son de interés: X_1 y X_2 , el número de partículas emitidas por la primera fuente durante la primera y la segunda hora, respectivamente; Y_1 y Y_2 , el número de partículas emitidas por la segunda fuente durante la primera y la segunda hora, respectivamente. Parece intuitivamente obvio que $(X_1 \text{ y } Y_1)$, o $(X_1 \text{ y } Y_2)$, o $(X_2 \text{ y } Y_1)$, o $(X_2 \text{ y } Y_2)$

y Y_2) son todas pares de variables aleatorias independientes. Las X dependen sólo de las características de la fuente 1 mientras que las Y dependen de las características de la fuente 2, y posiblemente no hay razón para suponer que una fuente influya de manera alguna en el comportamiento de la otra. Cuando consideramos la posible independencia de X_1 y X_2 , el asunto no es tan definido. ¿El número de partículas emitidas durante la segunda hora es influenciado por el número emitido durante la primera hora? Para responder a este interrogante tendríamos que obtener información adicional acerca del mecanismo de emisión. Ciertamente que no podríamos suponer a *priori*, que X_1 y X_2 son independientes.

Hagamos ahora más precisa la anterior noción intuitiva de independencia.

Definición. (a) Sea (X, Y) una variable aleatoria bidimensional discreta. Decimos que X y Y son variables aleatorias independientes si y sólo si $p(x_i, y_j) = p(x_i)q(y_j)$ para todo i y j . Esto es, $P(X = x_i, Y = y_j) = P(X = x_i)P(Y = y_j)$, para todo i y j .

(b) Sea (X, Y) una variable aleatoria bidimensional continua. Decimos que X y Y son variables aleatorias independientes si y sólo si $f(x, y) = g(x)h(y)$ para todo (x, y) , en donde f es la fdp conjunta, y g y h son las fdp marginales de X y Y respectivamente.

Observación: si comparamos la definición anterior con la dada para sucesos independientes, la semejanza es aparente: esencialmente necesitamos que la probabilidad conjunta (o fdp conjunta) pueda ser factorizada. El teorema siguiente indica que la definición anterior es equivalente a otro enfoque que podríamos haber tomado.

Teorema 6.1. (a) Sea (X, Y) una variable aleatoria bidimensional discreta.

Entonces X y Y son independientes si y sólo si $p(x_i | y_j) = p(x_i)$ para todo i y j (o lo que es equivalente, si y sólo si $q(y_j | x_i) = q(y_j)$ para todo i y j).

(b) Sea (X, Y) una variable aleatoria bidimensional continua. Entonces X y Y son independientes si y sólo si $g(x | y) = g(x)$, o lo que es equivalente, si y sólo si $h(y | x) = h(y)$, para todo (x, y) .

Demostración: ver problema 6.10.

EJEMPLO 6.10. Supongamos que una máquina se usa para un trabajo específico en la mañana y para uno diferente en la tarde. Representemos por X y Y el número de veces que la máquina falla en la mañana y en la tarde, respectivamente. La tabla 6.3 da la distribución de probabilidades conjunta de (X, Y) .

Un cálculo fácil revela que para *todas* las entradas de la tabla 6.3 tenemos

$$p(x_i, y_j) = p(x_i)q(y_j).$$

Así X y Y son variables aleatorias independientes. (Para comparar véase también el ejemplo 3.7.)

TABLA 6.3

$X \backslash Y$	0	1	2	$q(y_j)$
0	0,1	0,2	0,2	0,5
1	0,04	0,08	0,08	0,2
2	0,06	0,12	0,12	0,3
$p(x_i)$	0,2	0,4	0,4	1,0

EJEMPLO 6.11. Sean X y Y la duración de dos dispositivos electrónicos. Supóngase que su fdp conjunta está dada por

$$f(x, y) = e^{-(x+y)}, \quad x \geq 0, \quad y \geq 0.$$

Puesto que podemos factorizar $f(x, y) = e^{-x}e^{-y}$, se establece la independencia de X y Y .

EJEMPLO 6.12. Supóngase que $f(x, y) = 8xy$, $0 \leq x \leq y \leq 1$. (El dominio está dado por la región sombreada en la figura 6.7). Aunque f ya está escrita en forma factorizada, X y Y no son independientes, puesto que el dominio de definición

$$\{(x, y) \mid 0 \leq x \leq y \leq 1\}$$

es tal que para una x dada, y puede tomar sólo valores mayores que la x dada y menores que 1. Por tanto X y Y no son independientes.

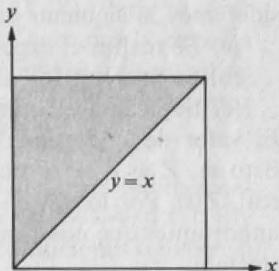


FIGURA 6.7

Observación: de la definición de la distribución marginal de probabilidades (en cualquiera de los casos discreto o continuo) es claro que la distribución conjunta de probabilidades determina, en forma única, la distribución marginal de probabilidades. Es decir, que conociendo la fdp f conjunta, podemos obtener las fdp marginales g y h . Sin embargo, lo inverso no es cierto. Es decir, en general, el conocimiento de las fdp marginales g y h no determinan la fdp conjunta f . Esto sólo es verdadero cuando X y Y son independientes, porque en este caso tenemos $f(x, y) = g(x)h(y)$.

El resultado siguiente indica que nuestra definición de variables aleatorias independientes está de acuerdo con nuestra definición previa de sucesos independientes.

Teorema 6.2. Sea (X, Y) una variable aleatoria bidimensional. Sean A y B sucesos cuya ocurrencia (o no ocurrencia) depende sólo de X y Y respectivamente. (Esto es, A es un subconjunto de R_X , el recorrido de X , mientras

que B es un subconjunto de R_Y , el recorrido de Y .) Entonces, si X y Y son variables aleatorias independientes, tenemos $P(A \cap B) = P(A)P(B)$.

Demostración: (sólo el caso continuo):

$$\begin{aligned} P(A \cap B) &= \iint_{A \cap B} f(x, y) dx dy = \iint_{A \cap B} g(x)h(y) dx dy \\ &= \int_A g(x) dx \int_B h(y) dy = P(A)P(B). \end{aligned}$$

6.4 Funciones de una variable aleatoria

Al definir una variable aleatoria X , señalamos, muy enfáticamente, que X es una función definida del espacio muestral S a los números reales. Al definir una variable aleatoria bidimensional (X, Y) , tratamos un par de funciones $X = X(s)$, $Y = Y(s)$, cada una de las cuales está definida en el espacio muestral de algún experimento y cada una de las cuales asigna un número real a cada $s \in S$, produciendo así el vector bidimensional $[X(s), Y(s)]$.

Consideremos ahora $Z = H_1(X, Y)$, una función de las dos variables aleatorias X y Y . Debe estar claro que $Z = Z(s)$ es otra vez una variable aleatoria. Consideremos la siguiente sucesión de pasos:

- (a) Se realiza el experimento ε y se obtiene el resultado s .
- (b) Se evalúan los números $X(s)$ y $Y(s)$.
- (c) Se evalúa el número $Z = H_1[X(s), Y(s)]$.

El valor de Z depende claramente de s , el resultado original del experimento. Esto es, $Z = Z(s)$ es una función que asigna a cada resultado $s \in S$ un número real, $Z(s)$. Por lo tanto, Z es una variable aleatoria. Algunas variables aleatorias importantes que nos van a interesar son $X + Y$, XY , X/Y , $\min(X, Y)$, $\max(X, Y)$, etcétera.

El problema que resolvimos en el capítulo anterior para la variable aleatoria unidimensional aparece otra vez: dada la distribución de probabilidades conjunta de (X, Y) , ¿cuál es la distribución de probabilidades de $Z = H_1(X, Y)$? (Debe haber quedado claro en las numerosas discusiones previas sobre este punto, que una distribución de probabilidades se induce en R_Z , el espacio muestral de Z .)

Si (X, Y) es una variable aleatoria discreta, este problema se resuelve muy fácilmente. Supóngase que (X, Y) tiene la distribución dada en los ejemplos 6.1 y 6.4. Las siguientes variables aleatorias (unidimensionales) podrían ser de interés:

$U = \min(X, Y)$ = el menor número de artículos producidos por las dos líneas;

$V = \max(X, Y)$ = el mayor número de artículos producido por las dos líneas;

$W = X + Y$ = número total de artículos producido por las dos líneas.

Para obtener la distribución de probabilidades de U , por ejemplo, procedemos como sigue. Los valores posibles de U son: 0, 1, 2 y 3. Para evaluar $P(U = 0)$ argumentamos que $U = 0$ si y sólo si ocurre uno de los casos siguientes: $X = 0$, $Y = 0$ o $X = 0$, $Y = 1$ o $X = 0$, $Y = 2$ o $X = 0$, $Y = 3$ o $X = 1$, $Y = 0$ o $X = 2$.

$Y = 0 \text{ o } X = 3, Y = 0 \text{ o } X = 4, Y = 0 \text{ o } X = 5, Y = 0$. Por lo tanto $P(U = 0) = 0,28$. El resto de las probabilidades asociadas con U se pueden obtener de una manera semejante. Por tanto, la distribución de probabilidades de U se puede resumir como sigue: $u: 0, 1, 2, 3; P(U = u): 0,28; 0,30; 0,25; 0,17$. La distribución de probabilidades de las variables aleatorias V y W como se definió anteriormente se pueden obtener de una manera semejante. (Ver problema 6.9).

Si (X, Y) es una variable bidimensional continua y si $Z = H_1(X, Y)$ es una función continua de (X, Y) , entonces Z será una variable aleatoria continua (unidimensional) y el problema de encontrar su fdp es algo más complicado. Para resolver este problema necesitamos un teorema que vamos a indicar y discutir a continuación. Antes de hacerlo, bosquejemos brevemente la idea básica.

Para encontrar la fdp de $Z = H_1(X, Y)$ es a menudo más simple introducir una segunda variable aleatoria, digamos $W = H_2(X, Y)$, y obtener primero la fdp conjunta de Z y W , digamos $k(z, w)$. Conociendo $k(z, w)$, podemos entonces obtener la fdp de Z , digamos $g(z)$, al integrar simplemente $k(z, w)$ con respecto a w . Esto es,

$$g(z) = \int_{-\infty}^{+\infty} k(z, w) dw.$$

Los problemas que quedan son (1) cómo encontrar la fdp conjunta de Z y W , y (2) cómo elegir la variable aleatoria apropiada $W = H_2(X, Y)$. Para resolver el último problema, simplemente, digamos que generalmente hacemos la elección más sencilla posible de W . En el contexto presente, W desempeña sólo un papel intermedio, y realmente no nos interesa en sí misma. Con el fin de encontrar la fdp conjunta de Z y W necesitamos el teorema 6.3.

Teorema 6.3. Supongamos que (X, Y) es una variable aleatoria bidimensional continua con fdp conjunta f . Sea $Z = H_1(X, Y)$ y $W = H_2(X, Y)$, y supongamos que las funciones H_1 y H_2 satisfacen las condiciones siguientes:

- (a) Las ecuaciones $z = H_1(x, y)$ y $w = H_2(x, y)$ se pueden resolver únicamente para x y y en función de z y w , digamos $x = G_1(z, w)$ y $y = G_2(z, w)$.
- (b) Las derivadas parciales $\partial x / \partial z$, $\partial x / \partial w$, $\partial y / \partial z$, y $\partial y / \partial w$ existen y son continuas.

Luego la fdp conjunta de (Z, W) , digamos $k(z, w)$, está dada por la expresión siguiente: $k(z, w) = f[G_1(z, w), G_2(z, w)] |J(z, w)|$, en donde $J(z, w)$ es el siguiente determinante 2×2 :

$$J(z, w) = \begin{vmatrix} \frac{\partial x}{\partial z} & \frac{\partial x}{\partial w} \\ \frac{\partial y}{\partial z} & \frac{\partial y}{\partial w} \end{vmatrix}.$$

Este determinante se llama *Jacobiano* de la transformación $(x, y) \rightarrow (z, w)$ y algunas veces se escribe así $\partial(x, y) / \partial(z, w)$. Observemos que $k(z, w)$ será distinta de cero para esos valores de (z, w) que corresponden a valores de (x, y) para los cuales $f(x, y)$ es distinta de cero.

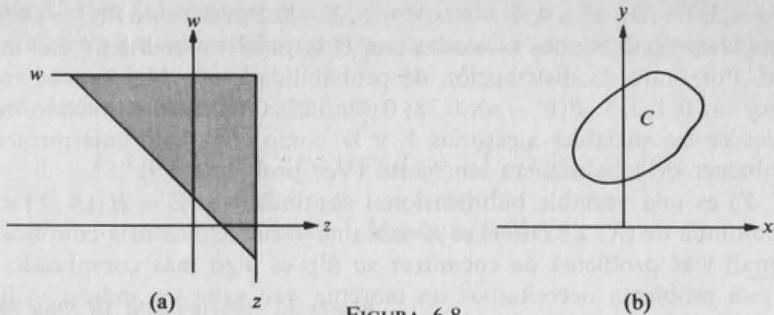


FIGURA 6.8

Observaciones: (a) Aunque no demostraremos este teorema, por lo menos indicaremos lo que necesita ser demostrado y dónde se encuentran las dificultades. Considérese la fda conjunta de la variable aleatoria bidimensional (Z, W) ,

$$K(z, w) = P(Z \leq z, W \leq w) = \int_{-\infty}^w \int_{-\infty}^z k(s, t) ds dt,$$

en donde k es la fdp buscada. Puesto que se supone que la transformación $(x, y) \rightarrow (z, w)$ es de uno a uno (véase la suposición (a) anterior), podemos encontrar el suceso, equivalente $\{Z \leq z, W \leq w\}$, en función de X y Y . Supóngase que este suceso se indica como C . (Ver figura 6.8). Esto es, $\{(X, Y) \in C\}$ si y sólo si $\{Z \leq z, W \leq w\}$. Por tanto

$$\int_{-\infty}^w \int_{-\infty}^z k(s, t) ds dt = \int_C f(x, y) dx dy.$$

Puesto que se supone que f es conocido, puede calcularse el segundo miembro de la integral. Diferenciándola con respecto a z y w obtenemos la fdp pedida. En la mayoría de los textos de cálculo avanzado se demuestra que estas técnicas conducen al resultado formulado en el teorema anterior.

(b) Nótese la gran semejanza entre el resultado anterior y el obtenido en el caso unidimensional tratado en el capítulo anterior. (Ver teorema 5.1). La condición de monotonía de la función $y = H(x)$ se sustituye por la condición de que la correspondencia entre (x, y) y (z, w) es uno a uno. La condición de diferencialidad es sustituida por ciertas suposiciones acerca de las derivadas parciales consideradas. La solución final obtenida es también muy semejante a la obtenida en el caso unidimensional: las variables x y y sencillamente se sustituyen por sus expresiones equivalentes en función de z y w , y el valor absoluto de dx/dy se sustituye por el valor absoluto del Jacobiano.

EJEMPLO 6.13. Supóngase que estamos apuntando a un blanco circular de radio uno, que ha sido colocado de modo que su centro esté en el origen de un sistema de coordenadas rectangulares (figura 6.9). Supóngase que las coordenadas (X, Y) de los puntos de impacto están distribuidas uniformemente en el círculo. Es decir,

$$f(x, y) = \begin{cases} 1/\pi & \text{si } (x, y) \text{ queda en el interior o sobre el círculo,} \\ 0, & \text{en otro punto.} \end{cases}$$

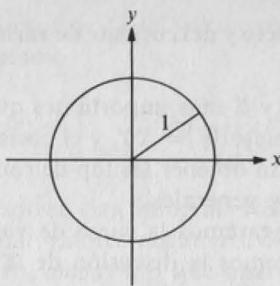


FIGURA 6.9

Supóngase que estamos interesados en la variable aleatoria R que representa la *distancia* del origen. (Ver figura 6.10). Es decir, $R = \sqrt{X^2 + Y^2}$. Encontraremos la fdp de R , digamos g , así: sea $\Phi = \tan^{-1}(Y/X)$. Por tanto, $X = H_1(R, \Phi)$ y $Y = H_2(R, \Phi)$ donde $x = H_1(r, \phi) = r \cos \phi$ y $y = H_2(r, \phi) = r \sin \phi$. (Simplemente hemos introducido coordenadas polares).

El Jacobiano es

$$J = \begin{vmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \phi} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \phi} \end{vmatrix} = \begin{vmatrix} \cos \phi & -r \sin \phi \\ \sin \phi & r \cos \phi \end{vmatrix} \\ = r \cos^2 \phi + r \sin^2 \phi = r.$$

Bajo la transformación anterior, el círculo unitario en el plano xy se aplica en el rectángulo en el plano $r\phi$ en la figura 6.11. Por tanto la fdp conjunta de (Φ, R) está dada por

$$g(\phi, r) = \frac{r}{\pi}, \quad 0 \leq r \leq 1, \quad 0 \leq \phi < 2\pi.$$

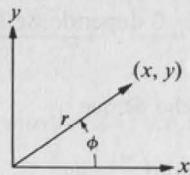


FIGURA 6.10

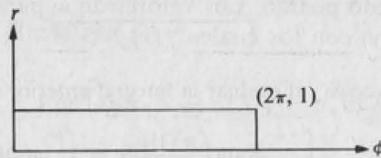


FIGURA 6.11

Así la fdp pedida de R , digamos h , está dada por

$$h(r) = \int_0^{2\pi} g(\phi, r) d\phi = 2r, \quad 0 \leq r \leq 1.$$

Observación: este ejemplo señala la importancia de obtener una representación precisa de la región con los valores posibles introducidos por la nueva variable aleatoria.

6.5 Distribuciones del producto y del cociente de variables aleatorias independientes

Entre las funciones de X y Y más importantes que deseamos considerar están la suma $S = X + Y$, el producto $W = XY$, y el cociente $Z = X/Y$. Podemos usar el método de esta sección para obtener las fdp de cada una de esas variables aleatorias bajo condiciones muy generales.

En el capítulo 12, investigaremos la suma de variables aleatorias con mayor detalle. Por tanto, postergaremos la discusión de la distribución de probabilidades de $X + Y$ hasta entonces. Sin embargo, consideraremos el producto y el cociente en los dos teoremas siguientes.

Teorema 6.4. Sea (X, Y) una variable aleatoria bidimensional continua y supongamos que X y Y son *independientes*. Por tanto la fdp f se puede escribir como $f(x, y) = g(x)h(y)$. Sea $W = XY$.

Luego la fdp de W , digamos p , está dada por

$$p(w) = \int_{-\infty}^{+\infty} g(u)h\left(\frac{w}{u}\right)\left|\frac{1}{u}\right| du. \quad (6.9)$$

Demostración: sea $w = xy$ y $u = x$. Así $x = u$ y $y = w/u$. El Jacobiano es

$$J = \begin{vmatrix} 1 & 0 \\ -w & 1 \\ \frac{-w}{u^2} & \frac{1}{u} \end{vmatrix} = \frac{1}{u}.$$

Por tanto la fdp conjunta de $W = XY$ y $U = X$ es

$$s(w, u) = g(u)h\left(\frac{w}{u}\right)\left|\frac{1}{u}\right|.$$

La fdp marginal de W se obtiene al integrar $s(w, u)$ con respecto a u , que da el resultado pedido. Los valores de w para los cuales $p(w) > 0$ depende de los valores (x, y) con los cuales $f(x, y) > 0$.

Observación: al evaluar la integral anterior se puede usar el hecho de que

$$\int_{-\infty}^{+\infty} g(u)h\left(\frac{w}{u}\right)\left|\frac{1}{u}\right| du = \int_0^{\infty} g(u)h\left(\frac{w}{u}\right)\frac{1}{u} du - \int_{-\infty}^0 g(u)h\left(\frac{w}{u}\right)\frac{1}{u} du.$$

EJEMPLO 6.14. Supóngase que tenemos un circuito en el cual varían de un modo aleatorio la corriente I y la resistencia R . Supóngase específicamente que I y R son variables aleatorias continuas independientes con las siguientes fdp.

$$I: \quad g(i) = 2i, \quad 0 \leq i \leq 1 \quad y 0 \text{ en otra parte};$$

$$R: \quad h(r) = r^2/9, \quad 0 \leq r \leq 3 \quad y 0 \text{ en otra parte}.$$

De interés es la variable aleatoria $E = IR$ (el voltaje en el circuito). Sea p la fdp de E . Según el teorema 6.4 tenemos

$$p(e) = \int_{-\infty}^{+\infty} g(i)h\left(\frac{e}{i}\right)\left|\frac{1}{i}\right| di.$$

Se debe tener cuidado al resolver esta integral. Nótese primero que la variable de integración no puede tomar valores negativos. Segundo, nótese con el fin de que el integrando sea positivo, *ambas* fdp que aparecen en el integrando deben ser positivas. Considerando los valores para los cuales g y h no son iguales a cero encontramos que deben satisfacerse las condiciones siguientes:

$$0 \leq i \leq 1 \quad \text{y} \quad 0 \leq e/i \leq 3.$$

Estas *dos* desigualdades son, a su vez, equivalentes a $e/3 \leq i \leq 1$. Por tanto, la integral anterior llega a ser

$$\begin{aligned} p(e) &= \int_{e/3}^1 2i \frac{e^2}{9i^2} \frac{1}{i} di \\ &= -\frac{2}{9}e^2 \frac{1}{i} \Big|_{e/3}^1 \\ &= \frac{2}{9}e(3-e), \quad 0 \leq e \leq 3. \end{aligned}$$

Un cálculo fácil demuestra que $\int_0^3 p(e)de = 1$.
(Ver figura 6.12.)

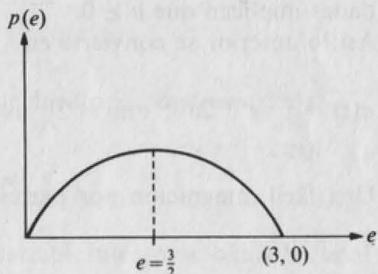


FIGURA 6.12

Teorema 6.5. Sea (X, Y) una variable aleatoria continua bidimensional y se supone que X y Y son independientes. [Por tanto, la fdp de (X, Y) puede escribirse como $f(x, y) = g(x)h(y)$.] Sea $Z = X/Y$. Luego la fdp de Z , digamos q , está dada por

$$q(z) = \int_{-\infty}^{+\infty} g(vz)h(v)|v| dv. \quad (6.10)$$

Demostración: sea $z = x/y$ y sea $v = y$. Por tanto, $x = vz$ y $y = v$. El Jacobiano es

$$J = \begin{vmatrix} v & z \\ 0 & 1 \end{vmatrix} = v.$$

Por tanto, la fdp conjunta de $Z = X/Y$ y $V = Y$ es igual a

$$t(z, v) = g(vz)h(v)|v|.$$

Integrando esta fdp conjunta con respecto a v , da la fdp marginal de Z .

EJEMPLO 6.15. Representemos por X y Y la duración de dos bombillas fabricadas mediante dos procedimientos distintos. Supongamos que X y Y son variables aleatorias independientes con fdp f y g respectivamente, en donde

$$f(x) = e^{-x}, \quad x \geq 0, \quad y 0 \text{ en otra parte};$$

$$g(y) = 2e^{-2y}, \quad y \geq 0, \quad y 0 \text{ en otra parte.}$$

De interés podría ser la variable aleatoria X/Y , que representa la razón de las dos duraciones. Sea q la fdp de Z .

Según el teorema 6.5 tenemos $q(z) = \int_{-\infty}^{+\infty} g(vz)h(v)|v| dv$. Puesto que X y Y pueden tomar sólo cantidades no negativas, la integración anterior sólo necesita ser calculada sobre los valores positivos de la variable de integración. Además, el integrando será positivo sólo cuando las dos fdp que aparecen sean positivas. Esto implica que debemos tener $v \geq 0$ y $vz \geq 0$. Puesto que $z > 0$, estas desigualdades implican que $v \geq 0$.

Así lo anterior se convierte en

$$q(z) = \int_0^{\infty} e^{-vz} 2e^{-2v} v dv = 2 \int_0^{\infty} ve^{-v(2+z)} dv.$$

Una fácil integración por partes da

$$q(z) = \frac{2}{(z+2)^2}, \quad z \geq 0.$$

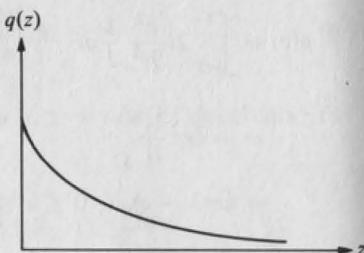


FIGURA 6.13

(Ver figura 6.13). De nuevo es un ejercicio fácil verificar que $\int_0^{\infty} q(z) dz = 1$.

6.6 Variables aleatorias n-dimensionales

Hasta ahora nuestra exposición se ha restringido a variables aleatorias bidimensionales. Sin embargo, como lo indicamos al comienzo de este capítulo, podemos interesarnos en tres o más características numéricas simultáneas.

Sólo haremos una brevíssima exposición de las variables aleatorias n -dimensionales. La mayor parte de los conceptos presentados anteriormente para el caso bidimensional se pueden extender al caso n -dimensional. Nos limitaremos al caso continuo. (Ver la nota al final de este capítulo).

Supongamos que (X_1, \dots, X_n) puede tomar todos los valores en una región del espacio n -dimensional. Esto es, el valor es un vector n -dimensional.

$$(X_1(s), \dots, X_n(s)).$$

Caracterizamos la distribución de probabilidades (X_1, \dots, X_n) como sigue: existe una función densidad de probabilidades conjunta que satisface las condiciones siguientes:

- $f(x_1, \dots, x_n) \geq 0$ para todo (x_1, \dots, x_n) .
- $\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f(x_1, \dots, x_n) dx_1 \cdots dx_n = 1$.

Con la ayuda de esta fdp *definimos*

$$P[(X_1, \dots, X_n) \in C] = \int_C \cdots \int f(x_1, \dots, x_n) dx_1 \cdots dx_n,$$

en donde C es un subconjunto del recorrido de (X_1, \dots, X_n) .

Con cada variable aleatoria n -dimensional podemos asociar un número de variables aleatorias de dimensión inferior. Por ejemplo, si $n = 3$, entonces

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x_1, x_2, x_3) dx_1 dx_2 = g(x_3),$$

en donde g es la fdp marginal de una variable aleatoria unidimensional X_3 , mientras que

$$\int_{-\infty}^{+\infty} f(x_1, x_2, x_3) dx_3 = h(x_1, x_2),$$

en donde h representa la fdp conjunta de la variable aleatoria bidimensional (X_1, X_2) , etc. El concepto de variables aleatorias independientes está también extendido de una manera natural. Decimos que (X_1, \dots, X_n) son variables aleatorias independientes si y sólo si sus fdp conjuntas $f(x_1, \dots, x_n)$ se pueden factorizar en

$$g_1(x_1) \cdots g_n(x_n).$$

Hay muchos casos en los cuales deseamos considerar las variables aleatorias n -dimensionales. Daremos unos pocos ejemplos.

(a) Supóngase que estudiamos el modelo de precipitación debido a un sistema particular de tormentas. Si tenemos una red de, digamos, 5 estaciones de observación y si X_i es la lluvia caída en la estación i debido a un sistema frontal particular, queríramos considerar la variable penta-dimensional $(X_1, X_2, X_3, X_4, X_5)$.

(b) Una de las aplicaciones más importantes de las variables aleatorias n -dimensionales aparece cuando tratamos con medidas repetidas de una variable aleatoria X . Supóngase que se pide la información acerca de la duración X , de cierto tubo electrónico. Un fabricante produce un gran número de estos tubos, de los cuales probamos n . Sea X_i la duración del i -ésimo tubo, $i = 1, \dots, n$. Por tanto (X_1, \dots, X_n) es una variable aleatoria n -dimensional. Si suponemos que cada X_i tiene la misma distribución de probabilidades (puesto que todos los tubos

se producen de la misma manera) y si suponemos que las X_i son todas variables aleatorias independientes (puesto que, posiblemente, la producción de un tubo no afecta la producción de los otros), podemos suponer que la variable aleatoria n -dimensional (X_1, \dots, X_n) está formada de las componentes X_1, \dots, X_n independientes e idénticamente distribuidas. (Debe ser obvio que aunque X_1 y X_2 tienen la misma distribución, no necesitan adoptar el mismo valor).

(c) Otra manera como aparecen las variables aleatorias n -dimensionales es la siguiente. Representemos por $X(t)$ la potencia requerida por cierta empresa industrial durante un tiempo t . Para t fijo, $X(t)$ es una variable aleatoria unidimensional. Sin embargo, podemos estar interesados en establecer la potencia pedida en determinados n tiempos específicos, $t_1 < t_2 < \dots < t_n$. Así deseamos estudiar la variable aleatoria n -dimensional.

$$[X(t_1), X(t_2), \dots, X(t_n)].$$

Este tipo de problemas se estudia a un nivel más avanzado. (Una referencia excelente para este tema es el libro *Stochastic Processes* por Emanuel Parzen. San Francisco: Holden-Day, 1962.)

Observación: en alguna parte de nuestra exposición nos hemos referido al concepto de «espacio- n .» Resumamos un poco algunas de las ideas básicas necesarias.

Con cada número real x , podemos asociar un punto sobre la recta de los números reales y reciprocamente. Análogamente, con cada par de números reales (x_1, x_2) , podemos asociar un punto en el plano rectangular de coordenadas y reciprocamente. Finalmente, con cada conjunto de tres números reales (x_1, x_2, x_3) , podemos asociar un punto en el espacio tridimensional de coordenadas y reciprocamente.

En muchos de los problemas que nos interesan, tratamos un conjunto de n números reales, (x_1, x_2, \dots, x_n) , también llamado un n -tuple. Aunque no podemos dibujar ningún gráfico, si $n > 3$, podemos continuar adoptando la terminología geométrica sugerida por los casos de dimensión menor, citados anteriormente. Así, hablaremos de un «punto» en un espacio n -dimensional determinado por el n -tuple (x_1, \dots, x_n) . Definiremos como espacio n (algunas veces llamado espacio n euclíadiano) el conjunto de todos los (x_1, \dots, x_n) , en donde x_i puede ser cualquier número real.

Aunque realmente no necesitaremos evaluar integrales n -dimensionales, vamos a encontrar que es un concepto muy útil y algunas veces necesitamos expresar una cantidad como una integral múltiple.

Si recordamos la definición de

$$\iint_A f(x, y) dx dy,$$

en donde A es una región del plano (x, y) , entonces la extensión de este concepto a

$$\int_R \cdots \int f(x_1, \dots, x_n) dx_1 \cdots dx_n,$$

en donde R es una región en el espacio n sería justificada. Si f representa la fdp conjunta de la variable aleatoria bidimensional (X, Y) entonces

$$\iint_A f(x, y) dx dy$$

representa $P[(X, Y) \in A]$. Análogamente, si f representa la fdp conjunta de (X_1, \dots, X_n) entonces

$$\int_R \cdots \int f(x_1, \dots, x_n) dx_1 \cdots dx_n$$

representa

$$P[(X_1, \dots, X_n) \in R].$$

PROBLEMAS

6.1. Suponga que la tabla siguiente representa la distribución de probabilidades conjunta de la variable aleatoria discreta (X, Y) . Calcular todas las distribuciones marginales y condicionales.

\backslash	X	1	2	3
Y				
1	$\frac{1}{12}$	$\frac{1}{6}$	0	
2	0	$\frac{1}{9}$	$\frac{1}{5}$	
3	$\frac{1}{18}$	$\frac{1}{2}$	$\frac{2}{15}$	

6.2. Supóngase que la variable aleatoria bidimensional (X, Y) tenga fdp conjunta

$$f(x, y) = kx(x - y), \quad 0 < x < 2, \quad -x < y < x, \\ = 0, \quad \text{en otra parte.}$$

- (a) Evaluar la constante k .
- (b) Encontrar la fdp marginal de X .
- (c) Encontrar la fdp marginal de Y .

6.3. Supóngase que la fdp conjunta de la variable aleatoria bidimensional (X, Y) está dada por

$$f(x, y) = x^2 + \frac{xy}{3}, \quad 0 < x < 1, \quad 0 < y < 2, \\ = 0, \quad \text{en otra parte.}$$

Calcular lo siguiente.

- (a) $P(X > \frac{1}{2})$; (b) $P(Y < X)$; (c) $P(Y < \frac{1}{2} | X < \frac{1}{2})$.

6.4. Suponga que se sacan dos cartas al azar de una baraja de cartas. Sea X el número de ases obtenidos y sea Y el número de reinas obtenido.

- (a) Obtener la distribución de probabilidades conjunta de (X, Y) .
- (b) Obtener la distribución marginal de X y de Y .
- (c) Obtener la distribución condicional de X (dado Y) y de Y (dado X).

6.5. ¿Para qué valores de k es $f(x, y) = ke^{-(x+y)}$ una fdp conjunta de (X, Y) en la región $0 < x < 1, 0 < y < 1$?

6.6. Supóngase que la variable aleatoria continua bidimensional (X, Y) está distribuida uniformemente en el cuadrado cuyos vértices son $(1, 0)$, $(0, 1)$, $(-1, 0)$ y $(0, -1)$. Encontrar la fdp marginales de X y de Y .

6.7. Supóngase que las dimensiones, X y Y , de una plancha metálica rectangular se pueden considerar como variables aleatorias continuas independientes con las siguientes fdp.

$$\begin{aligned} X: \quad g(x) &= x - 1, \quad 1 < x \leq 2, \\ &= -x + 3, \quad 2 < x < 3, \\ &= 0, \quad \text{en otra parte.} \end{aligned}$$

$$\begin{aligned} Y: \quad h(y) &= \frac{1}{2}, \quad 2 < y < 4, \\ &= 0, \quad \text{en otra parte.} \end{aligned}$$

Encontrar la fdp del área de la plancha, $A = XY$.

6.8. Representemos por X la duración de un dispositivo electrónico y supongamos que X es una variable aleatoria continua con fdp

$$\begin{aligned} f(x) &= \frac{1000}{x^2}, \quad x > 1000, \\ &= 0, \quad \text{en otra parte.} \end{aligned}$$

Sean X_1 y X_2 dos determinaciones independientes de la anterior variable aleatoria X . (Es decir, supóngase que estamos probando la duración de dos de tales dispositivos). Encontrar la fdp de la variable aleatoria $Z = X_1/X_2$.

6.9. Obtener la distribución de probabilidades de las variables aleatorias V y W presentadas en la p. 95.

6.10. Demostrar el teorema 6.1.

6.11. La fuerza magnética H en un punto P , a X unidades de un alambre que transporta una corriente I , está dada por $H = 2I/X$. (Ver figura 6.14). Supóngase que P es un punto variable. Es decir, X es una variable aleatoria continua distribuida uniformemente en $(3, 5)$. Supóngase que la corriente I es también una variable aleatoria continua, distribuida uni-

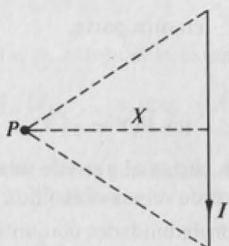


FIGURA 6.14

formemente en (10, 20). Supóngase, además, que las variables aleatorias X e I son independientes. Encontrar la fdp de la variable aleatoria H .

6.12. La intensidad de la luz en un punto dado está dada por la relación $I = C/D^2$, en donde C es la potencia luminosa de la fuente y D es la distancia de la fuente al punto dado. Supóngase que C está distribuida uniformemente en (1,2), mientras que D es una variable aleatoria continua, con fdp $f(d) = e^{-d}$, $d > 0$. Encontrar la fdp de I , si C y D son independientes. [Indicación: hallar primero la fdp de D^2 y luego aplicar los resultados de este capítulo.]

6.13. Cuando una corriente de I (ampères) pasa por una resistencia de R (ohms), la potencia generada es dada por $W = I^2R$ (watts). Supóngase que I y R son variables aleatorias independientes con las siguientes fdp.

$$\begin{aligned} I: \quad f(i) &= 6i(1 - i), \quad 0 \leq i \leq 1, \\ &= 0, \quad \text{en otra parte.} \end{aligned}$$

$$\begin{aligned} R: \quad g(r) &= 2r, \quad 0 < r < 1, \\ &= 0, \quad \text{en otra parte.} \end{aligned}$$

Determinar la fdp de la variable aleatoria W y dibujar su gráfico.

6.14. Supóngase que la fdp conjunta de (X, Y) es dada por

$$\begin{aligned} f(x, y) &= e^{-y}, \quad \text{para } x > 0, \quad y > x, \\ &= 0, \quad \text{en otra parte.} \end{aligned}$$

- (a) Encontrar la fdp marginal de X .
- (b) Encontrar la fdp marginal de Y .
- (c) Evaluar $P(X > 2 | Y < 4)$.

Otras características de las variables aleatorias

7.1 El valor esperado de una variable aleatoria

Consideremos la relación determinística $ax + by = 0$. Reconozcámola como una relación lineal entre x y y . Las constantes a y b son los *parámetros* de esta relación en el sentido de que para cualquier elección particular de a y b obtenemos una función lineal específica. En otros casos uno o más parámetros pueden representar la relación que se considera. Por ejemplo, si $y = ax^2 + bx + c$, son necesarios tres parámetros. Si $y = e^{-kx}$, es suficiente un parámetro. No sólo una relación particular se caracteriza por los parámetros sino, recíprocamente, por una cierta relación podemos definir varios parámetros pertinentes. Por ejemplo, si $ay + bx = 0$, entonces $m = -b/a$ representa la pendiente de la recta 0, si $y = ax^2 + bx + c$, entonces $-b/2a$ representa el valor para el cual existe un máximo relativo o un mínimo relativo.

En los modelos matemáticos no determinísticos o aleatorios que hemos venido considerando, los parámetros también pueden usarse para señalar la distribución de probabilidades. Con cada distribución de probabilidades podemos asociar ciertos parámetros que dan información valiosa acerca de la distribución (tal como la pendiente de una recta da una información útil acerca de la relación lineal que representa).

EJEMPLO 7.1. Supongamos que X es una variable aleatoria continua con fdp $f(x) = ke^{-kx}$, $x \geq 0$. Para verificar que esta es una fdp observe que $\int_0^\infty ke^{-kx} dx = 1$ para toda $k > 0$, y que $ke^{-kx} > 0$ para $k > 0$. Esta distribución se llama distribución exponencial, que estudiaremos más adelante con mucho detalle. Es una distribución muy útil para representar la duración, digamos X , de cierta clase de componentes de equipos. La interpretación de k , en este sentido, se discutirá también posteriormente.

EJEMPLO 7.2. Supongamos que en una línea indefinida de montaje se produce un artículo dado. La probabilidad de que un artículo sea defectuoso es p , y este valor es el mismo para todos los artículos. Suponiendo también que los artículos sucesivos son defectuosos (D) o no defectuosos (N), independientemente uno de otro. Sea la variable aleatoria X el número de artículos inspeccionados

hasta que se encuentra el primer artículo defectuoso. Así un resultado típico del experimento sería de la forma $NNNND$. Por tanto $X(NNNND) = 5$. Los valores posibles de X son: $1, 2, \dots, n, \dots$ Puesto que $X = k$ si y sólo si los primeros $(k - 1)$ artículos no son defectuosos y el k -ésimo artículo es defectuoso, asignamos la probabilidad siguiente al suceso $\{X = k\}$: $P(X = k) = p(1 - p)^{k-1}$, $k = 1, 2, \dots, n, \dots$ Para verificar que esta es una legítima distribución de probabilidades observemos que

$$\sum_{k=1}^{\infty} p(1 - p)^{k-1} = p[1 + (1 - p) + (1 - p)^2 + \dots]$$

$$= p \frac{1}{1 - (1 - p)} = 1 \quad \text{si } 0 < |p| < 1.$$

Así el parámetro p puede ser cualquier número que satisfaga $0 < p < 1$.

Supongamos que se especifican una variable aleatoria y su distribución de probabilidades. ¿Hay alguna manera de establecer esta distribución en función de escasos parámetros numéricos apropiados?

Antes de investigar la pregunta anterior, motivemos nuestra presentación al considerar el siguiente ejemplo.

EJEMPLO 7.3. Una máquina corta alambre de una longitud determinada. Debido a cierta imprecisión del mecanismo de corte, el largo del alambre cortado (en pulgadas), digamos X , se puede considerar como una variable aleatoria distribuida uniformemente en $[11,5, 12,5]$. El largo específico es de 12 pulgadas. Si $11,7 \leq X < 12,2$, el alambre se puede vender con una utilidad de \$0,25. Si $X \geq 12,2$, el alambre se puede cortar de nuevo, y por consiguiente se obtiene una utilidad de \$0,10. Y si $X < 11,7$, el alambre se descarta con una pérdida de \$0,02. Un cálculo sencillo indica que $P(X \geq 12,2) = 0,3$, $P(11,7 \leq X < 12,2) = 0,5$, y $P(X < 11,7) = 0,2$.

Supongamos que se corta un gran número de muestras de alambre, digamos N . Sea N_S el número de muestras para las cuales $X < 11,7$, N_R el número de muestras para las cuales $11,7 \leq X < 12,2$, y N_L el número de muestras para las cuales $X \geq 12,2$. Por tanto, la utilidad total obtenida de la producción de N muestras es igual $T = N_S(-0,02) + N_R(0,25) + N_L(0,10)$. La *utilidad total por alambre cortado*, digamos W , es igual $W = (N_S/N)(-0,02) + (N_R/N)(0,25) + (N_L/N)(0,1)$. (Observemos que W es una variable aleatoria, puesto que N_S , N_R , y N_L son variables aleatorias.)

Ya hemos mencionado que la frecuencia relativa de un suceso está próxima a la probabilidad de ese suceso si el número de repeticiones en las que se basa la frecuencia relativa es grande. (Discutiremos con más precisión esto en el capítulo 12.) Por tanto, si N es grande, esperaríamos que N_S/N fuera próximo a 0,2, N_R/N sea próximo a 0,5 y N_L/N sea próximo a 0,3. Luego para N grande, W podría aproximarse como sigue:

$$W \simeq (0,2)(-0,02) + 0,5(0,25) + (0,3)(0,1) = \$0,151.$$

Así, si se produjese un gran número de alambres, esperaríamos tener una utilidad de \$0,151 por alambre. El número 0,151 se llama valor esperado de una variable aleatoria X .

Definición. Sea X una variable aleatoria discreta con valores posibles x_1, \dots, x_n, \dots . Sea $p(x_i) = P(X = x_i), i = 1, 2, \dots, n, \dots$. El valor esperado de X (esperanza matemática de X), denotada por $E(X)$, se define como

$$E(X) = \sum_{i=1}^{\infty} x_i p(x_i) \quad (7.1)$$

si la serie $\sum_{i=1}^{\infty} x_i p(x_i)$ converge absolutamente, es decir, si $\sum_{i=1}^{\infty} |x_i| p(x_i) < \infty$. Este número también se designa como el valor promedio de X .

Observaciones: (a) Si X toma sólo un número finito de valores, la expresión anterior llega a ser $E(X) = \sum_{i=1}^n p(x_i)x_i$. Esta se puede considerar como un «promedio ponderado» de los valores posibles x_1, \dots, x_n . Si todos los valores posibles son igualmente probables, $E(X) = (1/n)\sum_{i=1}^n x_i$, lo que representa el promedio aritmético ordinario de los n valores posibles.

(b) Si se lanza un dado regular y la variable aleatoria X designa el número de puntos que salen, entonces $E(X) = 1/6(1 + 2 + 3 + 4 + 5 + 6) = 7/2$. Este ejemplo sencillo ilustra, categóricamente, que $E(X)$ no es el resultado que esperaríamos cuando X se observa una sola vez. De hecho, en la situación anterior, $E(X) = 7/2$ no es siquiera un valor *possible* de X ! Por el contrario, sucede que si obtenemos un gran número de observaciones independientes de X , tales como x_1, \dots, x_n , y calculamos el promedio aritmético de esos resultados, entonces, bajo condiciones generales regulares, el promedio aritmético estará cerca a $E(X)$ en un sentido probabilístico. Por ejemplo, en la situación anterior, si lanzáramos el dado un gran número de veces y luego calculáramos el promedio aritmético de los diversos resultados, esperaríamos que este promedio llegase a estar más próximo a $7/2$ cuanto más a menudo fuese lanzado el dado.

(c) Se debe observar la similitud entre la noción de valor esperado como se definió anteriormente (especialmente si X puede tomar sólo un número finito de valores) y la noción del promedio de un conjunto de números como, z_1, \dots, z_n . Corrientemente definimos $\bar{z} = 1/n \sum_{i=1}^n z_i$ como el promedio aritmético de los números z_1, \dots, z_n . Supóngase, además que tenemos los números z'_1, \dots, z'_k , donde z'_i ocurre n_i veces, $\sum_{i=1}^k n_i = n$. Haciendo $f_i = n_i/n$, $\sum_{i=1}^k f_i = 1$, definimos el promedio ponderado de los números z'_1, \dots, z'_k como

$$\frac{1}{n} \sum_{i=1}^k n_i z'_i = \sum_{i=1}^k f_i z'_i.$$

Aunque hay un fuerte parecido entre el promedio ponderado anterior y la definición de $E(X)$, es importante señalar que este último es un número (parámetro) asociado con una distribución de probabilidades teórica, mientras que el primero es simplemente el resultado de combinar un conjunto de números de una manera especial. Sin embargo, es más que un parecido superficial. Consideremos una variable aleatoria X y sean x_1, \dots, x_n los valores obtenidos cuando el experimento que da origen a X se realizó n veces independientes. (Es decir, x_1, \dots, x_n simple-

mente representa los resultados de n medidas repetidas de la característica numérica X .) Sea \bar{x} el promedio aritmético de esos n números. Entonces, como discutiremos mucho más precisamente en el capítulo 12, si n es suficientemente grande, \bar{x} estará «próxima» a $E(X)$ en un determinado sentido. Este resultado está muy relacionado con la idea (que también será discutida en el capítulo 12) de que la frecuencia relativa f_A asociada con n repeticiones de un experimento estará próxima a la probabilidad $P(A)$ si f_A está basada en un gran número de repeticiones de ε .

EJEMPLO 7.4. Un fabricante produce artículos de tal modo que el 10 por ciento son defectuosos y el 90 por ciento no lo son. Si se produce un artículo defectuoso, el fabricante pierde \$1, mientras que un artículo sin defectos le produce una utilidad de \$5. Si X es la utilidad neta por artículo, entonces X es una variable aleatoria cuyo valor esperado es calculado como $E(X) = -1(0,1) + 5(0,9) = \$4,40$. Supongamos que se produce un gran número de tales artículos. Entonces, puesto que el fabricante perderá \$1 alrededor del 10 por ciento de las veces y ganará \$5 alrededor del 90 por ciento de las veces, él esperará ganar alrededor de \$4,40 por artículo a la larga.

Teorema 7.1. Sea X una variable aleatoria distribuida binomialmente con parámetro p , basada en n repeticiones de un experimento. Entonces

$$E(X) = np.$$

Demostración: puesto que $P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$, tenemos

$$\begin{aligned} E(X) &= \sum_{k=0}^n k \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \\ &= \sum_{k=1}^n \frac{n!}{(k-1)!(n-k)!} p^k (1-p)^{n-k} \end{aligned}$$

(ya que el término con $k = 0$ es igual a cero). Sea $s = k - 1$ en la suma anterior. Como k toma valores de uno hasta n , s toma valores de cero hasta $(n-1)$. Sustituyendo k en todas partes por $(s+1)$ obtenemos

$$\begin{aligned} E(X) &= \sum_{s=0}^{n-1} n \binom{n-1}{s} p^{s+1} (1-p)^{n-s-1} \\ &= np \sum_{s=0}^{n-1} \binom{n-1}{s} p^s (1-p)^{n-1-s}. \end{aligned}$$

La suma en la última expresión es simplemente la suma de probabilidades binomiales con n sustituido por $(n-1)$ [esto es $(p + (1-p))^{n-1}$] que, por tanto, es igual a uno. Esto establece el resultado.

Observación: el resultado anterior corresponde ciertamente a nuestra noción intuitiva. Porque se supone que la probabilidad de algún suceso A es 0,3, por ejemplo, cuando se rea-

liza un experimento. Si repetimos este experimento 100 veces, por ejemplo, esperaríamos que A ocurriera alrededor de $100(0,3) = 30$ veces. El concepto de valor esperado, presentado anteriormente para variables discretas, se extenderá brevemente en el caso continuo.

EJEMPLO 7.5. Una máquina impresora tiene una probabilidad constante de 0,05 de fallar en cualquier día. Si la máquina no tiene fallas durante la semana, se obtiene una utilidad de S . Si ocurren 1 o 2 fallas, se obtiene una utilidad de R ($R < S$). Si ocurren 3 o más fallas, se obtiene una utilidad de $(-L)$. (Suponemos que R, S , y L son mayores que cero; también suponemos que si la máquina falla cualquier día, permanece fuera de uso durante el resto del día.) Sea X la utilidad obtenida en una semana de cinco días. Los valores posibles de X son R, S y $(-L)$. Sea B el número de fallas por semana. Tenemos

$$P(B = k) = \binom{5}{k} (0,05)^k (0,95)^{5-k}, \quad k = 0, 1, \dots, 5$$

Puesto que $X = S$ si y sólo si $B = 0$, $X = R$ si y sólo si $B = 1$ ó 2, y $X = (-L)$ si y sólo si $B = 3, 4$, ó 5, encontramos que,

$$\begin{aligned} E(X) &= SP(B = 0) + RP(B = 1 \text{ ó } 2) + (-L)P(B = 3, 4 \text{ ó } 5) \\ &= S(0,95)^5 + R[5(0,05)(0,95)^4 + 10(0,05)^2(0,95)^3] \\ &\quad + (-L)[10(0,05)^3(0,95)^2 + 5(0,05)^4(0,95) + (0,05)^5] \text{ dólares.} \end{aligned}$$

Definición. Sea X una variable aleatoria continua con fdp f . El *valor esperado* de X se define como

$$E(X) = \int_{-\infty}^{+\infty} xf(x)dx \tag{7.2}$$

Nuevamente puede suceder que esta integral (impropia) no converja. Por lo tanto decimos que $E(X)$ existe si y sólo si

$$\int_{-\infty}^{\infty} |x|f(x)dx$$

es finita.

Observación: debemos observar la analogía entre el valor esperado de una variable aleatoria y el concepto de «centro de masa» en mecánica. Si una masa unitaria está distribuida a lo largo de la recta en los puntos discretos x_1, \dots, x_n, \dots y si $p(x_i)$ es la masa en x_i , vemos entonces que $\sum_{i=1}^{\infty} x_i p(x_i)$ representa el centro de masa (respecto al origen). De modo semejante, si una masa unitaria está distribuida continuamente en una recta, y si $f(x)$ representa la densidad de masa en x , entonces $\int_{-\infty}^{+\infty} xf(x)dx$ se puede interpretar nuevamente como el centro de masa. En el sentido anterior, $E(X)$ puede representar «un centro» de la distribución de probabilidad. También, $E(X)$ se llama algunas veces medida de tendencia central y está en las *mismas unidades* que X .

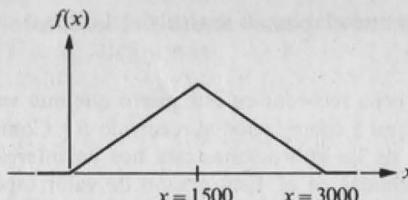


FIGURA 7.1

EJEMPLO 7.6. Vamos a definir la variable aleatoria X como sigue. Supongamos que X es el tiempo (en minutos) durante el cual un dispositivo eléctrico se utiliza a su máxima carga cierto período de tiempo determinado. Supongamos que X es una variable aleatoria continua con la siguiente fdp:

$$\begin{aligned}f(x) &= \frac{1}{(1500)^2} x, \quad 0 \leq x \leq 1500, \\&= \frac{-1}{(1500)^2} (x - 3000), \quad 1500 \leq x \leq 3000, \\&= 0, \quad \text{para cualquier otro valor.}\end{aligned}$$

Así

$$\begin{aligned}E(X) &= \int_{-\infty}^{+\infty} xf(x)dx \\&= \frac{1}{(1500)(1500)} \left[\int_0^{1500} x^2 dx - \int_{1500}^{3000} x(x - 3000) dx \right] \\&= 1500 \text{ minutos.}\end{aligned}$$

(Ver figura 7.1.)

EJEMPLO 7.7. El contenido de ceniza en el carbón (porcentaje), digamos X , se puede considerar como una variable aleatoria continua con la siguiente fdp: $f(x) = \frac{1}{4875} x^2$, $10 \leq x \leq 25$. Por lo tanto $E(X) = \frac{1}{4875} \int_{10}^{25} x^3 dx = 19,5$ por ciento. Así el contenido esperado de ceniza en la muestra particular de carbón que se considera es 19,5 por ciento.

Teorema 7.2. Supongamos a X distribuida uniformemente en el intervalo $[a, b]$. Entonces

$$E(X) = \frac{a + b}{2}.$$

Demostración: la fdp de X está dada por $f(x) = 1/(b - a)$, $a \leq x \leq b$. Por tanto

$$E(X) = \int_a^b \frac{x}{b - a} dx = \frac{1}{b - a} \frac{x^2}{2} \Big|_a^b = \frac{a + b}{2}.$$

(Nótese que esto representa el *punto medio* del intervalo $[a, b]$, como intuitivamente lo esperaríamos.)

Observación: valdría la pena recordar en este punto que una variable aleatoria X es una función de un espacio muestral S con relación al recorrido R_X . Como repetidamente lo hemos señalado, para la mayoría de las aplicaciones esta nos ha interesado sólo en el recorrido y en las probabilidades definidas en él. Esta noción de valor esperado fue completamente definida en función del recorrido [ver ecuaciones (7.1) y (7.2)]. Ocasionalmente, sin embargo, deberíamos observar la naturaleza funcional de X . Por ejemplo, ¿cómo expresamos la ecuación (7.1) en función de resultados $s \in S$, suponiendo que S es finito? Puesto que $x_i = X(s)$ para una $s \in S$ y puesto que

$$p(x_i) = P[s: X(s) = x_i],$$

podemos escribir

$$E(X) = \sum_{i=1}^n x_i p(x_i) = \sum_{s \in S} X(s)P(s), \quad (7.3)$$

en donde $P(s)$ es la probabilidad del suceso $\{s\} \subset S$. Por ejemplo, si el experimento consiste en clasificar tres artículos como defectuosos (D) o no defectuosos (N), un espacio muestral para este experimento sería

$$S = \{NNN, NND, NDN, DNN, NDD, DND, DDN, DDD\}.$$

Si X está definida como el número de defectuosos, y si se supone que todos los resultados anteriores son igualmente posibles, tenemos de acuerdo con la ecuación (7.3.)

$$\begin{aligned} E(X) &= \sum_{s \in S} X(s)P(s) \\ &= 0 \cdot \left(\frac{1}{8}\right) + 1 \cdot \left(\frac{1}{8}\right) + 1 \cdot \left(\frac{1}{8}\right) + 1 \cdot \left(\frac{1}{8}\right) + 2 \cdot \left(\frac{1}{8}\right) + 2 \cdot \left(\frac{1}{8}\right) + 3 \cdot \left(\frac{1}{8}\right) \\ &= \frac{3}{2}. \end{aligned}$$

Se habría podido obtener este resultado más fácilmente al aplicar directamente la ecuación (7.1). Sin embargo, es bueno recordar que para emplear la ecuación (7.1) necesitábamos conocer los valores $p(x_i)$, lo que a su vez significaba la necesidad de un cálculo tal como el utilizado anteriormente. Lo importante es que una vez que se conoce la distribución de probabilidades sobre R_X [en este caso los valores de $p(x_i)$], podemos suprimir la relación funcional entre R_X y S .

7.2 Esperanza de una función de una variable aleatoria

Como lo expusimos previamente, si X es una variable aleatoria y si $Y = H(X)$, es una función de X , entonces Y es también una variable aleatoria con una distribución de probabilidades. Por lo tanto será interesante y significativo evaluar $E(Y)$. Hay dos maneras de evaluar $E(Y)$ que resultan equivalentes. Demostrar que en general son equivalentes no es trivial y probaremos sólo un caso especial. Sin embargo, es importante que el lector comprenda los dos enfoques presentados a continuación.

Definición. Sea X una variable aleatoria y sea $Y = H(X)$.

(a) Si Y es una variable aleatoria discreta con valores posibles y_1, y_2, \dots y si $q(y_i) = P(Y = y_i)$, definimos

$$E(Y) = \sum_{i=1}^{\infty} y_i q(y_i). \quad (7.4)$$

(b) Si Y es una variable aleatoria continua con fdp g , definimos

$$E(Y) = \int_{-\infty}^{+\infty} yg(y)dy. \quad (7.5)$$

Observación: naturalmente estas definiciones son enteramente consistentes con la definición previa dada para el valor esperado de una variable aleatoria. De hecho, lo anterior representa simplemente una repetición en función de Y . Una «desventaja» de aplicar la definición anterior a fin de obtener $E(Y)$ es que la distribución de probabilidades de Y (es decir, la distribución de probabilidades en el recorrido R_Y) es necesaria. En los capítulos anteriores, expusimos métodos mediante los cuales podemos obtener o las probabilidades puntuales $q(y_i)$ o g , la fdp de Y . Sin embargo, el problema que aparece es si podemos obtener $E(Y)$ sin encontrar primero la distribución de probabilidades de Y , sólo a partir del conocimiento de la distribución de probabilidades de X . La respuesta es afirmativa como lo indica el siguiente teorema.

Teorema 7.3. Sea X una variable aleatoria y sea $Y = H(X)$.

(a) Si X es una variable aleatoria discreta y $p(x_i) = P(X = x_i)$, tenemos

$$E(Y) = E(H(X)) = \sum_{j=1}^{\infty} H(x_j)p(x_j). \quad (7.6)$$

(b) Si X es una variable aleatoria continua con fdp f , tenemos

$$E(Y) = E(H(X)) = \int_{-\infty}^{+\infty} H(x)f(x)dx. \quad (7.7)$$

Observación: este teorema hace mucho más sencilla la evaluación de $E(Y)$, porque quiere decir que no necesitamos encontrar la distribución de probabilidades de Y a fin de evaluar $E(Y)$. Es suficiente el conocimiento de la distribución de probabilidades de X .

Demostración: [sólo demostraremos la ecuación (7.6). La demostración de la ecuación (7.7) es algo más complicada]. Consideremos la suma $\sum_{j=1}^{\infty} H(x_j)p(x_j) = \sum_{j=1}^{\infty} (\sum_i H(x_i)p(x_i))$, en donde la suma interior se toma sobre todos los índices i para los cuales $H(x_i) = y_j$, para un y_j fijo. Por tanto todos los términos $H(x_i)$ son constantes en la suma interior. Por tanto

$$\sum_{j=1}^{\infty} H(x_j)p(x_j) = \sum_{j=1}^{\infty} y_j \sum_i p(x_i).$$

Pero,

$$\sum_i p(x_i) = \sum_i P[x|H(x_i) = y_j] = q(y_j).$$

Por tanto, $\sum_{j=1}^{\infty} H(x_j)p(x_j) = \sum_{j=1}^{\infty} y_j q(y_j)$, lo que establece la ecuación (7.6).

Observación: el método de demostración es equivalente esencialmente con el método de contar en el cual ponemos juntos todos los artículos que tienen el mismo valor. Así, si queremos encontrar la suma total de los valores 1, 1, 2, 3, 5, 3, 2, 1, 2, 2, 3, podemos sumarlos directamente o indicar que puesto que hay 3 unos, 4 doses, 3 treses y 1 cinco, la suma total es igual

$$3(1) + 4(2) + 3(3) + 1(5) = 25.$$

EJEMPLO 7.8. Sea V la velocidad del viento (kph) y supongamos que V está distribuida uniformemente en el intervalo $[0, 10]$. La presión, W (en lb/pie²), sobre la superficie del ala de un aeroplano está dada por la relación: $W = 0,003V^2$. Para encontrar el valor esperado de W , $E(W)$, podemos proceder de dos maneras:

(a) Usando el teorema 7.3., tenemos

$$\begin{aligned} E(W) &= \int_0^{10} 0,003v^2 f(v) dv \\ &= \int_0^{10} 0,003v^2 \frac{1}{10} dv \\ &= 0,1 \text{ lb/pie}^2 \end{aligned}$$

(b) Usando la definición de $E(W)$, necesitamos encontrar primero la fdp de W , g , y luego evaluar $\int_{-\infty}^{+\infty} wg(w)dw$. Para encontrar $g(w)$, observemos que $w = 0,003v^2$ es una función monótona de v , para $v \geq 0$. Podemos aplicar el teorema 5.1 y obtenemos

$$\begin{aligned} g(w) &= \frac{1}{10} \left| \frac{dv}{dw} \right| \\ &= \frac{1}{2} \sqrt{\frac{10}{3}} w^{-1/2}, \quad 0 \leq w \leq 0,3, \\ &= 0, \quad \text{para cualquier otro valor.} \end{aligned}$$

Por tanto

$$E(W) = \int_0^{0,3} wg(w)dw = 0,1$$

después de un cálculo sencillo. Así, como lo indicó el teorema las dos evaluaciones de $E(W)$ producen el mismo resultado.

EJEMPLO 7.9. En muchos problemas nos interesa sólo la *magnitud* de una variable aleatoria sin consideración a su signo algebraico. Es decir, nos interesa $|X|$. Supongamos que X es una variable aleatoria continua con la siguiente fdp:

$$\begin{aligned}f(x) &= \frac{e^x}{2} && \text{si } x \leq 0, \\&= \frac{e^{-x}}{2} && \text{si } x > 0.\end{aligned}$$

Sea $Y = |X|$. Para obtener $E(Y)$ debemos proceder de una de las dos maneras.

(a) Usando el teorema 7.3, tenemos

$$\begin{aligned}E(Y) &= \int_{-\infty}^{+\infty} |x| f(x) dx \\&= \frac{1}{2} \left[\int_{-\infty}^0 (-x) e^x dx + \int_0^{\infty} (x) e^{-x} dx \right] \\&= \frac{1}{2} [1 + 1] = 1.\end{aligned}$$

(b) Para evaluar $E(Y)$ usando la definición, necesitamos obtener la fdp de $Y = |X|$, g . Sea G la fda de Y . Luego

$$G(y) = P(Y \leq y) = P[|X| \leq y] = P[-y \leq X \leq y] = 2P(0 \leq X \leq y),$$

puesto que la fdp de X es simétrica respecto al cero. Por lo tanto

$$G(y) = 2 \int_0^y f(x) dx = 2 \int_0^y \frac{e^{-x}}{2} dx = -e^{-y} + 1.$$

Así tenemos para g , la fdp de Y , $g(y) = G'(y) = e^{-y}$, $y \geq 0$. Por lo tanto $E(Y) = \int_0^{\infty} yg(y) dy = \int_0^{\infty} ye^{-y} dy = 1$, como antes.

EJEMPLO 7.10. En muchos problemas podemos usar el valor esperado de una variable aleatoria a fin de hacer ciertas decisiones de una manera óptima.

Supóngase que un fabricante produce cierto tipo de aceite lubricante que pierde alguno de sus atributos especiales si no se usa dentro de cierto período de tiempo. Sea X el número de unidades de aceite pedidas al fabricante durante cada año. (Una unidad es igual a 1000 galones.) Supongamos que X es una variable aleatoria continua, distribuida uniformemente en $[2, 4]$. Por lo tanto la fdp f tiene la forma,

$$\begin{aligned}f(x) &= \frac{1}{2}, && 2 \leq x \leq 4, \\&= 0, && \text{para cualquier otro valor.}\end{aligned}$$

Supongamos que por cada una de las unidades vendidas se obtiene una utilidad de \$300, mientras que cada una de las unidades no vendidas (durante un año determinado) produce una pérdida de \$100, ya que una unidad no utilizada tendrá que ser descartada. Suponiendo que el fabricante debe decidir pocos meses antes del comienzo de cada año cuánto producirá, y que decide fabricar Y unidades. (Y no es una variable aleatoria, está especificada por el fabricante.) Sea Z la utilidad por año (en pesos). Aquí Z es evidentemente una variable aleatoria puesto que es una función de la variable aleatoria X . Específicamente, $Z = H(X)$ en donde

$$\begin{aligned}H(X) &= 300 Y && \text{si } X \geq Y, \\&= 300 X + (-100)(Y - X), && \text{si } X < Y.\end{aligned}$$

(La última expresión puede escribirse como $400X - 100Y$).

A fin de obtener $E(Z)$ aplicaremos el teorema 7.3 y escribimos

$$\begin{aligned}E(Z) &= \int_{-\infty}^{+\infty} H(x)f(x)dx \\&= \frac{1}{2} \int_2^4 H(x)dx.\end{aligned}$$

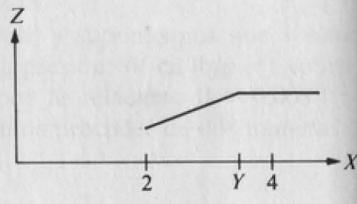


FIGURA 7.2

Para evaluar esta integral debemos considerar tres casos: $Y < 2$, $2 \leq Y \leq 4$, y $Y > 4$. Con ayuda de la figura 7.2 y después de algunas simplificaciones obtenemos

$$\begin{aligned}E(Z) &= 300 Y && \text{si } Y \leq 2 \\&= -100 Y^2 + 700 Y - 400 && \text{si } 2 < Y < 4 \\&= 1200 - 100 Y && \text{si } Y \geq 4.\end{aligned}$$

La pregunta siguiente es de interés. ¿Cómo elegiría el fabricante el valor de Y a fin de maximizar la utilidad esperada? Podemos responder fácilmente esta pregunta al poner simplemente $dE(Z)/dY = 0$. Esto produce $Y = 3,5$. (Ver figura 7.3.)

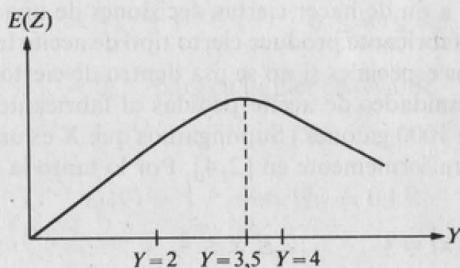


FIGURA 7.3

7.3 Variables aleatorias bidimensionales

Los conceptos discutidos anteriormente para el caso unidimensional también se mantienen para el caso de variables aleatorias de mayor dimensión. En particular para el caso bidimensional, hacemos la siguiente definición.

Definición. Sea (X, Y) una variable aleatoria bidimensional y sea $Z = H(X, Y)$ una función real de (X, Y) . Por lo tanto Z es una variable aleatoria (unidimensional) y definimos $E(Z)$ como sigue:

- (a) Si Z es una variable aleatoria discreta con valores posibles z_1, z_2, \dots y con

$$p(z_i) = P(Z = z_i),$$

entonces

$$E(Z) = \sum_{i=1}^{\infty} z_i p(z_i). \quad (7.8)$$

- (b) Si Z es una variable aleatoria continua con fdp f , tenemos

$$E(Z) = \int_{-\infty}^{+\infty} zf(z)dz. \quad (7.9)$$

Como en el caso unidimensional, se puede demostrar el teorema siguiente (análogo al teorema 7.3.)

Teorema 7.4. Sea (X, Y) una variable aleatoria bidimensional y sea $Z = H(X, Y)$.

- (a) Si (X, Y) es una variable aleatoria discreta y si

$$p(x_i, y_j) = P(X = x_i, Y = y_j), \quad i, j = 1, 2, \dots,$$

tenemos

$$E(Z) = \sum_{j=1}^{\infty} \sum_{i=1}^{\infty} H(x_i, y_j)p(x_i, y_j). \quad (7.10)$$

- (b) Si (X, Y) es una variable aleatoria continua con fdp conjunta f , tenemos

$$E(Z) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} H(x, y)f(x, y)dxdy. \quad (7.11)$$

Observación: no demostraremos el teorema 7.4. Otra vez, tal como en el caso unidimensional, este es un resultado muy útil puesto que indica que *no* necesitamos encontrar la distribución de probabilidades de la variable aleatoria Z a fin de evaluar su esperanza. Podemos encontrar directamente $E(Z)$ a partir del conocimiento de la distribución conjunta de (X, Y) .

EJEMPLO 7.11. Reconsideremos el ejemplo 6.14 y encontremos $E(E)$ en donde $E = IR$. Encontramos que I y R son variables aleatorias independientes con las siguientes fdp g y h respectivamente:

$$g(i) = 2i, \quad 0 \leq i \leq 1; \quad h(r) = r^2/9, \quad 0 \leq r \leq 3.$$

También encontramos que la fdp de E es $p(e) = \frac{2}{9}e(3 - e)$, $0 \leq e \leq 3$. Puesto que I y R son variables aleatorias independientes, la fdp conjunta de (I, R) es sencillamente el producto de las fdp de I y R : $f(i, r) = \frac{2}{9}ir^2$, $0 \leq i \leq 1$, $0 \leq r \leq 3$. Para evaluar $E(E)$ usando el teorema 7.4 tenemos

$$\begin{aligned} E(E) &= \int_0^3 \int_0^1 irf(i, r)di dr = \int_0^3 \int_0^1 ir^2 \cdot \frac{2}{9}ir^2 di dr \\ &= \frac{2}{9} \int_0^1 i^2 di \int_0^3 r^3 dr = \frac{3}{2}. \end{aligned}$$

Usando directamente la definición (7.9), tenemos

$$\begin{aligned} E(E) &= \int_0^3 ep(e)de = \int_0^3 e \cdot \frac{2}{9}e(3 - e)de \\ &= \frac{2}{9} \int_0^3 (3e^2 - e^3)de = \frac{3}{2}. \end{aligned}$$

7.4 Propiedades del valor esperado

Haremos una lista de propiedades del valor esperado de una variable aleatoria que será muy útil en el trabajo futuro. En cada caso supondremos que existen todos los valores esperados a los cuales nos referimos. Daremos las demostraciones sólo para el caso continuo. El lector debe ser capaz de dar el argumento para el caso discreto sustituyendo sencillamente las integrales por sumatorias.

Propiedad 7.1. Si $X = C$ en donde C es una constante, entonces $E(X) = C$.

Demostración

$$E(X) = \int_{-\infty}^{+\infty} Cf(x)dx = C \int_{-\infty}^{+\infty} f(x)dx = C.$$

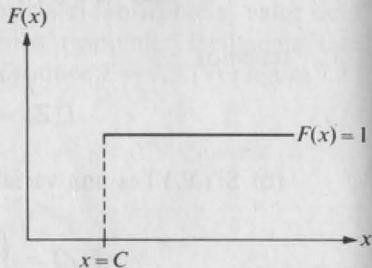


FIGURA 7.4

Observación: el significado de X igual C es el siguiente. Puesto que X es una función del espacio muestral a R_X , el significado de lo anterior es que R_X consta de un solo valor C . Por tanto X es igual a C si y sólo si $P[X(s) = C] = 1$. Esta noción se explica mejor en función de la fda de X . Llamada, $F(x) = 0$, si $x < C$; $F(x)$ es igual a 1, si $x \geq C$ (figura 7.4). Tal variable aleatoria se llama *degenerada* algunas veces.

Propiedad 7.2. Supongamos que C es una constante y X es una variable aleatoria. Entonces $E(CX) = CE(X)$.

$$\text{Demostración: } E(CX) = \int_{-\infty}^{+\infty} Cx f(x) dx = C \int_{-\infty}^{+\infty} x f(x) dx = CE(X).$$

Propiedad 7.3. Sea (X, Y) una variable aleatoria bidimensional con una distribución de probabilidades conjunta. Sean $Z = H_1(X, Y)$ y $W = H_2(X, Y)$. Entonces $E(Z + W) = E(Z) + E(W)$.

Demostración

$$\begin{aligned} E(Z + W) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} [H_1(x, y) + H_2(x, y)] f(x, y) dx dy \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} H_1(x, y) f(x, y) dx dy + \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} H_2(x, y) f(x, y) dx dy \\ &= E(Z) + E(W). \end{aligned} \quad [\text{en donde } f \text{ es la fdp conjunta de } (X, Y)]$$

Propiedad 7.4. Sean X y Y dos variables aleatorias cualquiera. Entonces $E(X + Y) = E(X) + E(Y)$.

Demostración: esta se deduce de inmediato de la propiedad 7.3 al hacer $H_1(X, Y) = X$, y $H_2(X, Y) = Y$.

Observaciones: (a) Combinando las propiedades 7.1, 7.2 y 7.4 observamos el siguiente hecho importante: si $Y = aX + b$, en donde a y b son constantes, entonces $E(Y) = aE(X) + b$. En palabras: la esperanza de una función lineal es esa misma función lineal de las esperanzas. Esto *no* es cierto a menos que esté implicada una función lineal y es un error común creer de otro modo. Por ejemplo, $E(X^2) \neq (E(X))^2$, $E(\ln X) \neq \ln E(X)$, etc. Así si X toma los valores -1 y $+1$, cada uno con probabilidad $\frac{1}{2}$, entonces $E(X) = 0$. Sin embargo,

$$E(X^2) = (-1)^2 \left(\frac{1}{2}\right) + (1)^2 \left(\frac{1}{2}\right) = 1 \neq 0^2.$$

(b) En general, es difícil obtener expresiones para $E(1/X)$ o $E(X^{1/2})$, por ejemplo, en función de $1/E(X)$ o $(E(X))^{1/2}$. Sin embargo, están disponibles algunas desigualdades, que son muy fáciles de derivar. (Ver los artículos de Fleiss, Murthy y Pillai, y Gurland en los números de febrero y diciembre de 1966 y abril de 1967, respectivamente, de *The American Statistician*.)

Por ejemplo, tenemos:

- (1) Si X toma solo valores positivos y tiene una esperanza finita, entonces $E(1/X) \geq 1/E(X)$.
- (2) Bajo la misma hipótesis que en (1), $E(X^{1/2}) \leq (E(X))^{1/2}$.

Propiedad 7.5. Sean X_1, \dots, X_n variables aleatorias. Entonces

$$E(X_1 + \dots + X_n) = E(X_1) + \dots + E(X_n).$$

Demostración: esta se deduce inmediatamente de la propiedad 7.4 al aplicar la inducción matemática.

Observación: combinando esta propiedad con la anterior, obtenemos

$$E\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n a_i E(X_i),$$

en donde las a_i son constantes.

Propiedad 7.6. Sea (X, Y) una variable aleatoria bidimensional y supongamos que X y Y son independientes. Entonces $E(XY) = E(X)E(Y)$.

Demostración:

$$\begin{aligned} E(XY) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xyf(x, y)dxdy \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xyg(x)h(y)dxdy \\ &= \int_{-\infty}^{+\infty} xg(x)dx \int_{-\infty}^{+\infty} yh(y)dy = E(X)E(Y). \end{aligned}$$

Observación: la hipótesis adicional de independencia es necesaria para establecer la propiedad 7.6, mientras que no se necesitó ninguna suposición para obtener la propiedad 7.4.

EJEMPLO 7.12. (Este ejemplo se basa en un problema de *An Introduction to Probability Theory and Its Applications* de W. Feller, página 225.)

Supóngase que necesitamos examinar un gran número de personas, buscando cierta característica, con resultados positivo o negativo. Más aún, supongamos que se toman muestras de varias personas y se prueba la muestra combinada como una unidad, tal como puede ser el caso en ciertos tipos de exámenes de sangre.

Suposición: la muestra combinada dará un resultado negativo si y sólo si todas las muestras componentes son negativas.

Así, en el caso de resultados positivos (de la muestra combinada), *todas* las muestras deben ser individualmente probadas de nuevo para determinar cuáles son positivas. Si las N personas son divididas en n grupos de k personas (suponiendo $N = kn$) aparecen las siguientes elecciones:

- (a) Examinar a todas las N personas individualmente, solicitando N pruebas.
- (b) Examinar grupos de k muestras, lo cual puede necesitar tan pocos como $n = N/k$ o tantos como $(k+1)n = N + n$ pruebas.

Nuestro objetivo será estudiar el número esperado de pruebas necesarias en (b) y luego compararlas con N .

Suposición: la probabilidad de que el resultado de la prueba sea positivo es

igual a p y es la misma para todas las personas. Aún más, los resultados de las pruebas para las personas del mismo grupo que se examina son independientes. Sea X = número de pruebas necesarias para determinar la característica que se estudia para todas las N personas, y sea X_i = número de pruebas necesarias para examinar personas en el i -ésimo grupo, $i = 1, \dots, n$.

Luego $X = X_1 + \dots + X_n$, y por lo tanto $E(X) = E(X_1) + \dots + E(X_n)$, lo cual es igual a $nE(X_1)$, puesto que todos los X_i tienen la misma esperanza. X_1 sólo toma dos valores: 1 y $k + 1$. Además,

$$\begin{aligned} P(X_1 = 1) &= P(\text{todas las } k \text{ personas en el grupo 1 son negativas}). \\ &= (1 - p)^k. \end{aligned}$$

Por tanto

$$P(X_1 = k + 1) = 1 - (1 - p)^k$$

y luego

$$\begin{aligned} E(X_1) &= 1 \cdot (1 - p)^k + (k + 1)[1 - (1 - p)^k] \\ &= k[1 - (1 - p)^k + k^{-1}]. \end{aligned}$$

Así

$$E(X) = nE(X_1) = N[1 - (1 - p)^k + k^{-1}].$$

(La fórmula anterior es válida sólo para $k > 1$, puesto que para $k = 1$ da $E(X) = N + pn$, lo cual obviamente es falso.)

Un asunto de interés es la elección de k para el cual el anterior $E(X)$ es más pequeño. Esto puede ser fácilmente manejado por algún procedimiento numérico. (Ver problema 7.11a.)

Finalmente, observemos que para que la «prueba en grupo» sea preferible a la prueba individual, deberíamos tener $E(X) < N$, esto es, $1 - (1 - p)^k + k^{-1} < 1$, lo cual es equivalente a $k^{-1} < (1 - p)^k$. Esto no puede ocurrir si $(1 - p) < \frac{1}{2}$. Porque en ese caso $(1 - p)^k < \frac{1}{2}^k < 1/k$, la última desigualdad proviene del hecho de que $2^k > k$. Así obtenemos la siguiente conclusión importante: si p , la probabilidad de una prueba positiva en cualquier persona determinada, es mayor que $\frac{1}{2}$, entonces no es aconsejable agrupar las muestras antes de examinar. (Ver problema 7.11b.)

EJEMPLO 7.13. Aplicemos algunas de las propiedades anteriores para derivar (nuevamente) la esperanza de una variable aleatoria distribuida binomialmente. El método usado puede aplicarse con ventaja en muchas situaciones semejantes.

Consideremos n repeticiones independientes y sea X el número de veces que ocurre un suceso A . Sea p igual a $P(A)$ y supongamos que este número es constante para todas las repeticiones consideradas.

Definamos las variables auxiliares Y_1, \dots, Y_n como sigue:

$$\begin{aligned} Y_i &= 1 && \text{si el suceso } A \text{ ocurre en la } i\text{-ésima repetición,} \\ &= 0, && \text{en cualquier otro caso.} \end{aligned}$$

Por lo tanto

$$X = Y_1 + Y_2 + \cdots + Y_n,$$

y aplicando la propiedad 7.5, obtenemos

$$E(X) = E(Y_1) + \cdots + E(Y_n).$$

Sin embargo,

$$E(Y_i) = 1(p) + 0(1 - p) = p, \quad \text{para todo } i.$$

Así $E(X) = np$, lo que concuerda con el resultado previo.

Observación: reinterpretamos este importante resultado. Consideremos la variable aleatoria X/n . Esta representa la frecuencia relativa del suceso A entre las n repeticiones de ε . Usando la propiedad 7.2, tenemos $E(X/n) = (np)/n = p$. Esto, intuitivamente, es como debería ser, porque expresa que la frecuencia relativa esperada del suceso A es p , donde $p = P(A)$. Representa la primera verificación teórica del hecho de que hay una relación entre la frecuencia relativa de un suceso y la probabilidad de ese suceso. En un capítulo posterior, obtendremos más resultados que dan una relación mucho más precisa entre la frecuencia relativa y la probabilidad.

EJEMPLO 7.14. Supóngase que la demanda D , por semana, de cierto producto es una variable aleatoria con determinada distribución de probabilidades, $P(D = n) = p(n)$, $n = 0, 1, 2, \dots$. Supóngase que el costo del fabricante es C_1 dólares por artículo, mientras que él vende el artículo por C_2 dólares. Cualquier artículo que no se venda al término de la semana debe almacenarse con un costo de C_3 dólares por artículo. Si el fabricante decide producir N artículos al comienzo de la semana, ¿cuál es su utilidad esperada por semana? ¿Para qué valor de N es máxima la utilidad esperada? Si T es la utilidad por semana, tenemos

$$\begin{aligned} T &= NC_2 - NC_1 && \text{si } D > N, \\ &= DC_2 - C_1N - C_3(N - D) && \text{si } D \leq N. \end{aligned}$$

Reescribiendo lo anterior, obtenemos

$$\begin{aligned} T &= N(C_2 - C_1) && \text{si } D > N, \\ &= (C_2 + C_3)D - N(C_1 + C_3) && \text{si } D \leq N. \end{aligned}$$

Por tanto la utilidad esperada se obtiene como sigue:

$$\begin{aligned} E(T) &= N(C_2 - C_1)P(D > N) + (C_2 + C_3) \sum_{n=0}^N np(n) \\ &\quad - N(C_1 + C_3)P(D \leq N) \end{aligned}$$

$$\begin{aligned}
 &= N(C_2 - C_1) \sum_{n=N+1}^{\infty} p(n) + (C_2 + C_3) \sum_{n=0}^N np(n) \\
 &\quad - N(C_1 + C_3) \sum_{n=0}^N p(n) \\
 &= N(C_2 - C_1) + (C_2 + C_3) \left[\sum_{n=0}^N np(n) - N \sum_{n=0}^N p(n) \right] \\
 &= N(C_2 - C_1) + (C_2 + C_3) \sum_{n=0}^N p(n)(n - N).
 \end{aligned}$$

Supongamos que se sabe que para D es apropiada la siguiente distribución de probabilidades: $P(D = n) = \frac{1}{5}, n = 1, 2, 3, 4, 5$. Por tanto

$$\begin{aligned}
 E(T) &= N(C_2 - C_1) + \frac{(C_2 + C_3)}{5} [N(N+1)/2 - N^2] \quad \text{si } N \leq 5, \\
 &= N(C_2 - C_1) + (C_2 + C_3)\frac{1}{5}(15 - 5N) \quad \text{si } N > 5.
 \end{aligned}$$

Supongamos que $C_2 = \$9$, $C_1 = \$3$, y $C_3 = \$1$. Por tanto

$$\begin{aligned}
 E(T) &= 6N + 2 \left[\frac{N(N+1)}{2} - N^2 \right] \\
 &\quad \text{si } N \leq 5, \\
 &= 6N + 2(15 - 5N) \quad \text{si } N > 5, \\
 &= 7N - N^2 \quad \text{si } N \leq 5, \\
 &= 30 - 4N \quad \text{si } N > 5.
 \end{aligned}$$

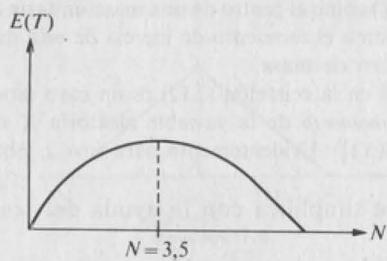


FIGURA 7.5

Luego el máximo ocurre para $N = 3.5$. (Ver figura 7.5.) Para $N = 3$ ó 4, tenemos $E(T) = 12$, el cual es el máximo obtenible puesto que N es un entero.

7.5 La varianza de una variable aleatoria

Supongamos que para una variable aleatoria X encontramos que $E(X)$ es igual a 2. ¿Cuál es la importancia de esto? Es importante que no atribuyamos más significado a esta información que la justificada. Significa, sencillamente, que si consideramos un gran número de valores de X , x_1, \dots, x_n , y promediamos esos valores de X , este promedio se aproximará a 2 si n es grande. Sin embargo, es muy importante que no demos mucha importancia a un valor esperado. Por ejemplo supóngase que X representa la duración de una bombilla que se recibe de un fabricante, y que $E(X) = 1000$ horas. Esto podría significar una de varias posibilidades. Podría significar que se espera que la mayor parte de las bombillas dure entre 900 y 1100 horas. También podría significar que las bombillas que se entregan están formadas por dos tipos de bombillas muy diferentes: alrededor de la mitad son de muy alta calidad y con duración de cerca de 1300 horas, mientras que la otra mitad son de muy mala calidad y durarán cerca de 700 horas.

Hay una necesidad obvia de presentar una medida cuantitativa que distinga entre tales situaciones. Varias medidas se sugieren por sí mismas, pero la siguiente es la cantidad usada más comúnmente.

Definición. Sea X una variable aleatoria. Definimos la *varianza* de X , denotada por $V(X)$ σ_X^2 , como sigue:

$$V(X) = E[X - E(X)]^2. \quad (7.12)$$

La raíz cuadrada positiva de $V(X)$ se llama *desviación estándar* de X y está denotada por σ_X .

Observaciones: (a) El *número* $V(X)$ está expresado en *unidades cuadradas* de X . Esto es, si X se mide en horas, entonces $V(X)$ está expresada en (horas)². Esta es una razón para considerar la desviación estándar. Se expresa en las *mismas unidades* que X .

(b) Otra medida posible podría haber sido $E|X - E(X)|$. Por diferentes razones, una de las cuales es que X^2 es una función «mejor comportada» que $|X|$, se prefiere la varianza.

(c) Si interpretamos a $E(X)$ como el centro de una masa unitaria distribuida sobre una recta, podemos interpretar $V(X)$ como el momento de inercia de esta masa, respecto a un eje perpendicular a través del centro de masa.

(d) $V(X)$ como se definió en la ecuación (7.12) es un caso especial del concepto más general siguiente. El *k-ésimo momento* de la variable aleatoria X respecto a su esperanza se define como $\mu_k = E[X - E(X)]^k$. Evidentemente para $k = 2$, obtenemos la varianza.

El cálculo de $V(X)$ se simplifica con la ayuda del resultado siguiente.

Teorema 7.5.

$$V(X) = E(X^2) - [E(X)]^2.$$

Demostración: desarrollando $E[X - E(X)]^2$ y usando las propiedades de la esperanza establecidas previamente, obtenemos

$$\begin{aligned}
 V(X) &= E[X - E(X)]^2 \\
 &= E\{X^2 - 2XE(X) + [E(X)]^2\} \\
 &= E(X^2) - 2E(X)E(X) + [E(X)]^2 \quad [\text{Recuérdese que } E(X) \text{ es una constante.}] \\
 &= E(X^2) - [E(X)]^2.
 \end{aligned}$$

EJEMPLO 7.15. La oficina meteorológica clasifica el tipo de cielo que es visible en relación con los «grados de nubosidad». Se usa una escala con 11 categorías: 0, 1, 2, ..., 10, en donde 0 representa un cielo perfectamente claro, 10 representa un cielo completamente cubierto, mientras que los otros valores representan diversas condiciones intermedias. Supongamos que tal clasificación se hace en una estación meteorológica determinada en un día y hora determinados. Sea X la variable aleatoria que toma uno de los 11 valores anteriores. Supongamos que la distribución de probabilidades de X es

$$\begin{aligned}
 p_0 &= p_{10} = 0,05; \\
 p_1 &= p_2 = p_8 = p_9 = 0,15; \\
 p_3 &= p_4 = p_5 = p_6 = p_7 = 0,06.
 \end{aligned}$$

Por tanto

$$\begin{aligned}
 E(X) &= 1(0,15) + 2(0,15) + 3(0,06) + 4(0,06) + 5(0,06) \\
 &\quad + 6(0,06) + 7(0,06) + 8(0,15) + 9(0,15) \\
 &\quad + 10(0,05) = 5,0.
 \end{aligned}$$

A fin de calcular $V(X)$ necesitamos calcular $E(X^2)$.

$$\begin{aligned}
 E(X^2) &= 1(0,15) + 4(0,15) + 9(0,06) + 16(0,06) + 25(0,06) \\
 &\quad + 36(0,06) + 49(0,06) + 64(0,15) + 81(0,15) \\
 &\quad + 100(0,05) = 35,6.
 \end{aligned}$$

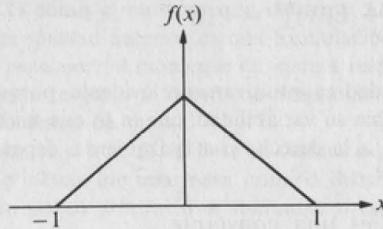


FIGURA 7.6

Luego

$$V(X) = E(X^2) - (E(X))^2 = 35,6 - 25 = 10,6,$$

y la desviación estándar $\sigma = 3,25$.

EJEMPLO 7.16. Supongamos que X es una variable aleatoria continua con fdp

$$\begin{aligned}f(x) &= 1 + x, & -1 \leq x \leq 0, \\&= 1 - x, & 0 \leq x \leq 1.\end{aligned}$$

(Ver figura 7.6.) Debido a la simetría de la fdp, $E(X) = 0$. (Ver la observación abajo.) Más aún,

$$\begin{aligned}E(X^2) &= \int_{-1}^0 X^2(1+x)dx \\&\quad + \int_0^1 x^2(1-x)dx = \frac{1}{6}.\end{aligned}$$

Por tanto $V(X) = \frac{1}{6}$.

Observación: supóngase que una variable aleatoria continua tiene una fdp que es simétrica respecto a $x = 0$. Es decir, $f(-x) = f(x)$ para todo x . Luego, siempre que exista $E(X)$, $E(X) = 0$, que es una consecuencia inmediata de la definición de $E(X)$. Esto puede extenderse a un punto arbitrario de simetría $x = a$, en tal caso $E(X) = a$ (Ver el problema 7.33).

7.6 Propiedades de la varianza de una variable aleatoria

Hay varias propiedades importantes, en parte análogas a las expuestas para la esperanza de una variable aleatoria, que se mantienen para la varianza.

Propiedad 7.7. Si C es una constante,

$$V(X + C) = V(X). \tag{7.13}$$

Demostración:

$$\begin{aligned}V(X + C) &= E[(X + C) - E(X + C)]^2 = E[(X + C) - E(X) - C]^2 \\&= E[X - E(X)]^2 = V(X).\end{aligned}$$

Observación: esta propiedad es intuitivamente evidente, porque al agregar una constante a un resultado X no cambia su variabilidad, que es lo que mide la varianza. Simplemente «desplaza» los valores de X a la derecha o a la izquierda, dependiendo del signo de C .

Propiedad 7.8. Si C es una constante,

$$V(CX) = C^2 V(X). \tag{7.14}$$

Demostración:

$$\begin{aligned}V(CX) &= E(CX)^2 - (E(CX))^2 = C^2 E(X^2) - C^2 (E(X))^2 \\&= C^2 [E(X^2) - (E(X))^2] = C^2 V(X).\end{aligned}$$

Propiedad 7.9. Si (X, Y) es una variable aleatoria bidimensional, y si X y Y son *independientes* entonces

$$V(X + Y) = V(X) + V(Y). \quad (7.15)$$

Demostración:

$$\begin{aligned} V(X + Y) &= E(X + Y)^2 - (E(X + Y))^2 \\ &= E(X^2 + 2XY + Y^2) - (E(X))^2 - 2E(X)E(Y) - (E(Y))^2 \\ &= E(X^2) - (E(X))^2 + E(Y^2) - (E(Y))^2 = V(X) + V(Y). \end{aligned}$$

Observación: es importante establecer que, en general, la varianza *no es aditiva como lo es el valor esperado*. Con la hipótesis adicional de independencia, la propiedad 7.9 es válida. La varianza no posee la propiedad de linealidad que dimos para la esperanza, es decir, $V(aX + b) \neq aV(X) + b$. En su lugar tenemos $V(aX + b) = a^2V(X)$.

Propiedad 7.10. Sean X_1, \dots, X_n n variables aleatorias independientes. Entonces

$$V(X_1 + \dots + X_n) = V(X_1) + \dots + V(X_n). \quad (7.16)$$

Demostración: ésta se deduce de la propiedad 7.9 con inducción matemática.

Propiedad 7.11. Sea X una variable aleatoria con varianza finita. Luego para cualquier número real α ,

$$V(X) = E[(X - \alpha)^2] - [E(X) - \alpha]^2. \quad (7.17)$$

Demostración: ver el problema 7.36.

Observaciones: (a) Esta es una extensión obvia del teorema 7.5, porque al hacer $\alpha = 0$ obtenemos el teorema 7.5.

(b) Si interpretamos $V(X)$ como el momento de inercia y $E(X)$ como el centro de una masa unitaria, entonces la propiedad anterior es una formulación del teorema muy conocido en mecánica de los *ejes paralelos*: el momento de inercia respecto a un punto arbitrario es igual al momento de inercia respecto al centro de masa más el cuadrado de la distancia de este punto arbitrario al centro de masa.

(c) $E[X - \alpha]^2$ es minimizado si $\alpha = E(X)$. Esto se deduce de inmediato de la propiedad anterior. Así el momento de inercia (de una masa unitaria distribuida en una recta) respecto a un eje que pasa por un punto arbitrario se minimiza si este punto se escoge como el centro de masa.

EJEMPLO 7.17. Calculemos la varianza de una variable aleatoria distribuida binomialmente con parámetro p .

Para calcular $V(X)$ podemos proceder de dos maneras. Puesto que ya conocemos que $E(X) = np$, sencillamente debemos calcular $E(X^2)$ y luego calcular $V(X)$ como $E(X^2) - (E(X))^2$. Para calcular $E(X^2)$ usamos el hecho de que

$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$, $k = 0, 1, \dots, n$. Por tanto $E(X^2) = \sum_{k=0}^n k^2 \binom{n}{k} p^k (1-p)^{n-k}$. Esta suma puede calcularse fácilmente, pero en vez de hacer esto, emplearemos un método más simple.

Nuevamente usaremos la representación de X presentada en el ejemplo 7.13, $X = Y_1 + Y_2 + \dots + Y_n$. Observemos que las Y_i son variables aleatorias *independientes* puesto que el valor de Y_i depende solo del resultado de la i -ésima repetición, y se supone que las repeticiones sucesivas son independientes. Por tanto podemos aplicar la propiedad 7.10 y obtener

$$V(X) = V(Y_1 + \dots + Y_n) = V(Y_1) + \dots + V(Y_n).$$

Pero $V(Y_i) = E(Y_i^2) - [E(Y_i)]^2$. Ahora

$$E(Y_i) = 1(p) + 0(1-p) = p, \quad E(Y_i^2) = 1^2(p) + 0^2(1-p) = p.$$

Por tanto

$$V(Y_i) = p - p^2 = p(1-p) \text{ para todo } i. \text{ Así } V(X) = np(1-p).$$

Observación: consideremos $V(X) = np(1-p)$ como una función de p para un n dado. Dibujemos un gráfico como se muestra en la figura 7.7.

Resolviendo $(d/dp)np(1-p) = 0$ encontramos que el valor máximo para $V(X)$ ocurre cuando $p = \frac{1}{2}$. El valor mínimo de $V(X)$ ocurre evidentemente en los extremos del intervalo en $p = 0$ y $p = 1$. Esto es intuitivamente como debería ser. Recordando que la varianza es una medida de la variación de la variable aleatoria X definida como el número de veces que ocurre el suceso A en n repeticiones, encontramos que esta variación es nula si $p = 0$ ó 1 (es decir, si A ocurre con probabilidad 0 ó 1) y es máxima cuando están «sin certeza como podemos estar» acerca de la ocurrencia o no ocurrencia de A , es decir, cuando $P(A) = \frac{1}{2}$.

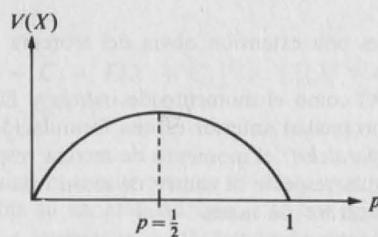


FIGURA 7.7

EJEMPLO 7.18. Supóngase que la variable aleatoria X está distribuida uniformemente en $[a, b]$. Como lo calculamos previamente, $E(X) = (a+b)/2$.

Para calcular $V(X)$ hallemos el valor de $E(X^2)$:

$$E(X^2) = \int_a^b x^2 \frac{1}{b-a} dx = \frac{b^3 - a^3}{3(b-a)}.$$

Por tanto

$$V(X) = E(X^2) - [E(X)]^2 = \frac{(b-a)^2}{12}$$

después de un cálculo sencillo.

Observaciones: (a) Este resultado es intuitivamente significativo. Indica que la varianza de X no depende individualmente de a y b sino sólo de $(b-a)^2$, es decir, del cuadrado de su diferencia. Por tanto dos variables aleatorias distribuidas cada una uniformemente en un intervalo (no necesariamente el mismo) tendrán iguales varianzas mientras las longitudes de los intervalos sean iguales.

(b) Es bien sabido el hecho de que los momentos de inercia de una barra delgada de masa M y longitud L respecto a un eje transversal que pasa por el centro están dados por $ML^2/12$.

7.7. Expresiones aproximadas para la esperanza y la varianza

Ya hemos observado que para evaluar $E(Y)$ o $V(Y)$, en donde $Y = H(X)$, no necesitamos conocer la distribución de probabilidades de Y , sino que podemos trabajar directamente con la distribución de probabilidades de X . De un modo semejante, si $Z = H(X, Y)$, podemos calcular $E(Z)$ y $V(Z)$ sin obtener primero la distribución de Z .

Si la función H es muy complicada, la evaluación de la esperanza y varianza anteriores puede conducir a integraciones (o sumas) que son muy difíciles. De aquí que sean muy útiles las aproximaciones siguientes.

Teorema 7.6. Sea una variable aleatoria X con $E(X) = \mu$, y $V(X) = \sigma^2$.

Supongamos que $Y = H(X)$. Luego

$$E(Y) \simeq H(\mu) + \frac{H''(\mu)}{2} \sigma^2, \quad (7.18)$$

$$V(Y) \simeq [H'(\mu)]^2 \sigma^2. \quad (7.19)$$

(A fin de hacer útiles las aproximaciones anteriores necesitamos evidentemente que H sea a lo menos diferenciable dos veces para $x = \mu$).

Demostración: (sólo un bosquejo): a fin de establecer la ecuación (7.18), desarrollemos la función H en una serie de Taylor para $x = \mu$ con dos términos. Así

$$Y = H(\mu) + (X - \mu)H'(\mu) + \frac{(X - \mu)^2 H''(\mu)}{2} + R_1,$$

en donde R_1 es un resto. Si descartamos el término resto R_1 , entonces, tomando el valor esperado en ambos miembros, tenemos

$$E(Y) \simeq H(\mu) + \frac{H''(\mu)}{2} \sigma^2,$$

puesto que $E(X - \mu) = 0$. A fin de establecer la ecuación (7.19), desarrollemos H en una serie de Taylor para $x = \mu$ con un término. Luego $Y = H(\mu) + (X - \mu)H'(\mu) + R_2$. Si descartamos el resto R_2 y tomamos la varianza en ambos lados, tenemos

$$V(Y) \simeq [H'(\mu)]^2\sigma^2.$$

EJEMPLO 7.19. Bajo ciertas condiciones la tensión superficial de un líquido (dina/cm) está dada por la fórmula $S = 2(1 - 0,005T)^{1.2}$ donde T es la temperatura del líquido (grados centígrados).

Supongamos que T es una variable aleatoria continua con la siguiente fdp,

$$\begin{aligned}f(t) &= 3000t^{-4}, & t \geq 10, \\&= 0, & \text{para cualquier otro valor.}\end{aligned}$$

Luego

$$E(T) = \int_{10}^{\infty} 3000t^{-3} dt = 15 \text{ (grados centígrados).}$$

Y

$$\begin{aligned}V(T) &= E(T^2) - (15)^2 \\&= \int_{10}^{\infty} 3000t^{-2} dt - 225 = 75 \text{ (grados centígrados)}^2.\end{aligned}$$

Para calcular $E(S)$ y $V(S)$ tenemos que calcular las integrales siguientes

$$\int_{10}^{\infty} (1 - 0,005t)^{1.2}t^{-4} dt$$

y

$$\int_{10}^{\infty} (1 - 0,005t)^{2.4}t^{-4} dt.$$

En vez de evaluar esas expresiones, obtendremos aproximaciones para $E(S)$ y $V(S)$ al usar las ecuaciones (7.18) y (7.19). A fin de usar esas fórmulas tenemos que calcular $H'(15)$ y $H''(15)$, en donde $H(t) = 2(1 - 0,005t)^{1.2}$. Tenemos

$$H'(t) = 2,4(1 - 0,005t)^{0.2}(-0,005) = -0,012(1 - 0,005t)^{0.2}.$$

Luego

$$H(15) = 1,82, H'(15) = 0,01.$$

De un modo semejante,

$$H''(t) = -0,0024(1 - 0,005t)^{-0.8}(-0,005) = 0,000012(1 - 0,005t)^{-0.8}.$$

Por lo tanto

$$H''(15) = \frac{0,000012}{(0,925)^{0.8}} = 0^+.$$

Así tenemos

$$E(S) \simeq H(15) + \frac{1}{2}75H''(15) = 1,82 \text{ (dinas/cm)}, \\ V(S) \simeq 75[H'(15)]^2 = 0,87 \text{ (dinas/cm)}^2.$$

Si Z es una función de dos variables, $Z = H(X, Y)$, se establece un resultado análogo.

Teorema 7.7. Sea (X, Y) una variable aleatoria bidimensional. Supongamos que $E(X) = \mu_x$, $E(Y) = \mu_y$; $V(X) = \sigma_x^2$ y $V(Y) = \sigma_y^2$. Sea $Z = H(X, Y)$. [Supondremos que existen las diversas derivadas de H para (μ_x, μ_y) .] Luego si X y Y son independientes, tenemos

$$E(Z) \simeq H(\mu_x, \mu_y) + \frac{1}{2} \left[\frac{\partial^2 H}{\partial x^2} \sigma_x^2 + \frac{\partial^2 H}{\partial y^2} \sigma_y^2 \right], \\ V(Z) \simeq \left[\frac{\partial H}{\partial x} \right]^2 \sigma_x^2 + \left[\frac{\partial H}{\partial y} \right]^2 \sigma_y^2,$$

en donde todas las derivadas parciales se evalúan en (μ_x, μ_y) .

Demostración: la demostración implica el desarrollo de H en una serie de Taylor en el punto (μ_x, μ_y) con uno y dos términos, descartando el resto, y luego tomando la esperanza y la varianza en ambos miembros como se hizo en la demostración del teorema 7.6. Dejaremos los detalles al lector. (Si X y Y no son independientes, se puede derivar una fórmula ligeramente más complicada.)

Observación: el resultado anterior puede extenderse a una función de n variables aleatorias independientes, $Z = H(X_1, \dots, X_n)$. Si $E(X_i) = \mu_i$, $V(X_i) = \sigma_i^2$, tenemos las siguientes aproximaciones, suponiendo que todas las derivadas existan.

$$E(Z) \simeq H(\mu_1, \dots, \mu_n) + \frac{1}{2} \sum_{i=1}^n \frac{\partial^2 H}{\partial x_i^2} \sigma_i^2, \\ V(Z) \simeq \sum_{i=1}^n \left(\frac{\partial H}{\partial x_i} \right)^2 \sigma_i^2,$$

en donde las derivadas parciales son evaluadas en el punto (μ_1, \dots, μ_n) .

EJEMPLO 7.20. Supóngase que tenemos un circuito simple para el cual el voltaje, M , se expresa por la ley de Ohm como $M = IR$, en donde I y R son la corriente y la resistencia del circuito respectivamente. Si I y R son variables aleatorias independientes, entonces M es una variable aleatoria y, usando el teorema 7.7, podemos escribir

$$E[M] \simeq E(I)E(R), \quad V[M] \simeq [E(R)]^2 V(I) + [E(I)]^2 V(R).$$

7.8 Desigualdad de Chebyshev

Hay una desigualdad muy conocida del matemático ruso Chebyshev que desempeña un papel importante en lo que resta de nuestra obra. Además, nos dará un medio para comprender cabalmente como la varianza mide la variabilidad respecto al valor esperado de una variable aleatoria.

Si conocemos la distribución de probabilidades de una variable aleatoria X (la fdp en el caso continuo o la probabilidad puntual en el caso discreto), podemos calcular $E(X)$ y $V(X)$, si existen. Sin embargo, lo recíproco *no* es verdadero. Esto es, conociendo $E(X)$ y $V(X)$ no podemos reconstruir la distribución de probabilidades de X y por tanto no podemos calcular cantidades tales como $P[|X - E(X)| \leq C]$. Sin embargo, resulta que aunque no podemos evaluar tales probabilidades [a partir de un conocimiento de $E(X)$ y $V(X)$], podemos dar una cota superior (o inferior) muy útiles para tales probabilidades. Este resultado está contenido en lo que se conoce como la desigualdad de Chebyshev.

Desigualdad de Chebyshev. Sea X una variable aleatoria con $E(X) = \mu$ y sea c un número real cualquiera. Entonces, si $E(X - c)^2$ es finita y ϵ es cualquier número positivo, tenemos

$$P[|X - c| \geq \epsilon] \leq \frac{1}{\epsilon^2} E(X - c)^2. \quad (7.20)$$

Las formas siguientes, equivalentes a (7.20), son inmediatas:

(a) Al considerar el suceso complementario obtenemos:

$$P[|X - c| < \epsilon] \geq 1 - \frac{1}{\epsilon^2} E(X - c)^2. \quad (7.20a)$$

(b) Al elegir $c = \mu$ obtenemos

$$P[|X - \mu| \geq \epsilon] \leq \frac{\text{Var } X}{\epsilon^2}. \quad (7.20b)$$

(c) Eligiendo $c = \mu$ y $\epsilon = k\sigma$, en donde $\sigma^2 = \text{Var } X > 0$, obtenemos

$$P[|X - \mu| \geq k\sigma] \leq k^{-2}. \quad (7.21)$$

Esta última forma (7.21) indica especialmente cómo la varianza mide el «grado de concentración» de la probabilidad próxima a $E(X) = \mu$.

Demostración (Demostraremos sólo 7.20 puesto que las otras se deducen como se indicó. Trataremos sólo el caso continuo. En el caso discreto, el argumento es muy parecido al de las integrales substituidas por sumas. Sin embargo, hay que tener cuidado con los puntos extremos de los intervalos):

Consideremos

$$P[|X - c| \geq \epsilon] = \int_{x:|x-c| \geq \epsilon} f(x) dx.$$

(Los límites de la integral dicen que estamos integrando entre $-\infty$ y $c - \epsilon$ y entre $c + \epsilon$ y $+\infty$.)

Ahora $|x - c| \geq \epsilon$ es equivalente a $(x - c)^2 / \epsilon^2 \geq 1$. Puesto que la integral anterior es

$$\leq \int_R \frac{(x - c)^2}{\epsilon^2} f(x) dx,$$

en donde

$$R = \{x : |x - c| \geq \epsilon\}.$$

Esta integral, a su vez, es

$$\leq \int_{-\infty}^{+\infty} \frac{(x - c)^2}{\epsilon^2} f(x) dx$$

lo que es igual

$$\frac{1}{\epsilon^2} E[X - c]^2,$$

como se pedia demostrar.

Observaciones: (a) Es importante darse cuenta de que el resultado anterior es notable precisamente debido a lo poco que se presupone acerca de la conducta probabilística de la variable aleatoria X .

(b) Como podríamos suponer, una información adicional respecto a la distribución de la variable aleatoria X nos permitirá mejorar la desigualdad que deducimos. Por ejemplo, si $C = \frac{3}{2}$, tenemos, de la desigualdad de Chebyshev,

$$P[|X - \mu| \geq \frac{3}{2}\sigma] \leq \frac{4}{9} = 0,44.$$

Supongamos que también sabemos que X está distribuida uniformemente en $(1 - 1/\sqrt{3}, 1 + 1/\sqrt{3})$. Por lo tanto $E(X) = 1$, $V(X) = \frac{1}{6}$ y así

$$\begin{aligned} P[|X - \mu| \geq \frac{3}{2}\sigma] &= P[|X - 1| \geq \frac{3}{2}] = 1 - P[|X - 1| < \frac{3}{2}] \\ &= 1 - P[\frac{1}{2} < X < \frac{5}{2}] = 1 - \frac{\sqrt{3}}{2} = 0,134. \end{aligned}$$

Nótese que aunque la relación obtenida de la desigualdad de Chebyshev es consecuente con este resultado, la última es una relación más precisa. Sin embargo, en muchos problemas ninguna suposición referente a la distribución específica de la variable aleatoria está justificada, y en tales casos la desigualdad de Chebyshev puede darnos una información importante acerca del comportamiento de la variable aleatoria.

Como lo observamos en la ecuación (7.21), si $V(X)$ es pequeña, la mayor parte de la distribución de probabilidades de X está «concentrada» próxima a $E(X)$. Esto se puede expresar más detalladamente en el teorema siguiente.

Teorema 7.8. Supongamos que $V(X) = 0$. Luego $P[X = \mu] = 1$, en donde $\mu = E(X)$. (Informalmente, $X = \mu$, con «probabilidad 1».)

Demostración: de la ecuación (7.20b) encontramos que

$$P[|X - \mu| \geq \epsilon] = 0 \text{ para cualquier } \epsilon > 0.$$

Luego

$$P[|X - \mu| < \epsilon] = 1 \text{ para cualquier } \epsilon > 0.$$

Puesto que puede elegirse ϵ arbitrariamente pequeño, el teorema queda demostrado.

Observaciones: (a) Este teorema demuestra que la varianza cero no implica que toda la probabilidad esté concentrada en un solo punto, a saber $E(X)$.

(b) Si $E(X) = 0$, luego $V(X) = E(X^2)$, y por tanto en este caso, $E(X^2) = 0$ implica la misma conclusión.

(c) En el sentido anterior decimos que una variable aleatoria X es *degenerada*: toma sólo un valor con probabilidad 1.

7.9 El coeficiente de correlación

Hasta ahora nos hemos interesado en los parámetros asociados con la distribución de variables aleatorias unidimensionales tales como $E(X)$ y $V(X)$. Estos parámetros miden, en el sentido ya descrito previamente, ciertas características de la distribución. Si tenemos una variable aleatoria bidimensional (X, Y) , se encuentra un problema análogo. Naturalmente, podemos presentar nuevamente las variables aleatorias unidimensionales X y Y asociadas con (X, Y) . Sin embargo, surge la pregunta de si hay un parámetro significativo que mida de alguna manera el «grado de asociación» entre X y Y . Esta es una noción más vaga que será precisada en breve. Demos la siguiente definición formal.

Definición. Sea (X, Y) una variable aleatoria bidimensional. Definimos ρ_{xy} , el coeficiente de correlación, entre X y Y , como sigue:

$$\rho_{xy} = \frac{E\{[X - E(X)][Y - E(Y)]\}}{\sqrt{V(X)V(Y)}}. \quad (7.22)$$

Observaciones: (a) Suponemos que todas las esperanzas existen y que ambas $V(X)$ y $V(Y)$ son distintas de cero. Cuando no hay duda de cuáles variables aleatorias están implicadas escribiremos simplemente ρ en vez de ρ_{xy} .

(b) El numerador de ρ , $E\{[X - E(X)][Y - E(Y)]\}$, se llama la *covarianza* de X y Y , y se denota algunas veces como σ_{xy} .

(c) El coeficiente de correlación es una cantidad adimensional.

(d) Antes de que la definición anterior pueda ser significativa, debemos establecer exactamente lo que mide ρ . Esto lo haremos al considerar un número de propiedades de ρ .

Teorema 7.9.

$$\rho = \frac{E(XY) - E(X)E(Y)}{\sqrt{V(X)V(Y)}}.$$

Demostración: consideremos

$$\begin{aligned} E\{(X - E(X))(Y - E(Y))\} &= E[XY - XE(Y) - YE(X) + E(X)E(Y)] \\ &= E(XY) - E(X)E(Y) - E(Y)E(X) + E(X)E(Y) \\ &= E(XY) - E(X)E(Y). \end{aligned}$$

Teorema 7.10. Si X y Y son independientes, entonces $\rho = 0$.

Demostración: se deduce inmediatamente del teorema 7.9, puesto que

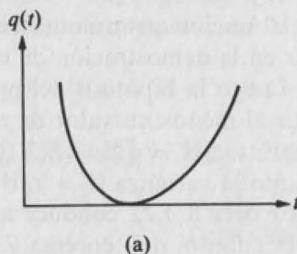
$$E(XY) = E(X)E(Y)$$

si X y Y son independientes.

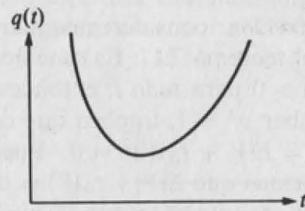
Observación: el recíproco del teorema 7.10 no es verdadero en general. (Ver el problema 7.39.) Esto es, podemos tener $\rho = 0$, y aún así X y Y no necesitan ser independientes. Si $\rho = 0$ decimos que X y Y son no correlacionados. Así, siendo no correlacionados e independientes no son, en general, equivalentes. El ejemplo siguiente ilustra este punto.*

Sean X y Y variables aleatorias cualesquiera que tienen la misma distribución. Sea $U = X - Y$ y $V = X + Y$. Luego $E(U) = 0$ y $\text{cov}(U, V) = E[(X - Y)(X + Y)] = E(X^2 - Y^2) = 0$. Luego U y V son no correlacionados. Aún si X y Y son independientes, U y V pueden ser dependientes, como lo indica la elección siguiente de X y Y . Sean X y Y los números que aparecen respectivamente en el primero y segundo dados, cuando se han lanzado un par de dados regulares. Ahora, por ejemplo, hallamos que $P[V = 4 | U = 3] = 0$ (puesto que si $X - Y = 3$, $X + Y$ no puede ser igual a 4), mientras que $P(V = 4) = 3/36$. Así U y V son dependientes.

Teorema 7.11. $-1 \leq \rho \leq 1$. (Esto es, ρ toma los valores entre -1 y $+1$ inclusive.)



(a)



(b)

FIGURA 7.8

Demostración: consideremos la siguiente función de la variable real t :

$$q(t) = E[V + tW]^2,$$

* El ejemplo en este párrafo está tomado de una exposición de un artículo titulado «Mutually Exclusive Events, Independence and Zero Correlation», por J. D. Gibbons, que apareció en *The American Statistician*, 22, núm. 5, diciembre 1968, págs. 31-32.

en donde $V = X - E(X)$ y $W = Y - E(Y)$. Puesto que $[V + tW]^2 \geq 0$, tenemos que $q(t) \geq 0$ para todo t . Desarrollando obtenemos

$$q(t) = E[V^2 + 2tVW + t^2W^2] = E(V^2) + 2tE(VW) + t^2E(W^2).$$

Así $q(t)$ es una expresión cuadrática en t . En general, si una expresión cuadrática $q(t) = at^2 + bt + c$ tiene la propiedad de que $q(t) \geq 0$ para todo t , significa que su gráfico corta el eje t en solo un punto o en ninguno, como se indicó en la figura 7.8. Esto, a su vez, indica que el discriminante $b^2 - 4ac$ debe ser ≤ 0 , puesto que $b^2 - 4ac > 0$ significaría que $q(t)$ tiene dos raíces reales distintas. Aplicando esta conclusión a la función $q(t)$ que consideramos anteriormente, obtenemos

$$4[E(VW)]^2 - 4E(V^2)E(W^2) \leq 0.$$

Esto implica que

$$\frac{[E(VW)]^2}{E(V^2)E(W^2)} \leq 1, \quad \text{y por tanto}$$

$$\frac{\{E[X - E(X)][Y - E(Y)]\}^2}{V(X)V(Y)} = \rho^2 \leq 1.$$

Así $-1 \leq \rho \leq 1$.

Teorema 7.12. Supongamos que $\rho^2 = 1$. Luego (con probabilidad 1 en el sentido del teorema 7.8), $Y = AX + B$, en donde A y B son constantes. En palabras: si el coeficiente de correlación ρ es ± 1 , entonces Y es una función lineal de X (con probabilidad 1).

Demostración: consideraremos nuevamente la función $q(t)$ descrita en la demostración del teorema 7.11. Es sencillo observar en la demostración de ese teorema que si $q(t) > 0$ para todo t , entonces $\rho^2 < 1$. Luego la hipótesis del presente teorema, a saber $\rho^2 = 1$, implica que debe existir al menos un valor de t , sea t_0 , tal que $q(t_0) = E(V + t_0W)^2 = 0$. Puesto que $V + t_0W = [X - E(X)] + t_0[Y - E(Y)]$, tenemos que $E(V + t_0W) = 0$ y por tanto la varianza $(V + t_0W) = E(V + t_0W)^2$. Así encontramos que la hipótesis del teorema 7.12 conduce a la conclusión de que la varianza de $(V + t_0W) = 0$. Por tanto, del teorema 7.8 podemos concluir que la variable aleatoria $(V + t_0W) = 0$ (con probabilidad 1). Por tanto $[X - E(X)] + t_0[Y - E(Y)] = 0$. Reescribiendo encontramos que $Y = AX + B$ (con probabilidad 1), como se iba a demostrar.

Observaciones: el reciproco del teorema 7.12 también se mantiene como se demostró en el teorema 7.13.

Teorema 7.13. Supongamos que X y Y son dos variables aleatorias para las cuales $Y = AX + B$, en donde A y B son constantes. Entonces $\rho^2 = 1$. Si $A > 0$, $\rho = +1$; si $A < 0$, $\rho = -1$.

Demostración: puesto que $Y = AX + B$, tenemos que $E(Y) = AE(X) + B$ y $V(Y) = A^2V(X)$.

También

$$E(XY) = E[X(AX + B)] = AE(X^2) + BE(X).$$

Luego

$$\begin{aligned}\rho^2 &= \frac{[E(XY) - E(X)E(Y)]^2}{V(X)V(Y)} \\ &= \frac{\{AE(X^2) + BE(X) - E(X)[AE(X) + B]\}^2}{V(X)A^2V(X)} \\ &= \frac{[AE(X^2) + BE(X) - A(E(X))^2 - BE(X)]^2}{A^2(V(X))^2} \\ &= \frac{A^2\{E(X^2) - [E(X)]^2\}^2}{A^2(V(X))^2} = 1.\end{aligned}$$

(La segunda afirmación del teorema se deduce al observar que $\sqrt{A^2} = |A|$.)

Observación: los teoremas 7.12 y 7.13 establecen las siguientes características importantes del coeficiente de correlación: el coeficiente de correlación es una medida del *grado de linealidad* entre X y Y . Los valores de ρ próximos a $+1$ ó -1 indican un alto grado de linealidad mientras que los valores de ρ próximos a 0 indican una ausencia de tal linealidad. Los valores positivos de ρ muestran que Y tiende a aumentar con valores crecientes de X , mientras que los valores negativos de ρ muestran que Y tiende a disminuir con valores crecientes de X . Existe una idea considerablemente errónea acerca de la interpretación del coeficiente de correlación. Un valor de ρ próximo a cero sólo indica la ausencia de una relación *lineal* entre X y Y . No impide la posibilidad de alguna relación *no lineal*.

EJEMPLO 7.21. Supongamos que la variable aleatoria bidimensional (X, Y) está distribuida uniformemente en la región triangular

$$R = \{(x, y) \mid 0 < x < y < 1\}.$$

(Ver la figura 7.9.) Luego la fdp está dada como

$$f(x, y) = 2, \quad (x, y) \in R, \\ = 0, \quad \text{para cualquier otro valor.}$$

Luego las fdp marginales de X y Y son

$$g(x) = \int_x^1 (2) dy = 2(1 - x), \quad 0 \leq x \leq 1; \\ h(y) = \int_0^y (2) dx = 2y, \quad 0 \leq y \leq 1.$$

Por tanto

$$E(X) = \int_0^1 x2(1 - x) dx = \frac{1}{3}, \quad E(Y) = \int_0^1 y2y dy = \frac{2}{3};$$

$$E(X^2) = \int_0^1 x^2 2(1 - x) dx = \frac{1}{6}, \quad E(Y^2) = \int_0^1 y^2 2y dy = \frac{1}{2};$$

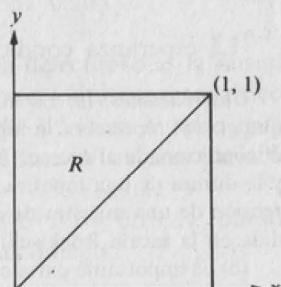


FIGURA 7.9

$$V(X) = E(X^2) - [E(X)]^2 = \frac{1}{8}, \quad V(Y) = E(Y^2) - [E(Y)]^2 = \frac{1}{8};$$

$$E(XY) = \int_0^1 \int_0^y xy^2 dx dy = \frac{1}{4}.$$

Luego

$$\rho = \frac{E(XY) - E(X)E(Y)}{\sqrt{V(X)V(Y)}} = \frac{1}{2}$$

Tal como hemos indicado, el coeficiente de correlación es una cantidad sin dimensión. Su valor no se afecta por un cambio de escala. Se puede demostrar el siguiente teorema fácilmente. (Ver problema 7.41.)

Teorema 7.14. Si ρ_{XY} es el coeficiente de correlación entre X y Y , y si $V = AX + B$ y $W = CY + D$, en donde A, B, C y D son constantes, entonces $\rho_{vw} = (AC/|AC|) \rho_{xy}$. (Suponemos que $A \neq 0, C \neq 0$.)

7.10 Esperanza condicional

Tal como definimos el valor esperado de una variable aleatoria X (en función de su distribución de probabilidades) como $\int_{-\infty}^{+\infty} xf(x) dx$ o $\sum_{i=1}^{\infty} x_i p(x_i)$, así podemos definir la esperanza condicional de una variable aleatoria (en función de su distribución condicional de probabilidades) como sigue.

Definición. (a) Si (X, Y) es una variable aleatoria continua bidimensional, definimos la *esperanza condicional* de X para $Y = y$ dado como

$$E(X | y) = \int_{-\infty}^{+\infty} xg(x | y) dx. \quad (7.23)$$

(b) Si (X, Y) es una variable aleatoria discreta bidimensional, definimos la esperanza condicional de X para $Y = y_j$ dado como

$$E(X | y_j) = \sum_{i=1}^{\infty} x_i p(x_i | y_j). \quad (7.24)$$

La esperanza condicional de Y para X dado se define análogamente.

Observaciones: (a) La interpretación de la esperanza condicional es como sigue. Puesto que $g(x | y)$ representa la fdp condicional de X para $Y = y$ dado, $E(X | y)$ es la esperanza de X condicionada al suceso $\{Y = y\}$. Por ejemplo, si (X, Y) representa la resistencia a la tensión y la dureza de una muestra de acero, entonces $E(X | y = 52.7)$ es la resistencia esperada a la tensión de una muestra de acero elegida al azar del universo de muestras cuya dureza (medida en la escala Rockwell) es 52.7.

(b) Es importante darse cuenta de que en general $E(X | y)$ es una función de y y por lo tanto es una variable aleatoria. Análogamente $E(Y | x)$ es una función de x y también es una variable aleatoria. [Estrictamente hablando, $E(X | y)$ es el *valor* de la variable aleatoria $E(X | Y)$.]

(c) Puesto que $E(Y | X)$ y $E(X | Y)$ son variables aleatorias, será preciso hablar de sus esperanzas. Así podemos considerar $E[E(X | Y)]$, por ejemplo. Es importante establecer que la esperanza interna se toma respecto a la distribución condicional de X dado que Y es igual a y , mientras que la esperanza exterior se toma respecto a la distribución de probabilidades de Y .

Teorema 7.15.

$$E[E(X | Y)] = E(X), \quad (7.25)$$

$$E[E(Y | X)] = E(Y). \quad (7.26)$$

Demostración: (caso continuo solamente): por definición,

$$E(X | y) = \int_{-\infty}^{+\infty} x g(x | y) dx = \int_{-\infty}^{+\infty} x \frac{f(x, y)}{h(y)} dx,$$

en donde f es la fdp conjunta de (X, Y) y h es la fdp marginal de Y .

Por tanto

$$E[E(X | Y)] = \int_{-\infty}^{+\infty} E(X | y) h(y) dy = \int_{-\infty}^{+\infty} \left[\int_{-\infty}^{+\infty} x \frac{f(x, y)}{h(y)} dx \right] h(y) dy.$$

Si todas las esperanzas existen, es posible escribir la integral iterada anterior con el orden de integración invertido. Así

$$E[E(X | Y)] = \int_{-\infty}^{+\infty} x \left[\int_{-\infty}^{+\infty} f(x, y) dy \right] dx = \int_{-\infty}^{+\infty} x g(x) dx = E(X).$$

[Se puede usar un argumento semejante para establecer la ecuación (7.26)]. Este teorema es muy útil como lo ilustra el siguiente ejemplo.

EJEMPLO 7.22. Supóngase que varios cargamentos que traen diverso número de repuestos llegan diariamente. Si N es el número de artículos en el cargamento, la distribución de probabilidades de la variable aleatoria N está dada como sigue:

$n:$	10	11	12	13	14	15
$P(N = n):$	0,05	0,10	0,10	0,20	0,35	0,20

La probabilidad de que cualquier repuesto particular sea defectuoso es la misma para todos los repuestos y es igual a 0,10. Si X es el número de repuestos defectuosos que llegan cada día, ¿cuál es el valor esperado de X ? Para N dado igual n , X tiene una distribución binomial. Puesto que el mismo N es una variable aleatoria, procedemos como sigue.

Tenemos que $E(X) = E[E(X | N)]$. Sin embargo, $E(X | N) = 0,10 N$ puesto que para un N dado, X tiene una distribución binomial. Luego

$$\begin{aligned} E(X) &= E(0,10N) = 0,10E(N) \\ &= 0,10[10(0,05) + 11(0,10) + 12(0,10) + 13(0,20) + 14(0,35) + 15(0,20)] \\ &= 1,33. \end{aligned}$$

Teorema 7.16. Supóngase que X y Y son variables aleatorias independientes. Entonces

$$E(X|Y) = E(X) \quad \text{y} \quad E(Y|X) = E(Y)$$

Demostración: ver problema 7.43.

EJEMPLO 7.23 Supóngase que el suministro de potencia (kilowatts) de una compañía hidroeléctrica durante un período de tiempo específico es una variable aleatoria X , la que supondremos que tiene una distribución uniforme en $[10, 30]$. La demanda de potencia (kilowatts), Y , también es una variable aleatoria que supondremos que está distribuida uniformemente en $[10, 20]$. (Luego, en promedio, se suministra más potencia que la que se pide puesto que $E(X) = 20$, mientras que $E(Y) = 15$.) Por cada kilowatt suministrado la compañía obtiene una utilidad de \$0,03. Si la demanda excede el suministro, la compañía obtiene una utilidad adicional de otra fuente y obtiene una utilidad con esta potencia de \$0,01 por kilowatt suministrado. ¿Cuál es la utilidad esperada durante el tiempo específico que se considera?

Sea T esta utilidad. Tenemos

$$\begin{aligned} T &= 0,03Y && \text{si } Y < X, \\ &= 0,03X + 0,01(Y - X) && \text{si } Y > X. \end{aligned}$$

Para evaluar $E(T)$ escribámosla como $E[E(T|X)]$. Tenemos

$$\begin{aligned} E(T|X) &= \begin{cases} \int_{10}^x 0,03y \frac{1}{20} dy + \int_x^{20} (0,01y + 0,02x) \frac{1}{20} dy & \text{si } 10 < x < 20, \\ \int_{10}^{20} 0,03y \frac{1}{20} dy & \text{si } 20 < x < 30, \end{cases} \\ &= \begin{cases} \frac{1}{20}[0,015x^2 - 1,5 + 2 + 0,4x - 0,005x^2 - 0,02x^2] & \text{si } 10 < x < 20, \\ \frac{9}{20} & \text{si } 20 < x < 30, \end{cases} \\ &= \begin{cases} 0,05 + 0,04x - 0,001x^2 & \text{si } 10 < x < 20, \\ 0,45 & \text{si } 20 < x < 30. \end{cases} \end{aligned}$$

Por tanto

$$E[E(T|X)] = \frac{1}{20} \int_{10}^{20} (0,05 + 0,04x - 0,001x^2) dx + \frac{1}{20} \int_{20}^{30} 0,45 dx = \$0,43.$$

7.11 Regresión del promedio

Como lo señalamos en la sección anterior, $E(X|y)$ es el valor de la variable aleatoria $E(X|Y)$ y es una *función de y*. El gráfico de esta función de y se conoce la *curva de regresión* (del promedio) de X sobre Y . Análogamente, el gráfico de la función de x , $E(Y|x)$ se llama la curva de regresión (del promedio) de Y sobre X . Para cada uno de los y fijos, $E(X|y)$ es el valor esperado de la variable aleatoria (unidimensional) cuya distribución de probabilidades está definida por la ecuación (6.5) ó (6.7). (Ver la figura 7.10). En general, el valor esperado dependerá de y . [Se pueden hacer interpretaciones análogas para $E(Y|x)$.]

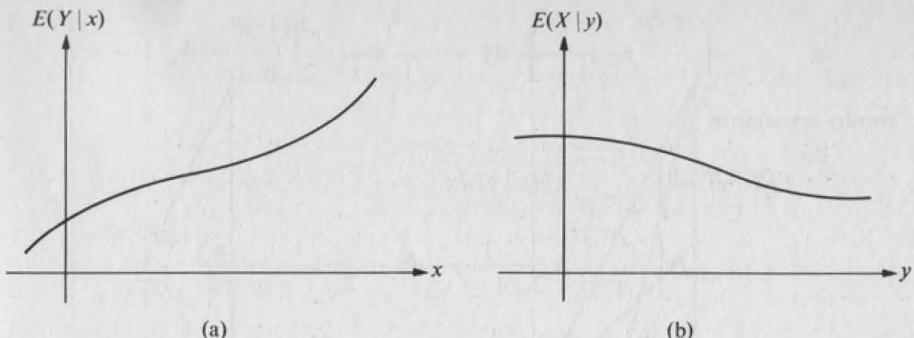


FIGURA 7.10

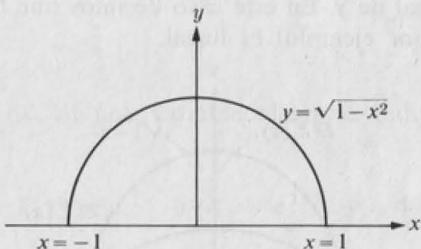


FIGURA 7.11

EJEMPLO 7.24. Supongamos que (X, Y) está distribuida uniformemente en la semicircunferencia indicada en la figura 7.11. Luego $f(x, y) = 2/\pi$, $(x, y) \in$ semicircunferencia. Así

$$g(x) = \int_0^{\sqrt{1-x^2}} \frac{2}{\pi} dy = \frac{2}{\pi} \sqrt{1-x^2}, \quad -1 \leq x \leq 1;$$

$$h(y) = \int_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}} \frac{2}{\pi} dx = \frac{4}{\pi} \sqrt{1-y^2}, \quad 0 \leq y \leq 1.$$

Por tanto

$$g(x | y) = \frac{1}{2\sqrt{1-y^2}}, \quad -\sqrt{1-y^2} \leq x \leq \sqrt{1-y^2};$$

$$h(y | x) = \frac{1}{\sqrt{1-x^2}}, \quad 0 \leq y \leq \sqrt{1-x^2}.$$

Luego

$$E(Y | x) = \int_0^{\sqrt{1-x^2}} y h(y | x) dy$$

$$= \int_0^{\sqrt{1-x^2}} y \frac{1}{\sqrt{1-x^2}} dy = \frac{1}{\sqrt{1-x^2}} \left. \frac{y^2}{2} \right|_0^{\sqrt{1-x^2}} = \frac{1}{2} \sqrt{1-x^2}.$$

De modo semejante

$$\begin{aligned} E(X | y) &= \int_{-\sqrt{1-y^2}}^{+\sqrt{1-y^2}} x g(x | y) dx \\ &= \int_{-\sqrt{1-y^2}}^{+\sqrt{1-y^2}} x \frac{1}{2\sqrt{1-y^2}} dx = \frac{1}{2\sqrt{1-y^2}} \left. \frac{x^2}{2} \right|_{-\sqrt{1-y^2}}^{+\sqrt{1-y^2}} \\ &= 0. \end{aligned}$$

Puede suceder que una o ambas curvas de regresión sean en realidad líneas rectas (figura 7.12). Es decir, $E(Y | x)$ puede ser una función *lineal* de x y/o $E(X | y)$ puede ser función lineal de y . En este caso decimos que la regresión del promedio de Y sobre X (por ejemplo) es lineal.

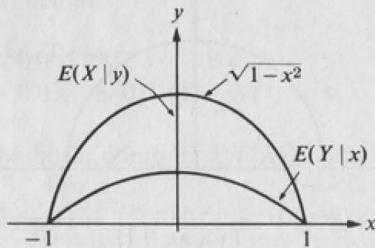


FIGURA 7.12

EJEMPLO 7.25. Supongamos que (X, Y) está distribuida uniformemente en el triángulo indicado en la figura 7.13. Por tanto $f(x, y) = 1$, $(x, y) \in T$. Las expresiones siguientes para las fdp marginal y condicional se verifican fácilmente:

$$g(x) = 2x, \quad 0 \leq x \leq 1; \quad h(y) = \frac{2-y}{2}, \quad 0 \leq y \leq 2.$$

$$g(x | y) = \frac{2}{2-y}, \quad y/2 \leq x \leq 1; \quad h(y | x) = \frac{1}{2x}, \quad 0 \leq y \leq 2x.$$

Así $E(Y | x) = \int_0^{2x} y h(y | x) dy = \int_0^{2x} y (1/2x) dy = x$. De modo semejante,

$$E(X | y) = \int_{y/2}^1 x g(x | y) dx = \int_{y/2}^1 x \frac{2}{2-y} dx = \frac{y}{4} + \frac{1}{2}.$$

Así *ambas* regresiones de Y sobre X y de X sobre Y son lineales (figura 7.14).

Resulta que si la regresión del promedio de Y sobre X es lineal, por ejemplo $E(Y | x) = \alpha x + \beta$, podemos expresar fácilmente los coeficientes α y β en función de ciertos parámetros de la distribución conjunta de (X, Y) . Tenemos el teorema siguiente.

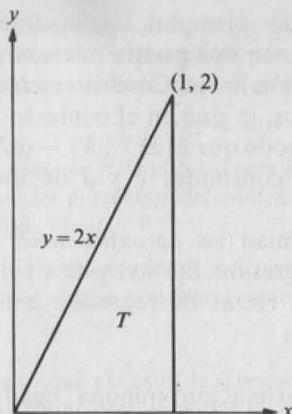


FIGURA 7.13

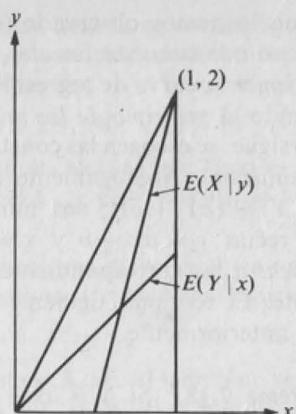


FIGURA 7.14

Teorema 7.17. Sea (X, Y) una variable aleatoria bidimensional y supongamos que

$$E(X) = \mu_x, \quad E(Y) = \mu_y, \quad V(X) = \sigma_x^2, \quad \text{y} \quad V(Y) = \sigma_y^2.$$

Sea ρ el coeficiente de correlación entre X y Y . Si la regresión de Y sobre X es lineal, tenemos

$$E(Y | x) = \mu_y + \rho \frac{\sigma_y}{\sigma_x} (x - \mu_x). \quad (7.27)$$

Si la regresión de X sobre Y es lineal, tenemos

$$E(X | y) = \mu_x + \rho \frac{\sigma_x}{\sigma_y} (y - \mu_y). \quad (7.28)$$

Demostración: la demostración de este teorema está bosquejada en el problema 7.44.

Observaciones: (a) Como se sugirió en la exposición anterior, es posible que *una* de las regresiones del promedio sea lineal mientras que la otra no lo sea.

(b) Nótese el importante papel desempeñado por el coeficiente de correlación en las expresiones interiores. Si la regresión de X sobre Y , por ejemplo, es lineal, y si $\rho = 0$, entonces (nuevamente) encontramos que $E(X | y)$ no depende de y . Nótese también que el signo algebraico de ρ determina el signo de la pendiente de regresión.

(c) Si ambas funciones de regresión son lineales, encontramos, al resolver las ecuaciones (7.27) y (7.28) simultáneamente, que las rectas de regresión se cortan en el «centro» de la distribución, (μ_x, μ_y) .

Como lo hemos observado (ejemplo 7.23, por ejemplo), las funciones de regresión no necesitan ser lineales. Sin embargo, aún nos podría interesar el tratar de *aproximar* la curva de regresión con una función lineal. Corrientemente se hace recurriendo al *principio de los mínimos cuadrados*, lo que en el contexto presente es como sigue: se escogen las constantes a y b de modo que $E[E(Y | X) - (aX + b)]^2$ sea minimizada. Análogamente se escogen las constantes c y d de modo que $E[E(X | Y) - (cY + d)]^2$ sea minimizada.

Las rectas $y = ax + b$ y $x = cy + d$ se llaman las *aproximaciones mínimas cuadráticas* a las correspondientes curvas de regresión $E(Y | x)$ y $E(X | y)$, respectivamente. El teorema siguiente relaciona esas rectas de regresión con las expuestas anteriormente.

Teorema 7.18. Si $y = ax + b$ es la aproximación mínima cuadrática a $E(Y | x)$ y si $E(Y | x)$ es en realidad una función lineal de x , es decir

$$E(Y | x) = a'x + b',$$

entonces $a = a'$ y $b = b'$. Para la regresión de X sobre Y , se mantiene una proposición análoga.

Demostración: ver el problema 7.45.

PROBLEMAS

7.1. Encontrar el valor esperado de las siguientes variables aleatorias.

- (a) La variable aleatoria X definida en el problema 4.1.
- (b) La variable aleatoria X definida en el problema 4.2.
- (c) La variable aleatoria T definida en el problema 4.6.
- (d) La variable aleatoria X definida en el problema 4.18.

7.2. Demostrar que $E(X)$ no existe para la variable aleatoria X definida en el problema 4.25.

7.3. Lo siguiente representa la distribución de probabilidades de D , la demanda diaria de cierto producto. Calcular $E(D)$.

$$\begin{aligned} d: & \quad 1, 2, 3, 4, 5, \\ P(D=d): & \quad 0,1, 0,1, 0,3, 0,3, 0,2. \end{aligned}$$

7.4. En la fabricación del petróleo, la temperatura de destilación T (grados centígrados), es crucial para determinar la calidad del producto final. Supongamos que T se considera como una variable aleatoria distribuida uniformemente en $(150, 300)$.

Supongamos que cuesta C_1 pesos producir un galón de petróleo. Si el aceite destila a una temperatura menor que $200^\circ C$, el producto se vende como nafta y se vende a C_2 pesos por galón. Si se destila a una temperatura mayor que $200^\circ C$, se conoce como aceite destilado refinado y se vende en C_3 pesos por galón. Encontrar la utilidad neta esperada (por galón).

7.5. Cierta aleación se forma al combinar la mezcla fundida de dos metales. La aleación que resulta contiene cierto porcentaje de plomo, X , que puede considerarse como una variable aleatoria. Supongamos que X tiene la siguiente fdp:

$$f(x) = \frac{3}{2}10^{-5}x(100 - x), \quad 0 \leq x \leq 100.$$

Supongamos que P , la utilidad neta obtenida al vender esta aleación (por libra), es la siguiente función del porcentaje del contenido de plomo: $P = C_1 + C_2X$. Calcular la utilidad esperada (por libra).

7.6. Supóngase que un instrumento electrónico tiene una duración X (en unidades de 1000 horas) que se considera como una variable aleatoria continua con la siguiente fdp:

$$f(x) = e^{-x}, \quad x > 0.$$

Suponiendo que el costo de fabricación de tal artículo es \$2.00. El fabricante vende el artículo por \$5.00, pero garantiza un reembolso total si $X \leq 0.9$. ¿Cuál es la utilidad esperada del fabricante por artículo?

7.7. Las 5 primeras repeticiones de un experimento cuestan \$10.00 cada una. Todas las repeticiones que siguen cuestan \$5.00 cada una. Suponiendo que el experimento se repite hasta obtener el primer resultado exitoso. Si la probabilidad de un resultado exitoso es siempre igual a 0.9 y si las repeticiones son independientes, ¿cuál es el costo esperado de la operación completa?

7.8. Se sabe que un lote contiene 2 artículos defectuosos y 8 no defectuosos. Si estos artículos se inspeccionan al azar, uno después de otro, ¿cuál es el número esperado de artículos que se deben escoger para inspección a fin de sacar todos los defectuosos?

7.9. Un lote de 10 motores eléctricos debe ser rechazado totalmente o bien vendido, según el resultado del siguiente proceso: se escogen al azar dos motores y se inspeccionan. Si uno o más son defectuosos, el lote es rechazado; de otro modo es aceptado. Supóngase que cada uno de los motores cuesta \$75 y se vende por \$100; si el lote contiene 1 motor defectuoso, ¿cuál es la utilidad esperada del fabricante?

7.10. Suponiendo que D , la demanda diaria de un artículo, es una variable aleatoria con la siguiente distribución de probabilidades:

$$P(D = d) = C2^d/d!, \quad d = 1, 2, 3, 4.$$

- (a) Evaluar la constante C .
- (b) Calcular la demanda esperada.

(c) Suponer que un artículo se vende por \$5.00. Un fabricante produce diariamente K artículos. Cualquier artículo que no se venda al término del día, debe descartarse con una pérdida de \$3.00. (i) Encontrar la distribución de probabilidades de la utilidad diaria, como una función de K . (ii) ¿Cuántos artículos deberían fabricarse para maximizar la utilidad diaria esperada?

7.11. (a) Con $N = 50$, $p = 0.3$, efectuar algunos cálculos para encontrar cuál es el valor de k que minimiza $E(X)$ en el ejemplo 7.12.

(b) Usando los valores anteriores de N y p y usando $k = 5, 10, 25$, determinar para cada uno de los valores de k si es preferible el «examen del grupo».

7.12. Suponiendo que X y Y son variables aleatorias independientes con las siguientes fdp:

$$f(x) = 8/x^3, \quad x > 2; \quad g(y) = 2y, \quad 0 < y < 1.$$

- (a) Encontrar la fdp de $Z = XY$.
 (b) Obtener $E(Z)$ de dos maneras: (i) usando la fdp de Z como se obtuvo en (a); (ii) directamente, sin usar la fdp de Z .

7.13. Suponiendo que X tiene fdp

$$f(x) = 8/x^3, \quad x > 2.$$

Sea $W = \frac{1}{3}X$.

- (a) Calcular $E(W)$, usando la fdp de W .
 (b) Calcular $E(W)$, sin usar la fdp de W .

7.14. Se lanza un dado regular 72 veces. Puesto que X es el número de veces que aparece el seis, evaluar $E(X^2)$.

7.15. Encontrar el valor esperado y la varianza de la variable aleatoria Y y Z del problema 5.2.

7.16. Encontrar el valor esperado y la varianza de la variable aleatoria Y del problema 5.3.

7.17. Encontrar el valor esperado y la varianza de las variables aleatorias Y y Z del problema 5.5.

7.18. Encontrar el valor esperado y la varianza de las variables aleatorias Y, Z y W del problema 5.6.

7.19. Encontrar el valor esperado y la varianza de las variables aleatorias V y S del problema 5.7.

7.20. Encontrar el valor esperado y la varianza de la variable aleatoria Y del problema 5.10 para cada uno de los tres casos.

7.21. Encontrar el valor esperado y la varianza de la variable aleatoria A del problema 6.7.

7.22. Encontrar el valor esperado y la varianza de la variable aleatoria H del problema 6.11.

7.23. Encontrar el valor esperado y la varianza de la variable aleatoria W del problema 6.13.

7.24. Supóngase que X es una variable aleatoria para la cual $E(X) = 10$ y $V(X) = 25$. ¿Para qué valores positivos de a y b tiene $Y = aX - b$ esperanza 0 y varianza 1?

7.25. Supóngase que S , un voltaje aleatorio, varía entre 0 y 1 volt y está distribuido uniformemente en ese intervalo. Suponiendo que la señal S es perturbada por un ruido aleatorio independiente aditivo N , que está distribuido uniformemente entre 0 y 2 volts.

- (a) Encontrar el voltaje esperado de la señal, tomando en cuenta el ruido.
 (b) Encontrar la potencia esperada cuando la señal perturbada se aplica a una resistencia de 2 ohms.

7.26. Supóngase que X está distribuida uniformemente en $[-a, 3a]$. Encontrar la varianza de X .

7.27. Un blanco está formado por tres círculos concéntricos de radios $1/\sqrt{3}$, 1 y $\sqrt{3}$ pies. Los disparos dentro del círculo interior valen 4 puntos, dentro del anillo siguiente 3 puntos, y dentro del tercer anillo 2 puntos. Los disparos fuera del blanco valen cero. Sea R la varia-

ble aleatoria que representa la distancia del impacto al centro. Suponiendo que la fdp de R sea $f(r) = 2/\pi(1 + r^2)$, $r > 0$. Calcular el valor esperado del puntaje después de 5 disparos.

7.28. Supóngase que la variable aleatoria continua X tiene la fdp

$$f(x) = 2xe^{-x^2}, \quad x \geq 0.$$

Sea $Y = X^2$. Calcular $E(Y)$:

- (a) directamente sin obtener primero la fdp de Y ,
- (b) obteniendo primero la fdp de Y .

7.29. Suponiendo que la variable aleatoria bidimensional (X, Y) está distribuida uniformemente en el triángulo de la figura 7.15. Evaluar $V(X)$ y $V(Y)$.

7.30. Suponiendo que (X, Y) está distribuida uniformemente en el triángulo de la figura 7.16. Evaluar la fdp marginal de X y de Y .

- (a) Obtener la fdp marginal de X y de Y .
- (b) Evaluar $V(X)$ y $V(Y)$.

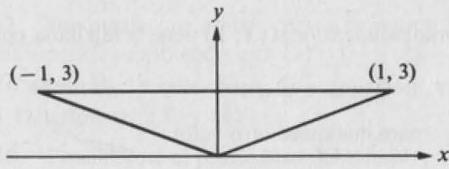


FIGURA 7.15

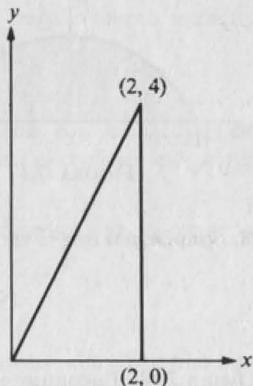


FIGURA 7.16

7.31. Supóngase que X y Y son variables aleatorias para las cuales $E(X) = \mu_x$, $E(Y) = \mu_y$, $V(X) = \sigma_x^2$ y $V(Y) = \sigma_y^2$. Usando el teorema 7.7 obtenga una aproximación para $E(Z)$ y $V(Z)$, en donde $Z = X/Y$.

7.32. Supóngase que X y Y son variables aleatorias independientes, distribuidas cada una uniformemente en $(1, 2)$. Sea $Z = X/Y$.

- (a) Usando el teorema 7.7 obtener expresiones aproximadas para $E(Z)$ y $V(Z)$.
- (b) Usando el teorema 6.5 obtener la fdp de Z y luego encontrar el valor exacto de $E(Z)$ y $V(Z)$. Comparar con (a).

7.33. Demostrar que si X es una variable aleatoria continua con fdp f que tiene la propiedad de que el gráfico de f es simétrico respecto a $x = a$, luego $E(X) = a$, dado que existe $E(X)$. (Ver el ejemplo 7.16.)

7.34. (a) Supóngase que la variable aleatoria X toma los valores -1 y 1 cada uno con probabilidad $1/2$. Considerando $P[|X - E(X)| \geq k\sqrt{V(X)}]$ como una función de k , $k > 0$.

Dibuja esta función de k y, en el mismo sistema de coordenadas, dibuja la cota superior de la probabilidad anterior como está dada por la desigualdad de Chebyshev.

(b) Lo mismo que en (a) excepto que $P(X = -1) = 1/3$, $P(X = 1) = 2/3$.

7.35. Comparar la cota superior de la probabilidad $P[|X - E(X)| \geq 2\sqrt{V(X)}]$ obtenida de la desigualdad de Chebyshev con la probabilidad exacta de X , si X está distribuida uniformemente en $(-1, 3)$.

7.36. Verificar la ecuación (7.17).

7.37. Supóngase que la variable aleatoria bidimensional (X, Y) está distribuida uniformemente en R , donde R está definida por $\{(x, y) | x^2 + y^2 \leq 1, y \geq 0\}$. (Ver la Fig. 7.17.) Evaluar ρ_{xy} , el coeficiente de correlación.

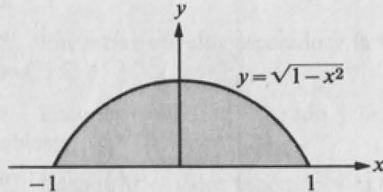


FIGURA 7.17

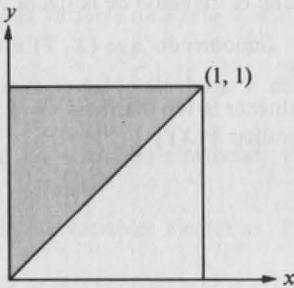


FIGURA 7.18

7.38. Supóngase que la variable aleatoria bidimensional (X, Y) tiene la fdp dada por

$$f(x, y) = ke^{-y}, \quad 0 < x < y < 1$$

$$= 0, \quad \text{para cualquier otro valor.}$$

(ver la figura 7.18). Encontrar el coeficiente de correlación ρ_{xy} .

7.39. El ejemplo siguiente ilustra que $\rho = 0$ no implica independencia. Suponiendo que (X, Y) tiene una distribución conjunta de probabilidades dada por la tabla 7.1.

- (a) Demostrar que $E(XY) = E(X)E(Y)$ y luego $\rho = 0$.
- (b) Indicar por qué X y Y no son independientes.
- (c) Demostrar que este ejemplo puede generalizarse como sigue. La elección del número $1/8$ no es crucial. Lo importante es que todos los valores dentro de un círculo son los mismos,

TABLA 7.1

$X \backslash Y$	-1	0	1
-1	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$
0	$\frac{1}{8}$	0	$\frac{1}{8}$
1	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$

7.40. Supóngase que A y B son dos sucesos asociados con un experimento ε . Supóngase que $P(A) > 0$ y $P(B) > 0$. Se definen las variables aleatorias X y Y como sigue

$$X = 1 \quad \text{si } A \text{ ocurre y } 0 \text{ en otro caso,}$$

$$Y = 1 \quad \text{si } B \text{ ocurre y } 0 \text{ en otro caso.}$$

Demostrar que $\rho_{xy} = 0$ implica que X y Y son independientes.

7.41. Demostrar el teorema 7.14.

7.42. Para la variable aleatoria (X, Y) definida en el problema 6.5, calcular $E(X | y)$, $E(Y | x)$ y verificar que $E(X) = E[E(X | Y)]$ y $E(Y) = E[E(Y | X)]$.

7.41. Demostrar el teorema 7.14.

7.42. Para la variable aleatoria (X, Y) definida en el problema 6.5, calcular $E(X | y)$, $E(Y | x)$, y verificar que $E(X) = E[E(X | Y)]$ y $E(Y) = E[E(Y | X)]$.

7.43. Demostrar el teorema 7.16.

7.44. Demostrar el teorema 7.17. [Indicación: en el caso continuo, multiplique la ecuación $E(Y|x) = Ax + B$ por $g(x)$, la fdp de X , e integre de $-\infty$ a ∞ . Haga lo mismo, usando $xg(x)$ y luego resuelva las dos ecuaciones resultantes para A y para B].

7.45. Demostrar el teorema 7.18.

7.46. Si X , Y y Z son variables aleatorias no correlacionadas con desviaciones estándar 5, 12 y 9, respectivamente y si $U = X + Y$ y $V = Y + Z$, evaluar el coeficiente de correlación entre U y V .

7.47. Supóngase que ambas curvas de regresión de los promedios son realmente lineales. Específicamente, suponiendo que $E(Y|x) = -\frac{3}{2}x - 2$ y $E(X|y) = -\frac{3}{5}y - 3$.

(a) Determinar el coeficiente de correlación ρ .

(b) Determinar $E(X)$ y $E(Y)$.

7.48. Considérese el pronóstico del tiempo en dos alternativas: «lluvia» o «no lluvia» en las próximas 24 horas. Suponiendo que $p = \text{Prob}(llovía \text{ en las próximas 24 horas}) > \frac{1}{2}$. El pronosticador anota 1 punto si acierta y 0 si no. Al hacer n pronósticos, un pronosticador sin destreza elige al azar r días cualesquiera ($0 \leq r \leq n$) para decir «llueve» y los $n - r$ días restantes para decir «no llueve». Su puntaje total anotado es S_n . Compute $E(S_n)$ y $\text{Var}(S_n)$ y encuentre el valor de r para el cual $E(S_n)$ es el mayor. [Indicación: sea $X_i = 1$ ó 0 dependiendo si el i -ésimo pronóstico es correcto o no. Luego $S_n = \sum_{i=1}^n X_i$. Nótese que los X_i no son independientes].

La variable aleatoria de Poisson y otras variables aleatorias

8.1 La distribución de Poisson

Tal como en los modelos determinísticos, en los cuales ciertas relaciones funcionales desempeñan un papel importante (tales como lineales, cuadráticas, exponenciales, trigonométricas, etc.) también, al elaborar modelos no determinísticos para fenómenos observables, encontramos que ciertas distribuciones de probabilidades aparecen más a menudo que otras. Una razón de esto es que, como en el caso determinístico, ciertos modelos matemáticos relativamente simples parecen ser capaces de describir un gran número de fenómenos.

En este capítulo expondremos con mucho detalle diversas variables aleatorias discretas. En el capítulo siguiente haremos lo mismo con variables aleatorias continuas.

Presentemos formalmente la siguiente variable aleatoria. Posteriormente indicaremos bajo qué condiciones esta variable aleatoria podría representar al resultado de un experimento aleatorio.

Definición. Sea X una variable aleatoria que toma los valores posibles: $0, 1, \dots, n, \dots$. Si

$$P(X = k) = \frac{e^{-\alpha} \alpha^k}{k!}, \quad k = 0, 1, \dots, n, \dots, \quad (8.1)$$

decimos que X tiene una *distribución de Poisson* con parámetro $\alpha > 0$.

Para verificar que la anterior representa a una legítima distribución de probabilidades, sencillamente observemos que $\sum_{k=0}^{\infty} P(X = k) = \sum_{k=0}^{\infty} (e^{-\alpha} \alpha^k / k!) = e^{-\alpha} e^{\alpha} = 1$.

Observación: puesto que estamos definiendo directamente la variable aleatoria en función de su recorrido y distribución de probabilidades, sin referencia a ningún espacio muestral original S , podemos suponer que el espacio muestral S ha sido identificado con R_X y que $X(s) = s$. Es decir, los resultados del experimento son sencillamente los números $0, 1, 2, \dots$ y las probabilidades asociadas con cada uno de esos resultados están dados por la ecuación (8.1).

Teorema 8.1. Si X tiene una distribución de Poisson con parámetro α , entonces $E(X) = \alpha$ y $V(x) = \alpha$.

$$\text{Demostración: } E(X) = \sum_{k=0}^{\infty} \frac{ke^{-\alpha}\alpha^k}{k!} = \sum_{k=1}^{\infty} \frac{e^{-\alpha}\alpha^k}{(k-1)!}.$$

Haciendo $s = k - 1$, encontramos que se transforma

$$E(X) = \sum_{s=0}^{\infty} \frac{e^{-\alpha}\alpha^{s+1}}{s!} = \alpha \sum_{s=0}^{\infty} \frac{e^{-\alpha}\alpha^s}{s!} = \alpha.$$

De modo semejante,

$$E(X^2) = \sum_{k=0}^{\infty} \frac{k^2 e^{-\alpha}\alpha^k}{k!} = \sum_{k=1}^{\infty} k \frac{e^{-\alpha}\alpha^k}{(k-1)!}.$$

Haciendo nuevamente $s = k - 1$, obtenemos

$$E(X^2) = \sum_{s=0}^{\infty} (s+1) \frac{e^{-\alpha}\alpha^{s+1}}{s!} = \alpha \sum_{s=0}^{\infty} s \frac{e^{-\alpha}\alpha^s}{s!} + \alpha \sum_{s=0}^{\infty} \frac{e^{-\alpha}\alpha^s}{s!} = \alpha^2 + \alpha$$

[puesto que la primera suma representa $E(X)$ mientras que la segunda suma es igual a uno]. Luego

$$V(X) = E(X^2) - (E(X))^2 = \alpha^2 + \alpha - \alpha^2 = \alpha.$$

Observación: nótese la interesante propiedad que posee la variable aleatoria de Poisson: su esperanza es igual a su varianza.

8.2 La distribución de Poisson como una aproximación a la distribución binomial

La distribución de Poisson desempeña un papel importante por derecho propio como un modelo probabilístico apropiado para un gran número de fenómenos aleatorios. Este punto se expondrá en la sección siguiente. Aquí nos interesa la importancia de esta distribución para aproximar las probabilidades binomiales.

EJEMPLO 8.1. Suponiendo que las llamadas telefónicas llegan a una gran central telefónica y que en un período especial de tres horas (180 minutos) se han recibido un total de 270 llamadas, o sea 1,5 llamadas por minuto. Supóngase que, con base en la evidencia anterior, queremos calcular la probabilidad de recibir 0, 1, 2, etc. llamadas durante los próximos tres minutos.

Al considerar el fenómeno de llamadas recibidas, podríamos llegar a la conclusión de que en *cualquier* instante es tan probable que ocurra una llamada telefónica como en *cualquier* otro instante. Es decir, la probabilidad permanece constante en un intervalo de tiempo. La dificultad es que aún en un intervalo de tiempo muy corto, el número de puntos no sólo es infinito sino que no puede ser enumerado. Esto nos lleva a una serie de aproximaciones que describiremos ahora.

Para empezar, podríamos considerar la subdivisión del intervalo de tres minutos en nueve subintervalos de 20 segundos cada uno. Podríamos considerar entonces cada uno de esos nueve intervalos como un experimento de Bernoulli durante el cual observamos una llamada (éxito) o ninguna llamada (fracaso) con $P(\text{éxito}) = (1,5) 20/60 = 0,5$. Así podríamos estar inclinados a decir que la pro-

babilidad de dos llamadas durante el intervalo de tres minutos (es decir, 2 éxitos en 9 experimentos con $P(\text{éxito}) = (0,5)$) es igual a $\binom{9}{2}(1/2)^9 = 9/128$.

La dificultad con esta aproximación es que ignoramos las posibilidades de, digamos, dos o tres, etc., llamadas durante uno de nuestros ensayos con períodos de 20 segundos. Si esta posibilidad se tomase en consideración, el uso anterior de la distribución binomial no sería legítimo ya que esa distribución es aplicable sólo cuando existe una dicotomía, una llamada o ninguna llamada.

Para evitar esta dificultad hacemos la aproximación siguiente y, realmente, nos lleva a una sucesión completa de aproximaciones. Una manera de estar más o menos seguro de que al menos se recibe una llamada en la central durante un pequeño intervalo de tiempo, es hacer que ese intervalo sea muy corto. Así, en vez de considerar nueve intervalos de 20 segundos de duración, consideraremos los 18 intervalos siguientes, cada uno de 10 segundos de largo. Podemos representar ahora nuestro experimento como 18 experimentos de Bernoulli con $P(\text{éxito}) = P(\text{recibir una llamada durante un subintervalo}) = (1,5)10/60 = 0,25$. Por tanto $P(\text{dos llamadas durante el intervalo de tres minutos}) = (\frac{1}{2}^8)(0,25)^2(0,75)^{16}$. Nótese que aunque ahora tratamos una distribución binomial diferente a la de antes (es decir, que tiene parámetros $n = 18$, $p = 0,25$ en vez de $n = 9$, $p = 0,5$), el valor esperado es el mismo, a saber $np = 18(0,25) = 9(0,5) = 4,5$.

Si continuamos de esta manera, aumentando el número de subintervalos (es decir, n), disminuiremos al mismo tiempo la probabilidad de recibir una llamada (es decir, p) de tal manera que np permanece constante.

Así, el ejemplo precedente nos conduce a formular la pregunta siguiente: ¿qué sucede a las probabilidades binomiales $\binom{n}{k}p^k(1-p)^{n-k}$ si $n \rightarrow \infty$ y $p \rightarrow 0$, de tal manera que np permanezca constante, es decir, $np = \alpha$?

Los cálculos siguientes dan la respuesta a esta pregunta muy importante.

Consideremos la expresión general para la probabilidad binomial,

$$\begin{aligned} P(X = k) &= \binom{n}{k} p^k (1-p)^{n-k} = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \\ &= \frac{n(n-1)(n-2)\cdots(n-k+1)}{k!} p^k (1-p)^{n-k}. \end{aligned}$$

Hagamos $np = \alpha$. Por tanto $p = \alpha/n$, y $1-p = 1-\alpha/n = (n-\alpha)/n$. Sustituyendo todos los términos que contienen p por sus expresiones equivalentes en función de α , obtenemos

$$\begin{aligned} P(X = k) &= \frac{n(n-1)\cdots(n-k+1)}{k!} \left(\frac{\alpha}{n}\right)^k \left(\frac{n-\alpha}{n}\right)^{n-k} \\ &= \frac{\alpha^k}{k!} \left[(1) \left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right) \right] \left[1 - \frac{\alpha}{n} \right]^{n-k} \\ &= \frac{\alpha^k}{k!} \left[(1) \left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right) \right] \\ &\quad \times \left(1 - \frac{\alpha}{n}\right)^n \left(1 - \frac{\alpha}{n}\right)^{-k}. \end{aligned}$$

Ahora hagamos $n \rightarrow \infty$ de tal manera que $np = \alpha$ permanezca constante. Esto significa obviamente que $p \rightarrow 0$ cuando $n \rightarrow \infty$, porque de otra manera np no podría permanecer constante. (Análogamente, podríamos necesitar que $n \rightarrow \infty$ y $p \rightarrow 0$ de tal manera que $np \rightarrow \alpha$.)

En la expresión anterior los términos de la forma $(1 - 1/n)$, $(1 - 2/n)$, ... tienden a uno cuando n tiende a infinito, tal como lo hace $(1 - \alpha/n)^{-k}$. Es bien sabido (de la definición del número e) que $(1 - \alpha/n)^n \rightarrow e^{-\alpha}$ cuando $n \rightarrow \infty$.

Así, $\lim_{n \rightarrow \infty} P(X = k) = e^{-\alpha} \alpha^k / k!$ Es decir, en el límite obtenemos la distribución de Poisson con parámetro α . Resumamos este importante resultado en el siguiente teorema.

Teorema 8.2. Sea X una variable aleatoria distribuida binomialmente con parámetro p (con base en n repeticiones del experimento). Esto es,

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}.$$

Supóngase que cuando $n \rightarrow \infty$, $np = \alpha$ (constante), o equivalente, cuando $n \rightarrow \infty$, $p \rightarrow 0$ tal que $np \rightarrow \alpha$. Bajo estas condiciones tenemos

$$\lim_{n \rightarrow \infty} P(X = k) = \frac{e^{-\alpha} \alpha^k}{k!},$$

la distribución de Poisson con parámetro α .

Observaciones: (a) El teorema anterior, esencialmente, dice que podemos aproximar las probabilidades binomiales con las probabilidades de la distribución de Poisson siempre que n sea grande y p pequeño.

(b) Ya hemos verificado que si X tiene una distribución binomial, $E(X) = np$, mientras que si X tiene una distribución de Poisson (con parámetro α), $E(X) = \alpha$.

(c) La distribución binomial está caracterizada por dos parámetros, n y p , mientras que la distribución de Poisson está caracterizada por un solo parámetro, $\alpha = np$, que representa al número esperado de éxitos por unidad de tiempo (o por unidad de espacio en algún otro caso). Este parámetro también se designa como la *intensidad* de la distribución. Es importante distinguir entre el número esperado de ocurrencias por *unidad* de tiempo y el número esperado de ocurrencias en el tiempo especificado. Así, en el ejemplo 8.1 la intensidad es 1,5 llamadas por minuto y por lo tanto el número esperado de llamadas es un período de 10 minutos, por ejemplo, sería 15.

(d) También podemos considerar el siguiente argumento para evaluar la varianza de una variable aleatoria X de Poisson con parámetro α : X se puede considerar un caso límite de una variable aleatoria Y distribuida binomialmente con parámetro n y p , en donde $n \rightarrow \infty$ y $p \rightarrow 0$ de tal manera que $np \rightarrow \alpha$. Puesto que $E(Y) = np$ y $\text{Var}(Y) = np(1 - p)$, observemos que en el límite $\text{Var}(Y) \rightarrow \alpha$.

Hay extensas tablas disponibles para la distribución de Poisson. (E. C. Molina, *Poisson's Exponential Binomial Limit*, D. Van Nostrand Company, Inc., New York, 1942). Una breve tabulación de esta distribución se da en el apéndice.

Consideremos otros tres ejemplos adicionales que ilustran las aplicaciones de la distribución de Poisson mencionadas previamente.

EJEMPLO 8.2. En una concurrida intersección de tráfico la probabilidad p de que un automóvil tenga un accidente es muy escasa, digamos $p = 0,0001$. Sin embargo, durante cierta parte del día, entre las 4 p.m., y las 6 pm., un gran número de automóviles pasa por la intersección, digamos 1000. Bajo dichas condiciones, ¿cuál es la probabilidad de que dos o más accidentes ocurran durante ese período?

Formulemos algunas hipótesis. Supongamos, en primer lugar, que el valor anterior de p es el mismo para cada uno de los automóviles. En segundo lugar, supongamos que si un automóvil tiene o no accidentes, no depende de lo que sucede a cualquier otro automóvil. (Esta suposición, obviamente, no es realista pero sin embargo la formularemos). Podemos suponer así que si X es el número de accidentes entre los 1000 automóviles que llegan, entonces X tiene una distribución binomial con $p = 0,0001$. (Otra hipótesis, no indicada explícitamente, es que n , el número de automóviles que pasa por la intersección entre 4p.m. y 6p.m., está predeterminada en 1000. Evidentemente, un enfoque más realista sería considerar n misma como una variable aleatoria cuyo valor depende de un mecanismo aleatorio. Sin embargo, no haremos esto aquí, sino que consideraremos n como fija). Por tanto podemos obtener el valor exacto de la probabilidad buscada:

$$\begin{aligned} P(X \geq 2) &= 1 - P(X = 0) - P(X = 1) \\ &= 1 - (0,9999)^{1000} - 1000(0,0001)(0,9999)^{999}. \end{aligned}$$

La evaluación de los valores anteriores da origen a una dificultad considerable. Puesto que n es grande y p es pequeño, aplicamos el teorema 8.2 y obtenemos la aproximación siguiente:

$$P(X = k) \cong \frac{e^{-0.1}(0.1)^k}{k!}.$$

Por tanto,

$$P(X \geq 2) \cong 1 - e^{0.1}(1 + 0.1) = 0,0045.$$

EJEMPLO 8.3. Supóngase que un proceso de fabricación produce artículos de tal manera que cierta proporción (constante) de artículos, digamos p , son defectuosos. Si se obtiene un lote n de tales artículos, la probabilidad de obtener exactamente k defectuosos puede ser calculada de la distribución binomial como $P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$, en donde X es el número de defectuosos en el lote. Si n es grande y p es pequeña (como a menudo es el caso), debemos aproximar la probabilidad anterior por

$$P(X = k) \cong \frac{e^{-np}(np)^k}{k!}.$$

Supóngase, por ejemplo, que un fabricante produce artículos de los cuales alrededor de 1 en 1000 son defectuosos. Esto es, $p = 0,001$. Por tanto, usando la dis-

tribución binomial, encontramos que en un lote de 500 artículos la probabilidad de que ninguno sea defectuoso es $(0,999)^{500} = 0,609$. Si aplicamos la aproximación de Poisson, esta probabilidad puede escribirse como $e^{-0,5} = 0,61$. La probabilidad de encontrar 2 o más artículos defectuosos es, de acuerdo con la aproximación de Poisson, $1 - e^{-0,5}(1 + 0,5) = 0,085$.

EJEMPLO 8.4. [Sugerido por una exposición en *Cálculo de Probabilidades* de A. Renyi (en alemán), VEB Deutscher Verlag der Wissenschaft, Berlin, 1962].

En la fabricación de botellas de vidrio, se encuentran partículas duras y pequeñas en el vidrio fundido del cual se hacen las botellas. Si aparece una sola partícula en una botella, la botella no puede ser usada y debe ser descartada. Puede suponerse que las partículas están espaciadas al azar en el vidrio fundido. Supondremos que el vidrio fundido se produce de tal manera que el número de partículas es (en promedio) el mismo para una cantidad constante de vidrio fundido. Supóngase en particular que en 100 kilos de vidrio fundido, se encuentran x de tales partículas y que es necesario 1 kilo de vidrio fundido para hacer una de estas botellas.

Pregunta: ¿qué porcentaje de botellas tendrá que ser descartado debido a que son defectuosas? A primera vista la «solución» de este problema podría ser como sigue. Puesto que el material para 100 botellas contiene x partículas, aproximadamente el x por ciento de las botellas deberá ser descartado. Una pequeña reflexión indicará, sin embargo, que la solución anterior no es correcta, ya que una botella defectuosa puede tener más de 1 partícula, bajando así el porcentaje de botellas defectuosas obtenidas del material que sobra.

A fin de obtener una solución «correcta», hagamos las siguientes hipótesis simplificantes: (a) cada una de las partículas puede aparecer en el material de cada una de las botellas con igual probabilidad y (b) la distribución de cualquier partícula es independiente de cualquier otra partícula. Con estas hipótesis, podemos reducir nuestro problema al siguiente modelo de «urna». Entre N urnas, se distribuyen al azar n esferas. ¿Cuál es la probabilidad de que en una urna elegida aleatoriamente, se encuentren exactamente k esferas? (Las urnas evidentemente corresponden a las partículas.)

Siendo Z el número de esferas encontradas en una urna escogida aleatoriamente, se deduce, de la hipótesis anterior, que Z está distribuida binomialmente con parámetro $1/N$. Luego,

$$P(Z = k) = \binom{n}{k} \left(\frac{1}{N}\right)^k \left(1 - \frac{1}{N}\right)^{n-k}.$$

Supóngase ahora que el vidrio fundido se prepara en cantidades muy grandes. En realidad, supongamos que se prepara en unidades de 100 kilos, y que se han suministrado M de tales unidades. Luego $N = 100M$ y $n = xM$. Sea $\alpha = x/100$, lo que iguala la proporción de partículas por botella. Así $N = n/\alpha$ y la probabilidad anterior puede escribirse como

$$P(Z = k) = \binom{n}{k} \left(\frac{\alpha}{n}\right)^k \left(1 - \frac{\alpha}{n}\right)^{n-k}.$$

Así, cuando el proceso de producción continúa (esto es, $M \rightarrow \infty$ y luego $n \rightarrow \infty$) obtenemos

$$P(Z = k) \simeq \frac{e^{-\alpha} \alpha^k}{k!} \quad \text{en donde } \alpha = \frac{x}{100}.$$

Calculemos la probabilidad de que deba descartarse una botella. Esta es igual a $1 - P(Z = 0)$. Luego $P(\text{botella defectuosa}) \simeq 1 - e^{-x/100}$. Si el número de botellas producidas es muy grande, debemos identificar obligadamente la probabilidad de una botella defectuosa con la frecuencia relativa de botellas defectuosas. Por tanto el porcentaje de botellas defectuosas es aproximadamente $100(1 - e^{-x/100})$. Si desarrollamos $100(1 - e^{-x/100})$ en una serie de Maclaurin, obtenemos

$$\begin{aligned} 100 \left[1 - \left(1 - \frac{x}{100} + \frac{x^2}{2(100)^2} - \frac{x^3}{3!(100)^3} + \dots \right) \right] \\ = x - \frac{x^2}{2(100)} + \frac{x^3}{6(100)^2} - \dots \end{aligned}$$

Así, si x es pequeña, la proporción de botellas descartadas es aproximadamente x como se sugirió primero. Sin embargo, para un x mayor esto no es válido. En caso que $x = 100$, el porcentaje de botellas descartadas *no* es 100, sino que $100(1 - e^{-1}) = 63,21$ por ciento. Este, naturalmente, es un caso extremo y no se encontraría en un proceso controlado razonablemente. Supongamos que $x = 30$ (un número más realista). Por tanto en vez de descartar el 30 por ciento (nuestra solución inicial nuevamente), descartaríamos sólo $100(1 - e^{-0,3}) = 25,92$ por ciento. Podríamos observar que si x es razonablemente grande, es más económico producir botellas más pequeñas. Por ejemplo, si necesitamos solo 0,25 kilos de vidrio fundido por botella en vez de 1 kilo, y si $x = 30$, entonces el porcentaje descartado se reduce de 25,92 por ciento a 7,22 por ciento.

8.3 El proceso de Poisson

En la sección anterior se usó la distribución de Poisson como un medio para aproximar una distribución conocida, a saber, binomial. Sin embargo, la distribución de Poisson desempeña un papel muy importante por derecho propio, puesto que representa un modelo probabilístico apropiado para un gran número de fenómenos observables.

Aunque no vamos a dar una deducción completamente rigurosa de algunos resultados que vamos a exponer, el enfoque general es de tal importancia que se deberá tomar en cuenta para comprenderlos, aún cuando no podamos justificar cada uno de los pasos.

Para referirnos a un ejemplo específico mientras concluimos los detalles matemáticos, consideremos una fuente de material radiactivo que emite partículas α . Sea definido X_t como el número de partículas emitidas durante un período especificado de tiempo $[0, t]$. Vamos a enunciar algunas hipótesis acerca de la variable aleatoria (discreta) X_t , que nos permitirán determinar la distribución de

probabilidades de X_t . La posibilidad de estas hipótesis (recordando lo que X_t representa) está comprobada por el hecho de que la evidencia empírica sostiene una cantidad muy considerable de resultados teóricos que vamos a derivar.

Puede ser útil señalar que en la deducción de cualquier resultado matemático, debemos aceptar algunos postulados o axiomas fundamentales. En la búsqueda de axiomas para describir fenómenos observables, algunos axiomas pueden ser más apropiados (y menos arbitrarios) que otros. Por ejemplo, al describir el movimiento de un objeto impulsado hacia arriba con cierta velocidad inicial, podríamos suponer que la distancia sobre el suelo, llamémosla s , es una función cuadrática del tiempo t ; es decir, $s = at^2 + bt + c$. Esta sería, difícilmente, una hipótesis intuitiva, de acuerdo con nuestra experiencia. En lugar, podríamos suponer que la aceleración es una constante y luego *deducir* de esto que s debe ser una función cuadrática de t . Lo importante, naturalmente, es que si debemos suponer algo con el objeto de elaborar nuestro modelo matemático, preferiríamos suponer lo que es apropiado, en vez de lo que no lo es.

El mismo objetivo nos guía a elaborar un modelo probabilístico para la emisión de partículas α de una fuente radiactiva. La variable aleatoria X_t , definida anteriormente puede tomar los valores $0, 1, 2, \dots$. Sea $p_n(t) = P[X_t = n]$, $n = 0, 1, 2, \dots$

Vamos a enunciar ahora las cinco *hipótesis* siguientes.

- A₁: El número de partículas emitidas durante intervalos de tiempo no sobrepuestos son variables aleatorias *independientes*.
- A₂: Si X_t se define como antes y si Y_t es igual al número de partículas emitidas durante $[t_1, t_1 + t]$, para cualquier $t_1 > 0$, las variables aleatorias X_t y Y_t tienen la misma distribución de probabilidades. (En otras palabras, *la distribución* del número de partículas emitidas durante cualquier intervalo depende sólo de la longitud del intervalo y no de los puntos extremos.)
- A₃: $p_1(\Delta t)$ es igual aproximadamente a $\lambda\Delta t$, si Δt es suficientemente pequeña, donde λ es una constante positiva. Esto lo escribimos como $p_1(\Delta t) \sim \lambda\Delta t$. En esta sección $a(\Delta t) \sim b(\Delta t)$ significa que $a(\Delta t)/b(\Delta t) \rightarrow 1$ cuando $\Delta t \rightarrow 0$. También supondremos que $\Delta t > 0$. (Esta hipótesis expresa que si el intervalo es suficientemente pequeño, la probabilidad de obtener exactamente una emisión durante ese intervalo es directamente proporcional a la longitud del intervalo).
- A₄: $\sum_{k=2}^{\infty} p_k(\Delta t) \sim 0$. (Esto implica que $p_k(\Delta t) \rightarrow 0$, $k \geq 2$.) Esto nos dice que la probabilidad de obtener dos o más emisiones en un intervalo suficientemente pequeño es despreciable.
- A₅: $X_0 = 0$, o equivalentemente, $p_0(0) = 1$. Esto equivale a una condición inicial para el modelo que estamos describiendo.

Como lo demostraremos en breve, las cinco hipótesis anotadas anteriormente harán posible que deduzcamos una expresión para $p_n(t) = P[X_t = n]$. Saquemos ahora ciertas conclusiones de las hipótesis anteriores.

(a) Las hipótesis A_1 y A_2 juntas implican que la variable aleatoria $X_{\Delta t}$ y $[X_{t+\Delta t} - X_t]$ son variables aleatorias independientes con la misma distribución de probabilidades. (Ver figura 8.1.)



FIGURA 8.1

(b) De las hipótesis A_3 y A_4 podemos concluir que

$$p_0(\Delta t) = 1 - p_1(\Delta t) - \sum_{k=2}^{\infty} p_k(\Delta t) \sim 1 - \lambda \Delta t. \quad (8.2)$$

(c) Podemos escribir

$$\begin{aligned} p_0(t + \Delta t) &= P[X_{t+\Delta t} = 0] \\ &= P[X_t = 0 \quad \text{y} \quad (X_{t+\Delta t} - X_t) = 0] \\ &= p_0(t)p_0(\Delta t). \quad [\text{Ver conclusión (a).}] \\ &\sim p_0(t)[1 - \lambda \Delta t]. \quad [\text{Ver ecuación (8.2).}] \end{aligned}$$

(d) Luego tenemos

$$\frac{p_0(t + \Delta t) - p_0(t)}{\Delta t} \sim -\lambda p_0(t).$$

Haciendo $\Delta t \rightarrow 0$, y observando que el lado izquierdo representa al cociente de la diferencia de la función p_0 y por tanto tiende a $p'_0(t)$ (más precisamente, a la derivada por la derecha, puesto que $\Delta t > 0$), tenemos

$$p'_0(t) = -\lambda p_0(t) \quad \text{o, equivalentemente,} \quad \frac{p'_0(t)}{p_0(t)} = -\lambda.$$

Integrando en ambos miembros con respecto a t , obtenemos $\ln p_0(t) = -\lambda t + C$, en donde C es una constante de integración. De la hipótesis A_5 encontramos, al hacer $t = 0$, que $C = 0$. Luego

$$p_0(t) = e^{-\lambda t}. \quad (8.3)$$

Así, nuestras hipótesis nos han conducido a una expresión para $P[X_t = 0]$. Utilizando esencialmente el mismo enfoque, obtendremos ahora $p_n(t)$ para $n \geq 1$.

(e) Considerando $p_n(t + \Delta t) = P[X_{t+\Delta t} = n]$.

Ahora $X_{t+\Delta t} = n$ si y sólo si $X_t = x$ y $[X_{t+\Delta t} - X_t] = n - x$, $x = 0, 1, 2, \dots, n$. Utilizando las hipótesis A_1 y A_2 , tenemos

$$\begin{aligned} p_n(t + \Delta t) &= \sum_{x=0}^n p_x(t)p_{n-x}(\Delta t) \\ &= \sum_{x=0}^{n-2} p_x(t)p_{n-x}(\Delta t) + p_{n-1}(t)p_1(\Delta t) + p_n(t)p_0(\Delta t). \end{aligned}$$

Utilizando las hipótesis A_3 y A_4 y también la ecuación (8.2), obtenemos

$$p_n(t + \Delta t) \sim p_{n-1}(t)\lambda \Delta t + p_n(t)[1 - \lambda \Delta t].$$

Luego

$$\frac{p_n(t + \Delta t) - p_n(t)}{\Delta t} \sim \lambda p_{n-1}(t) - \lambda p_n(t).$$

Nuevamente haciendo $\Delta t \rightarrow 0$, y observando nuevamente que el lado izquierdo representa el cociente diferencial de la función p_n , obtenemos

$$p'_n(t) = -\lambda p_n(t) + \lambda p_{n-1}(t), \quad n = 1, 2, \dots$$

Esta representa a un sistema lineal infinito de ecuaciones diferenciales de diferencias. El lector interesado puede verificar que si definimos la función q_n por la relación $q_n(t) = e^{\lambda t} p_n(t)$, el sistema anterior se transforma en $q'_n(t) = \lambda q_{n-1}(t)$, $n = 1, 2, \dots$. Puesto que $p_0(t) = e^{-\lambda t}$, encontramos que $q_0(t) = 1$. [Nótese también que $q_n(0) = 0$ para $n > 0$.] Así obtenemos, recursivamente,

$$\begin{aligned} q'_1(t) &= \lambda, & \text{y por tanto } q_1(t) &= \lambda t; \\ q'_2(t) &= \lambda q_1(t) = \lambda^2 t, & \text{y por tanto } q_2(t) &= \frac{(\lambda t)^2}{2}. \end{aligned}$$

En general, $q'_n(t) = \lambda q_{n-1}(t)$ y por tanto $q_n(t) = (\lambda t)^n / n!$ Recordando la definición de q_n , finalmente obtenemos

$$p_n(t) = e^{-\lambda t} (\lambda t)^n / n!, \quad n = 0, 1, 2, \dots \quad (8.4)$$

Hemos demostrado así que el número de partículas emitidas durante el intervalo de tiempo $[0, t]$ de una fuente radiactiva, bajo las suposiciones hechas anteriormente, es una variable aleatoria, con una distribución de Poisson con parámetro (λt) .

Observaciones: (a) Es importante darse cuenta de que la distribución de Poisson apareció como una consecuencia de ciertas suposiciones que hicimos. Esto significa que cada vez que dichas suposiciones sean válidas (o al menos lo sean aproximadamente) la distribución de Poisson debe usarse como un modelo apropiado. Resulta que hay una gran cantidad de fenómenos para los cuales es apropiado el modelo de Poisson.

(i) Representemos por X_t el número de llamadas telefónicas que llegan a una central telefónica durante un período de tiempo de longitud t . Las suposiciones anteriores se satisfacen aproximadamente, en especial durante el «período congestionado» del día. Luego X_t tiene una distribución de Poisson.

(ii) Representemos por X_t al número de electrones que salen del cátodo de un tubo al vacío. Nuevamente las suposiciones son apropiadas y, por tanto, X_t tiene una distribución de Poisson.

(iii) El ejemplo siguiente (de astronomía) indica que el razonamiento anterior puede aplicarse no sólo al número de ocurrencias de un suceso durante un período de tiempo fijo, sino también al número de ocurrencias de un suceso dentro de los límites fijados de un área o un volumen. Supóngase que un astrónomo investiga una parte de la Vía Láctea y supóngase que en la parte considerada la densidad de estrellas, llamémosla λ , es constante. (Esto significa que en un volumen de V (unidades cúbicas), se encontrarían, en promedio, λV estrellas). Sea X_V igual al número de estrellas encontradas en una parte de la Vía Láctea, que tiene un volumen V . Si se satisfacen las suposiciones anteriores (sustituyendo «volumen» por «tiempo»), entonces $P[X_V = n] = (\lambda V)^n e^{-\lambda V} / n!$ (Las suposiciones, interpretadas en el ejemplo presente, establecerían esencialmente que el número

de estrellas que aparecen en partes no sobreuestas del firmamento, representa variables aleatorias independientes y que la probabilidad de que aparezca más de una estrella en una porción pequeña del cielo es cero.)

(iv) Se nos ocurre otra aplicación en el campo biológico, si hacemos que X_A sea el número de células sanguíneas visibles bajo el microscopio, en donde el área de la superficie visible bajo el microscopio es dada por A unidades cuadradas.

(b) La constante λ apareció originalmente como una constante de proporcionalidad en la hipótesis A₃. Vale la pena mencionar las siguientes interpretaciones de λ : si X_t representa el número de ocurrencias de un suceso durante un intervalo de tiempo de longitud t , entonces, $E(X_t) = \lambda t$, y por tanto $\lambda = [E(X_t)]/t$ representa la razón esperada con la cual se emiten las partículas. Si X_V representa al número de ocurrencias de algún suceso dentro de un volumen especificado V , entonces $E(X_V) = \lambda V$, y por lo tanto $\lambda = [E(X_V)/V]$ representa la densidad esperada con la cual aparecen las estrellas.

(c) Es importante señalar que nuestra exposición en la sección 8.3 no se refirió sólo a una variable aleatoria X que posee una distribución de Poisson, sino que para cada $t > 0$, encontramos que X_t tenía una distribución de Poisson con un parámetro dependiente de t . Tal colección (infinita) de variables aleatorias también se conoce como un *proceso de Poisson*. (Recíprocamente, se genera un proceso de Poisson cada vez que ocurre un suceso en algún intervalo de tiempo de modo que se satisfagan las hipótesis A₁ hasta A₅.) De manera análoga podemos definir un *proceso de Bernoulli*: si $X_1, X_2, \dots, X_n, \dots$ son los números de ocurrencias de sucesos en 1, 2, ..., n , experimentos de Bernoulli, entonces la colección de variables aleatorias X_1, \dots, X_n, \dots se llaman un proceso de Bernoulli.

EJEMPLO 8.5. Una complicada maquinaria, cuando funciona perfectamente, puede producir una utilidad de C pesos por hora ($C > 2$) a una compañía. Sin embargo, esta máquina tiene una tendencia a fallar a horas inesperadas e impredecibles. Supóngase que el número de fallas durante cualquier período de longitud t horas es una variable aleatoria con una distribución de Poisson de periodo t . Si la máquina falla x veces durante t horas, la pérdida ocasionada (la improductividad de la máquina más la reparación) es igual a $(x^2 + x)$ pesos. Luego la utilidad total P durante cualquier período de t horas es igual a $P = Ct - (X^2 + X)$, en donde X es la variable aleatoria que indica el número de fallas de la máquina. Por tanto P es una variable aleatoria, y podría ser de interés elegir t (lo que está a nuestra voluntad) de manera tal que la utilidad esperada sea maximizada. Tenemos

$$E(P) = Ct - E(X^2 + X).$$

Según el teorema 8.1 encontramos que $E(X) = t$ y $E(X^2) = t + (t)^2$. Luego se deduce que $E(P) = Ct - 2t - t^2$. Para encontrar el valor de t para la cual $E(P)$ es maximizada, diferenciamos $E(P)$ e igualamos la expresión resultante a cero. Obtenemos $C - 2 - 2t = 0$, obteniendo $t = \frac{1}{2}(C - 2)$ horas.

EJEMPLO 8.6. Sea X_t igual al número de partículas emitidas por una fuente radiactiva durante un intervalo de tiempo de longitud t . Supóngase que X_t tiene una distribución de Poisson con parámetro αt . Se instala un instrumento para anotar el número de partículas emitidas. Supóngase que hay una probabilidad constante p de que cualquier partícula emitida no sea contada. Si R_t es igual al número de partículas contadas durante el intervalo específico, ¿cuál es la distribución de probabilidades de R_t ?

Para $X_t = x$, dado la variable aleatoria R_t tiene una distribución binomial con base en x repeticiones con parámetro $(1 - p)$. Esto es,

$$P(R_t = k | X_t = x) = \binom{x}{k} (1 - p)^k p^{x-k}.$$

Usando la fórmula de la probabilidad total, (ecuación 3.4), tenemos

$$\begin{aligned} P(R_t = k) &= \sum_{x=k}^{\infty} P(R_t = k | X_t = x) P(X_t = x) \\ &= \sum_{x=k}^{\infty} \binom{x}{k} (1 - p)^k p^{x-k} e^{-\alpha t} (\alpha t)^x / x! \\ &= \left(\frac{1 - p}{p} \right)^k \frac{e^{-\alpha t}}{k!} \sum_{x=k}^{\infty} \frac{1}{(x - k)!} (p\alpha t)^x. \end{aligned}$$

Sea $i = x - k$. Entonces

$$\begin{aligned} P(R_t = k) &= \left(\frac{1 - p}{p} \right)^k \frac{e^{-\alpha t}}{k!} \sum_{i=0}^{\infty} \frac{(p\alpha t)^{i+k}}{i!} \\ &= \left(\frac{1 - p}{p} \right)^k \frac{e^{-\alpha t}}{k!} (p\alpha t)^k e^{p\alpha t} \\ &= \frac{e^{-\alpha(1-p)t} [(1 - p)\alpha t]^k}{k!}. \end{aligned}$$

Así encontramos que R_t tiene una distribución de Poisson con parámetro $(1 - p)\alpha t$.

8.4 La distribución geométrica

Supóngase que efectuamos un experimento ε y estamos interesados sólo en la ocurrencia o no ocurrencia de algún suceso A . Supóngase como en la presentación de la distribución binomial, que repetidamente efectuamos ε , que las repeticiones son independientes, y que en cada una de las repeticiones $P(A) = p$ y $P(\bar{A}) = 1 - p = q$ permanecen constantes. Supóngase que repetimos el experimento hasta que A ocurre por primera vez. (Aquí nos sepáramos de las hipótesis que conducen a la distribución binomial. Allí el número de repeticiones era pre-determinado, mientras que aquí es una variable aleatoria.)

Definamos la variable aleatoria X como el número de repeticiones necesarias hasta incluir la primera ocurrencia de A . Así, X toma los valores posibles $1, 2, \dots$. Puesto que $X = k$ si y sólo si las primeras $(k - 1)$ repeticiones de ε resultan en \bar{A} mientras que la k -ésima resulta en A , tenemos

$$P(X = k) = q^{k-1} p, \quad k = 1, 2, \dots \tag{8.5}$$

Se dice que una variable aleatoria con una distribución de probabilidades ecuación (8.5), tiene una distribución *geométrica*.

Un cálculo sencillo indica que la ecuación (8.5) define una distribución de probabilidades legítima. Obviamente tenemos $P(X = k) \geq 0$. Y

$$\sum_{k=1}^{\infty} P(X = k) = p(1 + q + q^2 + \dots) = p \left[\frac{1}{1-q} \right] = 1.$$

Podemos obtener el valor esperado de X como sigue.

$$\begin{aligned} E(X) &= \sum_{k=1}^{\infty} kpq^{k-1} = p \sum_{k=1}^{\infty} \frac{d}{dq} q^k \\ &= p \frac{d}{dq} \sum_{k=1}^{\infty} q^k = p \frac{d}{dq} \left[\frac{q}{1-q} \right] = \frac{1}{p}. \end{aligned}$$

(El intercambio de diferenciales y sumatorias se justifica aquí puesto que la serie converge para $|q| < 1$). Un cálculo similar enseña que $V(X) = q/p^2$. (Nuevamente obtendremos los dos resultados en el capítulo 10, usando un enfoque distinto.) Para resumir lo anterior tenemos el teorema siguiente.

Teorema 8.3. Si X tiene una distribución geométrica como se da en la ecuación (8.5),

$$E(X) = 1/p \quad \text{y} \quad V(X) = q/p^2.$$

Observación: el hecho de que $E(X)$ sea el recíproco de p es interesante intuitivamente, puesto que dice que con valores pequeños de $p = P(A)$ se necesitan muchas repeticiones a fin de que ocurra A .

EJEMPLO 8.7. Supóngase que el costo de efectuar un experimento es \$ 1000. Si el experimento falla, se incurre en un costo adicional de \$ 300 debido a ciertos cambios que deben efectuarse antes de que se intente un nuevo ensayo. Si la probabilidad de éxito en cualquiera de los ensayos es 0,2, si los ensayos aislados son independientes y si los experimentos continúan hasta que se obtiene el primer resultado exitoso, ¿cuál es el costo esperado del procedimiento completo?

Si C es el costo y X el número de ensayos necesarios para obtener éxito, tenemos que $C = 1000X + 300(X - 1) = 1300X - 300$. Luego

$$E(C) = 1300E(X) - 300 = 1300 \frac{1}{0,2} - 300 = \$6200.$$

EJEMPLO 8.8. En cierta región la probabilidad de que una tormenta con truenos ocurra en un día cualquiera durante el verano (digamos enero y febrero) es igual a 0,1. Suponiendo la independencia de un día con otro, ¿cuál es la probabilidad de que la primera tormenta con truenos del verano ocurra el 3 de febrero?

Sea X el número de días (empezando el 1 de enero) hasta la primera tormenta y buscamos $P(X = 34)$ la cual es igual $(0,9)^{33}(0,1) = 0,003$.

EJEMPLO 8.9. Si la probabilidad de que cierto examen dé una reacción «positiva» igual a 0,4, ¿cuál es la probabilidad de que ocurran menos de 5 reacciones

«negativas» antes de la primera positiva? Haciendo que Y sea el número de reacciones negativas antes de la primera positiva, tenemos

$$P(Y = k) = (0.6)^k(0.4), \quad k = 0, 1, 2, \dots$$

Luego

$$P(Y < 5) = \sum_{k=0}^4 (0.6)^k(0.4) = 0.92.$$

Observación: Si X tiene una distribución geométrica como se describió en la ecuación (8.5) y si hacemos $Z = X - 1$, podemos interpretar a Z como el número de fallas que preceden al primer éxito. Tenemos que $P(Z = k) = q^k p$, $k = 0, 1, 2, \dots$, en donde $p = P(\text{éxito})$ y $q = P(\text{falla})$.

La distribución geométrica tiene una propiedad interesante que está resumida en el teorema siguiente.

Teorema 8.4. Supóngase que X tiene una distribución geométrica dada por la ecuación (8.5). Entonces para dos enteros positivos cualesquiera s y t ,

$$P(X \geq s + t \mid X > s) = P(X \geq t). \quad (8.6)$$

Demostración: ver el problema 8.18.

Observaciones: (a) El teorema anterior indica que la distribución geométrica «no tiene memoria» en el sentido siguiente. Supongamos que el suceso A no ha ocurrido durante las primeras s repeticiones de ε . Entonces la probabilidad de que no ocurra durante las próximas t repeticiones es la misma que la probabilidad de que no ocurra durante las primeras t repeticiones. En otras palabras, la información de ningún éxito es «olvidada» en lo que se refiere a cálculos subsecuentes.

(b) El recíproco del teorema anterior, es también cierto: si la ecuación (8.6) es válida para una variable aleatoria que toma sólo valores positivos, entonces la variable aleatoria *debe* tener una distribución geométrica. (No demostraremos esto aquí. Se puede encontrar tal exposición en *An Introduction to Probability Theory and Its Applications*, de Feller, John Wiley and Sons, Inc., 2.^a edición, New York, 1957, pág. 304.)

(c) Observaremos en el capítulo siguiente que existe una variable aleatoria continua con una distribución que posee una propiedad análoga a la ecuación (8.6), es decir la distribución exponencial.

EJEMPLO 8.10. Supongamos que un mecanismo es inspeccionado al finalizar cada día para ver si aún funciona regularmente. Sea $p = P$ [falla durante cualquier día dado]. Por tanto si X es el número de inspecciones necesarias para obtener la primera falla, X tiene una distribución de probabilidades geométricas y tenemos $P(X = n) = (1 - p)^{n-1} p$. Análogamente, $(1 - p)^{n-1} p = P$ [el artículo se encontrará que ha fallado en la n -ésima inspección y no en la $(n-1)$ ésmia inspección.]

El valor máximo de esta probabilidad se obtiene resolviendo

$$\frac{d}{dp} P(X = n) = 0.$$

Esto da

$$p(n-1)(1-p)^{n-2}(-1) + (1-p)^{n-1} = 0.$$

lo que equivale a

$$(1-p)^{n-2}[(1-p) - (n-1)p] = 0,$$

de la cual obtenemos $p = 1/n$.

8.5 La distribución de Pascal

Una generalización obvia de la distribución geométrica aparece si hacemos la siguiente pregunta. Supongamos que un experimento se continúa hasta que un suceso particular A ocurre por r -ésima vez. Si

$$P(A) = p, \quad P(\bar{A}) = q = 1 - p$$

en cada una de las repeticiones, definimos la variable aleatoria Y como sigue.

Y es el número de repeticiones necesarias a fin que A ocurra exactamente r veces.

Necesitamos la distribución de probabilidades de Y . Debe ser evidente que si $r = 1$, Y tiene la distribución geométrica dada por la ecuación (8.5).

Ahora $Y = k$ si y sólo si A ocurre en la k -ésima repetición y precisamente A ocurrió $(r-1)$ veces en las $(k-1)$ repeticiones previas. La probabilidad de este suceso es simplemente $p^r \binom{k-1}{r-1} p^{r-1} q^{k-r}$, puesto que lo que sucede en las primeras $(k-1)$ repeticiones es independiente de lo que sucede en la k -ésima repetición.

Luego

$$P(Y = k) = \binom{k-1}{r-1} p^r q^{k-r}, \quad k = r, r+1, \dots \quad (8.7)$$

Se ve muy sencillamente que para $r = 1$, lo anterior se reduce a la ecuación (8.5). Una variable aleatoria que tenga una distribución de probabilidades dada por la ecuación (8.7) tiene una distribución de *Pascal*.

Observación: la distribución de Pascal es también conocida comúnmente como la distribución *binomial negativa*. La razón para esto es que al verificar la condición

$$\sum_{k=r}^{\infty} P(Y = k) = 1$$

obtenemos

$$\sum_{k=r}^{\infty} \binom{k-1}{r-1} p^r q^{k-r} = p^r \sum_{k=r}^{\infty} \binom{k-1}{r-1} q^{k-r} = p^r (1-q)^{-r}$$

lo cual obviamente es igual a 1. La última igualdad resulta del desarrollo en serie de

$$(1-q)^{-r} = \sum_{n=0}^{\infty} \binom{-r}{n} (-q)^n,$$

es igual

$$\sum_{k=r}^{\infty} \binom{k-1}{r-1} q^{k-r}$$

después de algunas simplificaciones algebraicas y recordando la definición del coeficiente binomial (generalizado) (ver la observación antes del ejemplo 2.7). Debido al exponente negativo ($-r$) en la expresión anterior esta distribución se llama distribución binomial negativa. Para calcular $E(Y)$ y $V(Y)$ podemos proceder directamente, tratando de calcular las diversas sumas, o bien podemos proceder de la siguiente manera.

Sea

Z_1 = número de repeticiones necesarias hasta incluir la primera ocurrencia de A .

Z_2 = número de repeticiones necesarias entre la primera ocurrencia de A hasta incluir la segunda ocurrencia de A .

⋮

Z_r = número de repeticiones necesarias entre la $(r-1)$ ocurrencia hasta incluir la r -ésima ocurrencia de A .

Así vemos que todas las Z_i son variables aleatorias independientes, que tiene cada una, una distribución geométrica. También, $Y = Z_1 + \cdots + Z_r$. Por tanto usando el teorema 8.3, tenemos el siguiente teorema.

Teorema 8.5. Si Y tiene una distribución de Pascal dada por la ecuación (8.7), luego

$$E(Y) = r/p, \quad V(Y) = rq/p^2. \quad (8.8)$$

EJEMPLO 8.11. La probabilidad de que un experimento sea exitoso es 0,8. Si el experimento se repite hasta que ocurran cuatro resultados exitosos, ¿cuál es el número esperado de repeticiones necesarias? De lo anterior, tenemos $E(\text{número de repeticiones}) = 4/0,8 = 5$.

8.6 Relación entre las distribuciones binomial y de Pascal

Sea X una distribución binomial con parámetros n y p . (Es decir, $X = \text{número de éxitos en } n \text{ ensayos de Bernoulli con } P(\text{éxito}) = p$). Sea Y una distribución de Pascal con parámetros r y p . (Es decir, $Y = \text{número de ensayos de Bernoulli necesarios para obtener } r \text{ éxitos con } P(\text{éxito}) = p$.) Por tanto se establece la siguiente relación:

- (a) $P(Y \leq n) = P(X \geq r)$,
- (b) $P(Y > n) = P(X < r)$.

Demostración: (a) Si hay r o más éxitos en los primeros n ensayos, entonces es necesario n o menos ensayos para obtener los primeros r éxitos.

(b) Si hay menos que r éxitos en los primeros n ensayos entonces se necesitan más de n ensayos para obtener r éxitos.

Observaciones: (a) Las propiedades anteriores hacen posible el empleo de la distribución binomial tabulada para evaluar probabilidades asociadas con la distribución de Pascal. Por ejemplo, supóngase que deseamos calcular la probabilidad de que más de 10 repeticiones sean necesarias para obtener el tercer éxito cuando $p = P(\text{éxito}) = 0.2$. Tenemos, usando la notación anterior para X y Y ,

$$P(Y > 10) = P(X < 3) = \sum_{k=0}^2 \binom{10}{k} (0.2)^k (0.8)^{10-k} = 0.678$$

(de la tabla del apéndice).

(b) Comparemos brevemente las distribuciones binomial y de Pascal. En cada uno de los casos, nos interesan los ensayos repetidos de Bernoulli. La distribución binomial aparece cuando consideramos un número fijo (digamos n) de tales ensayos y estamos interesados en el número de éxitos que ocurren. La distribución de Pascal se encuentra cuando prefijamos el número de éxitos que debemos obtener y luego anotamos el número de ensayos de Bernoulli necesarios. Esto es particularmente útil para un problema estadístico que presentaremos con más detalles más tarde (ver ejemplo 14.1).

8.7 La distribución hipergeométrica

Supóngase que tenemos un lote de N artículos, de los cuales r son defectuosos y $(N - r)$ no son defectuosos. Supóngase que escogemos, al azar, n artículos del lote ($n \leq N$), sin sustitución. Sea X el número de artículos defectuosos encontrados. Puesto que $X = k$ si y sólo si obtenemos exactamente k artículos defectuosos (de los r defectuosos del lote) y exactamente $(n - k)$ no defectuosos [de los $(N - r)$ no defectuosos del lote], tenemos

$$P(X = k) = \frac{\binom{r}{k} \binom{N-r}{n-k}}{\binom{N}{n}}, \quad k = 0, 1, 2, \dots \quad (8.9)$$

Se dice que una variable aleatoria discreta que tiene la distribución de probabilidades de la ecuación (8.9) tiene una *distribución hipergeométrica*.

Observación: puesto que $\binom{a}{b} = 0$ cada vez que $b > a$, si a y b son enteros negativos, podemos definir las probabilidades anteriores para todo $k = 0, 1, 2, \dots$. Obviamente no podemos obtener más que r defectuosos, pero la probabilidad cero será asignada a ese suceso según ecuación (8.9).

EJEMPLO 8.12. Se embarcan motores eléctricos pequeños en lotes de 50. Antes de que tal cargamento sea aceptado, un inspector elige 5 motores y los inspecciona. Si ninguno de los motores probados es defectuoso, el lote es aceptado. Si se encuentra que uno o más son defectuosos, se inspecciona el cargamento completo. Supongamos que, en realidad, hay tres motores defectuosos en el lote. ¿Cuál es la probabilidad de que sea necesaria una inspección de 100 por ciento?

Si hacemos que X sea el número de motores defectuosos encontrados, se necesitará una inspección de 100 por ciento si y sólo si $X \geq 1$. Luego,

$$P(X \geq 1) = 1 - P(X = 0) = 1 - \frac{\binom{3}{0} \binom{47}{5}}{\binom{50}{5}} = 0.28.$$

Teorema 8.6. Sea X una distribución hipergeométrica dada por la ecuación (8.9). Sea $p = r/N$, $q = 1 - p$. Luego tenemos

$$(a) E(X) = np;$$

$$(b) V(X) = npq \frac{N-n}{N-1};$$

$$(c) P(X = k) \simeq \binom{n}{k} p^k (1-p)^{n-k},$$

para N grande.

Demostración: dejaremos los detalles de la demostración al lector. (Ver problema 8.19.)

Observación: la propiedad (c) del teorema 8.6 establece que si el tamaño N del lote es suficientemente grande, la distribución de X puede ser aproximada por la distribución binomial. Esto es razonablemente intuitivo. La distribución binomial se aplica cuando muestreamos con sustitución (puesto que en ese caso la probabilidad de obtener un artículo defectuoso permanece constante), mientras que la distribución hipergeométrica se aplica cuando muestreamos sin sustitución. Si el tamaño del lote es grande, no sería mucha la diferencia si un artículo en particular es devuelto o no al lote antes de escoger el siguiente. La propiedad (c) del teorema 8.6 es simplemente una expresión matemática de ese hecho. Nótese también que el valor esperado de una variable aleatoria hipergeométrica X es el mismo que el de la variable aleatoria correspondiente distribuida binomialmente, mientras que la varianza de X es un poco más pequeña que la correspondiente en el caso binomial. El «término de corrección» $(N-n)/(N-1)$ es aproximadamente igual a 1, para un N grande.

Podemos ilustrar el significado de (c) con el siguiente ejemplo sencillo. Supóngase que queremos evaluar $P(X = 0)$.

Para $n = 1$, obtenemos de la distribución hipergeométrica, $P(X = 0) = (N-r)/N = 1 - r/N = q$. De la distribución binomial obtenemos directamente $P(X = 0) = q$. Por tanto estas respuestas son las mismas, como deberían ser, además, para $n = 1$.

Para $n = 2$, obtenemos de la distribución hipergeométrica

$$P(X = 0) = \frac{N-r}{N} \cdot \frac{N-r-1}{N-1} = \left(1 - \frac{r}{N}\right) \left(1 - \frac{r}{N-1}\right).$$

De la distribución binomial obtenemos $P(X = 0) = q^2$. Debe observarse que $(1 - r/N) = q$, mientras que $[1 - r/(N-1)]$ es casi igual a q .

En general, la aproximación de la distribución hipergeométrica mediante la distribución binomial es muy buena si $n/N \leq 0,1$.

8.8 La distribución multinomial

Finalmente, consideraremos una variable aleatoria discreta importante de mayor dimensión, que puede ser considerada como una generalización de la distribución binomial. Consideraremos un experimento ε , su espacio muestral S , y una

partición de S en k sucesos mutuamente excluyentes A_1, \dots, A_k . (Es decir, cuando se efectúa ε uno y sólo uno de los sucesos A_i ocurre). Considérese n repeticiones independientes de ε . Sea $p_i = P(A_i)$ y supóngase que p_i permanece constante durante todas las repeticiones. Evidentemente tenemos $\sum_{i=1}^k p_i = 1$. Definamos las variables aleatorias X_1, \dots, X_k como sigue.

X_i es el número de veces que A_i ocurre entre las n repeticiones de ε , $i = 1, \dots, k$.

Las X_i no son variables aleatorias independientes puesto que $\sum_{i=1}^k X_i = n$. Entonces tan pronto como el valor de cualquiera de las $(k - 1)$ variables aleatorias es conocido, se determina el valor de la otra. Tenemos el resultado siguiente.

Teorema 8.7. Si X_i , $i = 1, 2, \dots, k$, están definidas como antes, tenemos

$$P(X_1 = n_1, X_2 = n_2, \dots, X_k = n_k) = \frac{n!}{n_1! n_2! \cdots n_k!} p_1^{n_1} \cdots p_k^{n_k}, \quad (8.10)$$

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

Demostración: el argumento es idéntico al utilizado para establecer las probabilidades binomiales. Simplemente debemos observar que el número de maneras de agrupar n objetos, n_1 de los cuales son de una clase, n_2 de los cuales son de segunda clase, \dots , n_k de los cuales son de una k -ésima clase está dado por

$$\frac{n!}{n_1! \cdots n_k!}.$$

Observaciones: (a) Si $k = 2$ lo anterior se reduce a la distribución binomial. En este caso designamos los dos sucesos posibles «éxito» y «fracaso».

(b) La distribución anterior se conoce como la *distribución multinomial de probabilidades*. Recordemos que los términos de la distribución binomial se obtuvieron del desarrollo de la expresión binomial $[p + (1 - p)]^n = \sum_{k=0}^n \binom{n}{k} p^k (1 - p)^{n-k}$. De una manera análoga, las probabilidades anteriores pueden obtenerse de un desarrollo de la expresión multinomial $(p_1 + p_2 + \cdots + p_k)^n$.

Teorema 8.8. Suponiendo que (X_1, \dots, X_k) tiene una distribución binomial dada por la ecuación (8.10). Entonces

$$E(X_i) = np_i \quad \text{y} \quad V(X_i) = np_i(1 - p_i), \quad i = 1, 2, \dots, k.$$

Demostración: ésta es una consecuencia inmediata de la observación que cada X_i como se definió anteriormente tiene una distribución binomial con probabilidad de éxito (es decir, la ocurrencia de A_i) igual a p_i .

EJEMPLO 8.13. Se fabrica una barra de un largo específico. Suponiendo que el largo verdadero X (pulgadas) es una variable aleatoria distribuida uniformemente en $[10, 12]$. Supóngase que sólo es de interés saber si ha ocurrido uno de los tres sucesos siguientes:

$$A_1 = \{X < 10,5\}, \quad A_2 = \{10,5 \leq X \leq 11,8\}, \quad \text{y} \quad A_3 = \{X > 11,8\}.$$

Tenemos

$$p_1 = P(A_1) = 0,25, \quad p_2 = P(A_2) = 0,65, \quad \text{y} \quad p_3 = P(A_3) = 0,1.$$

Así si se fabrican 10 de tales barras, la probabilidad de obtener exactamente 5 barras de longitud menor que 10,5 pulgadas y exactamente 2 de longitud mayor que 11,8 pulgadas está dada por

$$\frac{10!}{5!3!2!} (0,25)^5 (0,65)^3 (0,1)^2.$$

PROBLEMAS

8.1. Si X tiene una distribución de Poisson con parámetro β , y si $P(X = 0) = 0,2$ calcular $P(X > 2)$.

8.2. Sea X una distribución de Poisson con parámetro λ . Encontrar el valor de k para el cual $P(X = k)$ es la mayor. [Indicación: comparar $P(X = k)$ con $P(X = k - 1)$.]

8.3. (Este problema es tomado de *Probability and Statistical Inference for Engineers* por Derman y Klein, Oxford University Press, Londres, 1959.) El número de buques tanques, digamos N , que llegan cada día a cierta refinería tiene una distribución de Poisson con parámetro $\lambda = 2$. Las actuales instalaciones portuarias pueden despachar tres buques al día. Si más de tres buques tanques llegan en un día, los que están en exceso deben enviarse a otro puerto.

- (a) En un día determinado, ¿cuál es la probabilidad de tener que hacer salir buques tanques?
- (b) ¿En cuánto deben aumentarse las instalaciones actuales para permitir la atención a todos los buques tanques aproximadamente el 90 por ciento de los días?
- (c) ¿Cuál es el número esperado de buques tanques que llegan al día?
- (d) ¿Cuál es el número más probable de buques tanques que llegan diariamente?
- (e) ¿Cuál es el número esperado de buques tanques atendidos diariamente?
- (f) ¿Cuál es el número esperado de buques tanques devueltos diariamente?

8.4. Supóngase que la probabilidad de que un artículo producido por una máquina especial sea defectuoso es igual a 0,2. Si 10 artículos producidos, se seleccionan al azar, ¿cuál es la probabilidad de que no se encuentre más de un artículo defectuoso? Use las distribuciones binomial y de Poisson y compare las respuestas.

8.5. Una compañía de seguros ha descubierto que sólo alrededor del 0,1 por ciento de la población tiene cierto tipo de accidente cada año. Si los 10.000 asegurados fueran seleccionados aleatoriamente en la población, ¿cuál será la probabilidad de que no más de 5 de estos clientes tengan un accidente de ese tipo el próximo año?

8.6. Supóngase que X tiene una distribución de Poisson. Si

$$P(X = 2) = \frac{2}{3}P(X = 1),$$

calcular $P(X = 0)$ y $P(X = 3)$.

8.7. Un productor de películas produce 10 rollos de una película especialmente sensible cada año. Si la película no se vende dentro del año, debe descartarse. Experiencias pasadas indican que D , la demanda (pequeña) para la película es una variable aleatoria distribuida de Poisson con parámetro 8. Si se obtiene una utilidad de \$ 7 en cada rollo vendido, mientras que ocurre una pérdida de \$ 3 en cada rollo que debe ser descartado, calcular la utilidad esperada que el fabricante puede obtener con los 10 rollos que produce.

8.8 Supóngase que una fuente radiactiva emite partículas y que el número de tales partículas emitidas durante el período de una hora tiene una distribución de Poisson con parámetro λ . Se emplea un instrumento para contar y para anotar el número de las partículas emitidas. Si más de 30 partículas llegan durante cualquier período de una hora, el instrumento para anotar es incapaz de controlar el exceso y simplemente anota 30. Si Y es la variable aleatoria definida como el número de partículas *anotadas* por el instrumento que cuenta, obtenga la distribución de probabilidades de Y .

8.9. Supóngase que una fuente radiactiva emite partículas y que el número de las mismas emitidas durante un período de una hora tiene una distribución de Poisson con parámetro λ . Consideremos que el instrumento para contar esas emisiones falla ocasionalmente en anotar una partícula emitida. Supóngase, específicamente, que cualquier partícula emitida tiene una probabilidad p de ser anotada.

- (a) Si Y está definida como el número de partículas anotadas, ¿cuál es una expresión para la distribución de probabilidades de Y ?
- (b) Calcule $P(Y = 0)$ si $\lambda = 4$ y $p = 0,9$.

8.10. Supóngase que un depósito contiene 10.000 partículas. La probabilidad de que una de esas partículas salga del depósito es igual a 0,0004. ¿Cuál es la probabilidad de que ocurran más de 5 salidas? (Puede suponerse que las diversas salidas son independientes unas de otras.)

8.11. Supóngase que un libro de 585 páginas contiene 43 errores tipográficos. Si esos errores están distribuidos aleatoriamente en el libro, ¿cuál es la probabilidad de que 10 páginas, seleccionadas al azar, estén libres de errores? [Indicación: suponga que X = número de errores por página tiene una distribución de Poisson.]

8.12. Se observa una fuente radiactiva durante 7 intervalos de 10 segundos de duración cada uno y se cuenta el número de partículas emitidas durante cada período. Suponiendo que el número de partículas emitidas, digamos X , durante cada período observado tiene una distribución de Poisson con parámetro 5,0. (Es decir, las partículas son emitidas a razón de 0,5 partículas por segundo.)

- (a) ¿Cuál es la probabilidad de que en cada uno de los 7 intervalos de tiempo, sean emitidas 4 o más partículas?

- (b) ¿Cuál es la probabilidad de que al menos en uno de los 7 intervalos de tiempo, sean emitidas 4 o más partículas?

8.13. Se ha encontrado que el número de fallas de transistores en un computador electrónico en cualquier período de una hora puede considerarse como una variable aleatoria que tiene una distribución de Poisson con parámetro 0,1. (Es decir, en promedio hay una falla de un transistor cada 10 horas). Se inicia cierto proceso que necesita 20 horas de tiempo de cálculo.

- (a) Encuentre la probabilidad de que el proceso anterior pueda ser completado exitosamente sin una falla. (Se supone que la máquina llega a ser inoperante sólo si fallan 3 o más transistores).

(b) Lo mismo que en (a) excepto que la máquina llega a ser inoperante si fallan 2 o más transistores.

8.14. Al formar números binarios con n dígitos, la probabilidad de que aparezca un dígito incorrecto es, digamos, 0,002. Si los errores son independientes, ¿cuál es la probabilidad de encontrar cero, uno o más de un dígito incorrecto en un número binario de 25 dígitos? Si el computador forma 10^6 de tales números de 25 dígitos por segundo, ¿cuál es la probabilidad de que se forme un número incorrecto durante cualquier período de un segundo?

8.15. Se emplean dos procedimientos independientes en la operación de lanzamiento de cohetes. Supóngase que cada uno de los procedimientos se continúa hasta que se produce un lanzamiento exitoso. Se supone que al usar el procedimiento I, $P(S)$, la probabilidad de un lanzamiento exitoso es igual a p_1 , mientras que para el procedimiento II, $P(S) = p_2$. Se supone además, que cada semana se hace un intento con cada uno de los dos métodos. Representemos con X_1 y X_2 al número de semanas necesarias para obtener un lanzamiento exitoso por medio de I y II respectivamente. (Luego X_1 y X_2 son variables aleatorias independientes, que tiene cada una una distribución geométrica). Sea W el mínimo (X_1, X_2) y sea Z el máximo (X_1, X_2). Por tanto W representa al número de semanas necesarias para obtener un lanzamiento exitoso mientras que Z representa el número de semanas necesarias para obtener lanzamientos exitosos con ambos procedimientos. (Luego si el procedimiento I resulta en $\bar{S}\bar{S}\bar{S}S$, mientras que el procedimiento II resulta en SSS , tenemos $W = 3$, $Z = 4$).

(a) Obtener una expresión para la distribución de probabilidades de W . [Indicación: exprese en función de X_1 y X_2 , el suceso $\{W = k\}$.]

(b) Obtener una expresión para la distribución de probabilidades de Z .

(c) Escribir nuevamente las expresiones anteriores si $p_1 = p_2$.

8.16. Se arman cuatro componentes en un solo aparato. Las componentes originan fuentes independientes y $p_i = P(i\text{-ésima componente es defectuosa})$, $i = 1, 2, 3, 4$.

(a) Obtener una expresión de la probabilidad para que el aparato completo funcione.

(b) Obtener una expresión de la probabilidad para que al menos tres componentes funcionen.

(c) Si $p_1 = p_2 = 0,1$ y $p_3 = p_4 = 0,2$, calcule la probabilidad de que funcionen exactamente dos componentes.

8.17. Un mecánico mantiene un gran número de arandelas en un depósito. El 50 por ciento de esas arandelas son de $\frac{1}{4}$ pulgada de diámetro, el 30 por ciento son de $\frac{1}{8}$ pulgada de diámetro, y el 20 por ciento restante con $\frac{3}{8}$ pulgada de diámetro. Se supone que se eligen 10 arandelas.

(a) ¿Cuál es la probabilidad de que haya exactamente cinco arandelas de $\frac{1}{4}$ pulgada, cuatro de $\frac{1}{8}$ de pulgada, y una de $\frac{3}{8}$ de pulgada?

(b) ¿Cuál es la probabilidad de que sólo haya dos clases de arandelas entre las elegidas?

(c) ¿Cuál es la probabilidad de que las tres clases de arandelas estén entre las elegidas?

(d) ¿Cuál es la probabilidad de que haya tres de una clase, tres de otra clase, y cuatro de la tercera clase en una muestra de 10?

8.18. Demostrar el teorema 8.4.

8.19. Demostrar el teorema 8.6.

8.20. El número de partículas emitidas por una fuente radiactiva durante un período específico es una variable aleatoria con una distribución de Poisson. Si la probabilidad de ninguna emisión es igual a $\frac{1}{3}$, ¿cuál es la probabilidad de que ocurran 2 o más emisiones?

8.21. Supóngase que X_t , el número de partículas emitidas en t horas por una fuente radiactiva, tiene una distribución de Poisson con parámetro $20t$. ¿Cuál es la probabilidad de que exactamente 5 partículas sean emitidas durante un período de 15 minutos?

8.22. La probabilidad de un lanzamiento exitoso es igual a 0,8. Supongamos que se hacen ensayos de lanzamientos hasta que han ocurrido 3 lanzamientos exitosos. ¿Cuál es la probabilidad que sean necesarios 6 intentos? ¿Cuál es la probabilidad de que sean necesarios menos de 6 intentos?

8.23. En la situación descrita en el problema 8.22, supongamos que los ensayos de lanzamiento se hacen hasta que ocurren tres lanzamientos *consecutivos* exitosos. Responda las preguntas formuladas en el problema previo en este caso.

8.24. Considere nuevamente la situación descrita en el problema 8.22. Supóngase que cada uno de los ensayos de lanzamiento cuesta \$5.000. Además, un lanzamiento que fracase produce un costo adicional de \$500. Calcular el costo esperado para la situación descrita.

8.25. Con X y Y definidas como en la sección 8.6, probar o refutar lo siguiente:

$$P(Y < n) = P(X > r).$$

Algunas variables aleatorias continuas importantes

9.1 Introducción

En este capítulo proseguiremos la tarea que nos impusimos en el capítulo 8, y estudiaremos con mucho detalle varias variables aleatorias continuas importantes y sus características. Como lo señalamos antes, muchos problemas llegan a ser matemáticamente más sencillos al considerar un recorrido «idealizado» para una variable aleatoria X , en el cual *todos* los números reales posibles (en algún intervalo específico o conjuntos de intervalos) pueden considerarse como resultados posibles. De esta manera llegamos a las variables aleatorias continuas. Muchas de las variables aleatorias que presentaremos ahora tienen aplicaciones importantes y pospondremos hasta un capítulo posterior la exposición de algunas de sus aplicaciones.

9.2 La distribución normal

Una de las variables aleatorias continuas más importantes es la siguiente.

Definición. La variable aleatoria X , que toma todos los valores reales $-\infty < x < \infty$, tiene una distribución *normal* (o *Gaussiana*) si su fdp es de la forma

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2\right), \quad -\infty < x < \infty. \quad (9.1)$$

Los parámetros μ y σ deben satisfacer las condiciones $-\infty < \mu < \infty$, $\sigma > 0$. Puesto que tendremos muchas ocasiones de referirnos a la distribución anterior utilizaremos la notación siguiente: X tiene la distribución $N(\mu, \sigma^2)$ si y sólo si su distribución de probabilidades está dada por la ecuación (9.1). [Frecuentemente usamos la notación $\exp(t)$ para representar e^t .]

Esperaremos hasta el capítulo 12 para exponer la razón de la gran importancia de esta distribución. Indiquemos sencillamente ahora que *la distribución*

normal sirve como una aproximación excelente a una gran cantidad de distribuciones que tienen mucha importancia práctica. Aún más, esta distribución tiene varias propiedades matemáticas apreciables que hacen posible deducir importantes resultados teóricos.

9.3 Propiedades de la distribución normal

(a) Digamos que f es una fdp legítima. Evidentemente $f(x) \geq 0$. Además debemos verificar que $\int_{-\infty}^{+\infty} f(x) dx = 1$. Notemos que al hacer $t = (x - \mu)/\sigma$, podemos escribir $\int_{-\infty}^{+\infty} f(x) dx$ como $(1/\sqrt{2\pi}) \int_{-\infty}^{+\infty} e^{-t^2/2} dt = 1$.

El «truco» utilizado para evaluar esta integral (y esto es un truco) es considerar en vez de I , el cuadrado de esta integral, llamada I^2 . Así

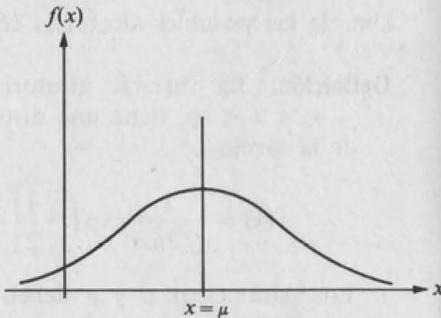
$$\begin{aligned} I^2 &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-t^2/2} dt \int_{-\infty}^{+\infty} e^{-s^2/2} ds \\ &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{-(s^2+t^2)/2} ds dt. \end{aligned}$$

Introduzcamos coordenadas polares para evaluar la integral doble:

$$s = r \cos \alpha, \quad t = r \sin \alpha.$$

Luego el elemento de área $ds dt$ se transforma en $r dr d\alpha$. Cuando s y t varían entre $-\infty$ y $+\infty$, mientras que r varía entre 0 y ∞ , mientras que α varía entre 0 y 2π . Luego

$$\begin{aligned} I^2 &= \frac{1}{2\pi} \int_0^{2\pi} \int_0^\infty r e^{-r^2/2} dr d\alpha \\ &= \frac{1}{2\pi} \int_0^{2\pi} -e^{-r^2/2} \Big|_0^\infty d\alpha \\ &= \frac{1}{2\pi} \int_0^{2\pi} d\alpha = 1. \end{aligned}$$



Por lo tanto $I = 1$ como se iba a demostrar.

FIGURA 9.1

(b) Consideremos la forma del gráfico de f . Tiene la muy conocida forma de campana indicada en la figura 9.1. Puesto que f depende sólo de x mediante la expresión $(x - \mu)^2$, es evidente que el gráfico de f será simétrico respecto a μ . Por ejemplo, si $x = \mu + 2$, $(x - \mu)^2 = (\mu + 2 - \mu)^2 = 4$, mientras que para $x = \mu - 2$, $(x - \mu)^2 = (\mu - 2 - \mu)^2 = 4$ también.

El parámetro σ puede interpretarse geométricamente. Observemos que para $x = \mu$ el gráfico de f es cóncavo hacia abajo. Cuando $x \rightarrow \pm\infty$, $f(x) \rightarrow 0$, asintóticamente. Puesto que $f(x) \geq 0$ para todo x , esto significa que para grandes valores de x (positivos o negativos), el gráfico de f es cóncavo hacia arriba. El punto en el cual cambia la concavidad se llama punto de inflexión, y se determina al resolver la ecuación $f''(x) = 0$. Cuando hacemos esto, encontramos que los puntos de inflexión ocurren en $x = \mu \pm \sigma$. Esto es, σ unidades a la derecha y a la izquierda de μ el gráfico de f cambia de concavidad. Así, si σ es relativamente grande, el gráfico de f , tiende a ser «achatado», mientras que si σ es pequeño el gráfico de f tiende a ser muy «aguzado».

(c) Además de la interpretación geométrica de los parámetros μ y σ , la siguiente interpretación probabilística puede asociarse con esas cantidades. Considerérese

$$E(X) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} x \exp\left(-\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2\right) dx.$$

Haciendo $z = (x - \mu)/\sigma$ y observando que $dx = \sigma dz$, obtenemos

$$\begin{aligned} E(X) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} (\sigma z + \mu) e^{-z^2/2} dz \\ &= \frac{1}{\sqrt{2\pi}} \sigma \int_{-\infty}^{+\infty} z e^{-z^2/2} dz + \mu \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-z^2/2} dz. \end{aligned}$$

La primera de las integrales anteriores es igual a cero puesto que el integrando, llamémoslo $g(z)$, tiene la propiedad de que $g(z) = -g(-z)$, y, por lo tanto, g es una función impar. La segunda integral (sin el factor μ) representa el área total bajo la fdp normal y, por lo tanto, es igual a la unidad. Luego $E(X) = \mu$.

A continuación consideremos

$$E(X^2) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} x^2 \exp\left(-\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2\right) dx.$$

Haciendo nuevamente $z = (x - \mu)/\sigma$, obtenemos

$$\begin{aligned} E(X^2) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} (\sigma z + \mu)^2 e^{-z^2/2} dz \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \sigma^2 z^2 e^{-z^2/2} dz + 2\mu\sigma \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} z e^{-z^2/2} dz \\ &\quad + \mu^2 \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-z^2/2} dz. \end{aligned}$$

La segunda integral nuevamente es igual a cero por el argumento usado anteriormente. La última integral (sin el factor μ^2) es igual a la unidad. Para calcular $(1/\sqrt{2\pi}) \int_{-\infty}^{+\infty} z^2 e^{-z^2/2} dz$, integramos por partes haciendo $ze^{-z^2/2} = dv$ y $z = u$. Luego $v = -e^{-z^2/2}$ mientras que $dz = du$. Se obtiene

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} z^2 e^{-z^2/2} dz = \frac{-ze^{-z^2/2}}{\sqrt{2\pi}} \Big|_{-\infty}^{+\infty} + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-z^2/2} dz = 0 + 1 = 1.$$

Luego, $E(X^2) = \sigma^2 + \mu^2$, y por tanto $V(X) = E(X^2) - (E(X))^2 = \sigma^2$. Así encontramos que los dos parámetros μ y σ^2 que caracterizan la distribución normal son la esperanza y la varianza de X , respectivamente. Para ponerlo en forma distinta, si sabemos que X está distribuido normalmente, sólo sabemos que su distribución de probabilidades es de cierto tipo (o pertenece a cierta familia). Si además, conocemos $E(X)$ y $V(X)$, la distribución de X está especificada completamente. Como anteriormente lo mencionamos, el gráfico de la fdp de una variable aleatoria distribuida normalmente es simétrica respecto a μ . El achatamiento del gráfico se determina por σ^2 en el sentido de que si X tiene una distribución $N(\mu, \sigma_1^2)$ y Y tiene distribución $N(\mu, \sigma_2^2)$, en donde $\sigma_1^2 > \sigma_2^2$, entonces sus fdp tendrían las formas relativas que se muestran en la figura 9.2.

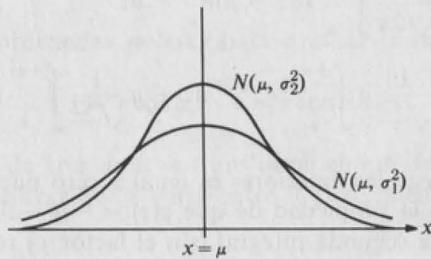


FIGURA 9.2

(d) Si X tiene una distribución $N(0, 1)$ decimos que X tiene una distribución *normal estandarizada*. Esto es, la fdp de X puede escribirse como

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}. \quad (9.2)$$

(Usaremos la letra φ sólo para la fdp de la variable aleatoria X anterior). La importancia de la distribución normal estandarizada se debe a que está *tabulada*. Cada vez que X tiene una distribución $N(\mu, \sigma^2)$ siempre podemos obtener la forma estandarizada, tomando simplemente una función *lineal* de X como lo indica el teorema siguiente.

Teorema 9.1. Si X tiene la distribución $N(\mu, \sigma^2)$ y si $Y = aX + b$, entonces Y tiene la distribución $N(a\mu + b, a^2\sigma^2)$.

Demostración: el hecho de que $E(Y) = a\mu + b$ y que $V(Y) = a^2\sigma^2$ se deduce inmediatamente de las propiedades de la esperanza y la varianza expuestas en el capítulo 7. Para demostrar que de hecho Y está distribuida normalmente, podemos aplicar el teorema 5.1, puesto que $aX + b$ es una función de X decreciente o creciente, según del signo de a . Luego, si g es la fdp de Y , tenemos

$$\begin{aligned} g(y) &= \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}\left[\frac{y-b}{a}-\mu\right]^2\right)\left|\frac{1}{a}\right| \\ &= \frac{1}{\sqrt{2\pi}\sigma|a|} \exp\left(-\frac{1}{2\sigma^2 a^2}[y-(a\mu+b)]^2\right), \end{aligned}$$

que representa la fdp de una variable aleatoria con distribución $N(a\mu + b, a^2\sigma^2)$.

Corolario: si X tiene distribución $N(\mu, \sigma^2)$ y si $Y = (X - \mu)/\sigma$, entonces Y tiene distribución $N(0, 1)$.

Demostración: es evidente que Y es una función lineal de X , y por tanto se aplica el teorema 9.1.

Observación: la importancia de este corolario es que al *cambiar las unidades* en las cuales se mide la variable podemos obtener la distribución estandarizada (ver d). Al hacer esto, obtenemos una distribución donde ningún parámetro es no especificado, lo cual es una situación muy propicia bajo el punto de vista de la tabulación de la distribución (ver la próxima sección).

9.4 Tabulación de la distribución normal

Supóngase que X tiene distribución $N(0, 1)$. Luego

$$P(a \leq X \leq b) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx.$$

Esta integral no puede evaluarse por métodos ordinarios. (La dificultad proviene del hecho que no podemos aplicar el Teorema Fundamental del Cálculo puesto que no podemos encontrar una función cuya derivada sea igual a $e^{-x^2/2}$). Sin embargo, los métodos de integración numérica pueden usarse para evaluar integrales de la forma anterior, y de hecho ha sido *tabulado* $P(X \leq s)$.

La fda de la distribución normal estandarizada se denominará permanentemente como Φ . Esto es,

$$\Phi(s) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^s e^{-x^2/2} dx. \quad (9.3)$$

(Ver figura 9.3). La función Φ ha sido tabulada ampliamente, y en el Apéndice se da un extracto de tal tabla. Podemos usar ahora la tabulación de la función Φ con el objeto de evaluar $P(a \leq X \leq b)$, en donde X tiene la distribución estandarizada $N(0, 1)$ puesto que

$$P(a \leq X \leq b) = \Phi(b) - \Phi(a).$$

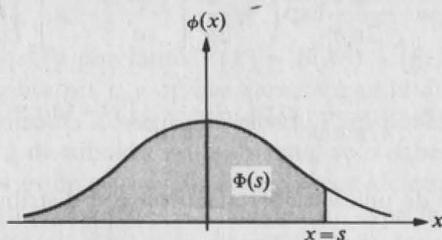


FIGURA 9.3

La importancia particular que tiene la tabulación anterior se debe al hecho de que si X tiene *cualquier* distribución normal $N(\mu, \sigma^2)$, la función tabulada Φ puede usarse para evaluar probabilidades asociadas con X .

Simplemente usamos el teorema 9.1 para observar que si X tiene distribución $N(\mu, \sigma^2)$, entonces $Y = (X - \mu)/\sigma$ tiene distribución $N(0, 1)$. Por tanto

$$\begin{aligned} P(a \leq X \leq b) &= P\left(\frac{a - \mu}{\sigma} \leq Y \leq \frac{b - \mu}{\sigma}\right) \\ &= \Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right). \end{aligned} \quad (9.4)$$

Es también evidente de la definición de Φ (ver figura 9.3) que

$$\Phi(-x) = 1 - \Phi(x). \quad (9.5)$$

Esta relación es muy útil puesto que en la mayoría de las tablas la función Φ se tabula sólo para valores positivos de x .

Calculemos finalmente $P(\mu - k\sigma \leq X \leq \mu + k\sigma)$, en donde X tiene distribución $N(\mu, \sigma^2)$. La probabilidad anterior puede expresarse en función de Φ escribiendo

$$\begin{aligned} P(\mu - k\sigma \leq X \leq \mu + k\sigma) &= P\left(-k \leq \frac{X - \mu}{\sigma} \leq k\right) \\ &= \Phi(k) - \Phi(-k). \end{aligned}$$

Usando la ecuación (9.5), para $k > 0$ tenemos,

$$P(\mu - k\sigma \leq X \leq \mu + k\sigma) = 2\Phi(k) - 1. \quad (9.6)$$

Nótese que la probabilidad anterior es independiente de μ y σ . En palabras: la probabilidad de que una variable aleatoria con distribución $N(\mu, \sigma^2)$ tome valores dentro de k desviaciones estándar del valor esperado depende sólo de k y está dado por la ecuación (9.6).

Observación: tendremos muchas oportunidades para referirnos a «funciones tabuladas». En un sentido dado, cuando una expresión puede escribirse en términos de funciones tabuladas, el problema está «resuelto». (Con la disponibilidad de medios modernos de computación, muchas funciones que no son tabuladas pueden evaluarse fácilmente. Aunque no esperamos que todos tengan fácil acceso a un computador, no parece ilógico suponer que ciertas tablas comunes estén disponibles). Así, nos deberíamos sentir tan cómodos con la función $\Phi(x) = (1/\sqrt{2\pi}) \int_{-\infty}^x e^{-s^2/2} ds$ como con la función $f(x) = \sqrt{x}$. Ambas funciones están tabuladas, y en algunos casos podríamos tener alguna dificultad para calcular directamente la función para $x = 0,43$, por ejemplo. Varias tablas de algunas de las funciones más importantes que encontraremos en nuestro trabajo aparecen en el Apéndice. Ocasionalmente se harán referencias a otras tablas que no aparecen en este texto.

EJEMPLO 9.1. Supóngase que X tiene distribución $N(3, 4)$. Deseamos encontrar un número c tal que

$$P(X > c) = 2P(X \leq c).$$

Observemos que $(X - 3)/2$ tiene distribución $N(0, 1)$. Por tanto

$$P(X > c) = P\left(\frac{X - 3}{2} > \frac{c - 3}{2}\right) = 1 - \Phi\left(\frac{c - 3}{2}\right).$$

También,

$$P(X \leq c) = P\left(\frac{X - 3}{2} \leq \frac{c - 3}{2}\right) = \Phi\left(\frac{c - 3}{2}\right).$$

La condición anterior puede escribirse, por tanto, como $1 - \Phi[(c - 3)/2] = 2\Phi[(c - 3)/2]$. Esto se reduce a $\Phi[(c - 3)/2] = \frac{1}{2}$. Por tanto (de la tabla de la distribución normal), encontramos que $(c - 3)/2 = -0,43$, dando $c = 2,14$.

EJEMPLO 9.2. Supóngase que la resistencia a romperse de un género de algodón (en libras), llamémosle X , está distribuida normalmente con $E(X) = 165$ y $V(X) = 9$. Suponiendo además que una muestra de este género se considera defectuosa si $X < 162$, ¿cuál es la probabilidad de que un género elegido al azar sea defectuoso?

Debemos calcular $P(X < 162)$. Sin embargo,

$$\begin{aligned} P(X < 162) &= P\left(\frac{X - 165}{3} < \frac{162 - 165}{3}\right) \\ &= \Phi(-1) = 1 - \Phi(1) = 0,159. \end{aligned}$$

Observación: una objeción inmediata al uso de la distribución normal puede encontrarse aquí. Es obvio que X , la resistencia del género de algodón, no puede tomar valores negativos, mientras que una variable aleatoria distribuida normalmente puede tomar todos los valores positivos y negativos. Sin embargo, el modelo anterior (aparentemente invalidado en vista de las objeciones encontradas) asigna una probabilidad despreciable al suceso $\{X < 0\}$. Esto es,

$$P(X < 0) = P\left(\frac{X - 165}{3} < \frac{0 - 165}{3}\right) = \Phi(-55) \approx 0.$$

La situación que aparece aquí ocurrirá con frecuencia: se supone que cierta variable aleatoria X que sabemos que no puede tomar valores negativos tiene una distribución normal, tomando así (teóricamente, a lo menos) ambos, valores positivos y negativos. Mientras que se escogen los parámetros μ y σ^2 de modo que $P(X < 0)$ sea prácticamente cero, tal representación es perfectamente válida.

El problema de encontrar la fdp de una función de una variable aleatoria, llamémosla $Y = H(X)$, como se expuso en el capítulo 5, aparece en el contexto presente en el cual la variable aleatoria X está distribuida normalmente.

EJEMPLO 9.3. Supóngase que el radio R de un cojinete de esferas está distribuido normalmente con valor esperado 1 y varianza 0,04. Encuentre la fdp del volumen del cojinete.

La fdp de la variable aleatoria R está dada por

$$f(r) = \frac{1}{\sqrt{2\pi(0,04)}} \exp\left(-\frac{1}{2}\left[\frac{r-1}{0,2}\right]^2\right).$$

Puesto que V es una función de R monótonamente creciente, podemos aplicar directamente el teorema 5.1 para la fdp de $V = 4/3R^3$, y obtener $g(v) = f(r)(dr/dv)$, en donde en todas partes r está expresado en función de v . De la relación anterior, obtenemos $r = \sqrt[3]{3v/4\pi}$. Luego $dr/dv = (1/4\pi)(3v/4\pi)^{-2/3}$. Al sustituir esas expresiones en la ecuación anterior, obtenemos la fdp deseada de V .

EJEMPLO 9.4. Supóngase que X , el diámetro interior (milímetros) de una tobera, es una variable aleatoria distribuida normalmente con esperanza μ y varianza 1. Si X no satisface ciertas especificaciones, le produce una pérdida al fabricante. Más exactamente, supóngase que la utilidad T (por tobera) es la siguiente función de X :

$$\begin{aligned} T &= C_1 \text{ (pesos)} && \text{si } 10 \leq X \leq 12, \\ &= -C_2 && \text{si } X < 10, \\ &= -C_3 && \text{si } X > 12. \end{aligned}$$

Por tanto, la utilidad esperada (por tobera) puede escribirse como

$$\begin{aligned} E(T) &= C_1 [\Phi(12 - \mu) - \Phi(10 - \mu)] \\ &\quad - C_2 [\Phi(10 - \mu)] - C_3 [1 - \Phi(12 - \mu)] \\ &= (C_1 + C_3)\Phi(12 - \mu) - (C_1 + C_2)\Phi(10 - \mu) - C_3. \end{aligned}$$

Supóngase que el proceso de fabricación puede alterarse de modo que se puedan obtener diferentes valores de μ . ¿Para qué valor de μ es máxima la utilidad esperada? Debemos calcular $dE(T)/(d\mu)$ e igualarla a cero. Denotando, como es corriente, la fdp de la distribución $N(0, 1)$ por φ , tenemos

$$\frac{dE(T)}{d\mu} = (C_1 + C_3)[- \varphi(12 - \mu)] - (C_1 + C_2)[- \varphi(10 - \mu)].$$

Luego

$$\begin{aligned} -(C_1 + C_3) \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}(12 - \mu)^2) \\ + (C_1 + C_2) \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}(10 - \mu)^2) = 0. \end{aligned}$$

O

$$e^{22 - 2\mu} = \frac{C_1 + C_3}{C_1 + C_2}.$$

Así

$$\mu = 11 - \frac{1}{2} \ln \left(\frac{C_1 + C_3}{C_1 + C_2} \right)$$

[Es fácil para el lector verificar que lo anterior da un valor máximo para $E(T)$].

Observaciones: (a) Si $C_2 = C_3$, es decir, si el diámetro X es muy grande o muy pequeño, igualmente son defectos importantes, entonces el valor μ para el cual se obtiene el valor máximo de $E(T)$ es $\mu = 11$. Si $C_2 > C_3$, el valor de μ es > 11 , mientras que si $C_2 < C_3$, el valor de μ es < 11 . Cuando $\mu \rightarrow +\infty$, $E(T) \rightarrow -C_3$, mientras que si $\mu \rightarrow -\infty$, $E(T) \rightarrow -C_2$.

(b) Considérense los valores de los costos siguientes: $C_1 = \$10$, $C_2 = \$3$, y $C_3 = \$2$. Entonces el valor de μ para el cual $E(T)$ se maximiza es igual a $\mu = 11 - \frac{1}{2} \ln \left[\frac{12}{11} \right] = \$11,04$. Luego el valor máximo obtenido por $E(T)$ es igual a $\$6,04$ por tobera.

9.5 La distribución exponencial

Definición. Se dice que una variable aleatoria continua X que toma todos los valores no negativos tiene una *distribución exponencial* con parámetro α positivo si su fdp está dada por

$$\begin{aligned} f(x) &= \alpha e^{-\alpha x}, \quad x > 0 \\ &= 0, \quad \text{para cualquier otro valor.} \end{aligned} \tag{9.7}$$

(Ver la figura 9.4) [Una integración inmediata indica que

$$\int_0^\infty f(x)dx = 1$$

y, por tanto, la ecuación (9.7) representa una fdp.]

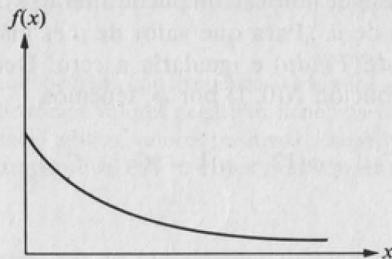


FIGURA 9.4

La distribución exponencial desempeña un papel importante en la descripción de una gran clase de fenómenos, especialmente en el área de la teoría de la confiabilidad. Dedicaremos el capítulo 11 a algunas de estas aplicaciones. Por el momento, investiguemos sencillamente algunas de las propiedades de la distribución exponencial.

9.6 Propiedades de la distribución exponencial

(a) La fda F de la distribución exponencial está dada por:

$$F(x) = P(X \leq x) = \int_0^x \alpha e^{-\alpha t} dt = 1 - e^{-\alpha x}, \quad x \geq 0 \quad (9.8)$$

$$= 0, \quad \text{para cualquier otro valor.}$$

[Por tanto $P(X > x) = e^{-\alpha x}$].

(b) El valor esperado de X se obtiene como sigue:

$$E(X) = \int_0^\infty x \alpha e^{-\alpha x} dx.$$

Integrando por partes y haciendo $\alpha e^{-\alpha x} dx = dv$, $x = u$, obtenemos $v = -e^{-\alpha x}$, $du = dx$. Luego,

$$E(X) = [-xe^{-\alpha x}]_0^\infty + \int_0^\infty e^{-\alpha x} dx = \frac{1}{\alpha}. \quad (9.9)$$

Así el valor esperado es igual al recíproco del parámetro α . [Simplemente redenominando el parámetro $\alpha = 1/\beta$, podríamos haber escrito la fdp de X como $f(x) = (1/\beta)e^{-x/\beta}$. De esta manera, el parámetro β es igual al valor esperado de X . Sin embargo, continuaremos usando la forma de la ecuación (9.7).]

(c) La varianza de X puede obtenerse con una integración semejante. Encontramos que $E(X^2) = 2/\alpha^2$ y por tanto

$$V(X) = E(X^2) - [E(X)]^2 = \frac{1}{\alpha^2}. \quad (9.10)$$

(d) La distribución exponencial tiene la importante propiedad siguiente, análoga a la ecuación (8.6) descrita para la distribución geométrica. Considerando para cualquier $s, t > 0$, $P(X > s + t | X > s)$. Tenemos

$$P(X > s + t | X > s) = \frac{P(X > s + t)}{P(X > s)} = \frac{e^{-\alpha(s+t)}}{e^{-\alpha s}} = e^{-\alpha t}.$$

Por tanto

$$P(X > s + t | X > s) = P(X > t). \quad (9.11)$$

Así hemos demostrado que la distribución exponencial también tiene la propiedad de «no tener memoria» como la distribución geométrica. (Véase la observación que sigue al teorema 8.4). Hacemos uso considerable de esta propiedad al aplicar la distribución exponencial a los modelos de fatiga en el capítulo 11.

Observación: como en el caso de la distribución geométrica, la inversa de la propiedad (d) también es cierta. La única variable aleatoria continua X que toma valores no negativos para los cuales $P(X > s + t | X > s) = P(X > t)$ para $s, t > 0$, es una variable aleatoria distribuida exponencialmente. [Aunque no demostraremos esto aquí, podría señalarse que la base del argumento implica el hecho de que la única función continua G que tiene la propiedad que $G(x + y) = G(x)G(y)$ para todo $x, y > 0$, es $G(x) = e^{-kx}$. Se ve fácilmente que si definimos $G(x) = 1 - F(x)$, en donde F es la fda de X , luego G satisfará esta condición.]

EJEMPLO 9.5. Supóngase que un fusible tiene una duración X que puede considerarse como una variable aleatoria continua con una distribución exponencial. Hay dos procesos mediante los cuales se puede fabricar el fusible. El proceso I da una duración esperada de 100 horas (esto es, el parámetro es igual a 100^{-1}), mientras que el proceso II da una duración esperada de 150 horas (esto es, el parámetro es igual a 150^{-1}). Suponiendo que el proceso II es dos veces más costoso (por fusible) que el proceso I, el cual cuesta C por fusible. Supóngase, además, que si un fusible dura menos de 200 horas, se carga una pérdida de K pesos en contra del fabricante. ¿Qué proceso se deberá usar? Calculemos el costo esperado para cada uno de los procesos. Para el proceso I, tenemos

$$\begin{aligned} C_I &= \text{costo (por fusible)} = C && \text{si } X > 200 \\ &= C + K && \text{si } X \leq 200. \end{aligned}$$

Por tanto,

$$\begin{aligned} E(C_I) &= CP(X > 200) + (C + K)P(X \leq 200) \\ &= Ce^{-(1/100)200} + (C + K)(1 - e^{-(1/100)200}) \\ &= Ce^{-2} + (C + K)(1 - e^{-2}) = K(1 - e^{-2}) + C. \end{aligned}$$

Con un cálculo semejante encontramos que

$$E(C_{II}) = K(1 - e^{-4/3}) + 2C.$$

Luego

$$E(C_{II}) - E(C_I) = C + K(e^{-2} - e^{-4/3}) = C - 0,13K.$$

Por tanto, preferimos el proceso I dado que $C > 0,13K$.

EJEMPLO 9.6. Supóngase que X tiene una distribución exponencial con parámetro α . Entonces $E(X) = 1/\alpha$. Calculemos la probabilidad de que X sobrepase su valor esperado (figura 9.5). Tenemos

$$\begin{aligned} P\left(X > \frac{1}{\alpha}\right) &= e^{-\alpha(1/\alpha)} \\ &= e^{-1} < \frac{1}{2}. \end{aligned}$$

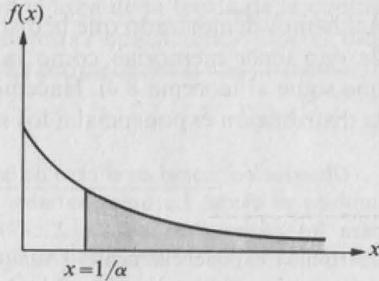


FIGURA 9.5

EJEMPLO 9.7. Supóngase que T , el tiempo para que falle un componente, está distribuido exponencialmente. Luego $f(t) = \alpha e^{-\alpha t}$. Si se instalan n de tales partes, ¿cuál es la probabilidad de que la mitad o más de esas partes funcionen aún al término de t horas? La probabilidad pedida es

$$\sum_{k=n/2}^n \binom{n}{k} (1 - e^{-\alpha t})^{n-k} (e^{-\alpha t k}) \quad \text{si } n \text{ es par};$$

$$\sum_{k=(n+1)/2}^n \binom{n}{k} (1 - e^{-\alpha t})^{n-k} (e^{-\alpha t k}) \quad \text{si } n \text{ es impar.}$$

EJEMPLO 9.8. Supóngase que la duración en horas, llamémosla T , de cierto tubo electrónico es una variable aleatoria con una distribución exponencial con parámetro β . Esto es, la fdp está dada por $f(t) = \beta e^{-\beta t}$, $t > 0$. Una máquina que

usa este tubo cuesta C_1 pesos/hora para funcionar. Mientras la máquina está funcionando, se obtiene una utilidad de C_2 pesos/hora. Debe contratarse un operador para un número *prefijado* de horas, digamos H , y obtiene un pago de C_3 pesos/hora. ¿Para qué valor de H es mayor la *utilidad esperada*?

Obtengamos primero una expresión para la utilidad, llamémosla R . Tenemos

$$\begin{aligned} R &= C_2 H - C_1 H - C_3 H \quad \text{si } T > H \\ &= C_2 T = C_1 T - C_3 H \quad \text{si } T \leq H. \end{aligned}$$

Nótese que R es una variable aleatoria puesto que es una función de T . Luego

$$\begin{aligned} E(R) &= H(C_2 - C_1 - C_3)P(T > H) - C_3 HP(T \leq H) \\ &\quad + (C_2 - C_1) \int_0^H t\beta e^{-\beta t} dt \\ &= H(C_2 - C_1 - C_3)e^{-\beta H} - C_3 H(1 - e^{-\beta H}) \\ &\quad + (C_2 - C_1)[\beta^{-1} - e^{-\beta H}(\beta^{-1} + H)] \\ &= (C_2 - C_1)[He^{-\beta H} + \beta^{-1} - e^{-\beta H}(\beta^{-1} + H)] - C_3 H. \end{aligned}$$

Para obtener el valor máximo de $E(R)$ lo diferenciamos con respecto a H y hacemos la derivada igual a cero. Tenemos

$$\begin{aligned} \frac{dE(R)}{dH} &= (C_2 - C_1)[H(-\beta)e^{-\beta H} + e^{-\beta H} - e^{-\beta H} + (\beta^{-1} + H)(\beta)e^{-\beta H}] - C_3 \\ &= (C_2 - C_1)e^{-\beta H} - C_3. \end{aligned}$$

Luego $dE(R)/dH = 0$ implica que

$$H = -\left(\frac{1}{\beta}\right) \ln \left[\frac{C_3}{C_2 - C_1} \right].$$

[A fin de que la solución anterior sea significativa, debemos tener $H > 0$ lo que ocurre si y sólo si $0 < C_3/(C_2 - C_1) < 1$, lo que a su vez es equivalente a $C_2 - C_1 > 0$ y $C_2 - C_1 - C_3 > 0$. Sin embargo, la última condición requiere sencillamente que las cifras del costo sean de una magnitud tal que pueda obtenerse una utilidad.]

Suponiendo en particular que $\beta = 0,01$, $C_1 = \$3$, $C_2 = \$10$, y $C_3 = \$4$. Luego $H = -100 \ln [\frac{4}{7}] = 55,9$ horas ≈ 56 horas. Luego, el operador debe ser contratado por 56 horas a fin de obtener la utilidad máxima. (Para una modificación ligera del ejemplo anterior, ver el problema 9.18.)

9.7 La distribución gama

Presentemos primero una función que es muy importante no sólo en la teoría de probabilidad, sino en muchas áreas de las matemáticas.

Definición. La función Gama denotada por Γ , se define como sigue:

$$\Gamma(p) = \int_0^{\infty} x^{p-1} e^{-x} dx, \quad \text{definida para } p > 0. \quad (9.12)$$

[Puede demostrarse que existe la integral impropia anterior (converge) si $p > 0$]. Si integramos por partes, haciendo $e^{-x} dx = dv$ y $x^{p-1} = u$, obtenemos

$$\begin{aligned}\Gamma(p) &= -e^{-x} x^{p-1} \Big|_0^{\infty} - \int_0^{\infty} [-e^{-x}(p-1)x^{p-2} dx] \\ &= 0 + (p-1) \int_0^{\infty} e^{-x} x^{p-2} dx \\ &= (p-1)\Gamma(p-1).\end{aligned}\quad (9.13)$$

Así hemos demostrado que la función Gama sigue una importante relación recursiva. Suponiendo que p es un *entero positivo*, hagamos $p = n$. Entonces aplicando la ecuación (9.13) repetidamente, obtenemos

$$\begin{aligned}\Gamma(n) &= (n-1)\Gamma(n-1) \\ &= (n-1)(n-2)\Gamma(n-2) = \cdots \\ &= (n-1)(n-2) \cdots \Gamma(1).\end{aligned}$$

Sin embargo, $\Gamma(1) = \int_0^{\infty} e^{-x} dx = 1$, por tanto tenemos

$$\Gamma(n) = (n-1)! \quad (9.14)$$

(si n es un entero positivo). (Luego podemos considerar que la función Gama es una generalización de la función factorial). También es fácil verificar que

$$\Gamma\left(\frac{1}{2}\right) = \int_0^{\infty} x^{-1/2} e^{-x} dx = \sqrt{\pi}. \quad (9.15)$$

(Ver problema 9.19).

Con la ayuda de la función Gama podemos presentar ahora la distribución Gama de probabilidades.

Definición. Sea X una variable aleatoria continua que toma sólo valores no negativos. Decimos que X tiene una *distribución de probabilidades Gama* si su fdp está dada por

$$\begin{aligned}f(x) &= \frac{\alpha^r}{\Gamma(r)} (\alpha x)^{r-1} e^{-\alpha x}, \quad x > 0 \\ &= 0, \quad \text{para cualquier otro valor}\end{aligned}\quad (9.16)$$

Esta distribución depende de los *dos parámetros*, r y α , de los cuales pedimos $r > 0$, $\alpha > 0$. [Debido a la definición de la función Gama, es fácil ver que $\int_{-\infty}^{+\infty} f(x) dx = 1$.] La figura 9.6 muestra el gráfico de la fdp de la ecuación (9.16) para diversos valores de r y $\alpha = 1$.

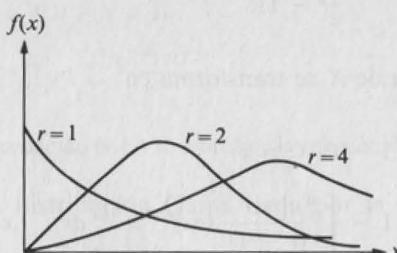


FIGURA 9.6

9.8 Propiedades de la distribución Gama

(a) Si $r = 1$, la ecuación (9.16) se transforma en $f(x) = \alpha e^{-\alpha x}$. Luego la *distribución exponencial es un caso especial de la distribución Gama*. (Si r es un entero positivo > 1 , la distribución gama está relacionada también con la distribución exponencial, pero de un modo ligeramente diferente. Nos referiremos a esto en el capítulo 10).

(b) En la mayoría de nuestras aplicaciones, el parámetro r será un *entero* positivo. En este caso, una relación interesante existe entre la fda de la distribución gama y la distribución de Poisson, que desarrollaremos ahora.

Considerando la integral $I = \int_a^{\infty} (e^{-y} y^r / r!) dy$, en donde r es un entero positivo y $a > 0$. Luego $r! I = \int_a^{\infty} e^{-y} y^r dy$. Integrando por partes, haciendo $u = y^r$ y $dv = e^{-y} dy$, obtendremos $du = ry^{r-1} dy$ y $v = -e^{-y}$. Luego $r! I = e^{-a} a^r + r \int_a^{\infty} e^{-y} y^{r-1} dy$. La integral en esta expresión es exactamente de la misma forma que la integral original con la sustitución de r por $(r - 1)$. Así, al continuar integrando por partes, puesto que r es un entero positivo, obtenemos

$$r! I = e^{-a} [a^r + ra^{r-1} + r(r-1)a^{r-2} + \cdots + r!].$$

Por tanto

$$I = e^{-a} [1 + a + a^2/2! + \cdots + a^r/r!]$$

$$= \sum_{k=0}^r P(Y = k),$$

en donde Y tiene una distribución de Poisson con parámetro a .

Consideremos ahora la fda de la variable aleatoria cuya fdp está dada por la ecuación (9.16). Puesto que r es un entero positivo, la ecuación (9.16) puede escribirse como

$$f(x) = \frac{\alpha}{(r-1)!} (\alpha x)^{r-1} e^{-\alpha x}, \quad 0 < x$$

y por consiguiente la fda de X se transforma en

$$\begin{aligned} F(x) &= 1 - P(X > x) \\ &= 1 - \int_x^{\infty} \frac{\alpha}{(r-1)!} (\alpha s)^{r-1} e^{-\alpha s} ds, \quad x > 0. \end{aligned}$$

Haciendo $(\alpha s) = u$, encontramos que esto se transforma en

$$F(x) = 1 - \int_{\alpha x}^{\infty} \frac{u^{r-1} e^{-u}}{(r-1)!} du, \quad x > 0.$$

Esta integral es precisamente de la forma considerada anteriormente, llamándola I (con $a = \alpha x$), luego

$$F(x) = 1 - \sum_{k=0}^{r-1} e^{-\alpha x} (\alpha x)^k / k!, \quad x > 0. \quad (9.17)$$

Por tanto, *la fda de la distribución Gama puede expresarse mediante la fda tabulada de la distribución de Poisson*. (Recordamos que esto es válido si el parámetro r es un entero positivo.)

Observación: el resultado indicado en la ecuación (9.17), que relaciona la fda de la distribución de Poisson con la fda de la distribución Gama, no es tan sorprendente como podría aparecer al principio, tal como lo indicará la siguiente exposición.

Primero que todo, recordemos la relación entre las distribuciones de Pascal y binomial (ver observación (b) sección 8.6). Una relación similar existe entre la distribución de Poisson y Gama considerando que la última es una distribución continua. Cuando tratamos una distribución de Poisson estamos interesados especialmente en el número de ocurrencias de algún suceso durante un período fijo de tiempo. Y, como se indicará, la distribución Gama aparece cuando pedimos la distribución del *tiempo* necesario para obtener un número especificado de ocurrencias del suceso.

Especificamente, supóngase que X = número de ocurrencias del suceso A durante $(0, t)$. Entonces bajo condiciones semejantes (es decir, satisfaciendo la hipótesis A_1 hasta A_5 en la sección 8.3) X tiene una distribución de Poisson con parámetro αt , en donde α es el número

esperado de ocurrencias de A durante un intervalo de tiempo unitario. Sea T = tiempo necesario para observar r ocurrencias de A . Tenemos

$$\begin{aligned} H(t) &= P(T \leq t) = 1 - P(T > t) \\ &= 1 - P(\text{menos que } r \text{ ocurrencias de } A \text{ acontecen en } (0, t]) \\ &= 1 - P(X < r) \\ &= 1 - \sum_{k=0}^{r-1} \frac{e^{-\alpha t} (\alpha t)^k}{k!}. \end{aligned}$$

Comparando esto con la ecuación (9.17) se establece la relación deseada.

(c) Si X tiene una distribución Gama dada por la ecuación (9.16), tenemos

$$E(X) = r/\alpha, \quad V(X) = r/\alpha^2. \quad (9.18)$$

Demostración: ver el problema 9.20.

9.9 La distribución χ -cuadrado

Un caso especial, muy importante, de la distribución Gama ecuación (9.16) se obtiene si hacemos $\alpha = \frac{1}{2}$ y $r = n/2$, en donde n es un entero positivo. Obtenemos una familia de distribuciones de un parámetro con fdp

$$f(z) = \frac{1}{2^{n/2}\Gamma(n/2)} z^{n/2-1} e^{-z/2}, \quad z > 0. \quad (9.19)$$

Una variable aleatoria Z que tiene fdp dada por la ecuación (9.19) se dice que tiene una *distribución χ -cuadrado con n grados de libertad* (denotada por χ_n^2). En la figura 9.7, la fdp para $n = 1, 2$ y $n > 2$. Una consecuencia inmediata de la ecuación (9.18) es que si Z tiene fdp de la ecuación (9.19), tenemos

$$E(Z) = n, \quad V(Z) = 2n. \quad (9.20)$$

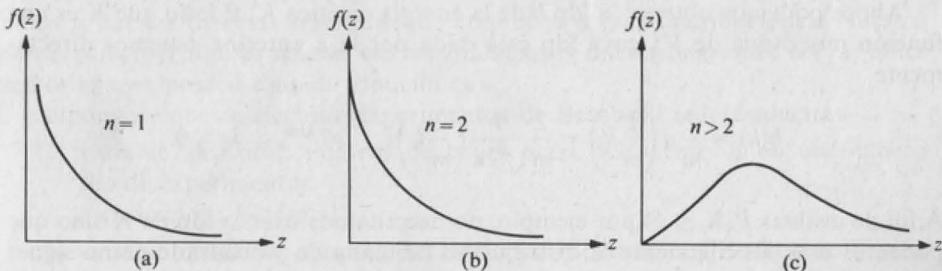


FIGURA 9.7

La distribución χ^2 -cuadrado tiene muchas aplicaciones importantes en inferencia estadística, algunas de las cuales citaremos posteriormente. Debido a su importancia, la distribución de χ^2 -cuadrado está tabulada para diversos valores del parámetro n . (Ver el Apéndice). Por tanto, podemos encontrar en la tabla qué valor, denotado por χ^2_α , satisface $P(Z \leq \chi^2_\alpha) = \alpha$, $0 < \alpha < 1$ (figura 9.8). El ejemplo 9.9 trata un caso especial de una caracterización general de la distribución de χ^2 -cuadrado que estudiaremos en un capítulo posterior.

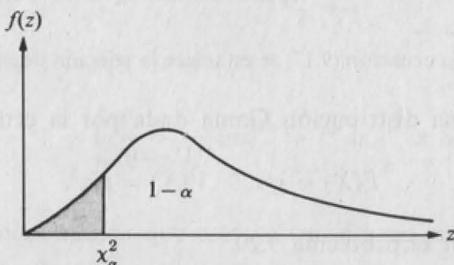


FIGURA 9.8

EJEMPLO 9.9. Supóngase que la velocidad V de un objeto tiene distribución $N(0, 1)$. Sea $K = mV^2/2$ la energía cinética del objeto. Para encontrar la fdp de K , busquemos primero la fdp de $S = V^2$. Aplicando el teorema 5.2 directamente tenemos

$$\begin{aligned} g(s) &= \frac{1}{2\sqrt{s}} [\varphi(\sqrt{s}) + (\varphi - \sqrt{s})] \\ &= s^{-1/2} \frac{1}{\sqrt{2\pi}} e^{-s/2}. \end{aligned}$$

Si comparamos esto con la ecuación (9.19) y recordamos que $\Gamma(\frac{1}{2}) = \sqrt{\pi}$, observamos que S tiene una distribución χ_1^2 . Luego encontramos que el cuadrado de una variable aleatoria con distribución $N(0, 1)$ tiene una distribución χ_1^2 . (Este es el resultado que generalizaremos posteriormente).

Ahora podemos obtener la fdp h de la energía cinética K . Puesto que K es una función monótona de V^2 cuya fdp está dada por la g anterior, tenemos directamente

$$h(k) = \frac{2}{m} g\left(\frac{2}{m} k\right) = \frac{2}{m} \frac{1}{\sqrt{2\pi}} \left(\frac{2}{m} k\right)^{-1/2} e^{-k/m}, \quad k > 0.$$

A fin de evaluar $P(K \leq 5)$ por ejemplo, no necesitamos usar la fdp de K sino que podemos usar sencillamente la distribución tabulada de χ^2 -cuadrado como sigue:

$$P(K \leq 5) = P((m/2)V^2 \leq 5) = P(V^2 \leq 10/m).$$

Esta última probabilidad puede obtenerse directamente de las tablas de la distribución de χ^2 -cuadrado (si m es conocido) puesto que V^2 tiene una distribución χ_1^2 . Puesto que $E(V^2) = 1$ y la varianza (V^2) = 2 [ver la ecuación (9.20)], encontramos directamente

$$E(K) = m/2 \quad \text{y} \quad V(K) = m^2/2.$$

Observación: la tabulación de la distribución de χ^2 -cuadrado que se da en el Apéndice, sólo da los valores para los cuales n , el número de grados de libertad, es menor que o igual a 45. La razón de esto es que si n es grande, podemos aproximar la distribución de χ^2 -cuadrado con la distribución normal, como se indica en el teorema siguiente.

Teorema 9.2. Supóngase que la variable aleatoria Y tiene una distribución χ_n^2 . Entonces para n suficientemente grande la variable aleatoria $\sqrt{2Y}$ aproximadamente tiene la distribución $N(\sqrt{2n-1}, 1)$. (La demostración no se da aquí).

Este teorema puede utilizarse como sigue. Suponiendo que necesitamos $P(Y \leq t)$, en donde Y tiene la distribución χ_n^2 y n es tan grande que la probabilidad anterior no puede obtenerse directamente de la tabla de la distribución de χ^2 -cuadrado. Usando el teorema 9.2 podemos escribir

$$\begin{aligned} P(Y \leq t) &= P(\sqrt{2Y} \leq \sqrt{2t}) \\ &= P(\sqrt{2Y} - \sqrt{2n-1} \leq \sqrt{2t} - \sqrt{2n-1}) \\ &\simeq \Phi(\sqrt{2t} - \sqrt{2n-1}). \end{aligned}$$

El valor de Φ puede obtenerse de las tablas de la distribución normal.

9.10 Comparación entre varias distribuciones

Hasta ahora hemos presentado diversas distribuciones de probabilidades importantes, tanto discretas como continuas: la binomial, Pascal y Poisson entre las discretas y la exponencial, geométrica y Gama entre las continuas. No volveremos a dar las diversas hipótesis que conducen a esas distribuciones. Nuestro interés principal aquí es señalar ciertas analogías (y diferencias) entre las variables aleatorias que poseen esas distribuciones.

1. Supóngase que se efectúan experimentos de Bernoulli independientes.
 - (a) variable aleatoria: número de ocurrencias del suceso A en un número fijo de experimentos.
distribución: binomial
 - (b) variable aleatoria: número necesario de experimentos de Bernoulli para obtener la primera ocurrencia de A
distribución: geométrica

- (c) variable aleatoria: número necesario de experimentos de Bernoulli para obtener la r -ésima ocurrencia de A
 distribución: Pascal
2. Supóngase un proceso de Poisson (ver la nota (c) del ejemplo 8.5 precedente).
- (d) variable aleatoria: número de ocurrencias de un suceso A en un intervalo de tiempo *determinado*
 distribución: Poisson
- (e) variable aleatoria: tiempo transcurrido hasta la primera ocurrencia de A
 distribución: exponencial
- (f) variable aleatoria: tiempo transcurrido hasta la r -ésima ocurrencia de A
 distribución: Gama.

Observación: nótese la similitud entre (a) y (d), (b) y (e), y finalmente entre (c) y (f).

9.11 La distribución normal bivariada

Todas las variables aleatorias que hemos presentado han sido variables aleatorias unidimensionales. Como lo mencionamos en el capítulo 6, las variables aleatorias de mayor dimensión desempeñan un papel importante para describir resultados experimentales. Una de las más importantes variables aleatorias bidimensionales continuas, una generalización directa de la distribución normal unidimensional, se define como sigue.

Definición. Sea (X, Y) una variable aleatoria continua, bidimensional que toma todos los valores en el plano euclíadiano. Decimos que (X, Y) tiene una *distribución normal bivariada* si su fdp conjunta está dada por la expresión siguiente:

$$f(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \times \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho\frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]\right\},$$

$$-\infty < x < \infty, \quad -\infty < y < \infty. \quad (9.21)$$

La fdp anterior depende de 5 parámetros. Para que f defina una fdp legítima [o sea, $f(x, y) \geq 0, \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) dx dy = 1$], debemos poner las siguientes restricciones a los parámetros: $-\infty < \mu_x < \infty$; $-\infty < \mu_y < \infty$; $\sigma_x > 0$; $\sigma_y > 0$; $-1 < \rho < 1$. Las siguientes propiedades de la distribución normal bivariada pueden verificarse fácilmente.

Teorema 9.3. Suponiendo que (X, Y) tiene la fdp como se da en la ecuación (9.21). Entonces

- las distribuciones marginales de X y de Y son $N(\mu_x, \sigma_x^2)$ y $N(\mu_y, \sigma_y^2)$, respectivamente;
- el parámetro ρ que aparece anteriormente es el coeficiente de correlación entre X y Y ;
- las distribuciones condicionales de X (dado que $Y = y$) y de Y (dado que $X = x$) son respectivamente

$$N\left[\mu_x + \rho \frac{\sigma_x}{\sigma_y}(y - \mu_y), \sigma_x^2(1 - \rho^2)\right], \quad N\left[\mu_y + \rho \frac{\sigma_y}{\sigma_x}(x - \mu_x), \sigma_y^2(1 - \rho^2)\right].$$

Demostración: ver el problema 9.21.

Observaciones: (a) El recíproco de (a) del teorema 9.3 no es verdadero. Es posible tener una fdp conjunta que no es normal bivariada aunque las fdp de X y de Y son normales unidimensionales.

(b) Observemos de la ecuación (9.21) que si $\rho = 0$, la fdp conjunta de (X, Y) puede factorizarse y, por tanto, X y Y son independientes. Luego en el caso de la distribución normal bivariada, encontramos que la correlación cero y la independencia son equivalentes.

(c) La parte (c) del teorema anterior demuestra que ambas funciones de regresión del promedio son *lineales*. También demuestra que la varianza de la distribución condicional se reduce en la misma proporción que $(1 - \rho^2)$. Esto es, si ρ está próxima a cero, la varianza condicional es esencialmente la misma que la varianza incondicional, mientras que si ρ está próximo a ± 1 , la varianza condicional está próxima a cero.

La fdp normal bivariada tiene varias propiedades interesantes. Estableceremos algunas en un teorema, dejando la demostración al lector.

Teorema 9.4. Considérese la superficie $z = f(x, y)$, en donde f es la fdp normal bivariada dada en la ecuación (9.3).

- $z = c$ (constante) corta la superficie en una *ellipse*. (Estas se llaman, algunas veces, contornos de una densidad de probabilidades constante).
- Si $\rho = 0$ y $\sigma_x = \sigma_y$, la ellipse anterior se transforma en una circunferencia. (¿Qué sucede a la ellipse anterior cuando $\rho \rightarrow \pm 1$?).

Demostración: ver el problema 9.22.

Observación: debido a la importancia de la distribución normal bivariada, se han tabulado diversas probabilidades asociadas a ella. (Ver D. B. Owen, *Handbook of Statistical Tables*, Addison-Wesley Publishing Company, Inc., Reading, Mass., 1962.)

9.12 Distribuciones truncadas

EJEMPLO 9.10. Supóngase que se fabrica cierto tipo de perno y su longitud, llamémosla Y , es una variable aleatoria con distribución $N(2,2,0,01)$. De un gran lote de tales pernos se saca un nuevo lote, descartando todos aquellos pernos para los cuales $Y > 2$. Por tanto, si X es la variable aleatoria que representa el largo de los pernos en el nuevo lote, y si F es su fda tenemos

$$\begin{aligned} F(x) &= P(X \leq x) \\ &= P(Y \leq x | Y \leq 2) = 1 \quad \text{si } x > 2, \\ &= P(Y \leq x)/P(Y \leq 2) \quad \text{si } x \leq 2. \end{aligned}$$

(Ver la figura 9.9). Luego f , la fdp de X , está dada por

$$\begin{aligned} f(x) &= F'(x) = 0 \quad \text{si } x > 2, \\ &= \frac{1}{\sqrt{2\pi}(0,1)} \exp\left(-\frac{1}{2}\left[\frac{x-2,2}{0,1}\right]^2\right) \quad \text{si } x \leq 2 \end{aligned}$$

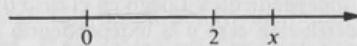


FIGURA 9.9

puesto que

$$P(Y \leq 2) = P\left(\frac{Y-2,2}{0,1} \leq \frac{2-2,2}{0,1}\right) = \Phi(-2).$$

[Φ , como es lo usual, es la fda de la distribución $N(0, 1)$].

La anterior es una ilustración de una *distribución normal truncada* (truncada a la derecha de $X = 2$, específicamente). Este ejemplo puede generalizarse como sigue.

Definición. Decimos que la variable aleatoria X tiene una distribución normal truncada a la derecha de $X = \tau$ si su fdp f es de la forma

$$\begin{aligned} f(x) &= 0 \quad \text{si } x > \tau, \\ &= K \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2\right) \quad \text{si } x \leq \tau. \end{aligned} \tag{9.22}$$

Nótese que si K se determina de la condición $\int_{-\infty}^{+\infty} f(x) dx = 1$ y por tanto

$$K = \frac{1}{\Phi[(\tau - \mu)/\sigma]} = \frac{1}{P(Z \leq \tau)}$$

en donde Z tiene distribución $N(\mu, \sigma^2)$. Análogamente a lo anterior tenemos la siguiente definición.

Definición. Decimos que la variable aleatoria X tiene una distribución normal truncada a la izquierda de $X = \gamma$, si su fdp f es de la forma

$$\begin{aligned} f(x) &= 0 \quad \text{si } x < \gamma, \\ &= \frac{K}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2\right) \quad \text{si } x \geq \gamma. \end{aligned} \quad (9.23)$$

Nuevamente, K se determina, de la condición $\int_{-\infty}^{+\infty} f(x) dx = 1$ y así

$$K = \left[1 - \Phi\left(\frac{\gamma - \mu}{\sigma}\right)\right]^{-1}.$$

Los conceptos presentados anteriormente para la distribución normal pueden extenderse de una manera evidente a otras distribuciones. Por ejemplo, una variable aleatoria X distribuida exponencialmente, truncada a la izquierda de $X = \gamma$, tendría la siguiente fdp:

$$\begin{aligned} f(x) &= 0 \quad \text{si } x < \gamma, \\ &= C\alpha e^{-\alpha x} \quad \text{si } x \geq \gamma. \end{aligned} \quad (9.24)$$

Nuevamente, C está determinada por la condición $\int_{-\infty}^{+\infty} f(x) dx = 1$ y por tanto

$$C = e^{\gamma\alpha}.$$

También podemos considerar una variable aleatoria truncada en el caso discreto. Por ejemplo, si una variable aleatoria X tiene una distribución de Poisson (con parámetro λ) está truncada a la derecha en $X = k + 1$, significa que X tiene la distribución siguiente:

$$\begin{aligned} P(X = i) &= 0 \quad \text{si } i \geq k + 1, \\ &= C \frac{\lambda^i}{i!} e^{-\lambda} \quad \text{si } i = 0, 1, \dots, k. \end{aligned} \quad (9.25)$$

De la condición $\sum_{i=0}^{\infty} P(X = i) = 1$ determinamos C y encontramos

$$C = \frac{1}{\sum_{j=0}^k (\lambda^j/j!)e^{-\lambda}}.$$

Así

$$P(X = i) = \frac{\lambda^i}{i!} \frac{1}{\sum_{j=0}^k (\lambda^j/j!)}, \quad i = 0, 1, \dots, k \text{ y } 0 \text{ para cualquier otro valor.}$$

Las distribuciones truncadas pueden aparecer en muchas aplicaciones importantes. Consideraremos algunos ejemplos a continuación.

EJEMPLO 9.11. Supóngase que X representa la duración de una componente. Si X está distribuida normalmente con

$$E(X) = 4 \quad \text{y} \quad V(X) = 4,$$

encontramos que

$$P(X < 0) = \Phi(-2) = 0,023.$$

Así, este modelo no es muy preciso puesto que asigna una probabilidad 0,023 a un suceso que sabemos que *no puede* ocurrir. En su lugar, podríamos considerar la variable aleatoria X anterior truncada a la izquierda en $X = 0$. Por tanto supondremos que la fdp de la variable aleatoria X está dada por

$$\begin{aligned} f(x) &= 0 \quad \text{si } x \leq 0, \\ &= \frac{1}{\sqrt{(2\pi)(2)}} \exp\left[-\frac{1}{2}\left(\frac{x-4}{2}\right)^2\right] \frac{1}{\Phi(2)} \quad \text{si } x > 0. \end{aligned}$$

Observación: hemos indicado que a menudo usamos la distribución normal para representar una variable aleatoria X de la cual *sabemos* que no puede tomar valores negativos. (Por ejemplo, el tiempo para fallar, el largo de una varilla, etc). Para ciertos valores de los parámetros $\mu = E(X)$ y $\sigma^2 = V(X)$ el valor de $P(X < 0)$ será despreciable. Sin embargo, si no es este el caso (como en el ejemplo 9.11) deberíamos considerar el uso de la distribución normal truncada a la izquierda en $X = 0$.

EJEMPLO 9.12. Supóngase que un sistema está formado por n componentes que funcionan independientemente y tiene cada uno la misma probabilidad p de funcionar correctamente. Cada vez que el sistema funciona mal, se inspecciona a fin de establecer cuántos y cuáles son los componentes que fallan. Definamos la variable aleatoria X como el número de componentes que fallan en un sistema descompuesto. Si suponemos que el sistema falla si y sólo si al menos un componente falla, entonces X tiene una *distribución binomial truncada a la izquierda* en $X = 0$. Por el hecho de que *haya* fallado el sistema impide la posibilidad de que $X = 0$. Específicamente tenemos

$$P(X = k) = \frac{\binom{n}{k}(1-p)^k p^{n-k}}{P(\text{sistema falla})}, \quad k = 1, 2, \dots, n.$$

Puesto que $P(\text{el sistema falla}) = 1 - p^n$, podemos escribir

$$P(X = k) = \frac{\binom{n}{k}(1-p)^k p^{n-k}}{1 - p^n}, \quad k = 1, 2, \dots, n.$$

EJEMPLO 9.13. Supóngase que una fuente radiactiva emite partículas de acuerdo con una distribución de Poisson con parámetro λ . Un dispositivo para contar esas emisiones funciona sólo si llegan menos de tres partículas. (Esto es,

si llegan más de tres partículas durante un período de tiempo especificado, el dispositivo deja de funcionar debido a que se produce un «cierre»). Por tanto, si Y es el número de partículas anotadas durante el intervalo de tiempo específico, Y tiene los valores posibles 0, 1 y 2. Así,

$$P(Y = k) = \frac{e^{-\lambda}}{k!} \frac{\lambda^k}{e^{-\lambda}[1 + \lambda + (\lambda^2/2)]}, \quad k = 0, 1, 2,$$

$$= 0, \quad \text{para cualquier otro valor.}$$

Puesto que la distribución normal truncada es muy importante, consideremos el problema siguiente asociado con esta distribución.

Supóngase que X es una variable aleatoria distribuida normalmente truncada hacia la derecha en $X = \tau$. Luego la fdp f es de la forma

$$f(x) = 0 \quad \text{si } x \geq \tau,$$

$$= \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] \frac{1}{\Phi[(\tau-\mu)/\sigma]} \quad \text{si } x \leq \tau.$$

Por tanto tenemos

$$E(X) = \int_{-\infty}^{+\infty} x f(x) dx = \frac{1}{\Phi[(\tau-\mu)/\sigma]} \int_{-\infty}^{\tau} \frac{x}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] dx$$

$$= \frac{1}{\Phi[(\tau-\mu)/\sigma]} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(\tau-\mu)/\sigma} (s\sigma + \mu) e^{-s^2/2} ds$$

$$= \frac{1}{\Phi[(\tau-\mu)/\sigma]} \left[\mu \Phi\left(\frac{\tau-\mu}{\sigma}\right) + \sigma \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(\tau-\mu)/\sigma} s e^{-s^2/2} ds \right]$$

$$= \mu + \frac{\sigma}{\Phi[(\tau-\mu)/\sigma]} \frac{1}{\sqrt{2\pi}} e^{-s^2/2} (-1)|_{-\infty}^{(\tau-\mu)/\sigma}$$

$$= \mu - \frac{\sigma}{\Phi[(\tau-\mu)/\sigma]} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\tau-\mu}{\sigma}\right)^2\right].$$

Nótese que la expresión obtenida para $E(X)$ se expresa mediante las funciones tabuladas. La función Φ naturalmente es la fda corriente de la distribución $N(0, 1)$, mientras que $(1/\sqrt{2\pi})e^{-x^2/2}$ es la ordenada de la fdp de la distribución $N(0, 1)$ y también está tabulada. En realidad, el cociente

$$\frac{(1/\sqrt{2\pi})e^{-x^2/2}}{\Phi(x)}$$

está tabulado. (Ver D. B. Owen, *Handbook of Statistical Tables*, Addison-Wesley Publishing Company, Inc., Reading, Mass., 1962.)

Utilizando el resultado anterior, podemos formular ahora la siguiente pregunta: para μ y σ dados, ¿dónde debería ocurrir la truncación (esto es, ¿cuál debería ser el valor de τ) de modo que el valor esperado *después* de la truncación tenga algún valor A pre asignado? Podemos responder esta pregunta con ayuda de la distribución normal tabulada. Supóngase que $\mu = 10$, $\sigma = 1$, y necesitamos que sea $A = 9.5$. Debemos resolver por tanto

$$9,5 = 10 - \frac{1}{\Phi(\tau - 10)} \cdot \frac{1}{\sqrt{2\pi}} e^{-(\tau - 10)^2/2}.$$

Esto se transforma en

$$\frac{1}{2} = \frac{(1/\sqrt{2\pi})e^{-(\tau - 10)^2/2}}{\Phi(\tau - 10)}.$$

Utilizando las tablas ya citadas, encontramos que $\tau - 10 = 0.52$, luego $\tau = 10.52$.

Observación: el problema presentado anteriormente sólo puede resolverse para ciertos valores de μ , σ , y A . Esto es, para μ y σ dados, puede no ser posible un valor específico de A . Consideremos la ecuación que puede resolverse:

$$\mu - A = \frac{\sigma}{\Phi[(\tau - \mu)/\sigma]} \frac{1}{\sqrt{2\pi}} \exp \left[-\frac{1}{2} \left(\frac{\tau - \mu}{\sigma} \right)^2 \right].$$

El segundo miembro de esta ecuación es evidentemente positivo. Luego debemos tener $(\mu - A) > 0$ con el objeto de que el problema anterior tenga solución. Esta condición no es muy inesperada puesto que sencillamente expresa que el valor esperado (después de una truncación a la derecha) debe ser menor que el valor esperado original.

PROBLEMAS

9.1. Supóngase que X tiene una distribución $N(2, 0, 16)$. Use la tabla de la distribución normal, para evaluar las probabilidades siguientes.

$$(a) P(X \geq 2,3) \quad (b) P(1,8 \leq X \leq 2,1)$$

9.2. El diámetro de un cable eléctrico está distribuido normalmente con promedio 0,8 y varianza 0,0004. ¿Cuál es la probabilidad de que el diámetro sobrepase 0,81 pulgadas?

9.3. Suponiendo que el cable en el problema 9.2 se considere defectuoso si el diámetro se diferencia de su promedio en más de 0,025. ¿Cuál es la probabilidad de obtener un cable defectuoso?

9.4. Se sabe que los errores en cierto instrumento para medir longitudes están distribuidos normalmente con valor esperado cero y desviación estándar de 1 pulgada. ¿Cuál es la probabilidad de que al medir los errores, sean mayores de 1 pulgada, 2 pulgadas, 3 pulgadas?

9.5. Suponiendo que la duración de los instrumentos electrónicos D_1 y D_2 tienen distribuciones $N(40, 36)$ y $N(45, 9)$ respectivamente. ¿Cuál debe preferirse para usarlo durante un periodo de 45 horas? ¿Cuál debe preferirse para usarlo durante un periodo de 48 horas?

9.6. Podemos interesarnos sólo en la magnitud de X , llamemos $Y = |X|$. Si X tiene una distribución $N(0, 1)$, determine la fdp de Y , y calcule $E(Y)$ y $V(Y)$.

9.7. Supóngase que estamos determinando la posición de un objeto en un plano. Sean X y Y los errores en las medidas de las coordenadas x y y respectivamente. Supóngase que X y Y son independientes y distribuidas idénticamente $N(0, \sigma^2)$. Encuentre la fdp de $R = \sqrt{X^2 + Y^2}$. (La distribución de R se conoce como la *distribución de Rayleigh*). [Indicación: sea $X = R \cos \psi$ y $Y = R \sin \psi$. Obtenga la fdp conjunta de (R, ψ) y luego obtenga la densidad marginal de R .]

9.8. Encuentre la fdp de la variable aleatoria $Q = X/Y$, en donde X y Y están distribuidas como en el problema 9.7. (La distribución de Q se conoce como la *distribución de Cauchy*). ¿Puede usted calcular $E(Q)$?

9.9. Una distribución muy relacionada con la distribución normal es la *distribución lognormal*. Supóngase que X está distribuida normalmente con promedio μ y varianza σ^2 . Sea $Y = e^X$. Entonces Y tiene la distribución log normal. (O sea Y es log normal si y sólo si $\ln Y$ es normal). Encuentre la fdp de Y . *Observación*: las variables aleatorias siguientes pueden representarse por la distribución anterior: el diámetro de partículas pequeñas después de un proceso de trituración, el tamaño de un organismo bajo la acción de pequeños impulsos y la duración de ciertos artículos.

9.10. Supóngase que X tiene una distribución $N(\mu, \sigma^2)$. Determine c (como una función de μ y σ) tal que $P(X \leq c) = 2P(X > c)$.

9.11. Supóngase que la temperatura (medida en grados centígrados) está distribuida normalmente con esperanza 50° y varianza 4. ¿Cuál es la probabilidad de que la temperatura T esté entre 48° y 53° centígrados?

9.12. Se especifica que el diámetro exterior de una flecha, llamémoslo D , debe ser de 4 pulgadas. Supóngase que D es una variable aleatoria distribuida normalmente con promedio 4 pulgadas y varianza 0,01 pulgada². Si el diámetro real se diferencia del valor especificado por más de 0,05 pulgada pero en menos de 0,08 pulgada, la pérdida del fabricante es de \$0,50. Si el diámetro real se diferencia del diámetro especificado en más de 0,08 pulgada, la pérdida es de \$1,00. La pérdida L , puede considerarse como una variable aleatoria. Encuentre la distribución de probabilidades de L y calcule $E(L)$.

9.13. Compare la *cota superior* de la probabilidad $P[|X - E(X)| \geq 2\sqrt{V(X)}]$ obtenida con la desigualdad de Chebyshev con la probabilidad exacta en cada uno de los casos siguientes:

- (a) X tiene distribución $N(\mu, \sigma^2)$.
- (b) X tiene distribución Poisson con parámetro λ .
- (c) X tiene distribución exponencial con parámetro α .

9.14. Supóngase que X es una variable aleatoria para la cual $E(X) = \mu$ y $V(X) = \sigma^2$. Suponiendo que Y está distribuida uniformemente en el intervalo (a, b) , determine a y b de modo que $E(X) = E(Y)$ y $V(X) = V(Y)$.

9.15. Supóngase que X , la resistencia a la ruptura de una cuerda (en libras), tiene distribución $N(100, 16)$. Cada 100 pies de alambre para cuerda produce una utilidad de \$25, si $X > 95$. Si $X \leq 95$, la cuerda puede utilizarse con un objetivo diferente y se obtiene una utilidad de \$10 por alambre. Encuentre la utilidad esperada por alambre.

9.16. Sean X_1 y X_2 variables aleatorias independientes cada una con una distribución

$N(\mu, \sigma^2)$. Sea $Z(t) = X_1 \cos \omega t + X_2 \operatorname{sen} \omega t$. Esta variable aleatoria es de interés en el estudio de señales aleatorias. Sea $V(t) = dZ(t)/dt$. (Se supone que ω es constante.)

(a) ¿Cuál es la distribución de probabilidades de $Z(t)$ y $V(t)$ para cualquier t fija?

(b) Demuestre que $Z(t)$ y $V(t)$ no están correlacionados. [Observación: puede demostrarse que $Z(t)$ y $V(t)$ son independientes pero es más difícil de hacer].

9.17. Un combustible para cohetes va a contener cierto porcentaje (llámémoslo X) de un compuesto particular. Las especificaciones exigen que X esté entre 30 y 35 por ciento. El fabricante tendrá una utilidad neta en el combustible (por galón) que es la siguiente función de X :

$$\begin{aligned} T(X) &= \$0,10 \text{ por galón} & \text{si } 30 < X < 35, \\ &= \$0,05 \text{ por galón} & \text{si } 35 \leq X < 40 \text{ ó } 25 < X \leq 30, \\ &= -\$0,10 \text{ por galón.} \end{aligned}$$

(a) Si X tiene distribución $N(33, 9)$, calcule $E(T)$.

(b) Supóngase que el fabricante desea aumentar su utilidad esperada $E(T)$ en 50 por ciento, aumentando su utilidad (por galón) en aquellas partidas de combustible que satisfacen las especificaciones, $30 < X < 35$. ¿Cuál debe ser su utilidad neta?

9.18. Consideremos el ejemplo 9.8. Suponiendo que se le pagan C_3 dólares/hora al operador mientras la máquina está funcionando y C_4 dólares/hora ($C_4 < C_3$) por el resto del tiempo durante el cual ha sido contratado después de que la máquina ha fallado. Determine nuevamente para qué valor de H (el número de horas que el operador es contratado), la utilidad esperada es máxima.

9.19. Demostrar que $\Gamma(\frac{1}{2}) = \sqrt{\pi}$. (Ver 9.15). [Indicación: hacer el cambio de variables $x = u^2/2$ en la integral $\Gamma(\frac{1}{2}) = \int_0^\infty x^{-1/2} e^{-x} dx$.]

9.20. Verificar las expresiones de $E(X)$ y $V(X)$, cuando X tiene una distribución Gama. [Ver ecuación (9.18)].

9.21. Demostrar el teorema 9.3.

9.22. Demostrar el teorema 9.4.

9.23. Supóngase que la variable aleatoria X tiene una distribución de χ^2 cuadrado con 10 grados de libertad. Si se nos pidiera encontrar dos números a y b tales que $P(a < x < b) = 0.85$, podríamos verificar que existen muchos pares de esa clase.

(a) Encuentre dos conjuntos diferentes de valores (a, b) que satisfagan la condición anterior.

(b) Supóngase que además de lo anterior, necesitemos que

$$P(X < a) = P(X > b).$$

¿Cuántos conjuntos de valores hay?

9.24. Supóngase que V , la velocidad (cm/seg) de un objeto que tiene una masa de 1 kilo, es una variable aleatoria que tiene una distribución $N(0, 25)$. Representemos por $K = 1000V^2/2 = 500V^2$ la energía cinética (KE) del objeto. Calcule $P(K < 200)$, $P(K > 800)$.

9.25. Supóngase que X tiene distribución $N(\mu, \sigma^2)$. Obtener una expresión aproximada para $E(Y)$ y $V(Y)$, usando el teorema 7.7. Si $Y = \ln X$.

9.26. Supóngase que X tiene una distribución normal truncada a la derecha como se da en la ecuación (9.22). Encontrar una expresión para $E(X)$ mediante funciones tabuladas.

9.27. Supóngase que X tiene una distribución exponencial truncada a la izquierda como está dada en la ecuación (9.24). Obtener $E(X)$.

9.28. (a) Encontrar la distribución de probabilidades de una variable aleatoria distribuida binomialmente (con base en n repeticiones de un experimento) truncado a la derecha en $X = n$; esto es, $X = n$ no puede ser observado.

(b) Encontrar el valor esperado y la varianza de la variable aleatoria descrita en (a).

9.29. Supóngase que una variable aleatoria distribuida normalmente con valor esperado μ y varianza σ^2 está truncada a la izquierda en $X = \tau$ y a la derecha en $X = \gamma$. Encontrar la fdp de esta variable aleatoria «truncada doblemente».

9.30. Supóngase que X , el largo de una varilla tiene distribución $N(10, 2)$. En vez de medir el valor de X , sólo se especifica si se cumplen ciertas exigencias. Específicamente, cada varilla fabricada se clasifica como sigue: $X < 8$, $8 \leq X < 12$, y $X \geq 12$. Si se fabrican 15 de tales varillas, ¿cuál es la probabilidad de que un número igual de varillas caiga en cada una de las categorías anteriores?

9.31. Se sabe que la lluvia anual que cae en cierta región es una variable aleatoria distribuida normalmente con promedio igual a 29.5 pulgadas y desviación standard 2.5 pulgadas. ¿Cuántas pulgadas de lluvia (anuales) caen en exceso alrededor del 5 por ciento de las veces?

9.32. Suponiendo que X tiene una distribución $N(0, 25)$, calcule $P(1 < X^2 < 4)$.

9.33. Sea X_t el número de partículas emitidas en t horas por una fuente radiactiva y se supone que X_t tiene una distribución de Poisson con parámetro βt . Sea T el número de horas hasta la primera emisión. Demuestre que T tiene una distribución exponencial con parámetro β [Indicación: encuentre el suceso equivalente (mediante X_t) del suceso $T > t$].

9.34. Supóngase que X_t está definida como en el problema 9.33 con $\beta = 30$. ¿Cuál es la probabilidad de que el tiempo entre emisiones sucesivas sea > 5 minutos? > 10 minutos? < 30 segundos?

9.35. En algunas tablas de la distribución normal $H(x) = (1/\sqrt{2\pi}) \int_0^x e^{-t^2/2} dt$ está tabulada para valores positivos de x (en vez de $\Phi(x)$ como aparece en el Apéndice). Si la variable aleatoria X tiene distribución $N(1, 4)$ expresar cada una de las probabilidades siguientes mediante valores tabulados de la función H .

(a) $P[|X| > 2]$ (b) $P[X < 0]$.

9.36. Supóngase que un dispositivo para telemedir satélites recibe dos clases de señales que pueden anotarse como números reales, X y Y , y que X y Y son variables aleatorias continuas independientes con fdp f y g , respectivamente. Supóngase que durante cualquier período de tiempo específico sólo puede recibirse una de esas señales, y luego retransmitida a la tierra, la señal que primero llega. Además la señal que origina X llega primero con probabilidad p y por tanto la señal que origina Y llega primero con probabilidad $1 - p$. Denotemos por Z la variable aleatoria cuyo valor realmente es recibido y transmitido.

(a) Expresar la fdp de Z mediante f y g .

(b) Expresar $E(Z)$ mediante $E(X)$ y $E(Y)$.

(c) Expresar $V(Z)$ mediante $V(X)$ y $V(Y)$.

(d) Supóngase que X tiene distribución $N(2, 4)$ y que Y tiene distribución $N(3, 3)$. Si $p = \frac{2}{3}$, calcule $P(Z > 2)$.

(e) Suponiendo que X y Y tienen distribuciones $N(\mu_1, \sigma_1^2)$ y $N(\mu_2, \sigma_2^2)$ respectivamente. Demuestre que si $\mu_1 = \mu_2$, la distribución de Z es «uni-modal», esto es la fdp de Z tiene un máximo relativo único.

9.37. Supóngase que el número de accidentes en una fábrica se puede representar por un proceso de Poisson con un promedio de 2 accidentes por semana. ¿Cuál es la probabilidad de que (a) el tiempo entre un accidente y el siguiente sea mayor de 3 días, (b) el tiempo de un accidente al tercero sea mayor de una semana? [Indicación: en (a), sea T = tiempo (en días) y calcule $P(T > 3)$].

9.38. Un proceso de fabricación produce en promedio un artículo defectuoso entre 300 fabricados. ¿Cuál es la probabilidad de que aparezca el *tercer* artículo defectuoso:

- antes de que sean producidos 1000 artículos?
- cuando se produce el 1000 *ésimo* artículo?
- después de que sea producido el 1000 *ésimo* artículo?

[Indicación: supóngase un proceso de Poisson].

La función generadora de momentos

10.1 Introducción

En este capítulo presentaremos un concepto matemático importante que tiene muchas aplicaciones en los modelos probabilísticos que estamos considerando. Con el objeto de presentar un desarrollo riguroso de este tema, habrían sido necesarias matemáticas de un nivel considerablemente mayor del que estamos suponiendo aquí. Sin embargo, si queremos evitar ciertas dificultades matemáticas que aparecen y si aceptamos que ciertas operaciones son válidas, entonces podemos obtener una comprensión suficiente de las principales ideas implicadas a fin de usarlas inteligentemente.

Con el objeto de motivar lo que sigue, recordemos nuestro primer encuentro con el logaritmo. Fue presentado sólo como una ayuda para calcular. Con cada número real positivo x , asociamos otro número, designado como $\log x$. (El valor de este número pudo obtenerse de tablas.) A fin de calcular xy , por ejemplo, obtenemos el valor de $\log x$ y $\log y$ y luego calculamos $\log x + \log y$, que representa $\log xy$. Conociendo $\log xy$, pudimos obtener luego el valor de xy (con la ayuda de tablas nuevamente). De manera semejante podemos simplificar con ayuda de los logaritmos la elaboración de otros cálculos aritméticos. El enfoque anterior es útil por las siguientes razones:

- (a) A cada número positivo x le corresponde exactamente un número, $\log x$, y este número se obtiene fácilmente de las tablas.
- (b) A cada valor de $\log x$ le corresponde exactamente un valor de x , y este valor nuevamente se obtiene de tablas. (Luego, la relación entre x y $\log x$ es uno a uno.)

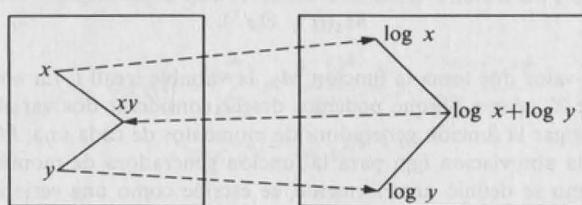


FIGURA 10.1

(c) Ciertas operaciones aritméticas que relacionan los números x y y , tales como multiplicación y división, pueden reemplazarse por operaciones más sencillas, tales como adición y sustracción, mediante los números «transformados» $\log x$ y $\log y$ (ver el esquema de la figura 10.1).

En vez de efectuar las operaciones directamente con los números x y y , obtenemos primero los números $\log x$ y $\log y$, hacemos nuestros cálculos con esos números, y luego los transformamos de nuevo.

10.2 La función generadora de momentos

Ahora consideraremos una situación más complicada. Supóngase que X es una variable aleatoria; es decir, X es una función del espacio muestral a los números reales. Al calcular diversas características de la variable aleatoria X , tales como $E(X)$ o $V(X)$, trabajamos directamente con la distribución de probabilidades de X [la distribución de probabilidades está dada por una *función*: la fdp en el caso continuo, o las probabilidades puntuales $p(x_i) = P(X = x_i)$ en el caso discreto. La última se puede también considerar como una función que toma valores distintos de cero sólo si $X = x_i, i = 1, 2, \dots$]. Posiblemente, podemos presentar otra *función* y hacer los cálculos necesarios mediante esta nueva función (tal como antes asociábamos con cada *número*, un nuevo *número*). Esto es, en realidad, lo que precisamente haremos. Primero daremos una definición formal.

Definición. Sea X una variable aleatoria discreta con distribución de probabilidades $p(x_i) = P(X = x_i), i = 1, 2, \dots$. La función M_X , llamada la *función generadora de momentos* de X , está definida por

$$M_X(t) = \sum_{j=1}^{\infty} e^{tx_j} p(x_j). \quad (10.1)$$

Si X es una variable aleatoria continua con fdp f , definimos la función generadora de momentos por

$$M_X(t) = \int_{-\infty}^{+\infty} e^{tx} f(x) dx. \quad (10.2)$$

Observaciones: (a) Tanto en el caso discreto o en el continuo, $M_X(t)$ es simplemente el valor esperado de e^{tX} . Luego podemos combinar las expresiones anteriores y escribir

$$M_X(t) = E(e^{tX}). \quad (10.3)$$

(b) $M_X(t)$ es el valor que toma la función M_X , la variable (real) t . La notación que indica la dependencia de X , se usa porque podemos desechar considerar dos variables aleatorias, X y Y , y luego investigar la función generadora de momentos de cada una, M_X y M_Y .

(c) Usaremos la abreviación fgm para la función generadora de momentos.

(d) La fgm como se definió anteriormente, se escribe como una serie infinita o integral (impropia), según que la variable aleatoria sea discreta o continua. Tal serie (o integral) puede que no exista siempre (es decir, que converja a un valor infinito) para todos los valores de t .

Por tanto pueden suceder que la fgm no esté definida para todos los valores de t . Sin embargo no nos interesaría esta posible dificultad. Cada vez que hagamos uso de la fgm, siempre supondremos que existe. (Para $t = 0$, la fgm existe siempre y es igual a 1.)

(e) Hay otra función muy relacionada con la fgm que a menudo se usa en su lugar. Se llama la *función característica*, se denota como C_X , y se define por $C_X(t) = E(e^{itX})$, en donde $i = \sqrt{-1}$, la unidad imaginaria. Por razones teóricas, hay una ventaja considerable al usar $C_X(t)$ en vez de $M_X(t)$. [Por esta razón, $C_X(t)$ existe siempre para todos los valores de t]. Sin embargo, a fin de evitar cálculos con números complejos restringiremos nuestra exposición a la función generadora de momentos.

(f) Postergaremos hasta la sección 10.4 la exposición de la razón para llamar a M_X la función generadora de momentos.

10.3 Ejemplos de funciones generadoras de momentos

Antes de considerar algunas aplicaciones importantes de la fgm a la teoría de la probabilidad, evaluemos algunas de estas funciones.

EJEMPLO 10.1. Supóngase que X está distribuida *uniformemente* en el intervalo $[a, b]$. Por tanto, la fgm está dada por

$$\begin{aligned} M_X(t) &= \int_a^b \frac{e^{tx}}{b-a} dx \\ &= \frac{1}{(b-a)t} [e^{bt} - e^{at}], \quad t \neq 0. \end{aligned} \quad (10.4)$$

EJEMPLO 10.2. Supóngase que X está distribuida *binomialmente* con parámetros n y p . Luego

$$\begin{aligned} M_X(t) &= \sum_{k=0}^n e^{tk} \binom{n}{k} p^k (1-p)^{n-k} \\ &= \sum_{k=0}^n \binom{n}{k} (pe^t)^k (1-p)^{n-k} \\ &= [pe^t + (1-p)]^n. \end{aligned} \quad (10.5)$$

(Esta última igualdad se deduce de una aplicación directa del teorema del binomio.)

EJEMPLO 10.3. Supóngase que X tiene una distribución de Poisson con parámetro λ . Así

$$\begin{aligned} M_X(t) &= \sum_{k=0}^{\infty} e^{tk} \frac{e^{-\lambda} \lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda e^t)^k}{k!} \\ &= e^{-\lambda} e^{\lambda e^t} = e^{\lambda(e^t - 1)}. \end{aligned} \quad (10.6)$$

(La tercera igualdad se deduce del desarrollo de e^y en $\sum_{n=0}^{\infty} (y^n / n!)$. Usamos esto con $y = \lambda e^t$).

EJEMPLO 10.4. Supóngase que X tiene una *distribución exponencial* con parámetro α . Por tanto,

$$M_X(t) = \int_0^\infty e^{tx} \alpha e^{-\alpha x} dx = \alpha \int_0^\infty e^{x(t-\alpha)} dx.$$

(Esta integral converge sólo si $t < \alpha$. Por tanto, la fgm existe sólo para esos valores de t . Suponiendo que se satisface esta condición, continuamos.) Luego

$$\begin{aligned} M_X(t) &= \frac{\alpha}{t - \alpha} e^{x(t-\alpha)} \Big|_0^\infty \\ &= \frac{\alpha}{\alpha - t}, \quad t < \alpha. \end{aligned} \tag{10.7}$$

Observación: puesto que la fgm es sólo un valor esperado de X , se puede obtener la fgm de una función de una variable aleatoria sin obtener su distribución de probabilidades primero (ver el teorema 7.3). Por ejemplo, si X tiene distribución $N(0,1)$ y deseamos encontrar la fgm de $Y = X^2$; podemos proceder sin obtener primero la fdp de Y . Simplemente podemos escribir

$$M_Y(t) = E(e^{tY}) = E(e^{tX^2}) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp(tx^2 - x^2/2) dx = (1 - 2t)^{-1/2}$$

después de una integración inmediata.

EJEMPLO 10.5 Supóngase que X tiene distribución $N(\mu, \sigma^2)$. Entonces

$$M_X(t) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} e^{tx} \exp\left(-\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2\right) dx.$$

Sea $(x - \mu)/\sigma = s$; luego $x = \sigma s + \mu$ y $dx = \sigma ds$. Por tanto

$$\begin{aligned} M_X(t) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp[t(\sigma s + \mu)] e^{-s^2/2} ds \\ &= e^{t\mu} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp(-\frac{1}{2}[s^2 - 2\sigma ts]) ds \\ &= e^{t\mu} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp\{-\frac{1}{2}[(s - \sigma t)^2 - \sigma^2 t^2]\} ds \\ &= e^{t\mu + \sigma^2 t^2/2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp(-\frac{1}{2}[s - \sigma t]^2) ds. \end{aligned}$$

Sea $s - \sigma t = v$; luego $ds = dv$ y obtenemos

$$\begin{aligned} M_X(t) &= e^{t\mu + \sigma^2 t^2/2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-v^2/2} dv \\ &= e^{(t\mu + \sigma^2 t^2/2)}. \end{aligned} \tag{10.8}$$

EJEMPLO 10.6. Sea X una *distribución Gama* con parámetros α y r (ver la ecuación (9.16)). Entonces

$$\begin{aligned} M_X(t) &= \frac{\alpha}{\Gamma(r)} \int_0^t e^{tx} (\alpha x)^{r-1} e^{-\alpha x} dx \\ &= \frac{\alpha^r}{\Gamma(r)} \int_0^\infty x^{r-1} e^{-x(\alpha-t)} dx. \end{aligned}$$

(Esta integral converge si $\alpha > t$). Sea $x(\alpha - t) = u$; luego

$$dx = (du)/(\alpha - t),$$

y obtenemos

$$\begin{aligned} M_X(t) &= \frac{\alpha^r}{(\alpha - t)\Gamma(r)} \int_0^\infty \left(\frac{u}{\alpha - t} \right)^{r-1} e^{-u} du \\ &= \left(\frac{\alpha}{\alpha - t} \right)^r \frac{1}{\Gamma(r)} \int_0^\infty u^{r-1} e^{-u} du. \end{aligned}$$

Puesto que la integral es igual a $\Gamma(r)$, tenemos

$$M_X(t) = \left(\frac{\alpha}{\alpha - t} \right)^r. \quad (10.9)$$

Observaciones: (a) Si $r = 1$, la función Gama se convierte en la distribución exponencial. Observemos que si $r = 1$, las ecuaciones (10.7) y (10.9) son iguales.

(b) Puesto que la distribución de χ^2 se obtiene como un caso especial de la distribución Gama al hacer $\alpha = 1/2$ y $r = n/2$ (n es un entero positivo), tenemos que si Z tiene distribución χ_n^2 , entonces

$$M_Z(t) = (1 - 2t)^{-n/2}. \quad (10.10)$$

10.4 Propiedades de la función generadora de momentos

Daremos ahora la razón para llamar M_X la función *generadora de momentos*. Recordemos el desarrollo en serie de Maclaurin de la función e^x :

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots + \frac{x^n}{n!} + \cdots$$

(Se sabe que esta serie converge para todos los valores de x .) Así

$$e^{tx} = 1 + tx + \frac{(tx)^2}{2!} + \cdots + \frac{(tx)^n}{n!} + \cdots$$

Ahora

$$M_X(t) = E(e^{tX}) = E \left(1 + tX + \frac{(tX)^2}{2!} + \cdots + \frac{(tX)^n}{n!} + \cdots \right).$$

Hemos demostrado que para una suma *finita*, el valor esperado de la suma es igual a la suma de los valores esperados. Sin embargo, considerando una suma

infinita como la anterior no podemos aplicar, inmediatamente, tal resultado. Sin embargo, resulta que bajo condiciones justamente generales esta operación es todavía válida. Supondremos que las condiciones pedidas se satisfacen y por consiguiente procedemos.

Recordemos que t es una constante en lo que respecta a la esperanza y podemos escribir

$$M_X(t) = 1 + tE(X) + \frac{t^2 E(X^2)}{2!} + \cdots + \frac{t^n E(X^n)}{n!} + \cdots$$

Puesto que M_X es una función de la variable real t , podemos considerar que tomamos la derivada de $M_X(t)$ con respecto a t , esto es, $[d/(dt)]M_X(t)$ o por brevedad, $M'(t)$. Nos encontramos nuevamente con una dificultad matemática. La derivada de una suma finita siempre es igual a la suma de las derivadas. (Suponiendo, naturalmente que todas las derivadas existen.) Sin embargo, para una suma infinita esto no siempre es así. Deben satisfacerse ciertas condiciones a fin de justificar esta operación. Simplemente supondremos que esas condiciones existen y continuaremos. (En la mayor parte de los problemas que encontraremos tal suposición está justificada.) Así,

$$M'(t) = E(X) + tE(X^2) + \frac{t^2 E(X^3)}{2!} + \cdots + \frac{t^{n-1} E(X^n)}{(n-1)!} + \cdots$$

Haciendo $t = 0$ encontramos que sólo subsiste el primer término y tenemos

$$M'(0) = E(X).$$

Luego la primera derivada de la fgm calculada en $t = 0$ da el valor esperado de la variable aleatoria. Si calculamos la segunda derivada de $M_X(t)$, nuevamente procederemos como antes, y obtenemos

$$M''(t) = E(X^2) + tE(X^3) + \cdots + \frac{t^{n-2} E(X^n)}{(n-2)!} + \cdots,$$

y haciendo $t = 0$, tenemos

$$M''(0) = E(X^2).$$

Continuando de esta manera obtenemos el siguiente teorema [suponiendo que $M^{(n)}(0)$ existe].

Teorema 10.1.

$$M^{(n)}(0) = E(X^n) \quad (10.11)$$

Esto es, la n -ésima derivada de $M_X(t)$ calculada en $t = 0$ da $E(X^n)$.

Observaciones: (a) Los números $E(X^n)$, $n = 1, 2, \dots$, se llaman los n -ésimos *momentos* de la variable aleatoria X respecto al cero. Por tanto hemos demostrado que conociendo la función M_X , pueden «generarse» los momentos. (De aquí el nombre de «función generadora de momentos».)

(b) Recordemos el desarrollo en serie de Maclaurin en general de una función h .

$$h(t) = h(0) + h'(0)t + \frac{h''(0)t^2}{2!} + \cdots + \frac{h^{(n)}(0)t^n}{n!} + \cdots,$$

en donde $h^{(n)}(0)$ es la n -ésima derivada de la función h calculada en $t = 0$. Aplicando este resultado a la función M_X , podemos escribir

$$\begin{aligned} M_X(t) &= M_X(0) + M'_X(0)t + \cdots + \frac{M_X^{(n)}(0)t^n}{n!} + \cdots \\ &= 1 + \mu_1 t + \mu_2 t^2/2! + \cdots + \frac{\mu_n t^n}{n!} + \cdots \end{aligned}$$

en donde $\mu_i = E(X^i)$, $i = 1, 2, \dots$. En particular,

$$V(X) = E(X^2) - (E(X))^2 = M''(0) - [M'(0)]^2.$$

(c) El lector puede preguntarse si el método anterior es del todo útil. ¿No sería más simple (y más elemental) calcular directamente los momentos de X en vez de obtener primero la fgm y luego diferenciarla? La respuesta es que para muchos problemas este método es más sencillo. El siguiente lo ilustrará.

EJEMPLO 10.7. Supóngase que X tiene una distribución binomial con parámetros n y p . Por tanto (ejemplo 10.2), $M_X(t) = [pe^t + q]^n$. Por tanto

$$M'(t) = n(pe^t + q)^{n-1}pe^t,$$

$$M''(t) = np[e^t(n-1)(pe^t + q)^{n-2}pe^t + (pe^t + q)^{n-1}e^t]$$

Por tanto $E(X) = M'(0) = np$, que concuerda con nuestro resultado anterior. También, $E(X^2) = M''(0) = np[(n-1)p + 1]$. Luego

$$V(X) = M''(0) - [M'(0)]^2 = np(1-p),$$

lo que nuevamente concuerda con lo encontrado previamente.

EJEMPLO 10.8. Supóngase que X tiene distribución $N(\alpha, \beta^2)$. Por tanto (ejemplo 10.5), $M_X(t) = \exp(\alpha t + \frac{1}{2}\beta^2 t^2)$. Así

$$M'(t) = e^{\alpha t + \beta^2 t^2/2}(\beta^2 t + \alpha),$$

$$M''(t) = e^{\beta^2 t^2/2 + \alpha t} \beta^2 + (\beta^2 t + \alpha)^2 e^{\beta^2 t^2/2 + \alpha t},$$

y $M'(0) = \alpha$, $M''(0) = \beta^2 + \alpha^2$, dado $E(X) = \alpha$ y $V(X) = \beta^2$ como antes.

Usemos el método de las fgm para calcular la esperanza y la varianza de una variable aleatoria con distribución de probabilidades geométrica, ecuación (8.5).

EJEMPLO 10.9. Sea X una distribución de probabilidades geométrica. O sea, $P(X = k) = q^{k-1}p$, $k = 1, 2, \dots$ ($p + q = 1$). Así

$$M_X(t) = \sum_{k=1}^{\infty} e^{tk} q^{k-1} p = \frac{p}{q} \sum_{k=1}^{\infty} (qe^t)^k.$$

Si nos restringimos a aquellos valores de t para los cuales $0 < qe^t < 1$ [esto es, $t < \ln(1/q)$] podemos entonces sumar la serie anterior como una serie geométrica y obtener

$$\begin{aligned} M_X(t) &= \frac{p}{q} qe^t [1 + qe^t + (qe^t)^2 + \dots] \\ &= \frac{p}{q} \frac{qe^t}{1 - qe^t} = \frac{pe^t}{1 - qe^t}. \end{aligned}$$

Por tanto

$$\begin{aligned} M'(t) &= \frac{(1 - qe^t) pe^t - pe^t(-qe^t)}{(1 - qe^t)^2} = \frac{pe^t}{(1 - qe^t)^2}; \\ M''(t) &= \frac{(1 - qe^t)^2 pe^t - pe^t 2(1 - qe^t)(-qe^t)}{(1 - qe^t)^4} = \frac{pe^t(1 + qe^t)}{(1 - qe^t)^3}. \end{aligned}$$

Por tanto

$$\begin{aligned} E(X) &= M'(0) = p/(1 - q)^2 = 1/p, \\ E(X^2) &= M''(0) = p(1 + q)/(1 - q)^3 = (1 + q)/p^2, \end{aligned}$$

y

$$V(X) = (1 + q)/p^2 - (1/p)^2 = q/p^2.$$

Tenemos así una verificación del teorema 8.5.

Los dos teoremas siguientes serán de mucha importancia en nuestras aplicaciones de la fgm.

Teorema 10.2. Supóngase que la variable aleatoria X tiene fgm M_X . Sea $Y = \alpha X + \beta$. Entonces M_Y , la fgm de la variable aleatoria Y está dada por

$$M_Y(t) = e^{\beta t} M_X(\alpha t). \quad (10.12)$$

En palabras: para encontrar la fgm de $Y = \alpha X + \beta$, calculamos la fgm de X en αt (en vez de t) y multiplicamos por $e^{\beta t}$.

Demostración:

$$\begin{aligned} M_Y(t) &= E(e^{Yt}) = E[e^{(\alpha X + \beta)t}] \\ &= e^{\beta t} E(e^{\alpha t X}) = e^{\beta t} M_X(\alpha t). \end{aligned}$$

Teorema 10.3. Sean X y Y dos variables aleatorias con fgm, $M_X(t)$ y $M_Y(t)$ respectivamente. Si $M_X(t) = M_Y(t)$ para todos los valores de t , entonces X y Y tienen la misma distribución de probabilidades.

Demostración: la demostración de este teorema es muy difícil para darla aquí. Sin embargo, es muy importante comprender exactamente lo que establece el teorema. Dice que si dos variables aleatorias tienen la misma fgm, tienen la misma distribución de probabilidades. Esto es, la fgm determina únicamente la distribución de probabilidades de la variable aleatoria.

EJEMPLO 10.10 Supóngase que X tiene distribución $N(\mu, \sigma^2)$. Sea $Y = \alpha X + \beta$. Luego Y está de nuevo distribuida normalmente. Del teorema 10.2 la fgm de Y es $M_Y(t) = e^{\beta t} M_X(\alpha t)$. Sin embargo, del ejemplo 10.5 tenemos que

$$M_X(t) = e^{\mu t + \sigma^2 t^2/2}.$$

Por tanto

$$\begin{aligned} M_Y(t) &= e^{\beta t} [e^{\alpha \mu t + (\alpha \sigma)^2 t^2/2}] \\ &= e^{(\beta + \alpha \mu)t} e^{(\alpha \sigma)^2 t^2/2}. \end{aligned}$$

Pero ésta es la fgm de una variable aleatoria distribuida normalmente con esperanza $\alpha\mu + \beta$ y varianza $\alpha^2\sigma^2$. Luego, de acuerdo con el teorema 10.3, la distribución de Y es normal.

El teorema siguiente también desempeña un papel vital en nuestro trabajo posterior.

Teorema 10.4. Supóngase que X y Y son variables aleatorias independientes.

Sea $Z = X + Y$. Sean $M_X(t)$, $M_Y(t)$ y $M_Z(t)$ las fgm de las variables aleatorias X , Y y Z , respectivamente. Luego

$$M_Z(t) = M_X(t)M_Y(t) \quad (10.13)$$

Demostración:

$$\begin{aligned} M_Z(t) &= E(e^{Zt}) = E[e^{(X+Y)t}] = E(e^{Xt}e^{Yt}) \\ &= E(e^{Xt})E(e^{Yt}) = M_X(t)M_Y(t). \end{aligned}$$

Observación: este teorema puede generalizarse como sigue: si X_1, \dots, X_n son variables aleatorias independientes con fgm $M_{X_i}(t)$, $i = 1, 2, \dots, n$, entonces la fgm de

$$Z = X_1 + \cdots + X_n,$$

está dada por

$$M_Z(t) = M_{X_1}(t) \cdots M_{X_n}(t). \quad (10.14)$$

10.5 Propiedades reproductivas

Hay varias distribuciones de probabilidades que tienen la notable y útil propiedad siguiente: si dos (o más) variables aleatorias independientes que tienen cierta distribución se suman, la variable aleatoria que resulta tiene una distribución del mismo tipo que la de los sumandos. Esta propiedad se llama *propiedad reproductiva*, y la estableceremos para varias distribuciones importantes con ayuda de los teoremas 10.3 y 10.4.

EJEMPLO 10.11. Supóngase que X y Y son variables aleatorias independientes con distribuciones $N(\mu_1, \sigma_1^2)$ y $N(\mu_2, \sigma_2^2)$ respectivamente. Sea $Z = X + Y$. Por tanto

$$\begin{aligned} M_Z(t) &= M_X(t)M_Y(t) = \exp(\mu_1 t + \sigma_1^2 t^2/2) \exp(\mu_2 t + \sigma_2^2 t^2/2) \\ &= \exp[(\mu_1 + \mu_2)t + (\sigma_1^2 + \sigma_2^2)t^2/2]. \end{aligned}$$

Sin embargo, esto representa la fgm de una variable aleatoria distribuida normalmente con valor esperado $\mu_1 + \mu_2$ y varianza $\sigma_1^2 + \sigma_2^2$. Así Z tiene esta distribución normal. (Ver el teorema 10.3.)

Observación: $E(Z) = \mu_1 + \mu_2$ y $V(Z) = \sigma_1^2 + \sigma_2^2$ pudo haberse obtenido inmediatamente de los resultados anteriores relacionados con las propiedades de la esperanza y la varianza. Pero establecer que Z es nuevamente distribuida *normalmente* necesitó el uso de la fgm. (Hay otro enfoque de este resultado que se mencionará en el capítulo 12.)

EJEMPLO 10.12. La longitud de una varilla es una variable aleatoria distribuida normalmente con medida 4 pulgadas y varianza 0,01 pulgada². Dos de tales varillas se ponen extremo con extremo y se ajustan en una muesca. El largo de esta muesca es de 8 pulgadas con una tolerancia de $\pm 0,1$ de pulgada. ¿Cuál es la probabilidad de que se ajusten las dos varillas?

Representando por L_1 y L_2 las longitudes de la varilla 1 y la varilla 2, tenemos que $L = L_1 + L_2$ está distribuida normalmente con $E(L) = 8$ y $V(L) = 0,02$. Por tanto

$$\begin{aligned} P[7,9 \leq L \leq 8,1] &= P\left[\frac{7,9 - 8}{0,14} \leq \frac{L - 8}{0,14} \leq \frac{8,1 - 8}{0,14}\right] \\ &= \Phi(+0,714) - \Phi(-0,714) = 0,526, \end{aligned}$$

de las tablas de la distribución normal.

Podemos generalizar el resultado anterior en el teorema siguiente.

Teorema 10.5. (*la propiedad reproductiva de la distribución normal*). Sean X_1, X_2, \dots, X_n n variables aleatorias independientes con distribución $N(\mu_i, \sigma_i^2)$, $i = 1, 2, \dots, n$. Sea $Z = X_1 + \dots + X_n$. Luego Z tiene distribución $N(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2)$.

La distribución de Poisson también posee una propiedad reproductiva.

Teorema 10.6. Sean X_1, \dots, X_n variables aleatorias independientes. Supóngase que X_i tiene una distribución de Poisson con parámetro α_i , $i = 1, 2, \dots, n$. Sea $Z = X_1 + \dots + X_n$. Luego Z tiene una distribución de Poisson con parámetro

$$\alpha = \alpha_1 + \dots + \alpha_n.$$

Demostración: consideremos primero el caso de $n = 2$.

$$M_{X_1}(t) = e^{\alpha_1(e^t - 1)}, \quad M_{X_2}(t) = e^{\alpha_2(e^t - 1)}.$$

Luego $M_Z(t) = e^{(\alpha_1 + \alpha_2)(e^t - 1)}$. Pero esta es la fgm de una variable aleatoria con una distribución de Poisson que tiene parámetro $\alpha_1 + \alpha_2$. Ahora podemos completar la demostración del teorema con ayuda de la inducción matemática.

EJEMPLO 10.13. Supóngase que el número de llamadas que llegan a una central telefónica entre las 9 a.m. y 10 a.m., es una variable aleatoria X_1 , con una distribución de Poisson con parámetro 3. Análogamente, el número de llamadas, que llegan entre las 10 a.m. y 11 a.m., digamos X_2 , también tiene una distribución de Poisson, con parámetro 5. Si X_1 y X_2 son independientes, ¿cuál es la probabilidad de que se reciban más de 5 llamadas entre las 9 a.m. y las 11 a.m.?

Sea $Z = X_1 + X_2$. Del teorema anterior, Z tiene una distribución de Poisson con parámetro $3 + 5 = 8$. Por tanto

$$\begin{aligned} P(Z > 5) &= 1 - P(Z \leq 5) = 1 - \sum_{k=0}^5 \frac{e^{-8}(8)^k}{k!} \\ &= 1 - 0,1912 = 0,8088. \end{aligned}$$

Otra distribución con una propiedad reproductiva es la distribución de chi cuadrado.

Teorema 10.7. Supóngase que la distribución de X es $\chi_{n_i}^2$, $i = 1, 2, \dots, k$, en donde los X_i son variables aleatorias independientes. Sea $Z = X_1 + \dots + X_k$. Entonces Z tiene distribución χ_n^2 , en donde $n = n_1 + \dots + n_k$.

Demostración: de la ecuación (10.10) tenemos $M_{X_i}(t) = (1 - 2t)^{-n_i/2}$, $i = 1, 2, \dots, k$.

Luego

$$M_Z(t) = M_{X_1}(t) \cdots M_{X_k}(t) = (1 - 2t)^{-(n_1 + \dots + n_k)/2}.$$

Pero esta es la fgm de una variable aleatoria que tiene la distribución χ_n^2 .

Ahora podemos dar una de las razones de la gran importancia de la distribución de chi cuadrado. En el ejemplo 9.9 encontramos que si X tiene distribución $N(0,1)$, X^2 tiene distribución χ_1^2 . Combinando esto con el teorema 10.7, tenemos el resultado siguiente.

Teorema 10.8. Supóngase que X_1, \dots, X_k son variables aleatorias independientes, donde cada una tiene distribución $N(0,1)$. Entonces $S = X_1^2 + X_2^2 + \dots + X_k^2$ tiene distribución χ_k^2 .

EJEMPLO 10.14. Supóngase que X_1, \dots, X_n son variables aleatorias independientes cada una con distribución $N(0,1)$. Sea $T = \sqrt{X_1^2 + \dots + X_n^2}$. De nuestra exposición previa sabemos que T^2 tiene distribución χ_n^2 .

Para encontrar la fdp de T , llamada h , proseguiremos como es corriente:

$$\begin{aligned} H(t) &= P(T \leq t) = P(T^2 \leq t^2) \\ &= \int_0^{t^2} \frac{1}{2^{n/2} \Gamma(n/2)} z^{n/2 - 1} e^{-z/2} dz. \end{aligned}$$

Por tanto, tenemos

$$\begin{aligned} h(t) &= H'(t) \\ &= \frac{2t}{2^{n/2}\Gamma(n/2)}(t^2)^{n/2-1}e^{-t^2/2} \\ &= \frac{2t^{n-1}e^{-t^2/2}}{2^{n/2}\Gamma(n/2)} \quad \text{si } t \geq 0. \end{aligned}$$

Observaciones: (a) Si $n = 2$, la distribución anterior se conoce como una *distribución de Rayleigh*. (Ver problema 9.7.)

(b) Si $n = 3$, la distribución anterior se conoce como distribución de Maxwell (o, algunas veces, distribución de velocidad de Maxwell) y tiene la siguiente interpretación importante. Supóngase que tenemos gas en un depósito cerrado. Representemos por (X, Y, Z) las componentes de la velocidad de una molécula escogida al azar. Supondremos que X, Y y Z son variables aleatorias independientes cada una con una distribución $N(0, \sigma^2)$. (Suponer la misma distribución para X, Y y Z , significa que la presión en el gas es la misma en todas direcciones. Suponer que las esperanzas son iguales a cero, significa que el gas no está escapándose.) Por lo tanto la *velocidad* de la molécula (es decir, la magnitud de su velocidad) está dada por $S = \sqrt{X^2 + Y^2 + Z^2}$. Observemos que $X/\sigma, Y/\sigma$ y Z/σ están distribuidas $N(0,1)$. Así $S/\sigma = \sqrt{(X/\sigma)^2 + (Y/\sigma)^2 + (Z/\sigma)^2}$, tiene la distribución de Maxwell. Por tanto g , la fdp de la velocidad S , está dada por

$$g(s) = \frac{2\sigma(\sigma s)^2}{\sqrt{2\pi}} e^{-\sigma^2 s^2/2}, \quad s \geq 0.$$

El gráfico de g se da en la figura 10.2 para $\sigma = 2$. Observe que valores muy grandes o muy pequeños de S son muy improbables. (Puede demostrarse que la constante σ que aparece como un parámetro en la distribución anterior, tiene la siguiente interpretación física: $\sigma = \sqrt{kT/M}$, en donde T es la temperatura absoluta, M es la masa de la molécula, y k se conoce como la constante de Boltzmann.)

Hemos expuesto algunas distribuciones que tienen la propiedad reproductiva. Consideraremos la distribución exponencial que, hablando estrictamente, no posee la propiedad reproductiva, pero que, sin embargo, tiene una propiedad análoga.

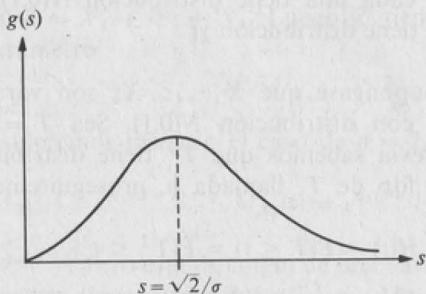


FIGURA 10.2

Sean $X_i, i = 1, 2, \dots, r$ es r variables aleatorias independientes, con idéntica distribución exponencial con parámetro α . Luego de la ecuación (10.7) tenemos

$$M_{X_i}(t) = \alpha/(\alpha - t).$$

Luego si $Z = X_1 + \dots + X_r$, tenemos $M_Z(t) = [\alpha/\alpha - t]^r$, que precisamente es la función generadora de momentos de la distribución Gama con parámetro α y r . (ecuación 10.9.) A menos que $r = 1$, esta no es una fgm de una distribución exponencial, luego esta distribución no posee una propiedad reproductiva. Pero tenemos una característica muy interesante de la distribución Gama, que resumimos en el teorema siguiente.

Teorema 10.9. Sea $Z = X_1 + \dots + X_r$, en donde los X_i son r variables aleatorias independientes e idénticamente distribuidas, donde cada una tiene una distribución exponencial con el mismo parámetro α . Se cumple entonces que Z tiene una distribución Gama con parámetros α y r .

Observaciones: (a) El teorema 10.9 no se cumple si los parámetros de las diversas distribuciones exponenciales son diferentes. Esto llega a ser evidente cuando consideramos la fgm de la suma de las variables aleatorias que resulta.

(b) El *corolario* siguiente del teorema anterior tiene mucha importancia en ciertas aplicaciones estadísticas: la variable aleatoria $W = 2\alpha Z$ tiene distribución χ^2_{2r} . Esta es una consecuencia inmediata del hecho de que $M_W(t) = M_Z(2\alpha t) = [\alpha/(\alpha - 2\alpha t)]^r = (1 - 2t)^{-2r/2}$. Comparando esto con la ecuación (10.10) da el corolario anterior. Luego podemos usar la distribución tabulada de χ cuadrado a fin de calcular ciertas probabilidades asociadas con Z . Por ejemplo, $P(Z \leq 3) = P(2\alpha Z \leq 6\alpha)$. Esta última probabilidad puede obtenerse directamente de las tablas de la distribución de χ cuadrado, si se dan α y r .

10.6 Sucesiones de variables aleatorias

Suponiendo que tenemos una sucesión de variables aleatorias $X_1, X_2, \dots, X_n, \dots$. Cada una de esas variables aleatorias puede describirse mediante F_i , su fda, siendo $F_i(t) = P(X_i \leq t), i = 1, 2, \dots$. Muy a menudo nos interesa lo que sucede a F_i cuando $i \rightarrow \infty$. Es decir, ¿hay alguna función de distribución límite F correspondiente a alguna variable aleatoria X tal que de alguna manera, las variables aleatorias X_i converjan en X ? La respuesta es afirmativa en muchos casos, y hay justamente un procedimiento directo para determinar F .

Tal situación puede aparecer cuando consideramos n observaciones independientes de una variable aleatoria X , llamémoslas X_1, \dots, X_n . Podríamos interesarnos por el promedio aritmético de esas observaciones $\bar{X}_n = (1/n)(X_1 + \dots + X_n)$. De nuevo \bar{X}_n es una variable aleatoria. Sea \bar{F}_n la fda de \bar{X}_n . Podría ser de interés aprender lo que sucede a la distribución de probabilidades de \bar{X}_n cuando n llega a ser grande. Así nuestro problema implica la conducta límite de \bar{F}_n cuando $n \rightarrow \infty$. El teorema siguiente, establecido sin demostración, nos permitirá resolver este y otros problemas semejantes.

Teorema 10.10. Sean X_1, \dots, X_n , una sucesión de variables aleatorias con fda F_1, \dots, F_n, \dots y fgm M_1, \dots, M_n, \dots . Supóngase que $\lim_{n \rightarrow \infty} M_n(t) = M(t)$, en donde $M(0) = 1$. Luego $M(t)$ es la fgm de la variable aleatoria X cuya fda F está dada por $\lim_{n \rightarrow \infty} F_n(t)$.

Observación: el teorema 10.10 dice que para obtener la distribución límite buscada, es suficiente estudiar las funciones generadoras de momentos de las variables aleatorias que se consideran. Obtenemos el valor límite de las sucesiones M_1, \dots, M_n, \dots , llamada $M(t)$. Debido a la propiedad de unicidad de la fgm, existe sólo una distribución de probabilidades que corresponde a la fgm $M(t)$. Podemos aceptar M como la fgm de una distribución conocida (tal como la normal, Poisson, etc.) o bien podemos usar métodos más avanzados para determinar la distribución de probabilidades de M . Tal como podemos obtener la fgm conociendo la fdp, también podemos obtener (en condiciones completamente generales) la fdp conociendo la fgm. Esto implicaría ciertos teoremas de inversión y no continuaremos más con esto.

10.7 Nota final

Hemos visto que la fgm puede ser una herramienta muy poderosa para estudiar diversos aspectos de las distribuciones de probabilidades. En particular, encontramos muy útil el uso de la fgm para estudiar sumas de variables aleatorias independientes e idénticamente distribuidas y obtener diversas leyes reproductivas. Estudiaremos nuevamente las sumas de variables aleatorias independientes en el capítulo 12, sin usar la fgm, pero con métodos semejantes a los que usamos cuando estudiamos el producto y el cociente de variables aleatorias en el capítulo 6.

PROBLEMAS

- 10.1. Suponiendo que X tiene fdp dada por

$$f(x) = 2x, \quad 0 \leq x \leq 1.$$

- (a) Determine la fgm de X .
 (b) Usando la fgm, calcule $E(X)$ y $V(X)$ y verifique su respuesta. (Ver observación, p. 232.)

- 10.2. (a) Encuentre la fgm del voltaje (*incluyendo* el ruido) tal como se expuso en el problema 7.25.

- (b) Usando la fgm, obtenga el valor esperado y la varianza de este voltaje.

- 10.3. Suponiendo que X tenga la fdp siguiente

$$f(x) = \lambda e^{-\lambda(x-a)}, \quad x \geq a.$$

(Esta es conocida como una *distribución exponencial con dos parámetros*.)

- (a) Encontrar la fgm de X .
 (b) Usando la fgm, encontrar $E(X)$ y $V(X)$.

- 10.4. Sea X el resultado cuando se lanza un dado regular.

- (a) Encontrar la fgm de X .
 (b) Usando la fgm, encontrar $E(X)$ y $V(X)$.

10.5. Encontrar la fgm de la variable aleatoria X del problema 6.7. Usando la fgm, encontrar $E(X)$ y $V(X)$.

10.6. Supóngase que la variable aleatoria continua X tenga fdp

$$f(x) = \frac{1}{2}e^{-|x|}, \quad -\infty < x < \infty.$$

(a) Obtener la fgm de X .

(b) Usando la fgm, encontrar $E(X)$ y $V(X)$.

10.7. Usando la fgm, demostrar que si X y Y son variables aleatorias independientes con distribuciones $N(\mu_x, \sigma_x^2)$ y $N(\mu_y, \sigma_y^2)$ respectivamente, entonces $Z = aX + bY$ está nuevamente distribuida normalmente, en donde a y b son constantes.

10.8. Suponiendo que la fgm de una variable aleatoria X es de la forma

$$M_X(t) = (0,4e^t + 0,6)^8.$$

(a) ¿Cuál es la fgm de la variable aleatoria $Y = 3X + 2$?

(b) Calcular $E(X)$.

(c) ¿Puede verificar su respuesta (b) por algún otro método? [Trate de «reconocer» $M_X(t)$].

10.9. Varias resistencias R_i , $i = 1, 2, \dots, n$, se ponen en serie en un circuito. Suponiendo que cada una de las resistencias está distribuida normalmente con $E(R_i) = 10$ ohms y $V(R_i) = 0,16$.

(a) Si $n = 5$, ¿cuál es la probabilidad de que la resistencia del circuito sobrepase los 49 ohms?

(b) ¿Cuál debe ser el valor de n de modo que la probabilidad de que la resistencia total sobrepase los 100 ohms sea aproximadamente 0,05?

10.10. En un circuito se ponen n resistencias en serie. Supóngase que cada una de las resistencias está distribuida uniformemente en $[0,1]$ y supóngase además, que todas las resistencias son independientes. Sea R la resistencia total.

(a) Encontrar la fgm de R .

(b) Usando la fgm, obtener $E(R)$ y $V(R)$. Verifique su respuesta con un cálculo directo.

10.11. Si X tiene una distribución χ_n^2 , utilizando la fgm, demuestre que $E(X) = n$ y $V(X) = 2n$.

10.12. Supóngase que V , la velocidad de un objeto (cm/seg), tiene distribución $N(0,4)$. Si $K = mV^2/2$ ergs es la energía cinética del objeto (donde m = masa), encuentre la fdp de K . Si $m = 10$ gramos, calcule $P(K \leq 3)$.

10.13. Supóngase que la duración de un artículo está distribuida exponencialmente con parámetro 0,5. Supóngase que 10 de tales artículos se instalan sucesivamente, de modo que el i -ésimo artículo se instala «inmediatamente» después de que el $(i-1)$ artículo ha fallado. Sea T_i el tiempo para fallar del i -ésimo artículo, $i = 1, 2, \dots, 10$, medido siempre desde el tiempo de instalación. Luego $S = T_1 + \dots + T_{10}$ representa el tiempo total de funcionamiento de los 10 artículos. Suponiendo que los T_i son independientes, calcule $P(S \geq 15,5)$.

10.14. Supóngase que X_1, \dots, X_{80} son variables aleatorias independientes, donde cada una tiene distribución $N(0,1)$. Calcular $P[X_1^2 + \dots + X_{80}^2 > 77]$. [Indicación: use el teorema 9.2.]

10.15. Demuestre que si X_i , $i = 1, 2, \dots, k$ representa el número de éxitos en n_i repeticiones de un experimento, en donde $P(\text{éxito}) = p$, para todo i , entonces $X_1 + \dots + X_k$ tiene una distribución binomial. (Esto es la distribución binomial tiene la propiedad reproductiva.)

10.16. (*La distribución de Poisson y la multinomial.*) Supóngase que X_i , $i = 1, 2, \dots, n$ son variables aleatorias independientes con una distribución de Poisson con parámetros α_i , $i = 1, \dots, n$. Sea $X = \sum_{i=1}^n X_i$. Entonces la distribución de probabilidades condicional conjunta de X_1, \dots, X_n dado $X = x$ está dada por una distribución multinomial. Esto es, $P(X_1 = x_1, \dots, X_n = x_n | X = x) = x!/(x_1! \dots x_n!) (\alpha_1/\sum_{i=1}^n \alpha_i)^{x_1} \dots (\alpha_n/\sum_{i=1}^n \alpha_i)^{x_n}$.

10.17. Obtener la fgm de una variable aleatoria que tiene una distribución geométrica. ¿Posee esta distribución una propiedad reproductiva bajo la adición?

10.18. Si la variable aleatoria X tiene una fgm dada por $M_X(t) = 3/(3 - t)$, obtener la desviación estándar de X .

10.19. Encontrar la fgm de una variable aleatoria que está distribuida uniformemente en $(-1, 2)$.

10.20. Cierta proceso industrial produce un gran número de cilindros de acero cuyas longitudes están distribuidas normalmente con promedio 3,25 pulgadas y desviación estándar 0,05 pulgada. Si se eligen al azar dos de tales cilindros y se ponen extremo con extremo, ¿cuál es la probabilidad de que la longitud combinada sea menor de 6,60 pulgadas?

Observación: al calcular $M'_X(t)$, en $t = 0$, puede aparecer una forma indeterminada. Es decir, $M'_X(0)$ puede ser de la forma $0/0$. En tal caso debemos tratar de aplicar la regla de L'Hôpital. Por ejemplo, si X está distribuida uniformemente en $[0, 1]$ encontramos fácilmente que $M_X(t) = (e^t - 1)/t$ y $M'_X(t) = (te^t - e^t + 1)/t^2$. Por tanto para $t = 0$, $M'_X(t)$ es indeterminada. Aplicando la regla de L'Hôpital, encontramos que $\lim_{t \rightarrow 0} M'_X(t) = \lim_{t \rightarrow 0} te^t/2t = \frac{1}{2}$. Esto concuerda, puesto que $M'_X(0) = E(X)$, que es igual a $\frac{1}{2}$ para la variable aleatoria descrita aquí.

Aplicaciones a la teoría de la confiabilidad

11.1 Conceptos básicos

En este capítulo investigaremos un área creciente y muy importante en la cual se aplican algunos de los conceptos presentados en los capítulos anteriores.

Supóngase que consideramos un componente (o un conjunto completo de componentes armados en un sistema) que se pone bajo una especie de «tensión». Podría ser una barra de acero bajo una carga, un fusible puesto en un circuito, un ala de aeroplano bajo la influencia de fuerzas, o un instrumento electrónico puesto en servicio. Supóngase que puede definirse un estado que designaremos como «falla» para cualquier componente (o el sistema). Es decir, la barra de acero puede agrietarse o romperse, el fusible puede quemarse, el ala puede doblarse, o el instrumento electrónico puede dejar de funcionar.

Si tal componente se pone bajo condiciones de tensión a un tiempo determinado, $t = 0$, y observamos hasta que falla (es decir, deja de funcionar correctamente bajo la tensión aplicada), el *tiempo para fallar* o la *duración* llamémoslo T , puede considerarse como una variable aleatoria continua con una fdp f . Hay mucha evidencia empírica para indicar que el valor de T no puede ser predicho por un modelo determinístico. Es decir, componentes «idénticos» sometidos a esfuerzos «idénticos» fallarán en tiempos diferentes e impredecibles. Algunos fallarán muy al comienzo de su servicio y otros en etapas posteriores. Naturalmente, que «la manera de fallar» dependerá del tipo de artículo que se considera. Por ejemplo, un fusible fallará de improviso en el sentido de que en un momento dado funciona perfectamente y al momento siguiente no funciona. Por otra parte, una barra de acero bajo una carga pesada se debilitará probablemente en el transcurso de un período largo de tiempo. En cualquier caso, el uso de un modelo probabilístico, considerando T como una variable aleatoria, parece ser el único enfoque realista. Presentamos ahora el siguiente concepto importante.

Definición. La *confiabilidad* de una componente (o sistema) en el tiempo t , llamémosla $R(t)$, está definida como $R(t) = P(T > t)$, en donde T es la duración de la componente. R se llama *función de confiabilidad*.

Observación: aunque el término «confiabilidad» tiene muchos significados técnicos diferentes, la acepción anterior se está aceptando más comúnmente. La definición dada aquí

expresa sencillamente que la confiabilidad de una componente es igual a la probabilidad de que la componente no falle durante el intervalo $[0, t]$ (o, equivalentemente, la confiabilidad es igual a la probabilidad de que la componente esté aún funcionando después de un tiempo t). Por ejemplo, si para un artículo particular, $R(t_1) = 0,90$ esto significa que aproximadamente el 90 por ciento de tales artículos, usados en ciertas condiciones, estarán todavía funcionando después de un tiempo t_1 .

Mediante la fdp de T , llamémosla f , tenemos

$$R(t) = \int_t^\infty f(s)ds.$$

Mediante la fda de T , llamémosla F , tenemos

$$R(t) = 1 - P(T \leq t) = 1 - F(t).$$

Además de la función confiabilidad R , otra función desempeña un papel importante para describir las partes que fallan de un artículo.

Definición. La *tasa de falla* (instantánea) Z (algunas veces llamadas *función de riesgo*) asociada con la variable aleatoria T está dada por

$$Z(t) = \frac{f(t)}{1 - F(t)} = \frac{f(t)}{R(t)}, \quad (11.1)$$

definida para $F(t) < 1$.

Observación: a fin de interpretar $Z(t)$ consideremos la probabilidad condicional

$$P(t \leq T \leq t + \Delta t | T > t),$$

es decir, la probabilidad de que el artículo falle durante las próximas Δt unidades de tiempo, dado que el artículo está funcionando correctamente en el instante t . Aplicando la definición de probabilidad condicional, podemos escribir lo anterior como

$$\begin{aligned} P(t \leq T \leq t + \Delta t | T > t) &= \frac{P(t < T \leq t + \Delta t)}{P(T > t)} \\ &= \int_t^{t+\Delta t} f(x)dx / P(T > t) = \Delta t f(\xi)/R(t), \end{aligned}$$

en donde $t \leq \xi \leq t + \Delta t$.

La última expresión (para un Δt pequeño y suponiendo que f es continua en t^+) es aproximadamente igual a $\Delta t Z(t)$. Así, en un lenguaje informal $\Delta t Z(t)$ representa la proporción de artículos que estará entre t y $t + \Delta t$, entre aquellos artículos que aún funcionan en el instante t .

De lo anterior observamos que f , la fdp de T , determina únicamente la tasa de fallas Z . Indicaremos ahora que lo recíproco también es válido: la tasa de fallas Z determina únicamente la fdp f .

Teorema 11. 1. Si T , el tiempo para fallar, es una variable aleatoria continua con fdp f y si $F(0) = 0$ en donde F es la fda de T , entonces f puede expresarse mediante la tasa de falla Z como sigue

$$f(t) = Z(t) e^{-\int_0^t Z(s)ds} \quad (11.2)$$

Demostración: puesto que $R(t) = 1 - F(t)$, tenemos $R'(t) = -F'(t) = -f(t)$. Luego

$$Z(t) = \frac{f(t)}{R(t)} = \frac{-R'(t)}{R(t)}.$$

Integrando ambos miembros de 0 a t :

$$\begin{aligned} \int_0^t Z(s) ds &= - \int_0^t \frac{R'(s)}{R(s)} ds = -\ln R(s)|_0^t \\ &= -\ln R(t) + \ln R(0) = -\ln R(t), \end{aligned}$$

dado que hay $\ln R(0) = 0$ lo que es válido si y sólo si $R(0) = 1$. [Esta última condición se satisface si $F(0) = 0$. Simplemente expresa que la probabilidad de una falla *inicial* es igual a cero; haremos esta suposición durante el resto de la exposición.] Por tanto

$$R(t) = e^{-\int_0^t Z(s) ds}$$

Luego

$$f(t) = F'(t) = \frac{d}{dt} [1 - R(t)] = Z(t) e^{-\int_0^t Z(s) ds}$$

Así hemos demostrado que la tasa de fallas Z determina unívocamente la fdp f .

Existe una relación interesante entre la función de confiabilidad R y el tiempo promedio de falla, $E(T)$.

Teorema 11.2. Si $E(T)$ es finito, entonces

$$E(T) = \int_0^\infty R(t) dt. \quad (11.3)$$

Demostración: considerando

$$\int_0^\infty R(t) dt = \int_0^\infty [\int_t^\infty f(s) ds] dt.$$

Integrando por partes, hacemos $\int_t^\infty f(s) ds = u$ y $dt = dv$. Luego $v = t$ y $du = -f(t) dt$. Así

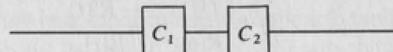
$$\int_0^\infty R(t) dt = t \int_t^\infty f(s) ds |_0^\infty + \int_0^\infty t f(t) dt.$$

La segunda integral del segundo miembro representa $E(T)$. Por tanto la demostración está completa si podemos demostrar que $t \int_t^\infty f(s) ds$ se anula en $t = 0$ y cuando $t \rightarrow \infty$. La anulación en $t = 0$ es inmediata. Sabiendo que $E(T)$ es finita, el lector puede completar la demostración.

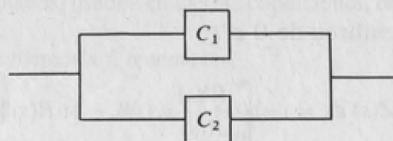
Los conceptos de confiabilidad y tasa de fallas están entre las herramientas necesarias más importantes para un estudio de los «modelos de falla». Brevemente vamos a ocuparnos de las siguientes preguntas:

(a) ¿Cuáles son las «leyes de falla» fundamentales que razonablemente se pueden suponer? (Es decir, ¿qué forma tendría la fdp de T ?)

(b) Supóngase que tenemos dos componentes, C_1 y C_2 , con leyes de falla conocidas. Suponiendo que dichas componentes se combinan en serie



o en paralelo



para formar un sistema. ¿cuál es la ley de fallas (o confiabilidad), del sistema?

La pregunta cuál es una ley de fallas «razonable» nos devuelve a un problema que hemos presentado antes: ¿cuál es un modelo matemático razonable para la descripción de algunos fenómenos observables? Desde un punto de vista estrictamente matemático, prácticamente podemos suponer cualquier fdp para T y luego estudiar sencillamente las consecuencias de esta suposición. Sin embargo, si estamos interesados en tener un modelo que represente (tan exactamente como sea posible) los datos de fallas realmente disponibles, nuestra elección del modelo debe tener esto en cuenta.

11.2 La ley normal de fallas

Hay muchos tipos de componentes cuya conducta de falla puede representarse por la distribución normal. Es decir, T es la duración de un artículo, su fdp está dada por

$$f(t) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\left[\frac{t-\mu}{\sigma}\right]^2\right).$$

[Nuevamente observamos que el tiempo para fallar T , debe ser mayor que (o igual) a cero. Por tanto, a fin de que el modelo anterior sea aplicable debemos insistir en que $P(T < 0)$ sea forzosamente cero.] Como la forma de la fdp normal lo indica, una ley normal de fallas implica que la mayor parte de los artículos fallan alrededor del tiempo promedio de falla, $E(T) = \mu$ y el número de fallas disminuye (simétricamente) cuando $|T - \mu|$ aumenta. Una ley normal de fallas significa que alrededor del 95,72 por ciento de las fallas tiene lugar para los valores de t que satisfacen $\{t \mid |t - \mu| < 2\sigma\}$. (Ver figura 11.1.)

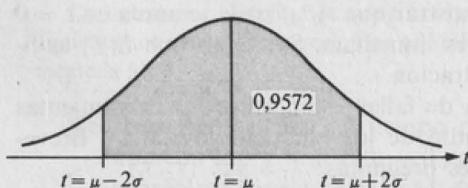


FIGURA 11.1

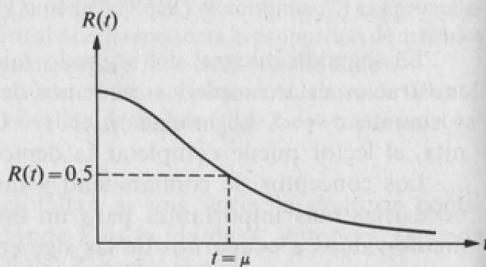


FIGURA 11.2

La función de confiabilidad de la ley normal de fallas puede expresarse mediante la función de distribución normal acumulativa tabulada Φ , como sigue:

$$\begin{aligned} R(t) &= P(T > t) = 1 - P(T \leq t) \\ &= 1 - \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^t \exp\left(-\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2\right) dx \\ &= 1 - \Phi\left(\frac{t-\mu}{\sigma}\right). \end{aligned}$$

La figura 11.2 muestra una curva general de confiabilidad para una ley normal de fallas. Nótese que a fin de obtener una confiabilidad alta (0,90, o mayor) el tiempo de operación debe ser considerablemente menor que μ , la duración esperada.

EJEMPLO 11.1. Suponiendo que la duración de una componente está distribuida normalmente con una desviación estándar igual a 10 (horas). Si la componente tiene una confiabilidad de 0,99 para un período de operación de 100 horas, ¿cuál debería ser su duración esperada?

La ecuación anterior se transforma en

$$0,99 = 1 - \Phi\left(\frac{100 - \mu}{10}\right).$$

De las tablas de la distribución normal $(100 - \mu)/10 = -2,33$. Luego $\mu = 123,3$ horas.

La ley normal de fallas representa un modelo apropiado para los componentes en los cuales la falla se debe a algunos efectos de «uso». Sin embargo, no está entre las más importantes leyes de fallas que existen.

11.3 La ley exponencial de fallas

Una de las leyes de fallas más importantes es aquella cuyo tiempo para fallar se describe mediante una distribución exponencial. Podemos describirla de varias maneras, pero probablemente la manera más sencilla es suponer que la tasa de fallas es constante. Es decir, $Z(t) = \alpha$. Una consecuencia inmediata de esta suposición es, según la ecuación (11.2), que la fdp asociada con el tiempo para fallar T , está dada por

$$f(t) = \alpha e^{-\alpha t}, \quad t > 0.$$

El recíproco es también inmediato: si f tiene la forma anterior, $R(t) = 1 - F(t) = e^{-\alpha t}$ y por tanto $Z(t) = f(t)/R(t) = \alpha$. Luego tenemos el siguiente resultado importante.

Teorema 11.3. Sea T , el tiempo para fallar, una variable aleatoria continua que toma todos los valores no negativos. Luego T tiene una distribución exponencial si y sólo si tiene una tasa constante de fallas.

Observación: la suposición de que una tasa constante de fallas puede ser interpretada como una indicación de que después de que el artículo ha sido usado, su probabilidad de fallar no ha cambiado. Dicho de una manera más informal, no hay efecto de «uso» cuando se estipula el modelo exponencial. Existe otra manera de expresar esto que hace esta parte aún más evidente.

Considérese para $\Delta t > 0$, $P(t \leq T \leq t + \Delta t | T > t)$. Esta representa la probabilidad de que el artículo falle durante las próximas Δt unidades, dado que no ha fallado en el instante t . Aplicando la definición de probabilidad condicional, encontramos que

$$P(t \leq T \leq t + \Delta t | T > t) = \frac{e^{-\alpha t} - e^{-\alpha(t + \Delta t)}}{e^{-\alpha t}} = 1 - e^{-\alpha \Delta t}.$$

Por tanto esta probabilidad condicional es independiente de t y sólo depende de Δt . Es en este sentido que podemos decir que una ley exponencial de fallas implica que la probabilidad de fallar es independiente del pasado. Es decir, mientras el artículo funcione es «tan bueno como nuevo».

Si desarrollamos el segundo miembro de la expresión anterior en una serie de Maclaurin obtenemos

$$\begin{aligned} P(t \leq T \leq t + \Delta t | T > t) &= 1 - \left[1 - \alpha \Delta t + \frac{(\alpha \Delta t)^2}{2!} - \frac{(\alpha \Delta t)^3}{3!} + \dots \right] \\ &= \alpha \Delta t + h(\Delta t). \end{aligned}$$

en donde $h(\Delta t)$ llega a ser despreciable para Δt pequeño. Luego para Δt suficientemente pequeño la probabilidad anterior es directamente proporcional a Δt .

Para muchos tipos de componentes la hipótesis que conduce a la ley exponencial de fallas no es sólo intuitivamente atractiva sino que en realidad está sostenida por evidencia empírica. Por ejemplo, es muy razonable suponer que un fusible o un cojinete de rubíes es «tan bueno como nuevo» mientras esté funcionando. Es decir, si el fusible no se ha fundido, prácticamente está como nuevo. Tampoco el cojinete cambiará mucho debido al uso. En estos casos, la ley exponencial de fallas representa un modelo apropiado para estudiar las características que fallan en un artículo.

Sin embargo, debemos dar aquí una palabra de precaución. Hay muchas situaciones que implican estudios de fallas para los cuales las hipótesis básicas que conducen a una ley exponencial no serán satisfechas. Por ejemplo, si una

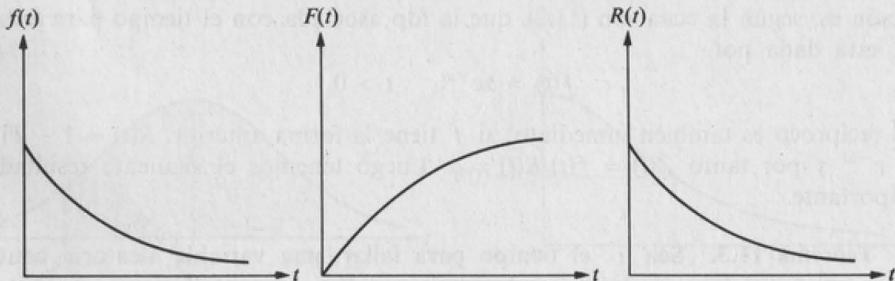


FIGURA 11.3

pieza de acero está expuesta a una tensión continua, evidentemente sufrirá un deterioro, y por tanto se debe considerar un modelo distinto al exponencial.

Aunque previamente discutimos las distintas propiedades de la distribución exponencial, resumámosla nuevamente a fin de tenerlas disponibles para el objetivo siguiente. (Ver la figura 11.3.) Si T , el tiempo para fallar, está distribuido exponencialmente (con parámetro α), tenemos

$$\begin{aligned} E(T) &= 1/\alpha; & V(T) &= 1/\alpha^2; \\ F(t) = P(T \leq t) &= 1 - e^{-\alpha t}; & R(t) &= e^{-\alpha t}. \end{aligned}$$

EJEMPLO 11.2. Si se da el parámetro α y se especifica $R(t)$, podemos encontrar t , el número de horas, de operación. Así, si $\alpha = 0,01$ y $R(t)$ es igual a 0,90, tenemos

$$0,90 = e^{-0,01t}$$

Luego $t = -100 \ln(0,90) = 10,54$ horas. Por tanto, si las 100 componentes funcionan durante 10,54 horas, aproximadamente 90 no fallarán durante ese período.

Observaciones: (a) Es muy importante darse cuenta de que en el caso excepcional podemos identificar el *tiempo* de operación (de algún valor inicial arbitrario fijo) con la *edad* para funcionar. Porque en el caso exponencial un artículo que no ha fallado es tan bueno como nuevo y, por tanto, su conducta durante cualquier periodo de servicio depende sólo de la longitud de ese periodo y no de su historia anterior. Sin embargo, cuando se supone una ley de falla no exponencial (tal como la ley normal o una de las distribuciones que consideraremos en breve), la historia pasada no tiene efecto sobre el comportamiento del artículo. Por tanto, mientras podamos definir T como el tiempo en servicio (hasta la falla) para el caso exponencial, debemos definir T como la duración *total* hasta la falla para los casos no exponenciales.

(b) La distribución exponencial que hemos presentado en relación con la duración de los componentes, tiene muchas otras aplicaciones importantes. En realidad, cada vez que una variable aleatoria continua T que toma valores no negativos satisface la hipótesis $P(T > s + t | T > s) = P(T > t)$ para todo s y t , entonces T tendrá una distribución exponencial. Así, si T representa el tiempo que demora un átomo radiactivo en desintegrarse, podemos suponer que T está distribuida exponencialmente, puesto que la suposición anterior parece satisfacerse.

EJEMPLO 11.3. No es irrazonable suponer que cuesta más producir un artículo con una gran duración esperada que uno con una esperanza pequeña de duración. Específicamente suponemos que el costo C para producir un artículo es la siguiente función de μ , el tiempo promedio para fallar,

$$C = 3\mu^2.$$

Supóngase que se obtiene una utilidad de D pesos por cada hora que el artículo funcione. Luego la utilidad por artículo está dada por

$$P = DT - 3\mu^2,$$

en donde T es el número de horas que el artículo funciona correctamente. Luego la utilidad esperada está dada por

$$E(P) = D\mu - 3\mu^2.$$

Para encontrar para qué valor de μ esta cantidad es máxima, sencillamente hacemos $dE(P)/d\mu$ igual a cero y resolvemos para μ . El resultado es $\mu = D/6$ y por tanto la utilidad esperada máxima por artículo es igual $E(P)_{\max} = D^2/12$.

EJEMPLO 11.4. Reconsideremos el ejemplo 11.3, haciendo las suposiciones adicionales siguientes. Supóngase que T , el tiempo para fallar, está distribuido exponencialmente con parámetro α . Luego μ , el tiempo esperado para fallar, está dado por $\mu = 1/\alpha$. Supóngase, además, que si el artículo no funciona correctamente al menos un número específico de horas, digamos t_0 , se fija un castigo de $K(t_0 - T)$ pesos, en donde $T(T < t_0)$ es el tiempo en el cual tiene lugar la falla. Por tanto, la utilidad por artículo está dada por

$$\begin{aligned} P &= DT - 3\mu^2 \quad \text{si } T > t_0, \\ &= DT - 3\mu^2 - K(t_0 - T) \quad \text{si } T < t_0. \end{aligned}$$

Luego la utilidad esperada (por artículo) puede expresarse como

$$\begin{aligned} E(P) &= D \int_{t_0}^{\infty} t \alpha e^{-\alpha t} dt - 3\mu^2 e^{-\alpha t_0} \\ &\quad + (D + K) \int_0^{t_0} t \alpha e^{-\alpha t} dt - (3\mu^2 + Kt_0)(1 - e^{-\alpha t_0}). \end{aligned}$$

Después de unas integraciones inmediatas lo anterior puede escribirse como

$$E(P) = D\mu - 3\mu^2 + K[\mu - \mu e^{-t_0/\mu} - t_0].$$

Nótese que si $K = 0$, esto se reduce al resultado obtenido en el ejemplo 11.3. Podríamos formularnos una pregunta análoga a la que apareció en el ejemplo previo: ¿para qué valor de μ toma $E(P)$ su valor máximo? No proseguiremos con los detalles de este problema puesto que implica la solución de una ecuación trascendental que debe resolverse numéricamente.

11.4 La ley exponencial de fallas y la distribución de Poisson

Hay una conexión muy próxima entre la ley exponencial de fallas descrita en la sección previa y un proceso de Poisson. Supóngase que la falla ocurre debido a la aparición de ciertos accidentes «aleatorios». Estos pueden deberse a fuerzas externas tales como una repentina ráfaga de viento o un aumento de voltaje o por causas internas tales como una desintegración química o un mal funcionamiento mecánico. Sea X , el número de accidentes que ocurren en un intervalo de tiempo de longitud t y supongamos que $X_t, t \geq 0$, determina un *proceso de Poisson*. Es decir, para cualquier t fijo, la variable aleatoria X_t tiene una distribución de Poisson

con parámetro αt . Supóngase que la falla en $[0, t]$ se produce si y sólo si al menos uno de tales accidentes ocurre. Sea T el tiempo para fallar, que supondremos que es una variable aleatoria continua. Luego,

$$F(t) = P(T \leq t) = 1 - P(T > t).$$

Ahora $T > t$ si y sólo si *ningún* accidente ocurre en $[0, t]$. Esto acontece si y sólo si $X_t = 0$. Luego

$$F(t) = 1 - P(X_t = 0) = 1 - e^{-\alpha t}.$$

Esto representa la fda de una ley exponencial de fallas. Encontramos así que la «causa» anterior de las fallas implica una ley exponencial de fallas.

Las ideas anteriores se pueden *generalizar* de dos maneras.

(a) Nuevamente suponemos que los accidentes aparecen de acuerdo con un proceso de Poisson. Suponemos, además, que cada vez que aparece tal accidente hay una probabilidad constante p de que no producirá fallas. Por tanto, si T es el tiempo para fallar, tenemos, como antes,

$$F(t) = P(T \leq t) = 1 - P(T > t).$$

Esta vez, $T > t$ si y sólo si (durante $[0, t]$) no ocurre ningón accidente, u ocurre un accidente y ninguna falla aparece, o dos accidentes ocurren y no aparece ninguna falla, o... Por tanto,

$$\begin{aligned} F(t) &= 1 - \left[e^{-\alpha t} + (\alpha t)e^{-\alpha t}p + (\alpha t)^2 \frac{e^{-\alpha t}}{2!} p^2 + \dots \right] \\ &= 1 - e^{-\alpha t} \sum_{k=0}^{\infty} \frac{(\alpha t p)^k}{k!} = 1 - e^{-\alpha t} e^{\alpha t p} = 1 - e^{-\alpha(1-p)t}. \end{aligned}$$

Luego T tiene una ley exponencial de fallas con parámetro $\alpha(1-p)$. (Nótese que si $p = 0$, tenemos el caso expuesto previamente.)

(b) Supóngase nuevamente que los accidentes aparecen de acuerdo con un proceso de Poisson. Esta vez supondremos que las fallas ocurren cada vez que r o más accidentes ($r \geq 1$) ocurren durante un intervalo de longitud t . Por tanto, si T es el tiempo para fallar, tenemos, como antes,

$$F(t) = 1 - P(T > t).$$

En este caso, $T > t$ si y sólo si ($r - 1$) o menos accidentes ocurren. Por tanto

$$F(t) = 1 - \sum_{k=0}^{r-1} \frac{(\alpha t)^k e^{-\alpha t}}{k!}$$

De acuerdo con la ecuación (9.17) lo anterior es igual a $\int_0^t [\alpha/(r-1)!](\alpha s)^{r-1} e^{-\alpha s} ds$ y por tanto representa la fda de una distribución Gama. Así, encontramos que la «causa» anterior de fallas sigue una *ley Gama de fallas*. (Si $r = 1$, naturalmente, esta se transforma en una distribución exponencial.)

11.5 La ley de fallas de Weisbull

Modifiquemos la noción de tasa constante de fallas que condujo a la ley exponencial de fallas. Supóngase que la tasa de fallas Z , asociada con T , la duración de un artículo, tiene la forma siguiente:

$$Z(t) = (\alpha\beta)t^{\beta-1}, \quad (11.4)$$

en donde α y β son constantes positivas. De la ecuación (11.2) obtenemos la expresión siguiente para la fdp de T :

$$f(t) = (\alpha\beta)t^{\beta-1}e^{-\alpha t^\beta}, \quad t > 0, \alpha, \beta > 0 \quad (11.5)$$

Se dice que la variable aleatoria que tiene la fdp dada por la ecuación (11.5) tiene una *distribución de Weisbull*. La figura 11.4 muestra la fdp para $\alpha = 1$ y $\beta = 1, 2, 3$. La función de confiabilidad R está dada por $R(t) = e^{-\alpha t^\beta}$ que es una función decreciente de t .

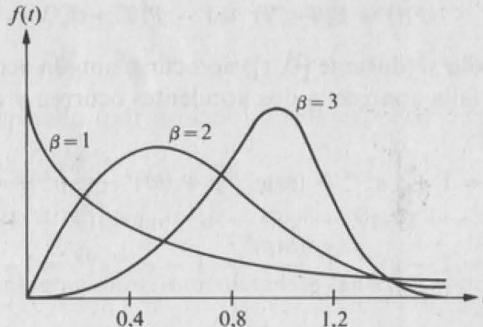


FIGURA 11.4

Observación: la distribución exponencial es un caso especial de distribución de Weisbull puesto que obtenemos la distribución exponencial si hacemos $\beta = 1$ en la ecuación (11.4). La suposición (ecuación 11.4) establece que $Z(t)$ no es una constante, sino que es proporcional a las potencias de t . Por ejemplo, si $\beta = 2$, Z es una función lineal de t ; si $\beta = 3$, Z es una función cuadrática de t ; etc. Luego Z es una función creciente, decreciente o constante de t , según el valor de β , como se indicó en la figura 11.5.

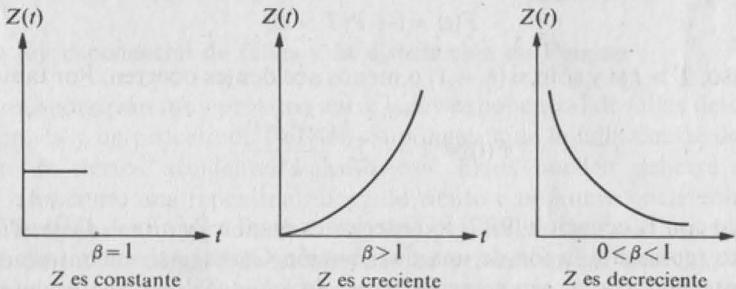


FIGURA 11.5

Teorema 11.4. Si la variable aleatoria T tiene una distribución de Weisbull con fdp dada por la ecuación (11.5), tenemos

$$E(T) = \alpha^{-1/\beta} \Gamma\left(\frac{1}{\beta} + 1\right), \quad (11.6)$$

$$V(T) = \alpha^{-2/\beta} \left\{ \Gamma\left(\frac{2}{\beta} + 1\right) - \left[\Gamma\left(\frac{1}{\beta} + 1\right) \right]^2 \right\}. \quad (11.7)$$

Demostración: ver el problema 11.8.

Observación: la distribución de Weisbull representa un modelo apropiado para una ley de falla siempre que el sistema esté compuesto por cierto número de componentes y la falla se debe principalmente al defecto «más grave» en un gran número de defectos del sistema. También utilizando la distribución de Weisbull, podemos obtener una tasa de fallas creciente y decreciente al hacer sencillamente una elección apropiada del parámetro β .

En ningún caso hemos agotado el número de leyes de falla razonables. Sin embargo, las que hemos mencionado son por cierto extremadamente importantes para representar modelos significativos para el estudio de las características que fallan en componentes o en sistemas de componentes.

11.6 Confiabilidad de los sistemas

Ahora que hemos considerado un número importante de distribuciones de fallas podemos volver a la segunda pregunta propuesta en la sección 11.1: ¿cómo podemos evaluar la confiabilidad de un sistema si conocemos la confiabilidad de sus componentes? Este puede ser un problema muy difícil y sólo discutiremos un caso más sencillo (pero relativamente importante).

Supóngase que las dos componentes estén acopladas en serie.



Esto significa que, para que el sistema anterior funcione, *ambas* componentes deben funcionar. Si, además, suponemos que las componentes funcionen *independientemente*, debemos conocer la confiabilidad del sistema, llamémoslo $R(t)$, mediante la confiabilidad de las componentes, llamémoslo $R_1(t)$ y $R_2(t)$, como sigue:

$$\begin{aligned}
 R(t) &= P(T > t) \quad (\text{en donde } T \text{ es el tiempo para fallar del sistema}) \\
 &= P(T_1 > t \text{ y } T_2 > t) \quad (\text{en donde } T_1 \text{ y } T_2 \text{ son los tiempos para fallar de los componentes } C_1 \text{ y } C_2, \text{ respectivamente}) \\
 &= P(T_1 > t)P(T_2 > t) = R_1(t)R_2(t).
 \end{aligned}$$

Encontramos así que $R(t) \leq \min[R_1(t), R_2(t)]$. Es decir, para un sistema formado por dos componentes independientes en serie, la confiabilidad del sistema es menor que la confiabilidad de cualquiera de sus partes.

La presentación anterior puede generalizarse naturalmente a n componentes y obtenemos el teorema siguiente.

Teorema 11.5. Si n componentes, que funcionan independientemente, están conectadas en serie, y si la i -ésima componente tiene confiabilidad $R_i(t)$, la confiabilidad del sistema completo, $R(t)$, está dada por

$$R(t) = R_1(t) \cdot R_2(t) \cdots R_n(t). \quad (11.8)$$

En particular, si T_1 y T_2 tienen leyes de falla exponenciales con parámetros α_1 y α_2 , la ecuación (11.8) se transforma en

$$R(t) = e^{-\alpha_1 t} e^{-\alpha_2 t} = e^{-(\alpha_1 + \alpha_2)t}.$$

Luego la fdp del tiempo para fallas del sistema, llamémoslo T , es dado por

$$f(t) = -R'(t) = (\alpha_1 + \alpha_2)e^{-(\alpha_1 + \alpha_2)t}.$$

Así hemos establecido el resultado siguiente.

Teorema 11.6. Si dos componentes que funcionan independientemente y tienen leyes de falla exponencial con parámetros α_1 y α_2 están conectadas en serie, la ley de falla del sistema resultante es nuevamente exponencial con parámetro igual a $\alpha_1 + \alpha_2$.

(Este teorema evidentemente puede generalizarse a n componentes en serie.)

EJEMPLO 11.5. (Tomado de I. Bazovsky, *Reliability Theory and Practice*, Prentice-Hall, Inc., Englewood Cliffs, N. J., 1961.) Se considera un circuito electrónico que consta de 4 transistores de silicón, 10 diodos de silicón, 20 resistencias compuestas, y 10 condensadores de cerámica en una operación continua en serie. Suponiendo que bajo ciertas condiciones de esfuerzo (es decir, voltaje prefijado, corriente y temperatura), cada uno de esos artículos tiene la siguiente tasa constante de fallas.

diodos de silicón:	0,000002
transistores de silicón:	0,00001
resistencias compuestas:	0,000001
condensadores de cerámica:	0,000002

Debido a la tasa constante de fallas supuesta, la distribución exponencial representa la ley de falla para cada una de las componentes anteriores. Debido a la conexión en serie, el tiempo para fallar del circuito completo está de nuevo distribuido exponencialmente con parámetro (tasa de falla) igual a:

$$10(0,000002) + 4(0,00001) + 20(0,000001) + 10(0,000002) = 0,0001.$$

Por tanto la confiabilidad del circuito está dada por $R(t) = e^{-0,00001t}$. Luego, para un período de 10 horas del funcionamiento, la probabilidad de que el circuito no falle está dada por $e^{-0,00001(10)} = 0,999$. El tiempo esperado para que el circuito falle es igual a $1/0,00001 = 10.000$ horas.

Otro sistema importante es un sistema en *paralelo* en el cual las componentes están conectadas de tal manera que el sistema deja de funcionar sólo si todas las componentes dejan de funcionar. Si sólo dos componentes están implicadas, el sistema puede dibujarse como en la figura 11.6. Nuevamente suponiendo que las componentes funcionan *independientemente* unas de otras, la confiabilidad del sistema, llamémosla $R(t)$, puede expresarse mediante la confiabilidad de las componentes, $R_1(t)$ y $R_2(t)$, como sigue:

$$\begin{aligned}
 R(t) &= P(T > t) = 1 - P(T \leq t) \\
 &= 1 - P[T_1 \leq t \text{ y } T_2 \leq t] \\
 &= 1 - P(T_1 \leq t)P(T_2 \leq t) \\
 &= 1 - \{[1 - P(T_1 > t)][1 - P(T_2 > t)]\} \\
 &= 1 - [1 - R_1(t)][1 - R_2(t)] \\
 &= R_1(t) + R_2(t) - R_1(t)R_2(t).
 \end{aligned}$$

La última fórmula indica que $R(t) \geq \max[R_1(t), R_2(t)]$. Es decir, un sistema compuesto por dos componentes que funcionan independientemente en paralelo será más confiable que cualquiera de las componentes.

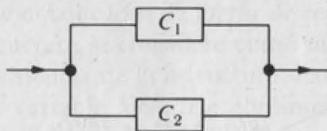


FIGURA 11.6

Todas las ideas presentadas anteriormente para dos componentes pueden generalizarse en el teorema siguiente.

Teorema 11.7. Si n componentes que funcionan independientemente actúan en paralelo, y si la i -ésima componente tiene confiabilidad $R_i(t)$, entonces la confiabilidad del sistema, llamada $R(t)$, está dada por

$$R(t) = 1 - [1 - R_1(t)][1 - R_2(t)] \cdots [1 - R_n(t)]. \quad (11.9)$$

A menudo sucede que todas las componentes tienen *igual* confiabilidad, digamos que $R_i(t) = r(t)$ para todo i . En este caso la expresión anterior se transforma en

$$R(t) = 1 - [1 - r(t)]^n. \quad (11.10)$$

Consideremos, en particular, dos componentes en paralelo, cada uno de ellos con tiempo de falla distribuido exponencialmente. Luego,

$$R(t) = R_1(t) + R_2(t) - R_1(t)R_2(t) = e^{-x_1 t} + e^{-x_2 t} - e^{-(x_1 + x_2)t}.$$

Así la fdp del tiempo de falla del sistema en paralelo, llamémoslo T , está dada por

$$f(t) = -R'(t) = \alpha_1 e^{-\alpha_1 t} + \alpha_2 e^{-\alpha_2 t} - (\alpha_1 + \alpha_2)e^{-(\alpha_1 + \alpha_2)t}.$$

Por tanto T no está distribuido exponencialmente. El valor esperado de T es igual a:

$$E(T) = \frac{1}{\alpha_1} + \frac{1}{\alpha_2} - \frac{1}{\alpha_1 + \alpha_2}.$$

Por cuanto a menudo una serie de operaciones es obligatoria (es decir, un número de componentes *debe* funcionar para que el sistema funcione), frecuentemente usamos una operación en paralelo a fin de aumentar la confiabilidad del sistema. El ejemplo siguiente ilustra esta parte.

EJEMPLO 11.6. Supongamos que tres unidades están trabajando en paralelo. Supóngase que cada una tiene la misma tasa constante de fallas $\alpha = 0,01$. (Es decir, el tiempo para fallar de cada una de las unidades está distribuido exponencialmente con parámetro $\alpha = 0,01$.) Por tanto, la confiabilidad para cada una de las unidades es $R(t) = e^{-0,01t}$, y así la confiabilidad en un período de operación de 10 horas es igual a $e^{-0,1} = 0,905$ o alrededor del 90 por ciento. ¿Cuánta ventaja puede obtenerse (mediante el aumento de la confiabilidad) al funcionar tres de tales unidades en paralelo?

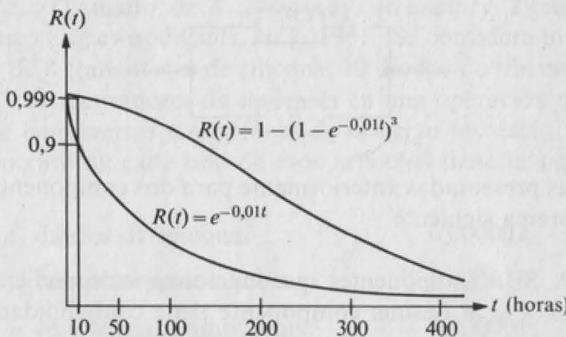


FIGURA 11.7

La confiabilidad de tres unidades que funcionan en paralelo durante 10 horas sería

$$\begin{aligned} R(10) &= 1 - [1 - 0,905]^3 = 1 - 0,00086 \\ &= 0,99914, \quad \text{o alrededor del 99,9 por ciento.} \end{aligned}$$

En la figura 11.7 vemos las curvas de confiabilidad de la unidad única versus las tres unidades en paralelo. Para la unidad sola, $R(t) = e^{-\alpha t}$, mientras que para las tres unidades en paralelo, $R(t) = 1 - (1 - e^{-\alpha t})^3$, con $\alpha = 0,01$.

Hemos considerado, hasta ahora, sólo las maneras más sencillas para combinar unidades individuales en un sistema, llamadas operaciones en serie y en paralelo de las componentes. Hay muchas otras maneras de combinar componentes, y

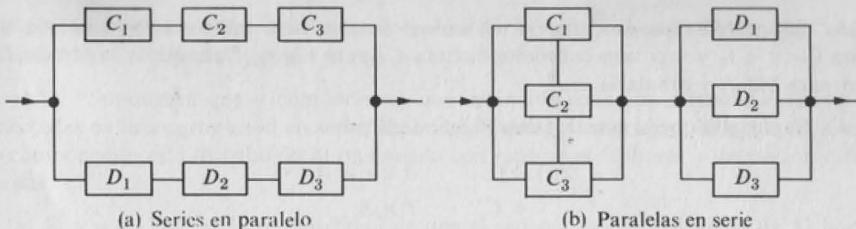


FIGURA 11.8

sencillamente nombraremos solo algunas. (Ver la figura 11.8.) Algunas de las preguntas que aparecen en relación con esas combinaciones se considerarán en los problemas al final del capítulo.

- (a) Series en paralelo. (Consideramos aquí grupos de componentes en paralelo como en un circuito que tiene, por ejemplo, m componentes en serie cada uno.)
 - (b) Paralelas en serie.
 - (c) Sistema sostenido. Consideramos aquí dos componentes de las cuales la segunda componente se «mantiene» y funciona si y sólo si la primera componente falla. En este caso, la segunda componente es requerida (instantáneamente) y funciona en el lugar de la primera componente.

Expongamos brevemente el concepto de *factor de seguridad*. Supóngase que la fuerza S aplicada a una estructura se considere como una variable aleatoria (continua). Análogamente, la resistencia de la estructura, llamémosla R , se puede considerar también como una variable aleatoria continua. Definamos el factor de seguridad de la estructura como la razón de R a S .

$$T = R/S.$$

Si R y S son variables aleatorias independientes con fdp g y h respectivamente, entonces la fdp de T está dada por

$$f(t) = \int_0^\infty g(ts)h(s)s\ ds.$$

(Ver el teorema 6.5.) La estructura fallará si $S > R$, es decir, si $T < 1$. Por tanto la probabilidad de fallar $P_F = \int_0^1 f(t)dt$.

PROBLEMAS

- 11.1. Supóngase que T , el tiempo para fallar, de un artículo está distribuido normalmente con $E(T) = 90$ horas y desviación estándar 5 horas. A fin de obtener una confiabilidad de 0,90, 0,95, 0,99, ¿cuántas horas de operación deben considerarse?

11.2. Suponiendo que la duración de un instrumento electrónico es distribuida exponencialmente. Se sabe que la confiabilidad del instrumento (para un período de operación de 100 horas) es 0,90. ¿Cuántas horas de operación deben considerarse para obtener una confiabilidad de 0,95?

11.3. Suponiendo que la duración de un instrumento tiene una tasa constante de falla C_0 para $0 < t < t_0$ y una tasa constante distinta C_1 para $t \geq t_0$. Determinar la fdp de T , el tiempo para fallar, y dibujarla.

11.4. Supóngase que la tasa de fallas Z está dada por

$$\begin{aligned} Z(t) &= 0, & 0 < t < A, \\ &= C, & t \geq A. \end{aligned}$$

(Esto implica que ninguna falla ocurre antes de $T = A$.)

- (a) Encontrar la fdp asociada con T , el tiempo para fallar.
- (b) Calcular $E(T)$.

11.5. Suponiendo que la ley de fallas de una componente tenga la siguiente fdp:

$$f(t) = (r + 1)A^{r+1}/(A + t)^{r+2}, \quad t > 0.$$

- (a) ¿Para qué valores de A y r es la anterior una fdp?
- (b) Obtener una expresión para la función de confiabilidad y la función de riesgo.
- (c) Demostrar que la función de riesgo es decreciente en t .

11.6. Supóngase que la ley de fallas de una componente es una combinación lineal de k leyes exponenciales de fallas. Es decir, la fdp del tiempo para fallar está dada por

$$f(t) = \sum_{j=1}^k c_j \beta_j e^{-\beta_j t}, \quad t > 0, \quad \beta_j > 0.$$

- (a) ¿Para qué valores de c_j es la anterior una fdp?
- (b) Obtener una expresión para la función de confiabilidad y la función de riesgo.
- (c) Obtener una expresión del promedio del tiempo para fallar.
- (d) Responder (b) y (c) si $\beta_j = \beta$ para todo j .

11.7. Cada uno de los seis tubos de un equipo de radio tiene una duración (en años) que puede considerarse una variable aleatoria. Supóngase que esos tubos funcionan independientemente uno de otro. ¿Cuál es la probabilidad de que ningún tubo tenga que ser reemplazado durante los primeros dos meses de servicio si:

- (a) La fdp del tiempo para fallar es $f(t) = 50te^{-25t^2}$, $t > 0$?
- (b) La fdp del tiempo para fallar es $f(t) = 25te^{-25t}$, $t > 0$?

11.8. Demostrar el teorema 11.4.

11.9. La duración de un satélite es una variable aleatoria distribuida exponencialmente con un tiempo de duración esperado de 1.5 años. Si tres satélites se lanzan simultáneamente, ¿cuál es la probabilidad que a lo menos dos estarán aún en órbita después de 2 años?

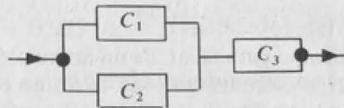


FIGURA 11.9

11.10. Tres componentes que funcionan independientemente están conectadas en un sistema aislado como se indica en la figura 11.9. Suponiendo que la confiabilidad de cada componente para un período de t horas de operación está dada por

$$R(t) = e^{-0.03t}.$$

Si T es el tiempo para fallar del sistema completo (en horas), ¿cuál es la fdp de T ? ¿Cuál es la confiabilidad del sistema?; ¿cómo se compara con $e^{-0.03t}$?

11.11. Supóngase que n componentes que están funcionando independientemente son conectadas en una agrupación en serie. Suponiendo que el tiempo para fallar de cada una de las componentes está distribuido normalmente con esperanza 50 horas y desviación estándar 5 horas.

(a) Si $n = 4$, ¿cuál es la probabilidad de que el sistema funcione después de 52 horas de operación?

(b) Si n componentes se conectan en paralelo, ¿cuál debería ser el valor de n a fin que la probabilidad de fallar durante las primeras 55 horas sea aproximadamente igual a 0.01?

11.12. (Sacado de Derman and Klein, *Probability and Statistical Inference*. Oxford University Press, New York, 1959.) La duración (L) en meses de cierto tubo al vacío usado en un equipo de radar está distribuida exponencialmente con parámetro $\beta = 2$. Al establecer su programa preventivo de mantención, una compañía desea decidir cuántos meses (m) después de la instalación debería remplazarse el tubo para minimizar el coste esperado por tubo. El costo por tubo en dólares está denotado por C . El tiempo útil más corto empleado entre la instalación y la sustitución es 0,01 mes. Obedeciendo a esta restricción, ¿qué valor de m minimizó $E(C)$, el coste esperado en cada una de las situaciones siguientes, en donde el costo C es la función dada de L y m ?

$$(a) C(L, m) = 3|L - m|.$$

$$(b) C(L, m) = 3 \quad \text{si } L < m,$$

$$= 5(L - m) \quad \text{si } L \geq m.$$

$$(c) C(L, m) = 2 \quad \text{si } L < m,$$

$$= 5(L - m) \quad \text{si } L \geq m.$$

(En cada uno de los casos, dibuje un gráfico de $E(C)$ como función de m .)

Observación: evidentemente C es una variable aleatoria puesto que es una función de L que es una variable aleatoria. $E(C)$ es una función de m , y el problema simplemente pide encontrar el valor de m que minimiza $E(C)$, sujeta a la restricción que $m \geq 0,01$.

11.13. Suponiendo que la tasa de fallas asociada con la duración T de un artículo está dada por la función siguiente:

$$\begin{aligned} Z(t) &= C_0, & 0 \leq t < t_0, \\ &= C_0 + C_1(t - t_0), & t \geq t_0. \end{aligned}$$

Observación: ésta representa otra generalización de la distribución exponencial. Lo anterior reduce a una tasa constante de fallas (y por tanto a la distribución exponencial) si $C_1 = 0$.

(a) Obtener la fdp de T , el tiempo para fallar.

(b) Obtener una expresión para la confiabilidad $R(t)$ y dibujar su gráfico.

11.14. Supóngase que cada uno de tres instrumentos electrónicos tiene una ley de falla dada por una distribución exponencial con parámetros β_1, β_2 , y β_3 . Supóngase que los tres instrumentos funcionan independientemente y están conectados en paralelo para formar un solo sistema.

(a) Obtener una expresión para $R(t)$, la confiabilidad del sistema.

(b) Obtener una expresión para la fdp de T , el tiempo para fallar el sistema. Dibujar la fdp.

(c) Encontrar el promedio de tiempo para fallar del sistema.

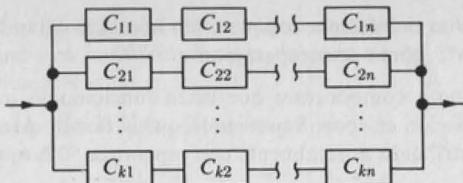


FIGURA 11.10

11.15. (a) Supóngase que n componentes están conectados en una agrupación en serie. Luego k de tales conexiones en serie se conectan en paralelo para formar un sistema completo. (Ver la figura 11.10.) Si cada una de las componentes tiene la misma confiabilidad, R , para un período determinado de operación, encuentre una expresión para la confiabilidad del sistema completo (para ese mismo período de operaciones).

(b) Supóngase que cada una de las componentes anteriores sigue una ley exponencial de fallas con tasa de falla 0,05. Supóngase además que el tiempo de operación es 10 horas y que $n = 5$. Determinar el valor de k a fin de que la confiabilidad del sistema completo sea igual a 0,99.

11.16. Supóngase que k componentes están conectados en paralelo. Luego n de tales conexiones en paralelo son unidas en serie en un solo sistema. (Ver la figura 11.11.) Responder (a) y (b) del problema 11.15 para esta situación.

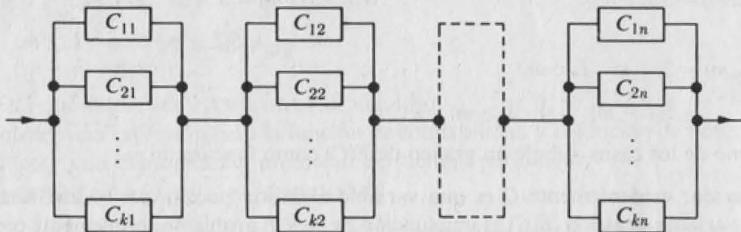


FIGURA 11.11

11.17. Supóngase que n componentes, que tienen cada una la misma tasa constante de fallas λ , están conectadas en paralelo. Encontrar una expresión para el tiempo promedio para fallar del sistema resultante.

11.18. (a) Un sistema de propulsión aéreo consta de tres motores. Supóngase que la tasa constante de fallas para cada uno de los motores es $\lambda = 0,0005$ y que los motores fallan independientemente uno de otro. Los motores están conectados en paralelo, ¿cuál es la confiabilidad de este sistema de propulsión para una misión que necesita 10 horas si al menos dos motores deben resistir?

(b) Responder la pregunta anterior para una misión que necesita 100 horas; 1000 horas. (Este problema sugerido por una exposición en I. Bazovsky, *Reliability Theory and Practice*. Prentice-Hall, Inc., Englewood Cliffs, N. J., 1961.)

11.19. Consideremos las componentes A, A', B, B' y C conectadas como se indica en las figuras 11.12 (a) y (b). (La componente C puede pensarse que representa un «seguro» si ambas A y B dejaran de funcionar.) Representando por $R_A, R_{A'}, R_B, R_{B'}$, y R_C las confiabilidades de las componentes individuales (y suponiendo que las componentes funcionan independientemente una de otra), obtener una expresión para la confiabilidad del sistema completo en cada uno de

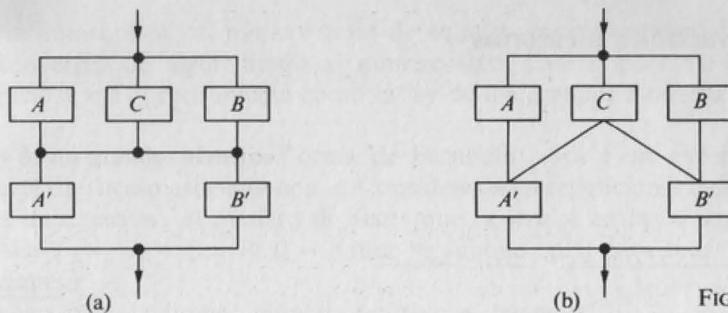


FIGURA 11.12

los casos. [Indicación: en el segundo caso (figura 11.12 (b)), use consideraciones de probabilidad condicional.]

11.20. Si todas las componentes consideradas en el problema 11.19 tienen la misma tasa constante de fallas λ , obtenga una expresión para la confiabilidad $R(t)$ del sistema indicado en la figura 11.12 (b). También encontrar el tiempo medio para que falle este sistema.

11.21. La componente A tiene una confiabilidad 0,9 cuando se utiliza con un propósito particular. La componente B que puede usarse en lugar de la componente A , tiene una confiabilidad de sólo 0,75, ¿cuál es el número mínimo de componentes del tipo B que tendrían que concretarse en paralelo a fin de obtener la confiabilidad que la componente A tiene por sí misma?

11.22. Supóngase que dos componentes que funcionan independientemente, cada una con la misma tasa constante de falla están conectadas en paralelo. Si T el tiempo para fallar del sistema resultante, obtenga la fgm de T . También determinar $E(T)$ y $V(T)$, usando la fgm.

11.23. Cada vez que hemos considerado un sistema formado por diversas componentes, siempre hemos supuesto que las componentes funcionan independientemente una de otra. Esta suposición ha simplificado considerablemente nuestros cálculos. Sin embargo, esto puede no ser siempre una suposición realista. En muchos casos es sabido que el comportamiento de una componente puede afectar el comportamiento de las otras. Esto es, en general, un problema muy difícil para enfrentar, y sólo consideraremos un caso especial aquí. Supongamos específicamente que dos componentes, C_1 y C_2 siempre fallan juntas. Es decir, falla C_1 si y sólo si falla C_2 . Demostrar que en este caso, $P(C_1 \text{ falla y } C_2 \text{ falla}) = P(C_1 \text{ falle}) = P(C_2 \text{ falle})$.

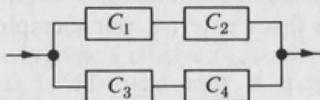


FIGURA 11.13

11.24. Considérense cuatro componentes C_1, C_2, C_3 y C_4 conectadas como se indicó en la figura 11.13. Suponiendo que las componentes funcionan independientemente una de otra con la excepción de C_1 y C_2 que siempre fallan juntas como se describió en el problema 11.23. Si T_i el tiempo para fallar de la componente C_i , es distribuido exponencialmente con parámetro β_i , obtener la confiabilidad $R(t)$ del sistema completo. Obtener también la fdp de T , el tiempo para fallar del sistema.

11.25. Considérese el mismo sistema tal como se describió en el problema 11.24, exceptuando esta vez que las componentes C_1 y C_3 fallan juntas. Responder a las preguntas del problema 11.24.

Sumas de variables aleatorias

12.1 Introducción

En este capítulo queremos precisar algo que hemos estado indicando a través del texto. Esto es, cuando el número de repeticiones de un experimento aumenta, f_A , la frecuencia relativa de un suceso A , converge (en un sentido probabilístico que describiremos) a la probabilidad teórica $P(A)$. Es este hecho lo que nos permite «identificar» la frecuencia relativa de un suceso, basada en un gran número de repeticiones, con la probabilidad del suceso.

Por ejemplo, si se produce un artículo nuevo y no tenemos conocimiento previo acerca de qué tan probable es ese artículo defectuoso, podríamos proceder a inspeccionar un gran número de esos artículos, digamos N , contar el número de artículos defectuosos entre ellos, sea n , y luego usar n/N como una aproximación para la probabilidad de que un artículo sea defectuoso. El número n/N es una variable aleatoria y su valor esencialmente depende de dos cosas. Primero, el valor n/N depende de la probabilidad fundamental p (posiblemente desconocida) de que un artículo sea defectuoso. Segundo, n/N depende de los n artículos que en particular hemos inspeccionado. Lo que demostraremos es que si el método de elegir los N artículos es «aleatorio», entonces el cociente n/N será próximo a p (en un sentido que se va a describir). (Evidentemente, la elección aleatoria de los N artículos es importante. Si fuéramos a elegir sólo esos artículos que presentan alguna característica física externa, por ejemplo, podríamos prejuiciar gravemente nuestro cálculo.)

12.2 La ley de los grandes números

Con la ayuda de la desigualdad de Chebyshev (ecuación 7.20), podemos derivar el resultado citado anteriormente. Consideraremos nuevamente un ejemplo. Supongamos que un cohete dirigido tiene una probabilidad de 0,95 de funcionar correctamente durante un cierto período de operaciones. Así, si disparamos N cohetes que tienen la confiabilidad anterior, y si X es el número de cohetes que no funcionan correctamente, tenemos $E(X) = 0,05N$, puesto que podemos suponer que X está distribuida binomialmente. Es decir, esperaríamos que fallara alrededor de un cohete entre 20. Cuando N , el número de cohetes

lanzado, se aumenta, X , el número total de cohetes que fallan dividido por N , debería converger de algún modo al número 0,05. Este importante resultado puede indicarse más precisamente como la ley de los grandes números.

La ley de los grandes números (forma de Bernoulli). Sea ε un experimento y sea A un suceso asociado con ε . Considerando n repeticiones independientes de ε , sea n_A el número de veces que ocurre A en las n repeticiones, y sea $f_A = n_A/n$. Sea $P(A) = p$ (que se supone igual para todas las repeticiones.)

Luego para cualquier número positivo ϵ , tenemos

$$\text{Prob} [|f_A - p| \geq \epsilon] \leq \frac{p(1-p)}{n\epsilon^2}$$

o, equivalente

$$\text{Prob} [|f_A - p| < \epsilon] \geq 1 - \frac{p(1-p)}{n\epsilon^2}. \quad (12.1)$$

Demostración: sea n_A el número de veces que ocurre el suceso A . Esta es una variable aleatoria distribuida binomialmente. Luego $E(n_A) = np$ y $V(n_A) = np(1-p)$. Como $f_A = n_A/n$, luego $E(f_A) = p$ y $V(f_A) = p(1-p)/n$.

Aplicando la desigualdad de Chebyshev a la variable aleatoria f_A , obtenemos

$$P \left[|f_A - p| < k \sqrt{\frac{p(1-p)}{n}} \right] \geq 1 - \frac{1}{k^2}.$$

Sea $\epsilon = k \sqrt{p(1-p)/n}$. Luego $k^2 = (n\epsilon^2)/[p(1-p)]$, y así

$$P [|f_A - p| < \epsilon] \geq 1 - \frac{p(1-p)}{n\epsilon^2}.$$

Observaciones: (a) El resultado anterior puede establecerse en otras maneras alternas equivalentes. Está claro que lo anterior implica inmediatamente que

$$\lim_{n \rightarrow \infty} P [|f_A - p| < \epsilon] = 1 \quad \text{para todo } \epsilon > 0.$$

En este sentido decimos que la frecuencia relativa f_A «converge» a $P(A)$.

(b) Es importante observar la diferencia entre la convergencia citada anteriormente (llamada *convergencia en probabilidad*) y el tipo de convergencia citada a menudo en cálculo. Cuando decimos que 2^{-n} converge a cero cuando $n \rightarrow \infty$ significa que para un n suficientemente grande 2^{-n} se transforma y permanece arbitrariamente próximo a cero. Cuando decimos que $f_A = n_A/n$ converge a $P(A)$ indicamos que la *probabilidad* del suceso

$$\{|n_A/n - P(A)| < \epsilon\}$$

puede hacerse arbitrariamente próxima a uno tomando un n suficientemente grande.

(c) Se obtiene aún otra forma de la ley de los grandes números cuando nos formulamos la siguiente pregunta. ¿Cuántas repeticiones de ε deberían hacerse a fin de tener una probabilidad de 0,95 a lo menos, de que la frecuencia relativa difiera de $p = P(A)$ en menos de 0,01?

Es decir, para $\epsilon = 0,01$ deseamos escoger n de modo que $1 - p(1 - p)/[n(0,01)^2] = 0,95$. Resolviendo para n obtenemos $n = p(1 - p)/(0,01)^2(0,05)$. Sustituyendo los valores específicos de 0,05 y 0,01 por δ y ϵ , respectivamente, tenemos

$$P[|f_A - p| < \epsilon] \geq 1 - \delta \quad \text{cuando } n \geq \frac{p(1 - p)}{\epsilon^2 \delta}.$$

Nuevamente debería insistirse en que tomando $n \geq p(1 - p)/\epsilon^2 \delta$ no garantiza nada acerca de $|f_A - p|$. Sólo hace probable que $|f_A - p|$ sea muy pequeño.

EJEMPLO 12.1. ¿Cuántas veces habría que lanzar un dado regular a fin de estar al menos 95 por ciento seguro de que la frecuencia relativa que salga un seis diste 0,01 de la probabilidad teórica $\frac{1}{6}$?

Aquí $p = \frac{1}{6}$, $1 - p = \frac{5}{6}$, $\epsilon = 0,01$ y $\delta = 0,05$. Luego de esta relación encontramos que $n \geq (\frac{1}{6})(\frac{5}{6})/(0,01)^2(0,05) = 27,778$.

Observaciones: (a) Recordemos que la f_A es una variable aleatoria y no precisamente un valor observado. Si lanzamos ahora 27,778 veces un dado y luego calculamos la frecuencia relativa que salga un seis, este número dista o no 0,01 de $1/6$. Lo importante del ejemplo anterior es que si lanzáramos 27,778 veces un dado en cada una de 100 habitaciones, en 95 de las habitaciones aproximadamente, la frecuencia relativa observada distaría 0,01 de $1/6$.

(b) En muchos problemas no conocemos el valor de $p = P(A)$ y por tanto no podemos usar la cota anterior de n . En ese caso podemos usar el hecho de que $p(1 - p)$ toma su valor máximo cuando $p = \frac{1}{2}$ y este valor máximo es igual a $1/4$. Así, ciertamente estariamos seguros si indicamos que para $n \geq 1/4\epsilon^2 \delta$ tenemos

$$P[|f_A - p| < \epsilon] \geq 1 - \delta.$$

EJEMPLO 12.2. Algunos artículos se producen de tal manera que la probabilidad de que un artículo sea defectuoso es p (supuesto desconocido). Un gran número de artículos, n , se clasifica como defectuosos o no defectuosos. ¿Cuál debe ser el tamaño de n de modo que podamos estar el 99 por ciento seguros de que la frecuencia relativa de los defectuosos se diferencia de p en menos de 0,05?

Puesto que no sabemos el valor de p debemos aplicar la última forma establecida de la ley de los grandes números. Luego con $\epsilon = 0,05$, $\delta = 0,01$ encontramos que si $n \geq 1/4(0,05)^2(0,01) = 10,000$, se satisface la condición pedida.

Como en nuestro ejemplo en la desigualdad de Chebyshev, encontraremos que un conocimiento adicional acerca de la distribución de probabilidades dará una proposición «mejorada». (Por ejemplo podríamos tener un número pequeño de repeticiones y todavía hacer la misma proposición referente a la proximidad de f_A a p .)

Observación: se puede obtener otra forma de la ley de los grandes números como sigue. Supongamos que X_1, \dots, X_n son variables aleatorias distribuidas independientes e idénticamente con promedio y varianza finita. Sea $E(X_i) = \mu$ y $V(X_i) = \sigma^2$. Definamos $\bar{X} = (1/n)(X_1 + \dots + X_n)$. Luego \bar{X} es una función de X_1, \dots, X_n , llamado su promedio aritmético y, por tanto, nuevamente es una variable aleatoria. (Estudiaremos esta variable aleatoria con más detalle en el capítulo 13. Por el momento digamos simplemente que podemos pensar que X_1, \dots, X_n como medidas independientes de una característica numérica X ,

que producen el promedio aritmético \bar{X}). De las propiedades de la esperanza y de la varianza inmediatamente tenemos, $E(\bar{X}) = \mu$ y $V(\bar{X}) = \sigma^2/n$. Apliquemos la desigualdad de Chebyshev a la variable aleatoria \bar{X} :

$$P\left[|\bar{X} - \mu| < \frac{k\sigma}{\sqrt{n}}\right] \geq 1 - \frac{1}{k^2}.$$

Sea $k\sigma/\sqrt{n} = \epsilon$. Luego $k = \sqrt{n}\epsilon/\sigma$ y podemos escribir

$$P[|\bar{X} - \mu| < \epsilon] \geq 1 - \frac{\sigma^2}{\epsilon^2 n}. \quad (12.2)$$

Cuando $n \rightarrow \infty$ el lado derecho de la desigualdad anterior se aproxima a uno. Esto en el sentido en el cual el promedio aritmético «converge» a $E(X)$.

EJEMPLO 12.3. Se prueba un gran número de tubos electrónicos. Sea T_i el tiempo para fallar del i -ésimo tubo. Supóngase, además, que todos los tubos provienen del mismo lote y que puede suponerse que todos están distribuidos exponencialmente con el mismo parámetro α .

Luego $E(T_i) = \alpha^{-1}$. Sea $T = (T_1 + \cdots + T_n)/n$. La forma anterior de la ley de los grandes números dice que si n es muy grande, sería «muy probable» que el valor obtenido para el promedio aritmético de un gran número de tiempos de fallas estuviera próximo a α^{-1} .

12.3 Aproximación normal a la distribución binomial

La ley de los grandes números como se estableció anteriormente se relaciona esencialmente con la variable aleatoria X distribuida binomialmente. X fue definida como el número de éxitos en n repeticiones independientes de un experimento y necesitamos asociar simplemente «éxito» con la ocurrencia del suceso A a fin de reconocer esta relación. Así el resultado anterior puede establecerse informalmente al decir que la frecuencia relativa de éxito, X/n , converge a la probabilidad de éxito p , en el sentido indicado previamente, cuando el número de repeticiones del experimento se aumenta.

Sin embargo, saber que X/n será «próxima» a p para un n grande, no nos indica cómo se obtiene esta «proximidad». A fin de investigar esto debemos estudiar la distribución de probabilidades de X cuando n es grande.

Por ejemplo, supóngase que un proceso de fabricación produce lavadoras, de las cuales alrededor del 5 por ciento son defectuosas (es decir, muchas). Si se inspeccionan 100 lavadoras, ¿cuál es la probabilidad de que menos de 4 sean defectuosas?

Siendo X el número de lavadoras defectuosas encontradas, la ley de los grandes números nos dice simplemente que $X/100$ debería ser «próximo» a 0,05. Sin embargo, no nos indica cómo calcular la probabilidad deseada. El valor *exacto* de esta probabilidad está dado por

$$P(X < 4) = \sum_{k=0}^3 \binom{100}{k} (0,05)^k (0,95)^{100-k}.$$

Esta probabilidad sería más difícil de calcular directamente. Ya hemos estudiado un método de aproximación para las probabilidades binomiales, tal como la aproximación de Poisson. Consideraremos ahora otra aproximación importante para tales probabilidades que se aplica cada vez que n es suficientemente grande.

Considerando que $P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$. Esta probabilidad depende de n de un modo muy complicado y no hay una indicación evidente de lo que sucede a la expresión anterior si n es grande. A fin de investigar esta probabilidad.

TABLA 12.1

n	$n!$	$\sqrt{2\pi} e^{-n} n^{n+1/2}$	Diferencia	Diferencia
				$n!$
1	1	0,922	0,078	0,08
2	2	1,919	0,081	0,04
5	120	118,019	1,981	0,02
10	$(3,6288)10^6$	$(3,5986)10^6$	$(0,0302)10^6$	0,008
100	$(9,3326)10^{157}$	$(9,3249)10^{157}$	$(0,0077)10^{157}$	0,0008

necesitamos usar la *fórmula de Stirling*, una aproximación muy conocida de $n!$. Esta fórmula establece que para un n grande,

$$n! \sim \sqrt{2\pi} e^{-n} n^{n+1/2}, \quad (12.3)$$

en el supuesto de que $\lim_{n \rightarrow \infty} (n!) / (\sqrt{2\pi} e^{-n} n^{n+1/2}) = 1$. (Una demostración de esta aproximación puede encontrarse en muchos textos de cálculo avanzado.) La tabla 12.1 puede dar una idea al lector de la exactitud de esta aproximación. Esta tabla está tomada de W. Feller, *Probability Theory and Its Applications*, 1.^a edición, New York: John Wiley and Sons, Inc., 1950.

Observación: aunque la diferencia entre $n!$ y su aproximación llega a ser mayor cuando $n \rightarrow \infty$, lo importante de observar en la tabla 12.1 es que el porcentaje de error (la última columna) llega a ser cada vez más pequeño.

Usando la fórmula de Stirling para los diversos factoriales que aparecen en la expresión de $P(X = k)$, puede demostrarse (después de muchas operaciones), que para un n grande,

$$\begin{aligned} P(X = k) &= \binom{n}{k} p^k (1 - p)^{n-k} \\ &\simeq \frac{1}{\sqrt{2\pi np(1-p)}} \exp\left(-\frac{1}{2}\left[\frac{k-np}{\sqrt{np(1-p)}}\right]^2\right) \end{aligned} \quad (12.4)$$

Finalmente puede demostrarse que para n grande,

$$\begin{aligned} P(X \leq k) &= P\left[\frac{X - np}{\sqrt{np(1-p)}} \leq \frac{k - np}{\sqrt{np(1-p)}}\right] \\ &\simeq \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(k-np)/\sqrt{np(1-p)}} e^{-t^2/2} dt. \end{aligned} \quad (12.5)$$

Así tenemos el siguiente resultado importante (conocido como la aproximación de DeMoivre-Laplace para la distribución binomial):

Aproximación normal a la distribución binomial. Si X tiene una distribución binomial con parámetro n y p y si

$$Y = \frac{X - np}{[np(1-p)]^{1/2}},$$

luego, para un n grande, Y tiene una distribución $N(0,1)$ aproximadamente en el supuesto que $\lim_{n \rightarrow \infty} P(Y \leq y) = \Phi(y)$. Esta aproximación es válida para valores de $n > 10$ suponiendo p próximo a $\frac{1}{2}$. Si p es próximo a 0 ó 1, n debería ser algo mayor para asegurar una buena aproximación.

Observaciones: (a) El resultado anterior no es sólo de considerable interés teórico sino también de mucha importancia práctica. Indica que podemos usar la distribución normal, muy tabulada, para calcular probabilidades que provienen de la distribución binomial.

(b) En la tabla 12.2 la exactitud de la aproximación (12.4) se demuestra para diversos valores de n, k y p .

TABLA 12.2

k	$n = 8, p = 0,2$		$n = 8, p = 0,5$		$n = 25, p = 0,2$	
	Aproximación	Exacto	Aproximación	Exacto	Aproximación	Exacto
0	0,130	0,168	0,005	0,004	0,009	0,004
1	0,306	0,336	0,030	0,031	0,027	0,024
2	0,331	0,294	0,104	0,109	0,065	0,071
3	0,164	0,147	0,220	0,219	0,121	0,136
4	0,037	0,046	0,282	0,273	0,176	0,187
5	0,004	0,009	0,220	0,219	0,199	0,196
6	0 +	0,001	0,104	0,109	0,176	0,163
7	0 +	0 +	0,030	0,031	0,121	0,111
8	0 +	0 +	0,005	0,004	0,065	0,062
9	0 +	0 +	0 +	0 +	0,027	0,029
10	0 +	0 +	0 +	0 +	0,009	0,012
11	0 +	0 +	0 +	0 +	0,002	0,004

Volviendo al ejemplo anterior, observamos que

$$E(X) = np = 100(0,05) = 5,$$

$$V(X) = np(1-p) = 4,75.$$

Luego podemos escribir

$$\begin{aligned} P(X \leq 3) &= P\left(\frac{0-5}{\sqrt{4,75}} \leq \frac{X-5}{\sqrt{4,75}} \leq \frac{3-5}{\sqrt{4,75}}\right) \\ &= \Phi(-0,92) - \Phi(-2,3) = 0,168, \end{aligned}$$

de las tablas de la distribución normal.

Observación: al usar la aproximación normal a la distribución binomial, estamos aproximando la distribución de una variable aleatoria discreta con la distribución de una variable aleatoria continua. Por tanto hay que tener algún cuidado con los puntos extremos del intervalo considerado. Por ejemplo, para una variable aleatoria continua, $P(X = 3) = 0$, mientras que para una variable aleatoria discreta esta probabilidad puede ser positiva.

Se ha encontrado que la siguiente *corrección para continuidad* mejora la aproximación anterior:

- (a) $P(X = k) \simeq P(k - \frac{1}{2} \leq X \leq k + \frac{1}{2})$.
- (b) $P(a \leq X \leq b) \simeq P(a - \frac{1}{2} \leq X \leq b + \frac{1}{2})$.

Usando esta última corrección para la evaluación anterior de $P(X \leq 3)$, tenemos

$$\begin{aligned} P(X \leq 3) &= P(0 \leq X \leq 3) = P(-\frac{1}{2} \leq X \leq 3\frac{1}{2}) \\ &\simeq \Phi(-0,69) - \Phi(-2,53) = 0,239. \end{aligned}$$

EJEMPLO 12.4. Supóngase que un sistema está formado por 100 componentes cada una de las cuales tiene una confiabilidad igual a 0,95. (Es decir, la probabilidad de que la componente funcione correctamente durante un tiempo específico es igual a 0,95.) Si esas componentes funcionan independientemente una de otra, y si el sistema completo funciona correctamente cuando al menos funcionan 80 componentes, ¿cuál es la confiabilidad del sistema?

Sea X el número de componentes que funcionan, debemos evaluar

$$P(80 \leq X \leq 100).$$

Tenemos

$$E(X) = 100(0,95) = 95; \quad V(X) = 100(0,95)(0,05) = 4,75.$$

Luego, usando la corrección para continuidad, obtenemos

$$\begin{aligned} P(80 \leq X \leq 100) &\simeq P(79,5 \leq X \leq 100,5) \\ &= P\left(\frac{79,5 - 95}{2,18} \leq \frac{X - 95}{2,18} \leq \frac{100,5 - 95}{2,18}\right) \\ &\simeq \Phi(2,52) - \Phi(-7,1) = 0,994. \end{aligned}$$

12.4 El teorema del límite central

La aproximación anterior representa sólo un caso especial de un resultado general. A fin de verificar esto, recordemos que la variable aleatoria X distribuida binomialmente se puede representar como la *suma* de las siguientes variables aleatorias independientes:

$$\begin{aligned} X_i &= 1 && \text{si el éxito ocurre en la } i\text{-ésima repetición;} \\ &= 0 && \text{si la falla ocurre en la } i\text{-ésima repetición.} \end{aligned}$$

Luego $X = X_1 + X_2 + \cdots + X_n$. (Ver el ejemplo 7.14.) Para esta variable aleatoria hemos demostrado que $E(X) = np$, $V(X) = np(1 - p)$ y, además, que para un n grande, $(X - np)/\sqrt{np(1 - p)}$ tiene la distribución aproximada $N(0, 1)$.

Si una variable aleatoria X puede representarse como una *suma de n variables aleatorias independientes cualesquiera* (satisfaciendo ciertas condiciones que son válidas en la mayor parte de las aplicaciones), entonces esta suma para un n suficientemente grande está distribuida aproximadamente normal. Este resultado notable se conoce como el teorema del límite central. Se puede establecer una forma de este teorema como sigue.

Teorema del límite central. Sea $X_1, X_2, \dots, X_n \dots$ una sucesión de variables aleatorias independientes con $E(X_i) = \mu_i$, y $V(X_i) = \sigma_i^2$, $i = 1, 2, \dots$. Sea $X = X_1 + X_2 + \cdots + X_n$. Luego bajo ciertas condiciones generales (que no se indicarán explicitamente aquí),

$$Z_n = \frac{X - \sum_{i=1}^n \mu_i}{\sqrt{\sum_{i=1}^n \sigma_i^2}}$$

tiene aproximadamente la distribución $N(0, 1)$. Es decir, si G_n es la fda de la variable aleatoria Z_n , tenemos $\lim_{n \rightarrow \infty} G_n(z) = \Phi(z)$.

Observaciones: (a) Este teorema representa una generalización obvia de la aproximación de DeMoivre-Laplace. Las variables aleatorias independientes X_i que toman sólo los valores 1 y 0 han sido sustituidas por variables aleatorias que poseen cualquier clase de distribución (mientras tengan esperanzas y varianzas finitas). El hecho de que las X_i pueden tener (evidentemente) cualquier clase de distribución y aún así la suma $X = X_1 + \cdots + X_n$ puede ser aproximada por una variable aleatoria distribuida normalmente, representa la razón básica para la importancia de la distribución normal en la teoría de probabilidades. Resulta que en muchos problemas, la variable aleatoria que se considera, puede ser representada como la suma de n variables aleatorias independientes y, por tanto, su distribución puede aproximarse por la distribución normal.

Por ejemplo, el consumo de electricidad en una ciudad en cualquier hora dada es la suma de los pedidos de un gran número de consumidores individuales. Puede considerarse la cantidad de agua de un embalse como la representación de la suma de un gran número de contribuciones individuales. Y el error de medida en un experimento físico está compuesto de muchos errores pequeños no observables que pueden considerarse aditivos. El bombardeo molecular sobre

una partícula suspendida en un líquido es la causa que la obliga a desplazarse en una dirección y una magnitud aleatorias, y su posición (después de un tiempo especificado) puede considerarse como una suma de desplazamientos individuales.

(b) Las condiciones generales citadas en la formulación anterior del teorema del límite central pueden resumirse informalmente como sigue: los términos individuales en la suma contribuyen con una cantidad despreciable a la variación de la suma y es muy improbable que cualquier tamaño individual haga una gran contribución a la suma. (Parece que los errores de medida tienen esta característica. El error final puede ser representado como una suma de muchas pequeñas contribuciones ninguna de las cuales contribuye mucho al error completo).

(c) Hemos establecido ya (teorema 10.5) que la suma de cualquier número finito de variables aleatorias independientes distribuida normalmente es de nuevo distribuida normalmente. El teorema del límite central establece que los sumandos no necesitan estar distribuidos normalmente a fin de que la suma sea aproximada a una distribución normal.

(d) No podemos demostrar el teorema anterior aquí, sin exceder el nivel de presentación programado. Sin embargo, hay un caso especialmente importante de este teorema que estableceremos y para el cual damos al menos un bosquejo de demostración.

Teorema 12.1. Sean X_1, \dots, X_n n variables aleatorias independientes que tienen la misma distribución. Sean $\mu = E(X_i)$ y $\sigma^2 = V(X_i)$ la esperanza y varianza común. Sea $S = \sum_{i=1}^n X_i$. Entonces $E(S) = n\mu$ y $V(S) = n\sigma^2$, y para un gran n , tenemos que $T_n = (S - n\mu)/\sqrt{n\sigma}$ tiene aproximadamente la distribución $N(0,1)$ considerando que $\lim_{n \rightarrow \infty} P(T_n \leq t) = \Phi(t)$.

Demostración (bosquejo): (El lector haría bien en revisar los conceptos básicos de la fgm presentados en el capítulo 10). Sea M la fgm (común) de las X_i . Puesto que las X_i son independientes, M_S , la fgm de S , está dada por $M_S(t) = [M(t)]^n$, y puesto que T_n es una función lineal de S , (usando el teorema 10.2) la fgm de T_n está dada por

$$M_{T_n}(t) = e^{-(\sqrt{n}\mu/\sigma)t} \left[M\left(\frac{t}{\sqrt{n}\sigma}\right)\right]^n.$$

Así

$$\ln M_{T_n}(t) = \frac{-\sqrt{n}\mu}{\sigma}t + n \ln M\left(\frac{t}{\sqrt{n}\sigma}\right).$$

(Observemos a esta altura que la idea de la demostración consiste en investigar $M_{T_n}(t)$ para grandes valores de n).

Desarrollemos $M(t)$ en una serie de Maclaurin:

$$M(t) = 1 + M'(0)t + \frac{M''(0)t^2}{2} + R,$$

en donde R es el término del resto. Recordando que $M'(0) = \mu$ y $M''(0) = \mu^2 + \sigma^2$, obtenemos

$$M(t) = 1 + \mu t + \frac{(\mu^2 + \sigma^2)t^2}{2} + R.$$

Luego

$$\ln M_{T_n}(t) = -\frac{\sqrt{n}\mu t}{\sigma} + n \ln \left[1 + \frac{\mu t}{\sqrt{n}\sigma} + \frac{(\mu^2 + \sigma^2)t^2}{2n\sigma^2} + R \right].$$

Usaremos ahora el desarrollo de Maclaurin para $\ln(1+x)$:

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} + \dots$$

(Este desarrollo es válido para $|x| < 1$. En nuestro caso

$$x = \frac{\mu t}{\sqrt{n}\sigma} + \frac{(\mu^2 + \sigma^2)t^2}{2n\sigma^2} + R;$$

y para n suficientemente grande, el valor absoluto de este desarrollo será menor que uno). Así, obtenemos

$$\begin{aligned} \ln M_{T_n}(t) = & -\frac{\sqrt{n}\mu}{\sigma} + n \left[\left(\frac{\mu t}{\sqrt{n}\sigma} + (\mu^2 + \sigma^2) \frac{t^2}{2n\sigma^2} + R \right) \right. \\ & \left. - \frac{1}{2} \left(\frac{\mu t}{\sqrt{n}\sigma} + (\mu^2 + \sigma^2) \frac{t^2}{2n\sigma^2} + R \right)^2 + \dots \right]. \end{aligned}$$

Puesto que simplemente estamos bosquejando los pasos principales de la demostración sin dar todos los detalles, omitamos algunos pasos algebraicos e indiquemos simplemente lo que estamos haciendo. Queremos investigar la expresión anterior ($\ln M_{T_n}(t)$) cuando $n \rightarrow \infty$. Cualquier término que tiene una potencia positiva de n en el denominador (tal como $n^{-1/2}$, por ejemplo) tenderá a cero cuando $n \rightarrow \infty$. Después de un proceso algebraico muy directo, pero tedioso, encontramos que

$$\lim_{n \rightarrow \infty} \ln M_{T_n}(t) = t^2/2.$$

Luego tenemos

$$\lim_{n \rightarrow \infty} M_{T_n}(t) = e^{t^2/2}.$$

Esta es la fgm de una variable aleatoria con distribución $N(0, 1)$. Debido a la propiedad de unicidad de la fgm (ver el teorema 10.3) podemos concluir que la variable aleatoria T_n converge en distribución (cuando $n \rightarrow \infty$) a la distribución $N(0, 1)$.

Observaciones: (a) Aunque la anterior no representa una demostración rigurosa, aún así debería dar alguna idea al lector para la deducción de este importante teorema. La forma más general del teorema del límite central (como se estableció originalmente) se puede demostrar usando un enfoque semejante al utilizado aquí.

(b) La forma especial del teorema del límite central como se estableció anteriormente expresa que el promedio aritmético $(1/n)\sum_{i=1}^n X_i$ de n observaciones de la misma variable aleatoria tiene aproximadamente una distribución normal, para una n grande.

(c) Aunque una demostración matemática establecería la validez de un teorema, puede que no contribuya mucho a la idea intuitiva del resultado. Por lo tanto presentamos el ejemplo que sigue para quienes están orientados más numéricamente.

EJEMPLO 12.5. Considérese una urna que contiene tres clases de objetos identificados como 0, 1 y 2. Suponiendo que hay 20 ceros, 30 unos y 50 doses. Se saca un artículo al azar y se anota su valor, digámoslo X . Supóngase que X tiene la distribución siguiente. (Ver la figura 12.1.)

X	0	1	2
$P(X = x)$	0,2	0,3	0,5

Supóngase que el artículo escogido primero se substituye y luego se escoge un

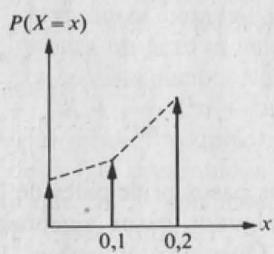


FIGURA 12.1

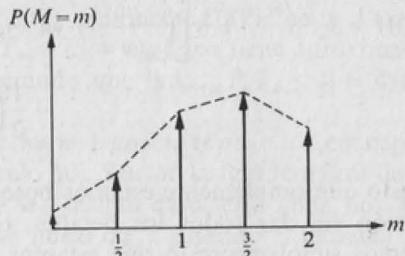


FIGURA 12.2

segundo artículo y se anota su valor, Y . Considérese la variable aleatoria $M = (X + Y)/2$ y su distribución (figura 12.2).

M	0	$\frac{1}{2}$	1	$\frac{3}{2}$	2
$P(M = m)$	0,04	0,12	0,29	0,30	0,25

Los valores anteriores de $P(M)$ los obtuvimos como sigue:

$$\begin{aligned} P(M = 0) &= P(X = 0, Y = 0) = (0,2)^2 = 0,04; \\ P(M = \frac{1}{2}) &= P(X = 0, Y = 1) + P(X = 1, Y = 0) \\ &= (0,2)(0,3) + (0,3)(0,2) = 0,12, \quad \text{etc.} \end{aligned}$$

Finalmente, supongamos que después de que el segundo artículo ha sido también remplazado, se escoge un tercer artículo y se anota su valor Z . Considerar la variable aleatoria $N = (X + Y + Z)/3$ y su distribución (figura 12.3):

N	0	$\frac{1}{3}$	$\frac{2}{3}$	1	$\frac{4}{3}$	$\frac{5}{3}$	2
$P(N = n)$	0,008	0,036	0,114	0,207	0,285	0,225	0,125

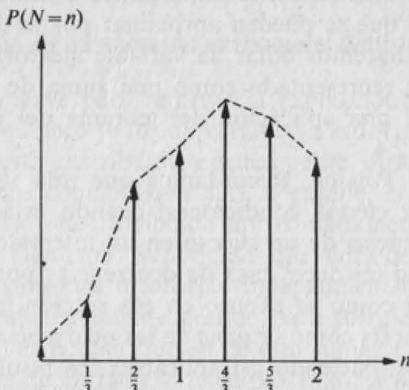


FIGURA 12.3

Las distribuciones de probabilidades de las variables aleatorias M y N ya están mostrando signos de «normalidad». Es decir, la aparición de la forma de campana de la curva de distribución empieza a ser evidente. Empezando con la distribución de X , que es muy simétrica, encontramos que el promedio de sólo tres observaciones tiene una distribución que ya muestra «signos de normalidad».

Naturalmente, el ejemplo anterior no *demuestra* nada. Sin embargo, representa una ilustración numérica de los resultados expuestos previamente de una manera más matemática. El lector debería continuar este ejemplo agregando una observación más y luego encontrar la distribución de probabilidades del promedio de las cuatro observaciones obtenidas. (Ver el problema 12.10.)

EJEMPLO 12.6. Supongamos que tenemos cierto número de voltajes, V_i , $i = 1, 2, \dots, n$, que se reciben en lo que se llama un «sumador». (Ver la figura 12.4.) Sea V la suma de los voltajes recibidos. Es decir, $V = \sum_{i=1}^n V_i$. Supóngase que cada una de las variables aleatorias V_i está distribuida uniformemente en el intervalo $[0, 10]$. Luego $E(V_i) = 5$ volts y $\text{var}(V_i) = 100/12$.

De acuerdo con el teorema del límite central, si n es suficientemente grande, la variable aleatoria

$$S = (V - 5n)\sqrt{12/10}\sqrt{n}$$

tiene aproximadamente la distribución $N(0, 1)$. Luego si $n = 20$ podemos calcular la probabilidad de que el voltaje total de entrada sobrepase 105 volts, como sigue:

$$\begin{aligned} P(V > 105) &= P\left(\frac{V - 100}{(10/\sqrt{12})\sqrt{20}} > \frac{105 - 100}{(10/\sqrt{12})\sqrt{20}}\right) \\ &\simeq 1 - \Phi(0,388) = 0,352. \end{aligned}$$

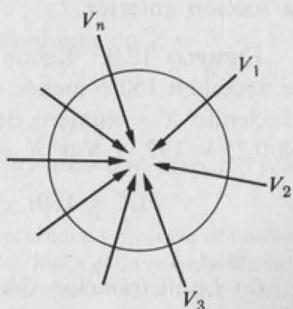


FIGURA 12.4

12.5 Otras distribuciones aproximadas por la distribución normal: Poisson, Pascal y Gama

Hay cierto número de distribuciones importantes distintas a la binomial expuestas en la sección 12.3 que se pueden aproximar por la distribución normal. Y en cada caso, como lo haremos notar, la variable aleatoria cuya distribución aproximaremos puede ser representada como una suma de variables aleatorias independientes, dando así una aplicación del teorema del límite central como se expuso en la sección 12.4.

(a) *La distribución de Poisson.* Recordamos que una variable aleatoria de Poisson aparece (sujeta a ciertas condiciones) cuando estamos interesados en el número total de ocurrencias de un suceso en un intervalo de tiempo de longitud t , con una intensidad (es decir, tasa de ocurrencias por unidad de tiempo) α (ver la sección 8.2). Tal como se expuso en esa sección, podemos considerar el número total de ocurrencias como la *suma* de las ocurrencias en intervalos muy pequeños, no sobrepuestos, haciendo así aplicables los resultados de la sección anterior.

EJEMPLO 12.7. Supóngase que las llamadas a una central telefónica particular ocurren a razón de 2 por minuto. ¿Cuál es la probabilidad de que se reciban 22 o menos llamadas en un período de 15 minutos? (Suponemos, naturalmente, que la intensidad permanece igual durante el período considerado.) Si X = número de llamadas recibidas, entonces $E(X) = 2(15) = 30$. Para calcular $P(X \leq 22)$, aplicando la corrección para la continuidad (ver la ecuación (b) que precede al ejemplo 12.4), tenemos

$$P(X \leq 22) \cong P\left(Y \leq \frac{22 + \frac{1}{2} - 30}{\sqrt{30}}\right)$$

en donde Y tiene distribución $N(0, 1)$. Luego la probabilidad anterior es igual a $\Phi(-1,37) = 0,0853$.

(b) *La distribución de Pascal.* Si Y = número de ensayos de Bernoulli necesarios para tener r éxitos, entonces Y tiene una distribución de Pascal y se puede representar como la suma de r variables aleatorias independientes (ver la sección 8.5). Luego para un r suficientemente grande, se aplican los resultados de la sección anterior.

EJEMPLO 12.8. Encontrar un valor aproximado de la probabilidad de que se necesiten 150 o menos ensayos para obtener 48 éxitos cuando $P(\text{éxito}) = 0,25$. Haciendo X = número de ensayos, tenemos (ver la ecuación 8.8) $E(X) = r/p = 48/0,25 = 192$, y $\text{Var } X = rq/p^2 = (48)(0,75)/(0,25)^2 = 576$. Luego

$$P(X \leq 150) \cong \Phi\left(\frac{150 + \frac{1}{2} - 192}{\sqrt{576}}\right) = \Phi(-1,73) = 0,0418.$$

(c) *La distribución Gama.* Como se indicó en el teorema 10.9 una variable aleatoria que tiene una distribución Gama (con parámetros α y r) se puede repre-

sentar como una suma de r variables aleatorias independientes distribuidas exponencialmente. Luego para un r grande, se aplica nuevamente el teorema del límite central.

12.6 La distribución de la suma de un número finito de variables aleatorias

El ejemplo 12.6 sirve para motivar la exposición siguiente. Sabemos que la suma de cualquier número finito de variables aleatorias independientes distribuidas normalmente está distribuida normalmente otra vez. Del teorema del límite central podemos concluir que para n grande, la suma de n variables aleatorias independientes está distribuida aproximadamente normal. Queda por resolver la pregunta siguiente: supóngase que consideramos $X_1 + \dots + X_n$, en donde las X son variables aleatorias independientes (no necesariamente normales) y n no es suficientemente grande para justificar el uso del teorema del límite central. ¿Cuál es la distribución de esta suma? Por ejemplo, ¿cuál es la distribución del voltaje de entrada V (ejemplo 12.6), si $n = 2$ o $n = 3$?

Primero consideraremos el importante caso de la suma de dos variables aleatorias. Puede establecerse el resultado siguiente.

Teorema 12.2 Supóngase que X y Y son variables aleatorias independientes con fdp g y h respectivamente. Sea $Z = X + Y$ y denótese la fdp de Z por s . Entonces,

$$s(z) = \int_{-\infty}^{+\infty} g(w)h(z-w) dw. \quad (12.6)$$

Demostración: Puesto que X y Y son independientes, su fdp conjunta f puede factorizarse:

$$f(x, y) = g(x)h(y).$$

Usando la transformación:

$$z = x + y, \quad w = x.$$

Luego $x = w$, $y = z - w$. El jacobiano de esta transformación es

$$J = \begin{vmatrix} 0 & 1 \\ 1 & -1 \end{vmatrix} = -1.$$

Luego el valor absoluto de J es 1 y por tanto la fdp conjunta de $Z = X + Y$ y $W = X$ es

$$k(z, w) = g(w)h(z-w).$$

La fdp de Z se obtiene ahora al integrar $k(z, w)$ de $-\infty$ a ∞ con respecto a w , de donde se obtiene el resultado anterior.

Observaciones: (a) La integral anterior que relaciona las funciones g y h ocurre en muchos temas matemáticos diferentes. A menudo se le menciona como la integral de *convolución* de g y h ; algunas veces se escribe como $g * h$.

(b) La evaluación de la integral anterior debe hacerse con mucho cuidado. En realidad,

la misma dificultad que apareció en la evaluación de la fdp de un producto o un cociente aparece nuevamente. Las funciones g y h a menudo serán distintas de cero sólo para ciertos valores de sus argumentos. Por tanto el integrando en la integral anterior será distinto de cero sólo para aquellos valores de la variable de integración w para los cuales *ambos* factores del integrando son distintos de cero.

(c) La fórmula anterior ecuación (12.6) puede usarse repetidamente (con dificultad creciente, sin embargo) para obtener la fdp de la suma de cualquier número finito de variables aleatorias continuas e independientes. Por ejemplo, si $S = X + Y + W$, podemos escribir esto como $S = Z + W$, en donde $Z = X + Y$. Luego podemos usar el enfoque anterior para obtener la fdp de Z , y luego, conociendo la fdp de Z , usar este método otra vez para obtener la fdp de S .

(d) Podemos derivar la ecuación (12.6) de otra manera sin usar la noción de jacobiano. Denotemos por S la fda de la variable aleatoria $Z = X + Y$. Luego

$$S(z) = P(Z \leq z) = P(X + Y \leq z) = \iint_R g(x)h(y)dx dy,$$

en donde

$$R = \{(x, y) \mid x + y \leq z\}.$$

(Ver la figura 12.5). Luego

$$\begin{aligned} S(z) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{z-x} g(x)h(y)dy dx \\ &= \int_{-\infty}^{+\infty} g(x) \left[\int_{-\infty}^{z-x} h(y)dy \right] dx. \end{aligned}$$

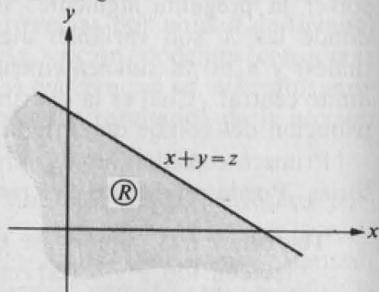


FIGURA 12.5

Diferenciando $S(z)$ con respecto a z (bajo el signo integral, lo cual puede justificarse) obtenemos

$$s(z) = S'(z) = \int_{-\infty}^{+\infty} g(x)h(z-x)dx,$$

lo que está de acuerdo con la ecuación (12.6).

(e) Puesto que la distribución de $X + Y$ debería posiblemente ser la misma que la distribución de $Y + X$, podríamos verificar que $\int_{-\infty}^{+\infty} g(x)h(z-x)dx = \int_{-\infty}^{+\infty} h(y)g(z-y)dy$. Haciendo sencillamente $z-x=y$ en la primera integral, producirá la segunda forma. Algunas veces indicamos esta propiedad al escribir $g * h = h * g$. Ver la observación (a).

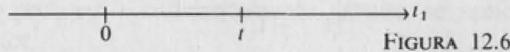


FIGURA 12.6

EJEMPLO 12.9. Se consideran dos instrumentos electrónicos, D_1 y D_2 . Supóngase que D_1 tiene una duración que puede representarse con una variable aleatoria T_1 que tiene una distribución exponencial con parámetro α_1 . Mientras que D_2 tiene una duración que puede representarse con una variable aleatoria T_2 que tiene una distribución exponencial con parámetro α_2 . Suponiendo que D_1 y D_2 están conectadas de tal manera que D_2 empieza a funcionar en el momento que D_1 deja de funcionar. Entonces $T = T_1 + T_2$ representa el tiempo total que el sistema formado por los dos instrumentos está funcionando. Suponiendo T_1 y T_2 son independientes, podemos aplicar el resultado anterior para obtener

$$g(t_1) = \alpha_1 e^{-\alpha_1 t_1}, \quad t_1 \geq 0,$$

$$h(t_2) = \alpha_2 e^{-\alpha_2 t_2}, \quad t_2 \geq 0.$$

(¡Para todos los otros valores de t_1 y t_2 , las funciones g y h se suponen igual a cero!). Luego usando la ecuación (12.6) encontramos que la fdp de $T_1 + T_2 = T$ está dada por

$$s(t) = \int_{-\infty}^{+\infty} g(t_1)h(t - t_1) dt_1, \quad t \geq 0.$$

El integrando es positivo si y sólo si ambos factores del integrando son positivos; es decir, cada vez $t_1 \geq 0$ y $t - t_1 \geq 0$. Esto es equivalente a $t_1 \geq 0$ y $t_1 \leq t$, lo cual a su vez es equivalente a $0 \leq t_1 \leq t$. (Ver la figura 12.6.) Luego la integral anterior llega a ser

$$\begin{aligned} s(t) &= \int_0^t \alpha_1 e^{-\alpha_1 t_1} \alpha_2 e^{-\alpha_2(t-t_1)} dt_1 \\ &= \alpha_1 \alpha_2 e^{-\alpha_2 t} \int_0^t e^{-t_1(\alpha_1 - \alpha_2)} dt_1 \\ &= \frac{\alpha_1 \alpha_2}{\alpha_2 - \alpha_1} (e^{-t\alpha_1} - e^{-t\alpha_2}), \quad \text{para } t \geq 0. \end{aligned}$$

Observaciones: (a) Nótese que la suma de dos variables aleatorias independientes, distribuidas exponencialmente, no está distribuida exponencialmente.

(b) Para $\alpha_1 > \alpha_2$, el gráfico de fdp de T se muestra en la figura 12.7.

(c) La expresión anterior para la fdp no está definida para $\alpha_1 = \alpha_2$, es decir, para el caso que T_1 y T_2 tengan la misma distribución exponencial. A fin de cuidar de este caso especial, consideremos la primera integral de $s(t)$ y hagamos $\alpha = \alpha_1 = \alpha_2$. Obtenemos

$$\begin{aligned} s(t) &= \int_0^t \alpha e^{-\alpha t_1} \alpha e^{-\alpha(t-t_1)} dt_1 \\ &= \alpha^2 e^{-\alpha t} \int_0^t dt_1 = \alpha^2 t e^{-\alpha t}, \quad t \geq 0. \end{aligned}$$

Esta representa a una distribución Gama (ver la ecuación (9.16)). El gráfico de esta fdp aparece en la figura 12.8. El máximo ocurre para $t = 1/\alpha = E(T_1) = E(T_2)$.

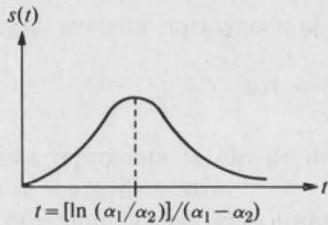


FIGURA 12.7

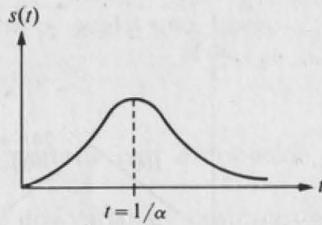


FIGURA 12.8

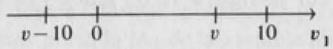
EJEMPLO 12.10. Reconsiderando el ejemplo 12.6, relacionado con la suma de dos voltajes aleatorios independientes V_1 y V_2 , cada uno de los cuales está distribuido uniformemente en $[0, 10]$. Luego

$$\begin{aligned} f(v_1) &= \frac{1}{10}, \quad 0 \leq v_1 \leq 10, \\ g(v_2) &= \frac{1}{10}, \quad 0 \leq v_2 \leq 10. \end{aligned}$$

(Recuérdese nuevamente que las funciones f y g son cero para cualquier otro valor). Si $V = V_1 + V_2$, tenemos

$$s(v) = \int_{-\infty}^{+\infty} f(v_1)g(v - v_1) dv_1.$$

Razonando como en el ejemplo 12.8, observamos que el integrando no es cero sólo para aquellos valores de v_1 que satisfacen $0 \leq v_1 \leq 10$ y $0 \leq v - v_1 \leq 10$. Esas condiciones son equivalentes a $0 \leq v_1 \leq 10$ y a $v - 10 \leq v_1 \leq v$.



(a)

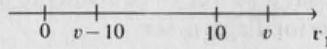


FIGURA 12.9

(v)

Existen dos casos, como se indicó en la figura 12.9.

- (a) $v - 10 \leq 10$ y $0 \leq v \leq 10$ que juntas implican que $0 \leq v \leq 10$.
- (b) $0 \leq v - 10 \leq 10$ y $v \geq 10$ que juntas implican que $10 \leq v \leq 20$.

En el caso (a), v_1 puede tomar los valores entre 0 y v , mientras que en el caso (b), v_1 puede tomar los valores entre $v - 10$ y 10.

Así obtenemos

$$\text{para } 0 \leq v \leq 10: \quad s(v) = \int_0^v \frac{1}{10} \frac{1}{10} dv_1 = \frac{v}{100},$$

$$\text{para } 10 \leq v \leq 20: \quad s(v) = \int_{v-10}^{10} \frac{1}{10} \frac{1}{10} dv_1 = \frac{20-v}{100}$$

Luego la fdp de V tiene el gráfico que se muestra en la figura 12.10.

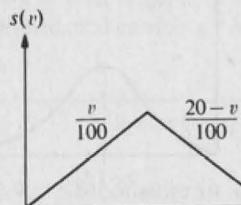


FIGURA 12.10

EJEMPLO 12.11. Como una ilustración final de las sumas de variables aleatorias, reconsideraremos un resultado que ya hemos probado, usando el método más indirecto de las funciones generadoras de momentos (ver el ejemplo 10.11), esto es, la suma de dos variables aleatorias normales e independientes está nuevamente distribuida normalmente. A fin de evitar algunas manipulaciones algebraicas complicadas, consideremos sólo un caso especial.

Suponiendo que $Z = X + Y$, en donde X y Y son variables aleatorias independientes, cada una con distribución $N(0, 1)$. Entonces

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad -\infty < x < \infty,$$

$$g(y) = \frac{1}{\sqrt{2\pi}} e^{-y^2/2}, \quad -\infty < y < \infty.$$

Luego la fdp de Z está dada por

$$\begin{aligned} s(z) &= \int_{-\infty}^{+\infty} f(x)g(z-x)dx = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-x^2/2} e^{-(z-x)^2/2} dx \\ &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-(1/2)[x^2+z^2-2zx+x^2]} dx \\ &= \frac{1}{2\pi} e^{-z^2/2} \int_{-\infty}^{+\infty} e^{-(x^2-zx)} dx \end{aligned}$$

Completando el cuadrado en el exponente del integrando, obtenemos

$$(x^2 - zx) = \left[\left(x - \frac{z}{2} \right)^2 - \frac{z^2}{4} \right].$$

Luego

$$s(z) = \frac{1}{2\pi} e^{-z^2/2} e^{z^2/4} \int_{-\infty}^{+\infty} e^{-(1/2)[\sqrt{2}(x-z/2)]^2} dx.$$

Sea $\sqrt{2}(x-z/2) = u$; entonces $dx = du/\sqrt{2}$ y obtenemos

$$s(z) = \frac{1}{\sqrt{2\pi}\sqrt{2}} e^{-z^2/4} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-u^2/2} du.$$

La integral anterior (incluyendo el factor $1/\sqrt{2\pi}$) es igual a uno. Luego

$$s(z) = \frac{1}{\sqrt{2\pi}\sqrt{2}} e^{-(1/2)(z/\sqrt{2})^2}.$$

Pero ésta representa la fdp de una variable aleatoria con distribución $N(0, 2)$ lo que se iba a demostrar.

Al presentar la distribución de la suma de dos variables aleatorias independientes, nos hemos reducido a variables aleatorias continuas. En el caso discreto el problema es algo más sencillo, al menos, en ciertos casos, como lo indica el teorema siguiente.

Teorema 12.3. Supóngase que X y Y son variables aleatorias independientes cada una de las cuales puede tomar sólo valores enteros no negativos. Sea $p(k) = P(X = k)$, $k = 0, 1, 2, \dots$ y sea $q(r) = P(Y = r)$, $r = 0, 1, 2, \dots$ Sea $Z = X + Y$ y sea $w(i) = P(Z = i)$. Entonces

$$w(i) = \sum_{k=0}^i p(k)q(i-k), \quad i = 0, 1, 2, \dots$$

Demostración:

$$\begin{aligned} w(i) &= P(Z = i) \\ &= P[X = 0, Y = i \text{ o } X = 1, Y = i - 1 \text{ o } \dots \text{ o } X = i, Y = 0] \\ &= \sum_{k=0}^i P[X = k, Y = i - k] = \sum_{k=0}^i p(k)q(i - k) \end{aligned}$$

puesto que X y Y son independientes

Observación: nótese la similitud entre esta suma y la integral de convolución derivada en el teorema 12.2.

EJEMPLO 12.12. X y Y representan al número de emisiones de partículas α de dos fuentes de material radiactivo durante un período de tiempo especificado de duración t . Suponiendo que X y Y tengan distribuciones de Poisson con parámetros β_1t y β_2t , respectivamente. $Z = X + Y$ representa el número total de partículas α emitidas por las dos fuentes. Usando el teorema 12.3, obtenemos

$$\begin{aligned} P(Z = k) &= \sum_{k=0}^i p(k)q(i - k) \\ &= e^{-(\beta_1 + \beta_2)t} \sum_{k=0}^i \frac{(\beta_1 t)^k (\beta_2 t)^{i-k}}{k!(i-k)!} \\ &= e^{-(\beta_1 + \beta_2)t} \frac{(\beta_1 t + \beta_2 t)^i}{i!}. \end{aligned}$$

(La última igualdad se obtiene aplicando el teorema del binomio a la suma anterior). La última expresión representa la probabilidad que una variable aleatoria que tiene una distribución de Poisson con parámetro $\beta_1t + \beta_2t$ tome el valor i . Así hemos verificado lo que ya sabíamos: la suma de dos variables aleatorias independientes de Poisson tiene una distribución de Poisson.

PROBLEMAS

12.1. (a) Una fábrica produce determinados artículos de tal manera que el 2 por ciento resulta defectuoso. Se inspecciona un gran número de tales artículos, n , y se anota la frecuencia relativa de defectuosos, f_D . ¿Cuál debería ser el tamaño de n a fin de que la probabilidad sea 0,98 de que f_D difiera de 0,02 en menos de 0,05?

(b) Conteste la (a) si 0,02, la probabilidad de tener un artículo defectuoso se sustituye por p que se supone desconocida.

12.2. Supóngase que se obtiene una muestra de tamaño de un gran conjunto de pernos, el 3 por ciento de los cuales son defectuosos. ¿Cuál es la probabilidad de que a lo más el 5 por ciento de los pernos elegidos sea defectuoso si:

- (a) $n = 6$? (b) $n = 60$? (c) $n = 600$?

12.3. (a) Un sistema compuesto está formado por 100 componentes que funcionan independientemente. La probabilidad de que cualquier componente falle durante el período de operación es igual a 0,10. A fin que funcione el sistema completo, al menos deben funcionar 85 componentes. Calcular esta probabilidad.

(b) Supóngase que el sistema anterior esté formado por n componentes, cada una con una confiabilidad de 0,90. El sistema funcionará si al menos el 80 por ciento de las componentes funciona correctamente. Determinar n de modo que el sistema tenga una confiabilidad de 0,95.

12.4. Supóngase que 30 instrumentos electrónicos, D_1, \dots, D_{30} , se usan de la manera siguiente: tan pronto como D_1 falla empieza a actuar D_2 . Cuando D_2 falla empieza a actuar D_3 , etc. Supóngase que el tiempo para fallar de D_i es una variable aleatoria distribuida exponencialmente con parámetro $\beta = 0,1$ hora $^{-1}$. Sea T el tiempo total de operación de los 30 instrumentos. ¿Cuál es la probabilidad de que T exceda 350 horas?

12.5. Al sumar números, un computador aproxima cada número al entero más próximo. Supongamos todos los errores de aproximación son independientes y distribuidos uniformemente entre $(-0,5, 0,5)$.

(a) Si se suman 1500 números, ¿cuál es la probabilidad de que la magnitud del error total exceda 15?

(b) ¿Cuántos números pueden sumarse juntos a fin de que la magnitud del error total sea menor que 10, con probabilidad 0,90?

12.6. Supóngase que $X_i, i = 1, 2, \dots, 50$, son variables aleatorias independientes que tienen cada una distribución de Poisson con parámetro $\lambda = 0,03$. Sea $S = X_1 + \dots + X_{50}$.

(a) Usando el teorema del límite central, calcular $P(S \geq 3)$.

(b) Comparar la respuesta en (a) con el valor exacto de esta probabilidad.

12.7. En un circuito simple se conectan dos resistencias R_1 y R_2 en serie. Por tanto, la resistencia total está dada por $R = R_1 + R_2$. Supóngase que R_1 y R_2 son variables aleatorias independientes que tiene cada una la fdp

$$f(r_i) = \frac{10 - r_i}{50}, \quad 0 < r_i < 10, \quad i = 1, 2.$$

Encontrar la fdp de R , la resistencia total y dibujar su gráfico.

12.8. Supóngase que las resistencias en el problema 12.7 están conectadas en paralelo. Encontrar la fdp de R , la resistencia total del circuito (establezca sólo, en forma de integral). [Indicación: la relación entre R , y R_1 , y R_2 está dada por $1/R = 1/R_1 + 1/R_2$.]

12.9. Al medir T , la duración de un artículo, se puede cometer un error que puede suponerse distribuido uniformemente entre $(-0,01, 0,01)$. Luego el tiempo anotado (en horas) se puede representar como $T + X$, en donde T tiene una distribución exponencial con parámetro 0,2 y X tiene la distribución uniforme ya descrita. Si T y X son independientes, encontrar la fdp de $T + X$.

12.10. Supóngase que X y Y son variables aleatorias independientes distribuidas idénticamente y que la fdp de X (y por tanto de Y) está dada por

$$\begin{aligned} f(x) &= a/x^2, & x > a, & a > 0, \\ &= 0, & \text{para cualquier otro valor.} \end{aligned}$$

Encuentre la fdp de $X + Y$. [Indicación: use fracciones parciales para la integración.]

12.11. Realizar los cálculos sugeridos al final del ejemplo 12.5.

12.12. (a) Un instrumento tiene un tiempo de falla T cuya distribución está dada por $N(100,4)$. Supóngase que al anotar T se comete un error cuyo valor puede representarse como una variable aleatoria X distribuida uniformemente en $(-1, 1)$. Si X y T son independientes, obtener la fdp de $S = X + T$ mediante Φ , la fdp de la distribución $N(0,1)$.

(b) Calcular $P(100 \leq S \leq 101)$. [Indicación: use la regla de Simpson para aproximar la integral.]

12.13. Supóngase que un nuevo aparato se prueba repetidamente bajo ciertas condiciones de esfuerzo hasta que falla. La probabilidad de fallar en cualquier ensayo es p_1 . Sea X igual al número de ensayos necesarios hasta incluir la primera falla. También se prueba un segundo aparato hasta que falla. Supóngase que una probabilidad constante de falla p_2 está asociada con él. Sea Y igual al número de ensayos necesarios hasta incluir su primera falla. Supóngase que X y Y son independientes y haga $Z = X + Y$. Por tanto, Z es igual al número de ensayos necesarios hasta que ambos aparatos hayan fallado.

(a) Encontrar la distribución de probabilidad de Z .

(b) Calcular $P(Z = 4)$ si $p_1 = 0,1$, $p_2 = 0,2$.

(c) Discutir (a) si $p_1 = p_2$.

Muestras y distribuciones muestrales

13.1 Introducción

Consideremos nuevamente un problema expuesto previamente. Supóngase que tenemos una fuente de material radiactivo que emite partículas α y que las hipótesis establecidas en el capítulo 8 son válidas, así que la variable aleatoria X definida como el número de partículas emitidas durante un período de tiempo especificado t tiene una distribución de Poisson con parámetro λt .

A fin de «usar» este modelo probabilístico para describir la emisión de partículas α necesitamos conocer el valor de λ . Las hipótesis que formulamos sólo conducen a la conclusión de que X tiene una distribución de Poisson con un parámetro λt . Pero si deseamos calcular $P(X > 10)$, por ejemplo, la respuesta será mediante λ a no ser que conozcamos su valor numérico. Análogamente, los parámetros importantes asociados con la distribución tales como $E(X)$ y $V(X)$ son funciones de λ .

Para buscar un valor numérico de λ debemos dejar el mundo de nuestros modelos matemáticos teóricos y entrar al mundo de las observaciones, al menos por ahora. Es decir, en este instante debemos observar la emisión de las partículas, obtener los valores numéricos de X y luego utilizar esos valores de una manera sistemática a fin de obtener una información atinada de λ .

Es importante para el lector tener una idea precisa acerca de la relación entre la verificación empírica y la deducción matemática que aparece en muchas áreas de matemáticas aplicadas. Esto es muy importante cuando construimos modelos probabilísticos para el estudio de los fenómenos observables.

Consideremos un ejemplo trivial de la trigonometría elemental. Un problema típico puede implicar el cálculo de la altura de un árbol. Un modelo matemático para este problema podría obtenerse al postular que la relación entre la altura desconocida h , el largo de la sombra s , y el ángulo α es de la forma $h = s \operatorname{tg} \alpha$. (Suponemos que el árbol permanece derecho y perpendicular al suelo, figura 13.1.)

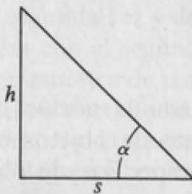


FIGURA 13.1

Por tanto, si s y α son conocidos, con ayuda de una tabla apropiada, podemos calcular h . El hecho importante que aquí formulamos es que s y α deben ser *conocidos* antes de que podamos calcular h . Es decir, alguien debe haber medido s y α . La deducción matemática que conduce a la relación $h = s \operatorname{tg} \alpha$ es completamente independiente de los medios mediante los cuales medimos s y α . Si esas mediciones son exactas, entonces $s \operatorname{tg} \alpha$ representará un valor exacto de h (suponiendo que el modelo es válido). Diciéndolo en forma diferente, no podemos deducir simplemente el valor de h con nuestros conocimientos de trigonometría y con la ayuda de las tablas trigonométricas. ¡Debemos dejar nuestro santuario (cuálquiera que sea) y hacer algunas mediciones! Y, precisamente, cómo se hicieron esas mediciones, de ninguna manera influyen en la *validez* de nuestra deducción matemática, aunque deba resolverse un problema importante.

Al usar los modelos probabilísticos necesitaremos entrar nuevamente al mundo empírico y hacer algunas mediciones. Por ejemplo, en el caso que consideramos se usa la distribución de Poisson como modelo probabilístico y por tanto necesitamos conocer el valor del parámetro λ . A fin de obtener alguna información acerca de λ debemos hacer algunas mediciones y luego usar esas mediciones de una manera sistemática con el objeto de calcular λ . En el capítulo 14 describiremos la manera de hacer este cálculo.

Finalmente, debemos insistir aquí en dos puntos. Primero, las mediciones necesarias para obtener información respecto a λ serán generalmente más fáciles de obtener que las que producirían mediciones para $e^{-\lambda t}(\lambda t)^k/k!$ (tal como es más fácil obtener mediciones para el largo de la sombra s y el ángulo α que para la altura h). Segundo, la manera como obtenemos mediciones para λ y la manera como usamos esas mediciones de ninguna manera invalida (o confirma) la aplicación del modelo de Poisson.

El anterior es un ejemplo típico de una gran cantidad de problemas. En muchos casos es relativamente natural (y apropiado) formular la hipótesis de que una variable aleatoria X tiene una distribución particular de probabilidades. Ya hemos visto un número de ejemplos que indican que hipótesis muy simples acerca de la conducta probabilística de X conducirán a un tipo determinado de distribuciones tales como la binomial, exponencial, normal, Poisson y otras. Cada una de esas distribuciones depende de ciertos parámetros. En algunos casos el valor de uno o más parámetros puede ser conocido. (Tal conocimiento puede provenir debido al estudio previo de las variables aleatorias). Muy a menudo, sin embargo, no conocemos el valor de los parámetros implicados. En tales casos debemos proceder como se sugirió anteriormente y obtener algunos valores empíricos de X y luego usar esos valores de alguna manera apropiada. Veremos cómo se hace esto en el capítulo 14.

13.2 Muestras aleatorias

Previamente hemos presentado la noción de muestreo aleatorio con o sin sustitución de un conjunto finito de objetos o *población* de objetos. Tenemos que considerar una población específica de objetos (personas, artículos manu-

facturados, etc.) acerca de la cual queremos hacer inferencia sin mirar cada objeto. Así «muestreamos». Es decir, tratamos de considerar algunos objetos «típicos» de los cuales esperamos extraer alguna información que de algún modo sea característica de la población completa. Seamos más precisos.

Supóngase que designamos consecutivamente cada uno de los elementos de una población finita, de modo que sin perder generalidad una población que consta de N objetos puede representarse como $1, 2, \dots, N$. Elijamos ahora n artículos, de la manera como vamos a describir a continuación. Definamos la siguiente variable aleatoria.

X_i = valor poblacional obtenido cuando se escoge el i ésimo artículo, $i = 1, 2, \dots, n$.

La distribución de probabilidades de las variables aleatorias X_1, \dots, X_n depende evidentemente de cómo vamos a muestrear. Si el muestreo es con sustitución, cada vez escogemos un objeto al azar, las variables aleatorias son independientes e idénticamente distribuidas. Es decir, para cada X_i , $i = 1, 2, \dots, n$ tenemos

$$P(X_i = j) = 1/N, \quad j = 1, 2, \dots, N.$$

Si el muestreo es sin sustitución, las variables aleatorias X_1, \dots, X_n ya no son independientes. En este caso, su distribución conjunta de probabilidades está dada por

$$P[X_1 = j_1, \dots, X_n = j_n] = \frac{1}{N(N-1)\cdots(N-n+1)},$$

en donde j_1, \dots, j_n son n valores cualesquiera de $(1, \dots, N)$. (Podemos demostrar que la distribución marginal de cualquier X_i , independientemente de los valores tomados por $X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n$, es la misma que la anterior cuando el muestreo se hace con sustitución).

En nuestra exposición hasta ahora, hemos supuesto que existe una población principal, $1, 2, \dots, N$ que es finita y acerca de la cual queremos tener una información basada en una muestra de tamaño $n < N$. En muchos casos no hay una población finita de la cual obtengamos muestras; en realidad podemos tener dificultades al definir una población principal de cualquier clase. Consideremos los ejemplos siguientes.

(a) Se lanza una moneda. Definamos la variable aleatoria X_1 = número de caras obtenidas. Bajo un solo aspecto podemos pensar que X_1 es una muestra de tamaño uno de la «población» de todos los lanzamientos posibles de esa moneda. Si lanzamos la moneda una segunda vez y definimos la variable aleatoria X_2 como el número de caras obtenidas con el segundo lanzamiento, X_1, X_2 posiblemente pudo considerarse como una muestra de tamaño dos de la *misma* población.

(b) La precipitación total anual en cierta localidad durante el año 1970 podría definirse como una variable aleatoria X_1 . Durante los años siguientes las variables aleatorias X_2, \dots, X_n pudieron definirse análogamente. Podemos considerar

nuevamente (X_1, \dots, X_n) como una muestra de tamaño n , obtenida de la población de todas las precipitaciones anuales posibles en esa localidad específica y podría suponerse realísticamente que las X_i son variables aleatorias independientes e idénticamente distribuidas.

(c) La duración de una bombilla fabricada mediante cierto procedimiento en una fábrica se estudia escogiendo n bombillas y midiendo su duración, T_1, \dots, T_n . Podemos considerar (T_1, \dots, T_n) como una muestra aleatoria de la población de todas las duraciones posibles de las bombillas fabricadas de esa manera específica.

Formalicemos esas nociones como sigue.

Definición. Sea X una variable aleatoria con cierta distribución de probabilidades. Sean X_1, \dots, X_n n variables aleatorias independientes que tiene cada una la misma distribución que X . Llamamos entonces a (X_1, \dots, X_n) una muestra aleatoria de la variable aleatoria X .

Observaciones: (a) Establezcamos más informalmente lo anterior: una muestra aleatoria de tamaño n de una variable aleatoria X corresponde a n mediciones repetidas de X , hechas básicamente bajo las mismas condiciones. Como lo hemos dicho anteriormente en otros párrafos, la noción matemáticamente idealizada de una muestra aleatoria se puede aproximar mejor sólo por las condiciones experimentales reales. A fin de que X_1 y X_2 tengan la misma distribución, todas las condiciones «relevantes» bajo las cuales se realiza el experimento deben ser las mismas cuando se observa X_1 que cuando se observa X_2 . Las condiciones experimentales nunca se pueden duplicar idénticamente, por supuesto. Lo importante es que esas condiciones, que son diferentes, deben tener poco o ningún efecto en el resultado del experimento. Sin embargo, algún cuidado deberá tenerse para asegurarnos de que realmente obtengamos una muestra aleatoria.

Por ejemplo, supóngase que la variable aleatoria que se considera es X , el número de llamadas que llegan a una central telefónica entre las 4 p.m. y las 5 p.m. el miércoles. A fin de obtener una muestra aleatoria de X , posiblemente deberíamos elegir n miércoles al azar y anotar el valor de X_1, \dots, X_n . Tendríamos que estar seguros de que todos los miércoles son miércoles «típicos». Por ejemplo, podríamos no incluir un miércoles particular si coincide con Navidad.

Nuevamente, tratemos de obtener una muestra aleatoria de la variable aleatoria X definida como la duración de un instrumento electrónico que se fabrica con determinadas especificaciones, desecharíamos asegurarnos de que no se ha obtenido un valor muestral de un artículo producido durante un momento en que el proceso de producción estaba fallando.

(b) Si X es una variable aleatoria continua con fdp f y si X_1, \dots, X_n es una muestra aleatoria de X , entonces g , la fdp conjunta de X_1, \dots, X_n , puede escribirse como $g(x_1, \dots, x_n) = f(x_1) \cdots f(x_n)$. Si X es una variable aleatoria discreta y $p(x_i) = P(X = x_i)$ entonces

$$P[X_1 = x_1, \dots, X_n = x_n] = p(x_1) \cdots p(x_n).$$

(c) Tal como lo hicimos anteriormente, usaremos letras mayúsculas para las variables aleatorias y letras minúsculas para el valor de la variable aleatoria. Luego los valores que toma una muestra (X_1, \dots, X_n) se denotarán por (x_1, \dots, x_n) . A menudo hablaremos del punto muestral (x_1, \dots, x_n) . Con esto indicaremos simplemente que consideraremos a (x_1, \dots, x_n) como las coordenadas de un punto en un espacio euclíadiano n dimensional.

13.3 Estadígrafos*

Una vez que hemos obtenido los valores de una muestra, habitualmente queremos usar esos valores muestrales con el objeto de hacer alguna inferencia de la población representada por la muestra que en el párrafo actual representa la distribución de probabilidades de la variable aleatoria que se está muestreando. Puesto que los diversos parámetros que caracterizan una distribución de probabilidades son números, es apenas natural que queramos calcular ciertas características numéricas específicas que se obtienen de los valores muestrales, lo que nos podría servir para hacer proposiciones apropiadas acerca de los valores de los parámetros que a menudo no son conocidos. Definamos el siguiente concepto importante.

Definición. Sea X_1, \dots, X_n una muestra aleatoria de una variable aleatoria X y sean x_1, \dots, x_n los valores tomados por la muestra. Sea H una función definida para el n tuplo (x_1, \dots, x_n) . Se define que $Y = H(X_1, \dots, X_n)$ sea un *estadígrafo*, que toma el valor $y = H(x_1, \dots, x_n)$.

En palabras: un estadígrafo es una función real de la muestra. Algunas veces usamos el término estadígrafo para referirnos al valor de la función. Luego podemos hablar del estadígrafo $y = H(x_1, \dots, x_n)$ cuando realmente deberíamos decir que y es el valor del estadígrafo $Y = H(X_1, \dots, X_n)$.

Observación: de acuerdo con la definición anterior, ¡un estadígrafo es una variable aleatoria! Es muy importante recordar esto. Será útil ahora considerar la distribución de probabilidades de un estadígrafo, su esperanza y su varianza. Cuando una variable aleatoria es realmente un estadígrafo, es decir, una función de una muestra, hablamos a menudo de su *distribución muestral* en vez de su distribución de probabilidades.

Como lo sugerimos al comienzo de este capítulo, usaremos la información obtenida de una muestra con el objeto de estimar ciertos parámetros desconocidos asociados con una distribución de probabilidades. Encontraremos que ciertos estadígrafos desempeñan un papel importante en la solución de este problema. Antes de que consideremos esto con más detalles (en el capítulo 14), estudiemos algunos estadígrafos importantes y sus propiedades.

13.4 Algunos estadígrafos importantes

Hay ciertos estadígrafos que se encuentran frecuentemente. Anotaremos algunos de ellos y expondremos algunas de sus propiedades importantes.

Definición. Sea (X_1, \dots, X_n) una muestra aleatoria de la variable aleatoria X . Los siguientes estadígrafos son de interés.

(a) $\bar{X} = (1/n) \sum_{i=1}^n X_i$ se llama el *promedio muestral*.

* N. del T. El autor emplea la palabra *statistics*, que hemos traducido por estadígrafo por ser la más comúnmente empleada en español.

- (b) $S^2 = [1/(n-1)] \sum_{i=1}^n (X_i - \bar{X})^2$ se llama la varianza muestral. (Explicaremos brevemente por qué dividimos por $(n-1)$ en vez de elegir simplemente n).
- (c) $K = \min(X_1, \dots, X_n)$ se llama el mínimo de la muestra. (K representa simplemente el valor observado más pequeño).
- (d) $M = \max(X_1, \dots, X_n)$ se llama el máximo de la muestra. (M representa el mayor valor observado).
- (e) $R = M - K$ se llama el recorrido muestral.
- (f) $X_n^{(j)} = j$ -ésima observación mayor en la muestra, $j = 1, 2, \dots, n$. (Tenemos que $X_n^{(1)} = M$ mientras que $X_n^{(n)} = K$).

Observaciones: (a) Las variables aleatorias $X_n^{(j)}, j = 1, 2, \dots, n$, se llaman estadígrafos de orden asociados con la variable aleatoria X_1, \dots, X_n . Si X es una variable aleatoria continua podemos suponer que $X_n^{(1)} > X_n^{(2)} > \dots > X_n^{(n)}$.

(b) Los valores extremos de la muestra (en la notación anterior, $X_n^{(1)}$ y $X_n^{(n)}$) son a menudo de interés considerable. Por ejemplo, en la construcción de represas para controlar inundaciones, la mayor altura que ha alcanzado un río en los 50 años anteriores puede ser muy importante.

Naturalmente hay muchos otros estadígrafos importantes como los que encontramos, pero evidentemente los mencionados desempeñan un papel importante en muchas aplicaciones estadísticas. Estableceremos (y demostraremos) algunos teoremas que se refieren a los estadígrafos anteriores.

Teorema 13.1. Sea X una variable aleatoria con esperanza $E(X) = \mu$ y varianza $V(X) = \sigma^2$. Sea \bar{X} el promedio muestral de una muestra aleatoria de tamaño n . Entonces

- $E(\bar{X}) = \mu$.
- $V(\bar{X}) = \sigma^2/n$.
- Para n grande, $(\bar{X} - \mu)/(\sigma/\sqrt{n})$ tiene aproximadamente la distribución $N(0, 1)$.

Demostración: (a) y (b) se deducen inmediatamente de las propiedades de la esperanza y de la varianza establecidas anteriormente:

$$E(\bar{X}) = E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \frac{1}{n} n\mu = \mu.$$

Puesto que los X_i son independientes

$$V(\bar{X}) = V\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} \sum_{i=1}^n V(X_i) = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}.$$

(c) se deduce de una aplicación directa del teorema del límite central. Podemos escribir $\bar{X} = (1/n)X_1 + \dots + (1/n)X_n$ como la suma de variables aleatorias independientemente distribuidas.

Observaciones: (a) Cuando el tamaño de muestra aumenta, el promedio muestral \bar{X} tiende a variar menos y menos. Esto es intuitivamente evidente y corresponde a nuestra experiencia con datos numéricos. Considérese, por ejemplo, el conjunto siguiente de 18 números

$$-1, 3, 2, -4, -5, 6, 7, 2, 0, 1, -2, -3, 8, 9, 6, -3, 0, 5.$$

Si calculamos el promedio de esos números, tomando dos a la vez en el orden anotado, obtenemos el siguiente conjunto de promedios:

$$1, -1, 0.5; 4.5; 0.5, -2.5; 8.5; 1.5; 2.5.$$

Si promediámos el conjunto original de números, tomando tres a la vez, obtenemos

$$1.3, -1.3, -1.3, 7, 7, 0.7.$$

Finalmente, si promediámos los números, tomando seis a la vez obtenemos

$$0.2, 0.8, 4.1.$$

La varianza en cada uno de esos conjuntos de promedios es menor que en el conjunto anterior, debido a que en cada caso el promedio está basado en un número mayor de valores. El teorema anterior indica precisamente cómo la variación de \bar{X} (medida mediante su varianza) disminuye cuando aumenta n . (En relación con esto véase la ley de los grandes números, sección 12.2 y en particular la ecuación 12.2).

(b) Si n no es suficientemente grande para asegurar la aplicación del Teorema del límite central, podemos tratar de encontrar la distribución exacta de probabilidades de \bar{X} por un medio directo (pero generalmente más complicado). En la sección 12.5, sugerimos un método mediante el cual podemos encontrar la distribución de probabilidades de la suma de variables aleatorias. Con una aplicación repetida de este método podemos obtener la distribución de probabilidades de \bar{X} , especialmente si n es relativamente pequeño.

(c) El teorema 13.1 indica que para un n suficientemente grande, el promedio muestral \bar{X} está distribuido aproximadamente normal (con esperanza μ y varianza σ^2/n).

Encontramos que no sólo \bar{X} sino la mayor parte de las funciones «con buen comportamiento» de \bar{X} tienen esta propiedad. En este nivel de presentación no podemos dar un desarrollo más explicativo de este resultado. Sin embargo, el resultado es de mucha importancia en muchas aplicaciones para sostener a lo menos un argumento heurístico, intuitivo.

Supóngase que $Y = r(\bar{X})$ y que r puede desarrollarse en una serie de Taylor respecto a μ . Luego $r(\bar{X}) = r(\mu) + (\bar{X} - \mu)r'(\mu) + R$, en donde R es el término del resto y puede expresarse como $R = [(\bar{X} - \mu)^2/2]r''(z)$, en donde z es un valor comprendido entre \bar{X} y μ . Si n es suficientemente grande, \bar{X} estará próxima a μ , y por tanto $(\bar{X} - \mu)^2$ será pequeña comparada con $(\bar{X} - \mu)$. Para n grande, podemos por tanto considerar que el resto es despreciable y *aproximar* $r(\bar{X})$ como sigue:

$$r(\bar{X}) \simeq r(\mu) + r'(\mu)(\bar{X} - \mu).$$

Vemos que para un n suficientemente grande, $r(\bar{X})$ se puede aproximar por una función lineal de \bar{X} . Puesto que \bar{X} será aproximadamente normal (para un n grande), encontramos que $r(\bar{X})$ también será aproximadamente normal, puesto que una función lineal de una variable aleatoria distribuida normalmente está también distribuida normalmente.

De la expresión anterior de $r(\bar{X})$ encontramos que

$$E(r(\bar{X})) \simeq r(\mu), \quad V(r(\bar{X})) \simeq \frac{[r'(\mu)]^2 \sigma^2}{n}.$$

Así para un n suficientemente grande, vemos que la distribución de $r(\bar{X})$ es aproximadamente $N(r(\mu), [r'(\mu)]^2 \sigma^2/n)$ (bajo condiciones muy generales de la función r).

Teorema 13.2. Sea X una variable aleatoria continua con fdp f y fda F .

Sea X_1, \dots, X_n una muestra aleatoria de X y sean K y M el mínimo y el máximo de la muestra respectivamente. Luego:

- (a) la fdp de M está dada por $g(m) = n[F(m)]^{n-1}f(m)$,
- (b) la fdp de K está dada por $h(k) = n[1 - F(k)]^{n-1}f(k)$.

Demostración: sea $G(m) = P(M \leq m)$ la fda de M . Ahora $\{M \leq m\}$ es equivalente al suceso $\{X_i \leq m, \text{ para todo } i\}$. Luego puesto que los X_i son independientes, encontramos

$$G(m) = P[X_1 \leq m \text{ y } X_2 \leq m \cdots \text{ y } X_n \leq m] = [F(m)]^n.$$

Por tanto

$$g(m) = G'(m) = n[F(m)]^{n-1}f(m).$$

La obtención de la fdp de K se deja al lector. (Ver el problema 13.1).

EJEMPLO 13.1. Un instrumento electrónico tiene una duración T que está distribuida exponencialmente con parámetro $\alpha = 0,001$; es decir, su fdp es $f(t) = 0,001e^{-0,001t}$. Supóngase que se prueban 100 de tales instrumentos, que dan valores observados T_1, \dots, T_{100} .

(a) ¿Cuál es la probabilidad de que $950 < \bar{T} < 1100$? Puesto que el tamaño de muestra es muy grande, podemos aplicar el teorema del límite central y proceder como sigue:

$$E(\bar{T}) = \frac{1}{0,001} = 1000, \quad V(\bar{T}) = \frac{1}{100}(0,001)^{-2} = 10.000.$$

Como $(\bar{T} - 1000)/100$ tiene aproximadamente la distribución $N(0,1)$. Por tanto

$$\begin{aligned} P(950 < \bar{T} < 1100) &= P\left(-0,5 < \frac{\bar{T} - 1000}{100} < 1\right) \\ &= \Phi(1) - \Phi(-0,5) \\ &= 0,532. \end{aligned}$$

de las tablas de la distribución normal.

Observación: en el caso presente podemos obtener inmediatamente la distribución exacta de \bar{T} sin recurrir al teorema del límite central. Demostramos en el teorema 10.9 que la suma de variables aleatorias independientes distribuidas exponencialmente tiene una distribución gama; es decir,

$$g(s) = \frac{(0,001)^{100} s^{99} e^{-0,001s}}{99!},$$

en donde g es la fdp de $T_1 + \dots + T_{100}$. Luego la fdp de \bar{T} está dada por

$$f(\bar{t}) = \frac{(0,1)^{100} \bar{t}^{99} e^{-0,1\bar{t}}}{99!}.$$

Luego \bar{T} tiene una distribución Gama con parámetros 0,1 y 100.

(b) ¿Cuál es la probabilidad de que el mayor valor observado sobrepase 7200 horas? Necesitamos que $P(M > 7200) = 1 - P(M \leq 7200)$. El valor máximo será menor que 7200 si y sólo si cada valor de muestra es menor que 7200. Luego

$$P(M > 7200) = 1 - [F(7200)]^{100}.$$

Para calcular $F(7200)$, recordemos que la variable aleatoria distribuida exponencialmente con parámetro 0,001, $F(t) = 1 - e^{-0,001t}$. Luego,

$$F(7200) = 1 - e^{0,001(7200)} = 1 - e^{-7,2} = 0,99925.$$

Así la probabilidad pedida es $1 - (0,99925)^{100}$, que es igual a 0,071.

(c) ¿Cuál es la probabilidad de que el tiempo más corto para fallar sea menor que 10 horas? Pedimos que $P(K < 10) = 1 - P(K \geq 10)$.

El mínimo de la muestra ahora es mayor que o igual a 10 si y sólo si cada valor muestral es mayor que o igual a 10. Luego

$$P(K < 10) = 1 - [1 - F(10)]^{100}.$$

Usando la expresión de F como se dio en el anterior (b), tenemos

$$1 - F(10) = e^{-0,001(10)} = e^{-0,01} = 0,99005.$$

Luego

$$P(K < 10) = 1 - (0,99005)^{100} = 0,63.$$

La última parte del ejemplo anterior puede generalizarse como lo indica el teorema siguiente.

Teorema 13.3. Sea X distribuida exponencialmente con parámetro α y sea (X_1, \dots, X_n) una muestra aleatoria de X . Sea $K = \min(X_1, \dots, X_n)$.

Entonces K está también distribuida exponencialmente con parámetro $n\alpha$.

Demostración: sea H la fda de K . Luego

$$H(k) = P(K \leq k) = 1 - P(K > k) = 1 - [1 - F(k)]^n,$$

en donde F es la fda de X . Como $F(x) = 1 - e^{-\alpha x}$. Luego $H(k) = 1 - e^{-n\alpha k}$. Derivando $H(k)$ respecto a k da, $h(k) = n\alpha e^{-n\alpha k}$.

Observación: este teorema puede generalizarse como sigue. Si X_1, \dots, X_n son variables aleatorias independientes y si X_i tiene una distribución exponencial con parámetro α_i , $i = 1, \dots, n$, entonces $K = \min(X_1, \dots, X_n)$ tiene una distribución exponencial con parámetro $\alpha_1 + \dots + \alpha_n$. Para una demostración de esto, ver el problema 13.2.

El teorema 13.4 nos da información acerca de la estadística S^2 .

Teorema 13.4. Suponiendo que X_1, \dots, X_n es una muestra aleatoria de una variable aleatoria X con esperanza μ y varianza σ^2 . Sea

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2,$$

en donde \bar{X} es el promedio muestral. Entonces se tiene lo siguiente:

- (a) $E(S^2) = \sigma^2$,
- (b) Si X está distribuida normalmente, $[(n-1)/\sigma^2]S^2$ tiene una distribución χ^2 cuadrado con $(n-1)$ grados de libertad.

Demostración: (a) Escribiendo

$$\begin{aligned} \sum_{i=1}^n (X_i - \bar{X})^2 &= \sum_{i=1}^n (X_i - \mu + \mu - \bar{X})^2 \\ &= \sum_{i=1}^n [(X_i - \mu)^2 + 2(\mu - \bar{X})(X_i - \mu) + (\mu - \bar{X})^2] \\ &= \sum_{i=1}^n (X_i - \mu)^2 + 2(\mu - \bar{X}) \sum_{i=1}^n (X_i - \mu) + n(\mu - \bar{X})^2 \\ &= \sum_{i=1}^n (X_i - \mu)^2 - 2n(\mu - \bar{X})^2 + n(\mu - \bar{X})^2 \\ &= \sum_{i=1}^n (X_i - \mu)^2 - n(\bar{X} - \mu)^2. \end{aligned}$$

Luego

$$E\left(\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2\right) = \frac{1}{n-1} \left[n\sigma^2 - n\frac{\sigma^2}{n} \right] = \sigma^2.$$

Observación: si hubiésemos dividido por n en vez de $(n-1)$ al definir S^2 , la propiedad anterior no habría sido válida.

(b) No demostraríamos (b) sino que sólo hacemos posible su validez al considerar el caso especial siguiente: Considerando $\sum_{i=1}^n (X_i - \bar{X})^2$ para $n = 2$. Luego

$$\begin{aligned} (X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 &= [X_1 - \frac{1}{2}(X_1 + X_2)]^2 + [X_2 - \frac{1}{2}(X_1 + X_2)]^2 \\ &= \frac{1}{4}[2X_1 - X_1 - X_2]^2 + \frac{1}{4}[2X_2 - X_1 - X_2]^2 \\ &= \frac{1}{4}[(X_1 - X_2)^2 + (X_2 - X_1)^2] = \frac{[X_1 - X_2]^2}{2}. \end{aligned}$$

Puesto que X_1 y X_2 están distribuidos independientemente con distribución $N(\mu, \sigma^2)$, encontramos que $(X_1 - X_2)$ tiene distribución $N(0, 2\sigma^2)$. Luego

$$\left[\frac{X_1 - X_2}{\sqrt{2\sigma}} \right]^2 = \frac{1}{\sigma^2} \sum_{i=1}^2 (X_i - \bar{X})^2 = \frac{1}{\sigma^2} S^2$$

tiene distribución χ cuadrado con un grado de libertad. (Ver el teorema 10.8). La demostración para un n general sigue una línea semejante. Debemos demostrar que $\sum_{i=1}^n (X_i - \bar{X})^2 / \sigma^2$ puede descomponerse en la suma de $(n - 1)$ cuadrados de variables aleatorias independientes, cada una con distribución $N(0, 1)$.

Observación: aunque S^2 se define como la suma de los cuadrados de n términos, esos n términos no son independientes. En realidad, $(X_1 - \bar{X}) + (X_2 - \bar{X}) + \dots + (X_n - \bar{X}) = \sum_{i=1}^n X_i - n\bar{X} = 0$. Luego hay una relación lineal entre los n términos lo que indica que tan pronto como $(n - 1)$ sean conocidos, el n ésimo está determinado.

Finalmente, establezcamos (sin demostración) un resultado que se refiere a la distribución de probabilidades del recorrido de muestra R .

Teorema 13.5. Sea X una variable aleatoria continua con fdp f . Sea $R = M - K$ el recorrido de muestra basado en una muestra aleatoria de tamaño n . Luego la fdp de R está dada por

$$g(r) = n(n - 1) \int_{s=-\infty}^{+\infty} \left[\int_{x=s}^{s+r} f(x) dx \right]^{n-2} f(s)f(s+r) ds, \quad \text{para } r \geq 0.$$

EJEMPLO 13.2. Un voltaje aleatorio V está distribuido uniformemente en $[0, 1]$. Se obtiene una muestra de tamaño n , V_1, \dots, V_n , y se calcula el recorrido muestral R . Encontramos que la fdp de R es

$$g(r) = n(n - 1) \int_{s=-\infty}^{+\infty} r^{n-2} f(s)f(s+r) ds.$$

Tenemos $f(s) = f(s+r) = 1$ para cualquier $0 \leq s \leq 1$ y $0 \leq s+r \leq 1$, las que juntas implican que $0 \leq s \leq 1 - r$. Luego

$$\begin{aligned} k(r) &= n(n - 1) \int_0^{1-r} r^{n-2} ds \\ &= n(n - 1)r^{n-2}(1 - r), \quad 0 \leq r \leq 1. \end{aligned}$$

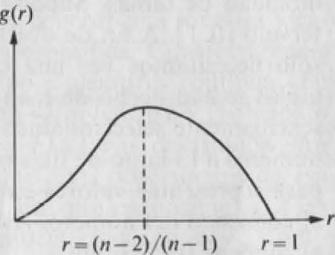


FIGURA 13.2

Para $n > 2$, el gráfico de la fdp de R tiene la forma que se indica en la figura 13.2. Observe que cuando $n \rightarrow \infty$ el valor de r en el cual ocurre un máximo se desplaza a la derecha. Luego, cuando el tamaño de la muestra se aumenta, llega a ser más y más posible que el recorrido R sea próximo a 1, que es intuitivamente lo que esperaríamos.

13.5 La transformación integral

Una muestra de una variable aleatoria X puede usarse para obtener información acerca de parámetros desconocidos asociados con la distribución de probabilidades de X . Sin embargo, podemos usar una muestra con un objeto diferente. Podríamos tomar algunas observaciones de una variable aleatoria cuya distribución está completamente especificada y luego usar esos valores muestrales para aproximar ciertas probabilidades que serían muy difíciles de obtener con un cálculo matemático directo. Por ejemplo, suponiendo que X tiene una distribución $N(0, 1)$ y queremos estudiar la variable aleatoria $Y = e^{-X} \operatorname{sen} X$. En especial, supongamos que queremos calcular $P(0 \leq Y \leq \frac{1}{2})$. A fin de obtener la respuesta exacta, necesitamos encontrar G , la fda de Y y luego calcular $G(\frac{1}{2}) - G(0)$. Encontraríamos mucha dificultad para hacer esto. Sin embargo, podemos usar otro enfoque que está basado en la idea de *simular* el experimento que da origen a la variable aleatoria Y . Entonces usamos la frecuencia relativa como una aproximación a la probabilidad buscada. Si esta frecuencia relativa se basa en un número de observaciones suficientemente grande, la ley de los grandes números justifica nuestro procedimiento.

Especificamente, supóngase que tenemos una muestra aleatoria de la variable aleatoria anterior, cuya distribución está completamente especificada, X_1, \dots, X_n . Para cada X_i definimos la variable aleatoria $Y_i = e^{-X_i} \operatorname{sen} X_i$. Luego, evaluamos la frecuencia relativa n_A/n , en donde n_A es igual al número de valores Y_i , sean y_i , que satisfacen $0 \leq y_i \leq \frac{1}{2}$. Luego n_A/n es la frecuencia relativa $0 \leq Y \leq \frac{1}{2}$, y si n es grande, esta frecuencia relativa estará «próxima» a $P[0 \leq Y \leq \frac{1}{2}]$ según la ley de los grandes números.

Para aplicar el procedimiento anterior, debemos encontrar un medio de «generar» una muestra aleatoria X_1, \dots, X_n de la variable aleatoria cuya distribución es $N(0, 1)$. Antes de indicar cómo se hace esto, expongamos brevemente una distribución para la cual este trabajo ya se ha realizado debido a la disponibilidad de tablas. Supóngase que X está distribuida uniformemente en el intervalo $[0, 1]$. A fin de obtener una muestra aleatoria para tal variable aleatoria, sólo necesitamos ver una tabla de números aleatorios (ver el Apéndice). Esas tablas se han hecho de manera que sean útiles para este propósito. Para usarlas, sencillamente seleccionamos una ubicación al azar en la tabla y luego obtenemos números a lo largo de filas o columnas. Si queremos usar esos números tabulados para representar valores entre 0 y 1, sólo necesitamos poner una coma decimal al comienzo del número. Así, el número anotado 4573 se usaría para representar al número 0,4573, etc.

La disponibilidad de esas tablas de números aleatorios hace la tarea de obtener una muestra aleatoria de una distribución arbitraria relativamente sencilla debido al resultado contenido en el teorema siguiente.

Teorema 13.6. Sea X una variable aleatoria con fdp f y fda F . [Se supone que $f(x) = 0$, $x \notin (a, b)$]. Sea Y la variable aleatoria definida por $Y = F(X)$. Luego Y está distribuida uniformemente en $[0, 1]$ (Y se designa como la transformación integral de X).

Demostración: puesto que X es una variable aleatoria continua, la fda F es una función continua estrictamente monótona con una inversa, F^{-1} . Es decir, $Y = F(X)$ puede resolverse para X mediante $Y: X = F^{-1}(Y)$. (Ver la figura 13.3.) [Si $F(x) = 0$ para $x \leq a$, definimos $F^{-1}(0) = a$. Análogamente, si $F(x) = 1$ para $x \geq b$, definimos $F^{-1}(1) = b$.]

Sea G la fda de la variable aleatoria Y definida anteriormente. Luego

$$G(y) = P(Y \leq y) = P(F(X) \leq y) = P(X \leq F^{-1}(y)) = F(F^{-1}(y)) = y.$$

Por tanto la fdp de Y , $g(y) = G'(y) = 1$. Esto establece nuestra demostración.

Observaciones: (a) Comentemos brevemente cómo se examina realmente el valor de una variable aleatoria Y . Dado un valor x , de la variable aleatoria X , calculamos luego el valor de $Y = F(X)$ como $y = F(x)$, en donde F es la fda de X (conocida).

(b) El teorema 13.6 enunciado y demostrado anteriormente para variables aleatorias continuas también es válido para variables aleatorias discretas. Hay que hacer un pequeño cambio en la demostración puesto que la fda de una variable aleatoria discreta es una función escalonada y no tiene una inversa única.

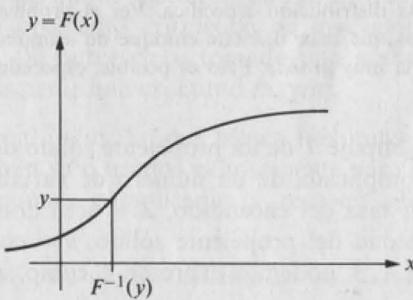


FIGURA 13.3

Podemos usar ahora el resultado anterior a fin de generar una muestra aleatoria de una variable aleatoria con una distribución específica. Consideraremos nuevamente sólo el caso continuo. Sea X una variable aleatoria continua con fda F de la cual se pide una muestra aleatoria. Sea y_1 un valor (entre 0 y 1) obtenido de una tabla de números aleatorios. Puesto que $Y = F(X)$ está distribuida uniformemente en $[0, 1]$, podemos considerar a y_1 como una observación de esa variable aleatoria. Resolviendo $y_1 = F(x_1)$ para x_1 (lo que es posible si X es continua), obtenemos un valor de una variable aleatoria cuya fda es F . Continuando este procedimiento con los números y_2, \dots, y_n obtenidos de una tabla de números aleatorios, tenemos $x_i, i = 1, \dots, n$, como solución de la ecuación $y_i = F(x_i)$, y por tanto, tenemos nuestros valores muestrales buscados.

EJEMPLO 13.3. Supóngase que queremos obtener una muestra de tamaño cinco de una variable aleatoria con distribución $N(2, 0,09)$ y supóngase que obtenemos los valores siguientes de una tabla de números aleatorios: 0,487; 0,722; 0,661; 0,194; 0,336. Definamos x_1 como sigue:

$$\begin{aligned} 0,487 &= \frac{1}{\sqrt{2\pi}(0,3)} \int_{-\infty}^{x_1} \exp\left[-\frac{1}{2}\left(\frac{t-2}{0,3}\right)^2\right] dt \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(x_1-2)/0,3} \exp\left(-\frac{s^2}{2}\right) ds = \Phi\left(\frac{x_1-2}{0,3}\right). \end{aligned}$$

De las tablas de la distribución normal encontramos que $(x_1 - 2)/0,3 = -0,03$. Luego $x_1 = (-0,03)(0,3) + 2 = 1,991$. Este número representa nuestro primer valor de muestra de la distribución especificada. Continuando de la misma manera con los otros valores obtenemos los siguientes valores adicionales de muestra: 2,177; 2,124; 1,742; 1,874.

El procedimiento anterior puede generalizarse como sigue. Para obtener un valor de muestra x_1 de la distribución $N(\mu, \sigma^2)$, obtenemos un valor muestral y_1 (entre 0 y 1) de una tabla de números aleatorios. El valor pedido x_1 está definido por la ecuación $\Phi((x_1 - \mu)/\sigma) = y_1$.

Observaciones: (a) Hay métodos alternos al sugerido anteriormente, para generar muestras aleatorias de una distribución específica. Ver el problema 13.8.

(b) Uno de los aspectos que hace útil este enfoque de «simulación» es la posibilidad de tener una muestra aleatoria muy *grande*. Esto es posible, especialmente, si se dispone de un computador electrónico.

EJEMPLO 13.4. El empuje T de un propelente sólido de un sistema de cohetes es una función muy complicada de un número de variables. Si X = sección de la garganta, Y = factor tasa del encendido, Z = área donde se quema el propelente sólido, W = densidad del propelente sólido, n = coeficiente de encendido, y C_i = constante, $i = 1, 2, 3$, podemos expresar T como sigue:

$$T = C_1 \left(\frac{X}{YWZ} \right)^{1/(n-1)} X + C_2 \left(\frac{X}{YWZ} \right)^{1/(n-1)} + C_3.$$

En exposiciones anteriores las siguientes hipótesis posibles: X , Y , W y Z son variables aleatorias independientes distribuidas normalmente con promedios y varianzas conocidos. Cualquier intento para obtener la distribución de probabilidades de la variable aleatoria T o aún de expresiones exactas para $E(T)$ y $V(T)$ fracasará debido a la relación compleja entre X , Y , W , Z y T .

Si pudiéramos generar una muestra aleatoria (grande) de X , Y , Z y W , por ejemplo obtener cuaternas (X_i, Y_i, Z_i, W_i) , podríamos luego generar una gran muestra de T , llamémosla (T_1, \dots, T_n) e intentar estudiar las características de la variable aleatoria T empíricamente (mediante la muestra). Supóngase, por ejemplo, que deseamos calcular $P(a \leq T \leq b)$. Para aplicar la ley de los grandes números, simplemente necesitamos obtener la frecuencia relativa del suceso $\{a \leq T \leq b\}$ de nuestra muestra (grande), y luego poder tener una certeza razonable de que la frecuencia relativa se diferencia muy poco de la probabilidad teórica que está siendo buscada.

Hasta ahora nos hemos interesado sólo en el problema de aproximar una probabilidad exacta con una frecuencia relativa basada en un gran número de observaciones. Sin embargo, el método que hemos sugerido puede usarse para obtener soluciones aproximadas de problemas que son de naturaleza completamente no probabilística. Indicaremos sólo uno de los muchos tipos de problemas que pueden considerarse de esta manera. El enfoque general se designa como el método de «Monte-Carlo». Una descripción muy buena de este método aparece en *Modern Mathematics for the Engineer*, de E. F. Beckenbach, publicado por McGraw-Hill Book Co., Inc., New York, 1956, capítulo 12. El ejemplo 13.5 se obtuvo de este libro.

EJEMPLO 13.5. Supóngase que deseamos calcular la integral $\int_0^1 x \, dx$ sin recurrir a los procedimientos triviales para obtener su valor $\frac{1}{2}$. Se procede como se indica. Se obtiene, de una tabla de números aleatorios, una muestra aleatoria de la variable aleatoria distribuida uniformemente en $[0, 1]$. Suponiendo que los valores muestrales son 0,69; 0,37; 0,39; 0,97; 0,66; 0,51; 0,60; 0,41; 0,76 y 0,09. Puesto que la integral pedida representa a $E(X)$, en donde X es la variable aleatoria uniformemente distribuida que se muestrea, parece razonable que podamos aproximar $E(X)$ usando el promedio aritmético de los valores muestrales. Encontramos que $\bar{X} = 0,545$. (Si hubiésemos tomado una muestra mayor, tendríamos una buena razón para esperar una exactitud mayor).

Esta ilustración trivial indica la idea básica tras muchos métodos de Monte-Carlo. Estos métodos han sido usados exitosamente para evaluar integrales múltiples sobre ciertas regiones complicadas y resolver algunas ecuaciones diferenciales.

Observación: los medios para obtener muestras de una distribución arbitraria como se describió en la sección 13.5, pueden llegar a ser muy complicados. Debido a la gran importancia de la distribución normal, existen tablas disponibles (Ver la tabla 7 en el Apéndice) que eliminan una gran parte de los cálculos descritos anteriormente. La tabla 7 da, directamente, muestras de la distribución $N(0, 1)$. Estos valores muestrales se llaman *desviaciones normales*. Si se necesitan n valores muestrales x_1, \dots, x_n de la distribución $N(0, 1)$, se toman directamente de la tabla 7 (escogiendo el punto de partida de alguna manera aleatoria posible, como se describió para la tabla de los números aleatorios).

De una manera sencilla, también puede usarse esta tabla para obtener muestras de una distribución normal arbitraria $N(\mu, \sigma^2)$. Sencillamente se multiplica x_i , el valor escogido de la tabla, por σ y luego se le suma μ . Es decir, se forma $y_i = \sigma x_i + \mu$. Luego y_i será un valor muestral de la distribución pedida.

PROBLEMAS

- 13.1. Deducir la expresión para la fdp del mínimo de una muestra. (Ver el teorema 13.2.)
- 13.2. Demostrar que si X_1, \dots, X_n son variables aleatorias independientes, que tiene cada una distribución exponencial con parámetro α_i , $i = 1, 2, \dots, n$ y si $K = \min(X_1, \dots, X_n)$, entonces K tiene una distribución exponencial con parámetro $\alpha_1 + \dots + \alpha_n$. (Ver el teorema 13.3.)

13.3. Supóngase que X tiene una distribución geométrica con parámetro p . Sea X_1, \dots, X_n una muestra aleatoria de X y sea $M = \max(X_1, \dots, X_n)$ y $K = \min(X_1, \dots, X_n)$. Encontrar la distribución de probabilidades de M y de K . [Indicación: $P(M = m) = F(m) - F(m - 1)$, en donde F es la fda de M .]

13.4. Se obtiene una muestra de tamaño 5 de una variable aleatoria con distribución $N(12, 4)$.

- (a) ¿Cuál es la probabilidad de que el promedio muestral exceda 13?
- (b) ¿Cuál es la probabilidad de que el mínimo de la muestra sea menor que 10?
- (c) ¿Cuál es la probabilidad de que el máximo de la muestra exceda 15?

13.5. La duración de un artículo (en horas) está distribuida exponencialmente con parámetro $\beta = 0.001$. Se prueban seis artículos y se anotan sus tiempos para fallar.

(a) ¿Cuál es la probabilidad de que ningún artículo falle antes de que hayan transcurrido 800 horas?

- (b) ¿Cuál es la probabilidad de que ningún artículo dure más de 3000 horas?

13.6. Supóngase que X tiene distribución $\Lambda(0, 0.09)$. Se obtiene una muestra de tamaño 25 de X , sea X_1, \dots, X_{25} . ¿Cuál es la probabilidad de que $\sum_{i=1}^{25} X_i^2$ excede 1.5?

13.7. Utilizando una tabla de números aleatorios, obtener una muestra aleatoria de tamaño 8 de una variable aleatoria que tiene las distribuciones siguientes:

- (a) exponencial, con parámetro 2,
- (b) χ cuadrado con 7 grados de libertad,
- (c) $N(4, 4)$.

13.8. En la sección 13.5 hemos dado un método con el cual puede generarse una muestra aleatoria de una distribución especificada. Hay muchos métodos distintos con los cuales puede hacerse esto, algunos de los cuales pueden preferirse al dado, especialmente si hay instrumentos de cómputo disponibles. Uno de tales métodos es el siguiente. Supóngase que deseamos tener una muestra aleatoria de una variable aleatoria que tiene una distribución de χ cuadrado con $2k$ grados de libertad. Procederemos como sigue: se obtiene una muestra aleatoria de tamaño k (con ayuda de la tabla de números aleatorios) de una variable aleatoria que está distribuida uniformemente en $(0, 1)$, sea U_1, \dots, U_k . Luego calculemos $X_1 = -2 \ln(U_1)$, $U_2 \cdots U_k) = -2 \sum_{i=1}^k \ln(U_i)$. La variable aleatoria X_1 tendrá entonces la distribución pedida, como lo indicaremos a continuación. Luego, continuando con este esquema, obtenemos otra muestra de tamaño k de una variable aleatoria distribuida uniformemente y encontramos así el segundo valor muestral X_2 . Nótese que este procedimiento necesita k observaciones de una variable aleatoria distribuida uniformemente para cada observación de χ_{2k}^2 . Para verificar la proposición que se hizo anteriormente procederemos como sigue.

(a) Obtener la función generadora de momentos de la variable aleatoria $-2 \ln(U_i)$, en donde U_i es distribuida uniformemente en $(0, 1)$.

(b) Obtener la función generadora de momentos de la variable aleatoria $-2 \sum_{i=1}^k \ln(U_i)$, en donde las U_i son variables aleatorias independientes cada una con la distribución anterior. Se compara esta fgm con la de la distribución χ cuadrado y luego se obtiene la conclusión buscada.

13.9. Usando el esquema bosquejado en el problema 13.8, obténgase una muestra aleatoria de tamaño 3 de la distribución χ_8^2 .

13.10. Una variable aleatoria continua X está distribuida uniformemente en $(-\frac{1}{2}, \frac{1}{2})$. Se obtiene una muestra de tamaño n de X y se calcula el promedio muestral \bar{X} . ¿Cuál es la desviación estándar de \bar{X} ?

13.11. Se toman muestras independientes de tamaño 10 y 15 de una variable aleatoria distribuida normalmente con esperanza 20 y varianza 3. ¿Cuál es la probabilidad de que el promedio de las dos muestras se diferencie (en valor absoluto) en más de 0,3?

13.12. (Para este ejercicio y los tres siguientes, lea la observación al final del capítulo 13.) Use la tabla de las desviaciones normales (tabla 7 del Apéndice) y obtenga una muestra de tamaño 30 de una variable aleatoria X que tiene una distribución $N(1, 4)$. Usar esta muestra para responder lo siguiente:

- Comparar $P(X \geq 2)$ con la frecuencia relativa de ese suceso.
- Comparar el promedio muestral \bar{X} y la varianza muestral S^2 con 1 y 4 respectivamente.
- Construir un gráfico de $F(t) = P(X \leq t)$. Usando el mismo sistema de coordenadas, obtener el gráfico de la función de distribución empírica F_n definida como sigue:

$$\begin{aligned} F_n(t) &= 0 && \text{si } t < X^{(1)} \\ &= k/n && \text{si } X^{(k)} \leq t < X^{(k+1)} \\ &= 1 && \text{si } t \geq X^{(n)}, \end{aligned}$$

en donde $X^{(i)}$ es la i -ésima mayor observación en la muestra (es decir, $X^{(i)}$ es el estadigrafo de i -ésimo orden). [La función F_n se usa frecuentemente para aproximar la fda F . Puede demostrarse que bajo condiciones muy generales $\lim_{n \rightarrow \infty} F_n(t) = F(t)$.]

13.13. Tenga X una distribución $N(0, 1)$. De la tabla 7 en el Apéndice, obtener una muestra de tamaño 20 de esta distribución. Sea $Y = |X|$.

- Usar esta muestra para comparar $P[1 < Y \leq 2]$ con la frecuencia relativa de ese suceso.
- Comparar $E(Y)$ con el promedio muestral \bar{Y} .
- Comparar la fda de Y , $F(t) = P(Y \leq t)$ con F_n la fda empírica de Y .

13.14. Supóngase que X tiene distribución $N(2, 9)$. Sea X_1, \dots, X_{20} una muestra aleatoria de X obtenida con ayuda de la tabla 7. Calcular

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

y comparar con $E(S^2) = 9$.

13.15. Tenga X una distribución $N(0, 1)$. Sea X_1, \dots, X_{30} una muestra aleatoria de X obtenida, usando la tabla 7. Calcular $P(X^2 \geq 0,10)$ y comparar este valor con la frecuencia relativa de ese suceso.

Estimación de los parámetros

14.1 Introducción

En el capítulo anterior sugerimos que una muestra de una variable aleatoria X puede usarse con el objeto de estimar uno o varios parámetros (desconocidos) asociados con la distribución de probabilidades de X . En este capítulo consideraremos en detalle este problema.

A fin de desarrollar un ejemplo específico, consideremos la situación siguiente. Un fabricante nos ha enviado 100.000 pequeños remaches. Un empalme perfectamente remachado necesita que cada remache ajuste perfectamente en un hueco y por consiguiente se tendrá alguna dificultad cuando el remache lo exceda. Antes de aceptar este cargamento, queremos tener alguna idea acerca de la magnitud de p , la proporción de remaches defectuosos (es decir, los que exceden el hueco), y procedemos como sigue. Se inspeccionan n remaches escogidos al azar en el lote. Debido al gran tamaño del lote, podemos suponer que escogemos con sustitución aunque realmente no procederíamos así. Se definen las siguientes variables aleatorias: $X_i = 1$, si el i -ésimo artículo es defectuoso, y 0 de otro modo $i = 1, 2, \dots, n$. Luego podemos considerar que X_1, \dots, X_n sea una muestra de la variable aleatoria X cuya distribución está dada por $P(X = 1) = p$, $P(X = 0) = 1 - p$.

La distribución de probabilidades de X depende del parámetro desconocido p de una manera muy sencilla. ¿Podemos usar la muestra X_1, \dots, X_n de alguna manera con el objeto de estimar el valor de p ? ¿Hay algún estadígrafo H tal que $H(X_1, \dots, X_n)$ puede usarse como un estimador (puntual) de p ?

Debería ser evidente que una muestra de tamaño n , en donde $n < 100.000$, nunca puede permitirnos reconstruir la verdadera composición del cargamento sin importar cuán hábilmente usemos la información obtenida de la muestra. En otras palabras, a no ser que inspeccionemos cada uno de los artículos (es decir, tomemos $n = 100.000$) nunca podremos conocer el valor verdadero de p . (Esta última frase se refiere evidentemente a muestreo sin sustitución.)

Luego, cuando proponemos a \hat{p} como un estimador de p realmente no esperamos que \hat{p} sea igual a p . (Recordamos que \hat{p} es una variable aleatoria y, por tanto, puede tomar muchos valores.) Este dilema da origen a dos preguntas importantes:

- (1) ¿Qué características queremos que posea un «buen» estimador?
- (2) ¿Cómo decidimos que un estimador es «mejor» que otro?

Puesto que esta puede ser la primera vez que el lector encuentre preguntas de esta clase, vale la pena comentar brevemente la naturaleza general de este problema. A muchas preguntas matemáticas, existe una respuesta definida. Esta puede ser muy difícil de encontrar, puesto que implica muchos problemas técnicos y debemos contentarnos con una aproximación. Con todo, habitualmente es evidente cuando tenemos una respuesta y cuando no. (Por ejemplo, supongamos que nos piden encontrar una raíz real de la ecuación $3x^5 - 4x^2 + 13x - 7 = 0$. Una vez que hemos encontrado una solución, es muy sencillo verificar si es la correcta: sólo necesitamos sustituirla en la ecuación dada. Si tenemos dos respuestas aproximadas, r_1 y r_2 , es también sencillo decidir cuál aproximación es mejor.)

Sin embargo, el problema actual, particularmente la estimación de p , no admite un análisis tan sencillo. En primer lugar, puesto que nunca podemos conocer el valor verdadero de p (en cualquier situación real al menos), no tiene sentido decir que nuestra estimación \hat{p} es «correcta». Segundo, si tenemos dos estimaciones de p , llamémoslas \hat{p}_1 y \hat{p}_2 , debemos encontrar algún medio para decidir cuál es «mejor». Esto significa que debemos establecer algún criterio que podamos aplicar para decidir si una estimación es preferible a otra.

14.2 Criterios para estimadores

Definiremos ahora algunos conceptos importantes que nos ayudarán a resolver el problema sugerido anteriormente.

Definición. Sea X una variable aleatoria con una distribución de probabilidades que depende de un parámetro desconocido θ . Sea X_1, \dots, X_n una muestra de X y sean x_1, \dots, x_n los valores muestrales correspondientes. Si $g(X_1, \dots, X_n)$ es una función de la muestra que va a ser usada para estimar θ nos referimos a g como un *estimador* de θ . El valor que toma g , es decir $g(x_1, \dots, x_n)$, será mencionado como una *estimación* de θ y habitualmente es escrito como $\hat{\theta} = g(x_1, \dots, x_n)$. (Ver la observación siguiente.)

Observación: en este capítulo violaremos una regla que hemos observado muy cuidadosamente hasta ahora: hacer una distinción cuidadosa entre una variable aleatoria y su valor. Es decir, a menudo hablaremos de $\hat{\theta}$, contra la estimación de θ , cuando realmente deberíamos hablar del estimador $g(X_1, \dots, X_n)$. También escribiremos $E(\hat{\theta})$ cuando, realmente, indicamos $E[g(X_1, \dots, X_n)]$. Sin embargo, en la parte en la cual nos tomemos esta libertad eliminaríamos cualquier ambigüedad posible.

Definición. Sea $\hat{\theta}$ una estimación del parámetro desconocido θ asociado con la distribución de la variable aleatoria X . Entonces $\hat{\theta}$ es un *estimador insesgado* (o estimación insesgada; ver la observación anterior) para θ si $E(\hat{\theta}) = \theta$ para cualquier θ .

Observación: cualquier estimación buena debería ser «próxima» al valor que está estimando. «Insesgadura» significa principalmente que el valor promedio de la estimación estará próximo al valor verdadero del parámetro. Por ejemplo, si el mismo estimador se usa repetidamente y

promediamos esos valores, esperaríamos que el promedio fuera próximo al valor verdadero del parámetro. Aunque es deseable que un estimador sea insesgado, puede haber ocasiones en las cuales podríamos preferir estimadores sesgados (ver más abajo). Es posible (y realmente es fácil) encontrar más de un estimador insesgado para un parámetro desconocido. A fin de hacer una buena elección en tales casos presentamos el concepto siguiente.

Definición. Sea $\hat{\theta}$ una estimación insesgada de θ . Decimos que $\hat{\theta}$ es una *estimación insesgada de varianza mínima* de θ si para todas las estimaciones θ^* tales que $E(\theta^*) = \theta$, tenemos $V(\hat{\theta}) \leq V(\theta^*)$ para cualquier θ . Es decir, entre todas las estimaciones insesgadas de θ , $\hat{\theta}$ tiene la varianza más pequeña.

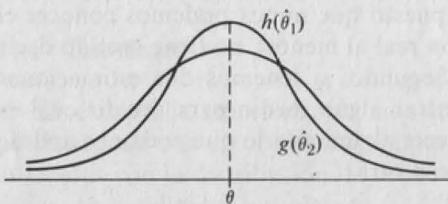


FIGURA 14.1

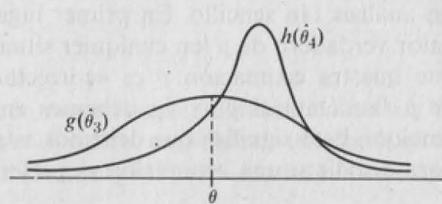


FIGURA 14.2

Observaciones: (a) La varianza de una variable aleatoria mide la variabilidad de la variable aleatoria respecto a su valor esperado. Por tanto, es intuitivamente atractivo pedir que un estimador insesgado tenga una varianza pequeña, porque así la variable aleatoria tiende a aproximarse a su promedio, lo cual en el caso de un estimador insesgado, significa aproximarse al valor verdadero del parámetro. Luego si $\hat{\theta}_1$ y $\hat{\theta}_2$ son dos estimaciones de θ , cuya fdp está bosquejada en la figura 14.1, posiblemente preferiríamos $\hat{\theta}_1$ a $\hat{\theta}_2$. Ambas estimaciones son insesgadas, y $V(\hat{\theta}_1) < V(\hat{\theta}_2)$.

En el caso de las estimaciones $\hat{\theta}_3$ y $\hat{\theta}_4$ la decisión no es tan evidente (figura 14.2) ya que $\hat{\theta}_3$ es insesgado mientras que $\hat{\theta}_4$ no es. Sin embargo, $V(\hat{\theta}_3) > V(\hat{\theta}_4)$. Esto significa que mientras en promedio $\hat{\theta}_3$ estará próxima a θ , su mayor varianza indica que no serían sorprendentes desviaciones considerables de θ , por otra parte $\hat{\theta}_4$ tendería a ser algo mayor que θ , en promedio y podría aún estar más próxima a θ que $\hat{\theta}_3$ (ver figura 14.2).

(b) Existen técnicas para encontrar estimaciones insesgadas de varianza mínima en general. Sin embargo no podremos presentarlas aquí. Haremos uso de este concepto principalmente con el objeto de elegir entre dos o más estimaciones insesgadas disponibles. Es decir, si $\hat{\theta}_1$ y $\hat{\theta}_2$ son ambas estimaciones insesgadas de θ , y si $V(\hat{\theta}_1) < V(\hat{\theta}_2)$, preferiríamos $\hat{\theta}_1$.

Otro criterio para discernir entre estimaciones es algo más difícil de formular y se basa en la siguiente definición.

Definición. Sea $\hat{\theta}$ una estimación (basada en una muestra X_1, \dots, X_n) del parámetro θ . Se dice que $\hat{\theta}$ es una estimación convergente de θ , si

$$\lim_{n \rightarrow \infty} \text{Prob} [|\hat{\theta} - \theta| > \epsilon] = 0 \quad \text{para todo } \epsilon > 0$$

o equivalentemente, si

$$\lim_{n \rightarrow \infty} \text{Prob} [|\hat{\theta} - \theta| \leq \epsilon] = 1 \quad \text{para todo } \epsilon > 0.$$

Observaciones: (a) Esta definición establece que una estimación es convergente si aumenta el tamaño de muestra, $\hat{\theta}$ converge en el sentido probabilístico anterior a θ . Nuevamente, esta es una propiedad intuitivamente atractiva que debe poseer una estimación. Porque dice que cuando aumenta el tamaño de muestra (lo que significaría, en circunstancias muy razonables, que se dispone de más información) la estimación llega a ser «mejor» en el sentido indicado.

(b) Es relativamente fácil verificar si una estimación es insesgada o no. También es muy elemental comparar las varianzas de dos estimaciones insesgadas. Sin embargo, verificar la convergencia aplicando la definición anterior, no es tan sencillo. El teorema siguiente es algunas veces muy útil.

Teorema 14.1. Sea $\hat{\theta}$ una estimación de θ basada en una muestra de tamaño n .

Si $\lim_{n \rightarrow \infty} E(\hat{\theta}) = \theta$, y si $\lim_{n \rightarrow \infty} V(\hat{\theta}) = 0$, entonces $\hat{\theta}$ es una estimación convergente de θ .

Demostración: usaremos la desigualdad de Chebyshev, ecuación (7.20). Luego escribamos:

$$\begin{aligned} P[|\hat{\theta} - \theta| \geq \epsilon] &\leq \frac{1}{\epsilon^2} E[(\hat{\theta} - \theta)^2] = \frac{1}{\epsilon^2} E[\hat{\theta} - E(\hat{\theta}) + E(\hat{\theta}) - \theta]^2 \\ &= \frac{1}{\epsilon^2} E\{[\hat{\theta} - E(\hat{\theta})]^2 + 2[\hat{\theta} - E(\hat{\theta})][E(\hat{\theta}) - \theta] \\ &\quad + [E(\hat{\theta}) - \theta]^2\} \\ &= \frac{1}{\epsilon^2} \{ \text{Var } \hat{\theta} + 0 + [E(\hat{\theta}) - \theta]^2 \}. \end{aligned}$$

Luego haciendo que $n \rightarrow \infty$ y utilizando la hipótesis del teorema, encontramos que $\lim_{n \rightarrow \infty} P[|\hat{\theta} - \theta| \geq \epsilon] \leq 0$ e igualamos así a 0.

Observación: si el estimador $\hat{\theta}$ es insesgado, la primera condición se satisface automáticamente.

Un criterio final que se aplica a menudo para estimar puede formularse como sigue. Suponiendo que X_1, \dots, X_n es una muestra de X y θ es un parámetro desconocido. Sea $\hat{\theta}$ una función de (X_1, \dots, X_n) .

Definición. Decimos que $\hat{\theta}$ es el mejor estimador insesgado de θ si:

- (a) $E(\hat{\theta}) = \theta$.
- (b) $\hat{\theta} = \sum_{i=1}^n a_i X_i$. Es decir, $\hat{\theta}$ es una función lineal de la muestra.
- (c) $V(\hat{\theta}) \leq V(\theta^*)$, en donde θ^* es cualquier otro estimador de θ que satisface las relaciones (a) y (b) anteriores.

Posteriormente, consideraremos un método muy general que nos dará buenas estimaciones para un gran número de problemas, dado que ellas satisfarán uno o

más de los criterios anteriores. Antes de hacer esto, consideraremos sencillamente algunos estimadores que intuitivamente son muy razonables, y luego verificaremos precisamente qué tan buenos o qué tan malos son mediante los criterios anteriores.

14.3 Algunos ejemplos

Los criterios anteriores de insesgados, varianza mínima, convergencia y linealidad nos dan al menos una pauta para juzgar un estimador. Consideremos ahora algunos ejemplos.

EJEMPLO 14.1. Reconsideremos el problema anterior. Hemos muestreado n remaches y encontrado que la muestra (X_1, \dots, X_n) produce exactamente k defectuosos; es decir, $Y = \sum_{i=1}^n X_i = k$. Debido a nuestras hipótesis, Y es una variable aleatoria distribuida binomialmente.

El estimador intuitivamente más sugestivo del parámetro p es $\hat{p} = Y/n$, la proporción de defectuosos encontrados en la muestra. Apliquemos algunos de los criterios anteriores para ver qué tan bueno es un estimador \hat{p} .

$$E(\hat{p}) = E\left(\frac{Y}{n}\right) = \frac{1}{n}(np) = p.$$

Luego \hat{p} es un estimador insesgado de p .

$$V(\hat{p}) = V\left(\frac{Y}{n}\right) = \frac{1}{n^2}(np)(1-p) = \frac{p(1-p)}{n}.$$

Luego $V(\hat{p}) \rightarrow 0$ cuando $n \rightarrow \infty$, y por tanto \hat{p} es estimador convergente. Como lo hemos señalado anteriormente, puede haber muchos estimadores insesgados para un parámetro, algunos de los cuales pueden ser muy malos, además. Por ejemplo, considerando en el contexto del ejemplo presente el estimador p^* definido así: $p^* = 1$ si el primer artículo elegido es defectuoso, y 0 en otro caso. Es decir, es evidente que no es muy buen estimador cuando observamos que su valor es una función sólo de X_1 , en vez de X_1, \dots, X_n . Sin embargo, p^* es insesgado, porque $E(p^*) = 1P(X=1) + OP(X=0) = p$. La varianza de p^* es $p(1-p)$, la que se compara muy mal con la varianza de p considerada anteriormente, esto es $p(1-p)/n$, especialmente si n es grande.

El resultado obtenido en el ejemplo anterior es un caso especial de la siguiente proposición general.

Teorema 14.2. Sea X una variable aleatoria con esperanza finita μ y varianza σ^2 . Sea \bar{X} el promedio muestral, obtenido en una muestra de tamaño n . Luego \bar{X} es un estimador insesgado y convergente de μ .

Demostración: ésta se deduce inmediatamente del teorema 13.1 donde demostramos que $E(\bar{X}) = \mu$ y $V(\bar{X}) = \sigma^2/n$ que tiende a 0 cuando $n \rightarrow \infty$.

Observación: que el resultado demostrado en el ejemplo 14.1 es un caso especial del teorema 14.2, se deduce cuando observamos que la proporción de defectuosos en la muestra, Y/n , puede escribirse como $(1/n)(X_1 + \dots + X_n)$, en donde los X_i toman los valores uno y cero, según que el artículo inspeccionado sea o no defectuoso.

El promedio muestral citado en el teorema 14.2 es una función lineal de la muestra. Es decir, es de la forma $a_1X_1 + a_2X_2 + \dots + a_nX_n$, con $a_1 = \dots = a_n = 1/n$. Fácilmente se observa que $\hat{\mu} = \sum_{i=1}^n a_iX_i$ es un estimador insesgado de μ para cualquier elección de los coeficientes que satisfacen la condición $\sum_{i=1}^n a_i = 1$. Surge la siguiente pregunta interesante. ¿Para qué elección de los a_i (sujetas a $\sum_{i=1}^n a_i = 1$) es más pequeña la varianza de $\sum_{i=1}^n a_iX_i$? Resulta que la varianza es minimizada si $a_i = 1/n$ para todo i . Es decir, \bar{X} , es el estimador lineal insesgado de varianza mínima.

Para verificar esto, consideremos

$$\hat{\mu} = \sum_{i=1}^n a_iX_i, \quad \sum_{i=1}^n a_i = 1.$$

Luego

$$\text{Var } \hat{\mu} = \sigma^2 \sum_{i=1}^n a_i^2$$

puesto que los X_i son variables aleatorias independientes con varianza común σ^2 . Escribimos

$$\begin{aligned} \sum_{i=1}^n a_i^2 &= (a_1 - 1/n)^2 + \dots + (a_n - 1/n)^2 + (2/n)(a_1 + \dots + a_n) - n(1/n^2) \\ &= (a_1 - 1/n)^2 + \dots + (a_n - 1/n)^2 + 1/n \left(\text{puesto que } \sum_{i=1}^n a_i = 1 \right). \end{aligned}$$

Luego esta expresión es minimizada evidentemente si $a_i = 1/n$ para todo i .

EJEMPLO 14.2. Supóngase que T , el tiempo para fallar de una componente, está distribuido exponencialmente. Es decir, la fdp de T está dada por $f(t) = \beta e^{-\beta t}$, $t \geq 0$. Supóngase que probamos n componentes, anotando el momento en que falla cada una, sean T_1, \dots, T_n . Deseamos una estimación insesgada del tiempo esperado para fallar $E(T) = 1/\beta$, basada en la muestra (T_1, \dots, T_n) . Uno de tales estimadores es $\bar{T} = (1/n) \sum_{i=1}^n T_i$. Del teorema 14.2 sabemos que $E(\bar{T}) = 1/\beta$. Puesto que $V(T) = 1/\beta^2$, el teorema 13.1 nos afirma que $V(\bar{T}) = 1/\beta^2 n$. Sin embargo, \bar{T} no es el único estimador insesgado de $1/\beta$. Consideremos, en efecto, el mínimo de la muestra, $Z = \min(T_1, \dots, T_n)$. De acuerdo con el teorema 13.3, nuevamente Z está distribuida exponencialmente con parámetro $n\beta$. Luego el estimador nZ también es un estimador insesgado de $1/\beta$.

Para evaluar la varianza calculamos

$$V(nZ) = n^2 V(Z) = n^2 \frac{1}{(n\beta)^2} = \frac{1}{\beta^2}.$$

Así, aunque los dos estimadores nZ y \bar{T} son insesgados, el último tiene una varianza más pequeña y por tanto debería preferirse.

Sin embargo, en esta situación especial hay otra consideración que podría influir en nuestra elección entre los dos estimadores sugeridos. Las n componentes podrían probarse simultáneamente. (Por ejemplo, podríamos poner n bombillas en n portalámparas y anotar el tiempo en que se queman.) Cuando usamos nZ como estimador, la prueba puede terminarse tan pronto como la primera componente falle. Al usar \bar{T} como estimador debemos esperar hasta que todas las componentes hayan fallado. Es muy posible que transcurra mucho tiempo entre la primera y la última falla. Diciéndolo de una manera distinta, si L es el tiempo necesario para probar los n artículos y calculamos la estimación para $1/\beta$, usando nZ , tenemos $L = \min(T_1, \dots, T_n)$, mientras que al usar \bar{T} , tenemos $L = \max(T_1, \dots, T_n)$. Luego, si el tiempo necesario para efectuar la prueba es de cualquier efecto serio (digamos en términos de costo), podríamos preferir al estimador con mayor varianza.

EJEMPLO 14.3. Suponiendo que se desea un estimador insesgado de la varianza σ^2 de una variable aleatoria, basada en una muestra X_1, \dots, X_n .

Aunque podríamos considerar $(1/n) \sum_{i=1}^n (\bar{X}_i - \bar{X})^2$, resulta que este estadígrafo tiene un valor esperado igual a $[(n-1)/n]\sigma^2$. (Ver el teorema 13.4.) Luego un estimador insesgado de σ^2 se obtiene al tomar

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2.$$

Observaciones: (a) Aunque dividir por $(n-1)$ en vez de n es distinto cuando n es relativamente pequeño, para un n grande la diferencia es pequeña cualquiera que sea el estimador que se use.

(b) El ejemplo 14.3 ilustra una situación muy común: puede suceder que un estimador $\hat{\beta}$, de un parámetro β , sea sesgado en el sentido $E(\hat{\beta}) = k\beta$; en tal caso consideraremos simplemente al nuevo estimador $\hat{\beta}/k$, que será insesgado.

(c) Puede demostrarse, aunque no lo haremos aquí, que el estimador anterior de σ^2 es convergente.

EJEMPLO 14.4. En la tabla 14.1 reproducimos los datos obtenidos en el famoso experimento realizado por Rutherford [Rutherford y Geiger, *Phil. Mag.* S6, 20, 698 (1910)] sobre la emisión de partículas α por una fuente radiactiva.

TABLA 14.1

k	0	1	2	3	4	5	6	7	8	9	10	11	Total
n_k	57	203	383	525	532	408	273	139	49	27	10	6	2612

En la tabla, k es el número de partículas observadas en una unidad de tiempo (unidad = $\frac{1}{8}$ minuto), mientras que n_k es el número de intervalos en los cuales k partículas fueron observadas. Si hacemos que X sea el número de partículas emitidas durante el intervalo de tiempo de $\frac{1}{8}$ (minuto) de duración, y suponemos que X sigue la distribución de Poisson, tenemos

$$P(X = k) = \frac{e^{-(1/8)\lambda} (\frac{1}{8}\lambda)^k}{k!}.$$

Puesto que $E(\bar{X}) = \frac{1}{8}\lambda$, podemos usar el promedio muestral para obtener un estimador insesgado de $\frac{1}{8}\lambda$. Para λ , obtenemos entonces el estimador $\hat{\lambda} = 8\bar{X}$.

Para calcular \bar{X} simplemente evaluamos

$$\frac{\sum_{k=0}^{11} kn_k}{\sum_{k=0}^{11} n_k} = 3,87 \text{ partículas.}$$

Por tanto un estimador insesgado de λ , basado en el promedio muestral es igual a 30,96 (que puede interpretarse como el número esperado de partículas emitidas por minuto).

EJEMPLO 14.5. En la fabricación de explosivos pueden ocurrir cierto número de inflamaciones al azar. Sea X el número de inflamaciones por día. Suponiendo que X tiene una distribución de Poisson con parámetro λ . La tabla 14.2 da algunos datos que se van a emplear para la estimación de λ .

TABLA 14.2

Número de inflamaciones, k	0	1	2	3	4	5	6	Total
Número de días con k inflamaciones, n_k	75	90	54	22	6	2	1	250

Usando nuevamente el promedio muestral para la estimación de λ , obtenemos

$$\hat{\lambda} = \frac{\sum_k kn_k}{\sum_k n_k} = 1,22 \text{ número de inflamaciones por día.}$$

EJEMPLO 14.6. Se ha establecido que el contenido de ceniza en el carbón está distribuido normalmente con parámetros μ y σ^2 . Los datos de la tabla 14.3 representan el contenido de ceniza en 250 muestras de carbón analizadas. [Datos obtenidos de E. S. Grummel and A. C. Dunningham (1930); *British Standards Institution 403, 17.*] n_x es el número de muestras que tienen un porcentaje x de ceniza.

A fin de estimar los parámetros μ y σ^2 , usaremos los estimadores insesgados presentados antes:

$$\hat{\mu} = \frac{\sum_x xn_x}{250} = 16,998, \quad \hat{\sigma}^2 = \frac{1}{249} \sum_x n_x(x - \hat{\mu})^2 = 7,1.$$

TABLA 14.3

Contenido de ceniza en el carbón											
x	9,25	9,75	10,25	10,75	11,25	11,75	12,25	12,75	13,25	13,75	
n_x	1	0	2	1	1	2	5	4	7	6	
x	14,25	14,75	15,25	15,75	16,25	16,75	17,25	17,75	18,25	18,75	
n_x	13	14	15	13	24	15	19	23	22	12	
x	19,25	19,75	20,25	20,75	21,25	21,75	22,25	22,75	23,25	23,75	
n_x	12	7	6	8	6	4	2	2	0	3	
x	24,25	24,75	25,25								
n_x	0	0	1								
Total de muestras	250										

Observación: supóngase que tenemos un estimador insesgado $\hat{\theta}$, de un parámetro θ . Puede suceder que estemos interesados sólo en estimar una función $g(\theta)$, de θ . [Por ejemplo, si X está distribuida exponencialmente con parámetro θ , posiblemente estaremos interesados en $1/\theta$, es decir, $E(X)$]. Debería suponerse que todo lo que necesitamos hacer es considerar $1/\hat{\theta}$ o $(\hat{\theta})^2$, por ejemplo, como el estimador insesgado apropiado de $1/\theta$ o $(\theta)^2$. Enfáticamente esto no es así. En realidad, una de las desventajas del criterio de insesgamiento es que si hemos encontrado un estimador insesgado de θ , en general debemos partir del comienzo para encontrar una estimación para $g(\theta)$. Sólo si $g(\theta) = a\theta + b$, es decir, si g es una función lineal de θ , es cierto que $E[g(\hat{\theta})] = g[E(\hat{\theta})]$. En general, $E[g(\hat{\theta})] \neq g[E(\hat{\theta})]$. Supóngase por ejemplo, que X es una variable aleatoria con $E(X) = \mu$ y $V(X) = \sigma^2$. Hemos visto que el promedio muestral \bar{X} es un estimador insesgado de μ . ¿Es \bar{X}^2 un estimador insesgado de $(\mu)^2$? La respuesta es «no» como lo indica el cálculo siguiente. Puesto que $V(\bar{X}) = E(\bar{X})^2 - (E(\bar{X}))^2$, tenemos

$$E(\bar{X})^2 = V(\bar{X}) + (E(\bar{X}))^2 = \sigma^2/n + (\mu)^2 \neq (\mu)^2.$$

Aunque los ejemplos anteriores demuestran muy convincentemente, que en general

$$E[g(\hat{\theta})] \neq g[E(\hat{\theta})],$$

resulta que en muchos casos la igualdad es válida, aproximadamente al menos, especialmente si el tamaño de muestra es grande. Así, en el ejemplo anterior encontramos que $E(\bar{X}) = \mu$ y $E(\bar{X})^2 = \mu^2 + \sigma^2/n$, [con $\hat{\theta} = \bar{X}$ y $g(z) = z^2$], que aproximadamente es igual a μ^2 si n es grande.

14.4 Estimadores de máxima verosimilitud

Sólo hemos considerado ciertos criterios con los cuales podemos juzgar un estimador. Es decir, conocido un estimador que se propone para un parámetro desconocido, podemos verificar si es insesgado y convergente, y podemos calcular

(al menos en principio) su varianza y compararla con la varianza de otro estimador. Sin embargo, no tenemos aún un procedimiento general con el cual podemos encontrar un estimador «razonable». Varios de tales procedimientos existen, y expondremos uno de ellos, llamado el método de la máxima verosimilitud. En muchos casos este método da estimadores razonables.

A fin de evitar la repetición de nuestra exposición para el caso discreto y el continuo, convengamos en la terminología siguiente para los propósitos de la exposición presente.

Escribiremos $f(x; \theta)$ tanto para la fdp de X (calculada en x) como para $P(X = x)$ si X es discreta. Incluimos θ (en la notación) a fin de acordarnos que la distribución de probabilidades de X depende del parámetro θ en el cual estamos interesados.

Sea X_1, \dots, X_n una muestra aleatoria de la variable aleatoria X y sean x_1, \dots, x_n los valores muestrales. Definamos la *función de probabilidad* L como la siguiente función de la muestra y de θ .

$$L(X_1, \dots, X_n; \theta) = f(X_1; \theta)f(X_2; \theta) \cdots f(X_n; \theta). \quad (14.1)$$

Si X es discreta, $L(x_1, \dots, x_n; \theta)$ representa $P[X_1 = x_1, \dots, X_n = x_n]$ mientras que si X es continua, $L(x_1, \dots, x_n; \theta)$ representa la fdp conjunta de (X_1, \dots, X_n) .

Si se ha obtenido la muestra (X_1, \dots, X_n) , los valores muestrales (x_1, \dots, x_n) son conocidos. Puesto que θ es desconocido, podríamos hacernos la siguiente pregunta. ¿Para qué valor de θ será mayor $L(x_1, \dots, x_n; \theta)$? Poniendo distintamente, supongamos que tenemos dos valores de θ , sean θ_1 y θ_2 . Supongamos además que $L(x_1, \dots, x_n; \theta_1) < L(x_1, \dots, x_n; \theta_2)$. Preferiríamos entonces θ_2 a θ_1 para los valores muestrales dados (x_1, \dots, x_n) . Porque, si θ_2 es realmente el valor verdadero de θ , entonces la probabilidad de obtener valores muestrales tales como los que teníamos es mayor que si θ_1 fuese el valor verdadero de θ . Informalmente, preferimos el valor del parámetro que hace tan probable como sea posible que el suceso en realidad ocurrió. Es decir, deseamos elegir el valor más probable de θ después de obtener los datos, suponiendo que cada valor de θ fuese igualmente posible antes que los datos fuesen obtenidos. Hagamos la siguiente definición formal.

Definición. El estimador de máxima verosimilitud de θ , llamado $\hat{\theta}$, basado en una muestra aleatoria X_1, \dots, X_n es el valor de θ que maximiza a $L(X_1, \dots, X_n; \theta)$, considerando como una función de θ para una muestra dada X_1, \dots, X_n , en donde L está definida por la ecuación (14.1). (Este habitualmente se designa como el estimador *ML*.)

Observaciones: (a) $\hat{\theta}$ será naturalmente un estadígrafo y por tanto una variable aleatoria puesto que su valor dependerá de la muestra (X_1, \dots, X_n) . (No consideraremos como solución una constante.)

(b) En la mayor parte de nuestros ejemplos θ representará un solo número real. Sin embargo, puede suceder que la distribución de probabilidades de X dependa de dos o más valores paramétricos (como se hace en la distribución normal, por ejemplo). En tal caso θ puede representar un vector, $\theta = (\alpha, \beta)$ o $\theta = (\alpha, \beta, \gamma)$, etc.

(c) A fin de encontrar el estimador ML debemos determinar el valor máximo de una función. Así, en muchos problemas podemos aplicar algunas de las técnicas estándar del cálculo para encontrar este máximo. Puesto que $\ln x$ es una función creciente de x .

$$\ln L(X_1, \dots, X_n; \theta)$$

obtendrá su valor máximo para el mismo valor de θ que lo obtendrá $L(X_1, \dots, X_n; \theta)$. Luego bajo condiciones muy generales, suponiendo que θ es un número real y que $L(X_1, \dots, X_n; \theta)$ es una función diferenciable de θ , podemos obtener el estimador ML $\hat{\theta}$ al resolver lo que se conoce como la *ecuación de verosimilitud*.

$$\frac{\partial}{\partial \theta} \ln L(X_1, \dots, X_n; \theta) = 0 \quad (14.2)$$

Si $\theta = (\alpha, \beta)$, la ecuación anterior debe sustituirse por las ecuaciones de probabilidad simultáneas

$$\begin{aligned} \frac{\partial}{\partial \alpha} \ln L(X_1, \dots, X_n; \alpha, \beta) &= 0, \\ \frac{\partial}{\partial \beta} \ln L(X_1, \dots, X_n; \alpha, \beta) &= 0. \end{aligned} \quad (14.3)$$

Nuevamente se insistirá en que el enfoque anterior no siempre es útil. Sin embargo, en un gran número de ejemplos importantes (algunos de los cuales presentaremos brevemente) este método proporciona el estimador ML pedido con relativa facilidad.

Propiedades de los estimadores de máxima verosimilitud:

(a) El estimador ML puede ser sesgado. Muy a menudo tal sesgo puede evitarse multiplicando por una constante apropiada.

(b) Bajo condiciones muy generales, los estimadores ML son convergentes. Es decir, si los tamaños de muestra sobre los cuales se basan es grande, el estimador ML estará «próximo» al valor del parámetro que se estima. (Los estimadores ML poseen otra propiedad muy importante, la propiedad del «gran tamaño de muestra» que se expondrá posteriormente.)

(c) Los estimadores ML poseen la muy importante *propiedad de invarianza*. Supóngase que $\hat{\theta}$ es el estimador ML de θ . Puede demostrarse que el estimador ML de $g(\theta)$ es $g(\hat{\theta})$. Es decir, si el estadístico A toma sus medidas en pies² y el estadístico B mide en pies y si el estimador ML de A es $\hat{\theta}$, entonces el de B sería $\sqrt{\hat{\theta}}$. Recordamos que esta propiedad no la tienen los estimadores insesgados.

Consideraremos ahora ciertas aplicaciones importantes de los estimadores ML.

EJEMPLO 14.7. Suponiendo que el tiempo para fallar T , de una componente tiene una distribución exponencial con parámetro β . La fdp de T por lo tanto está dada por

$$f(t) = \beta e^{-\beta t}, \quad t \geq 0.$$

Supóngase que se prueban n de tales componentes, dando los tiempos de falla T_1, \dots, T_n .

La función de verosimilitud de esta muestra por lo tanto está dada por

$$L(T_1, \dots, T_n; \beta) = \beta^n \exp\left(-\beta \sum_{i=1}^n T_i\right).$$

Luego $\ln L = n \ln \beta - \beta \sum_{i=1}^n T_i$. Por tanto,

$$\frac{\partial \ln L}{\partial \beta} = \frac{n}{\beta} - \sum_{i=1}^n T_i \quad \text{y} \quad \frac{\partial \ln L}{\partial \beta} = 0$$

da $\hat{\beta} = 1/\bar{T}$ donde \bar{T} es el promedio muestral de los tiempos de falla. Puesto que el valor esperado de T , el tiempo promedio para fallar, está dado por $1/\beta$, usando la propiedad de invarianza de los estimadores ML, encontramos que el estimador ML de $E(T)$ está dado por \bar{T} , el promedio muestral. Sabemos que $E(\bar{T}) = 1/\beta$ y por tanto \bar{T} , el estimador ML de $E(T)$, es insesgado.

Observación: en general, no es fácil encontrar la distribución de probabilidades de los estimadores ML, especialmente si el tamaño de muestra es pequeño. (Si n es grande, encontraremos que es posible una solución general.) Sin embargo, en el ejemplo actual podemos obtener la distribución de los estimadores ML. Del corolario del teorema 10.9 encontramos que $2n\beta\bar{T}$ tiene distribución χ^2_{2n} . Luego $P(\bar{T} \leq t) = P(2n\beta\bar{T} \leq 2n\beta t)$. Esta probabilidad se lee directamente de la tabla de la distribución de χ^2 cuadrado si n, β , y t son conocidos.

EJEMPLO 14.8. Se sabe que cierta proporción (fija), p , de detonantes, es defectuosa. De una gran partida, se eligen n al azar y se prueban. Definamos las variables aleatorias siguientes.

$X_i = 1$ si el i -ésimo detonante es defectuoso y 0 en otro caso, $i = 1, 2, \dots, n$

Por tanto (X_1, \dots, X_n) es una muestra aleatoria de la variable aleatoria X que tiene la distribución de probabilidades $P(X = 0) = f(0; p) = 1 - p$, $P(X = 1) = f(1; p) = p$. Es decir, $f(x, p) = p^x(1 - p)^{1-x}$, $x = 0, 1$. Luego

$$L(X_1, \dots, X_n; p) = p^k(1 - p)^{n-k},$$

en donde $k = \sum_{i=1}^n X_i$ = número total de defectuosos. Luego $\ln L(X_1, \dots, X_n; p) = k \ln p + (n - k) \ln(1 - p)$. Por tanto

$$\frac{\partial \ln L}{\partial p} = \frac{k}{p} + \frac{n - k}{1 - p}(-1) = \frac{k}{p} - \frac{n - k}{1 - p}.$$

Si $k = 0$ o n encontramos directamente, al considerar la expresión de L , que el valor máximo de L se obtiene cuando $p = 0$ ó 1, respectivamente. Para $k \neq 0$ o n , hacemos $\partial \ln L / \partial p = 0$ y encontramos como solución $\hat{p} = k/n = \bar{X}$, el pro-

medio muestral. Luego nuevamente encontramos que el estimador ML da un estimador insesgado del parámetro buscado.

EJEMPLO 14.9. Supóngase que la variable aleatoria X está distribuida normalmente con esperanza μ y varianza 1. Es decir, la fdp de X está dada por

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-(1/2)(x-\mu)^2}.$$

Si (X_1, \dots, X_n) es una muestra aleatoria de X , la función de verosimilitud de la muestra es

$$L(X_1, \dots, X_n; \mu) = \frac{1}{(2\pi)^{n/2}} \exp \left[-\frac{1}{2} \sum_{i=1}^n (X_i - \mu)^2 \right].$$

Por lo tanto

$$\ln L = -\frac{n}{2} \ln(2\pi) - \frac{1}{2} \sum_{i=1}^n (X_i - \mu)^2 \quad \text{y} \quad \frac{\partial \ln L}{\partial \mu} = \sum_{i=1}^n (X_i - \mu).$$

Luego $\partial \ln L / \partial \mu = 0$ da $\hat{\mu} = \bar{X}$, el promedio muestral.

EJEMPLO 14.10. Hasta ahora hemos considerado situaciones en las cuales pudimos encontrar el valor máximo de L al derivar simplemente L (o $\ln L$) con respecto al parámetro y hacer esta derivada igual a cero. Que esto no siempre es efectivo, se ilustra con el ejemplo siguiente.

Supóngase que la variable aleatoria X está distribuida uniformemente en el intervalo $[0, \alpha]$ en donde α es un parámetro desconocido. La fdp de X está dada por

$$f(x) = \begin{cases} 1/\alpha, & 0 \leq x \leq \alpha, \\ 0, & \text{para cualquier otro valor.} \end{cases}$$

Si (X_1, \dots, X_n) es una muestra de X , su función de verosimilitud está dada por

$$L(X_1, \dots, X_n; \alpha) = \begin{cases} (1/\alpha)^n, & 0 \leq X_i \leq \alpha \text{ para todo } i, \\ 0, & \text{para cualquier otro valor.} \end{cases}$$

Considerando L como una función de α para (X_1, \dots, X_n) dada, se observa que debemos tener $\alpha \geq X_i$ para todo i a fin de que L sea distinta de cero. Esto es equivalente a pedir que $\alpha \geq \max(X_1, \dots, X_n)$. Luego, si dibujamos L como una función de α obtenemos el gráfico que se muestra en la figura 14.3. De este gráfico, es inmediatamente evidente qué valor de α maximiza L , es decir

$$\hat{\alpha} = \max(X_1, \dots, X_n).$$

Estudiemos algunas propiedades de este estimador. Del teorema 13.2 obtenemos la fdp de $\hat{\alpha}$: $g(\hat{\alpha}) = n[F(\hat{\alpha})]^{n-1} f(\hat{\alpha})$. Pero $F(x) = x/\alpha$, $0 \leq x \leq \alpha$, y $f(x)$ están dados anteriormente. Luego obtenemos

$$g(\hat{\alpha}) = n \left[\frac{\hat{\alpha}}{\alpha} \right]^{n-1} \left(\frac{1}{\alpha} \right) = \frac{n(\hat{\alpha})^{n-1}}{\alpha^n}, \quad 0 \leq \hat{\alpha} \leq \alpha.$$

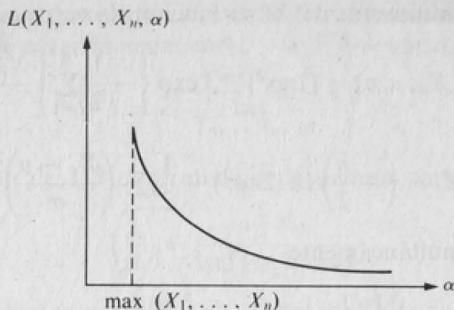


FIGURA 14.3

Para encontrar $E(\hat{\alpha})$ calculamos

$$\begin{aligned} E(\hat{\alpha}) &= \int_0^{\hat{\alpha}} \hat{\alpha} g(\hat{\alpha}) d\hat{\alpha} = \int_0^{\hat{\alpha}} \hat{\alpha} \frac{n \hat{\alpha}^{n-1}}{\alpha^n} d\hat{\alpha} \\ &= \frac{n}{\alpha^n} \frac{\hat{\alpha}^{n+1}}{n+1} \Big|_0^{\hat{\alpha}} = \frac{n}{n+1} \hat{\alpha}. \end{aligned}$$

Luego $\hat{\alpha}$ no es un estimador insesgado de α ; tiende a «subestimar» α . Si queremos un estimador insesgado, podemos usar

$$\hat{\alpha} = \frac{n+1}{n} \max(X_1, \dots, X_n).$$

Observemos que aunque $E(\hat{\alpha}) \neq \alpha$ tenemos que $\lim_{n \rightarrow \infty} E(\hat{\alpha}) = \alpha$. Así, para verificar la convergencia debemos demostrar aún que $V(\hat{\alpha}) \rightarrow 0$ cuando $n \rightarrow \infty$. (Ver el teorema 14.1.) Debemos calcular $E(\hat{\alpha})^2$:

$$\begin{aligned} E(\hat{\alpha})^2 &= \int_0^{\hat{\alpha}} (\hat{\alpha})^2 g(\hat{\alpha}) d\hat{\alpha} = \int_0^{\hat{\alpha}} (\hat{\alpha})^2 \frac{n(\hat{\alpha})^{n-1}}{\alpha^n} d\hat{\alpha} \\ &= \frac{n}{\alpha^n} \frac{(\hat{\alpha})^{n+2}}{n+2} \Big|_0^{\hat{\alpha}} = \frac{n}{n+2} \alpha^2. \end{aligned}$$

Luego

$$V(\hat{\alpha}) = E(\hat{\alpha})^2 - (E(\hat{\alpha}))^2 = \frac{n}{n+2} \alpha^2 - \left[\frac{n}{n+1} \alpha \right]^2 = \alpha^2 \left[\frac{n}{(n+2)(n+1)^2} \right].$$

Luego $V(\hat{\alpha}) \rightarrow 0$ cuando $n \rightarrow \infty$ y por tanto la convergencia está demostrada.

EJEMPLO 14.11. Consideremos un ejemplo en el cual los dos parámetros (ambos desconocidos) especifican la distribución. Supóngase que X tiene distribución $N(\mu, \sigma^2)$. Luego la fdp de X es

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2\right).$$

Si (X_1, \dots, X_n) es una muestra de X , su función de verosimilitud está dada por

$$L(X_1, \dots, X_n; \mu, \sigma) = (2\pi\sigma^2)^{-n/2} \exp \left\{ -\frac{1}{2} \sum_{i=1}^n \left[\frac{X_i - \mu}{\sigma} \right]^2 \right\}.$$

Luego

$$\ln L = \left(-\frac{n}{2} \right) \ln (2\pi\sigma^2) - \frac{1}{2} \sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma} \right)^2.$$

Debemos resolver simultáneamente

$$\frac{\partial \ln L}{\partial \mu} = 0 \quad \text{y} \quad \frac{\partial \ln L}{\partial \sigma} = 0.$$

Tenemos

$$\frac{\partial \ln L}{\partial \mu} = \sum_{i=1}^n \frac{(X_i - \mu)}{\sigma^2} = 0,$$

lo que da $\hat{\mu} = \bar{X}$, el promedio muestral. Y

$$\frac{\partial \ln L}{\partial \sigma} = -\frac{n}{\sigma} + \sum_{i=1}^n \frac{(X_i - \mu)^2}{\sigma^3} = 0,$$

que da

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2.$$

Observe que el método ML produce un estimador sesgado de σ^2 , puesto que ya hemos visto que un estimador insesgado es de la forma $1/(n-1)\sum_{i=1}^n (X_i - \bar{X})^2$.

EJEMPLO 14.12. Anteriormente consideramos (Ejemplo 14.7) el problema de estimar el parámetro β en una ley exponencial de falla, al probar n artículos y anotar sus tiempos de falla, T_1, \dots, T_n . Otro método podría ser el siguiente. Supóngase que sacamos n artículos, los probamos, y después que ha transcurrido cierto tiempo, T_0 horas, sencillamente contamos el número X de artículos que han fallado. Nuestra muestra consiste en X_1, \dots, X_n , en donde $X_i = 1$ si el i -ésimo artículo ha fallado en el período especificado, y 0 de otra manera. Luego la función de verosimilitud de la muestra es

$$L(X_1, \dots, X_n; \beta) = p^k (1-p)^{n-k},$$

en donde $k = \sum_{i=1}^n X_i$ = número de artículos que han fallado y $p = \text{Prob}$ (el artículo falla). Ahora p es una función del parámetro que se está estimando; es decir,

$$p = P(T \leq T_0) = 1 - e^{-\beta T_0}.$$

Utilizando el resultado del ejemplo 14.8, encontramos que el estimador ML de p es $\hat{p} = k/n$. Aplicando la propiedad de invarianza del estimador ML (ob-

servando que p es una función creciente de β , obtenemos el estimador ML de β sencillamente al resolver la ecuación $1 - e^{-\beta T_0} = k/n$. Un cálculo fácil da

$$\hat{\beta} = -\frac{1}{T_0} \ln\left(\frac{n-k}{n}\right),$$

o, para la estimación de $1/\beta$ el promedio de tiempo para fallar,

$$\left(\frac{1}{\beta}\right) = \frac{-T_0}{\ln[(n-k)/n]}.$$

En todos los ejemplos presentados, el método ML da ecuaciones que eran relativamente sencillas de resolver. No es lo mismo en muchos problemas y a menudo debemos recurrir a métodos numéricos (aproximados) a fin de obtener los estimadores. El ejemplo siguiente ilustra tales dificultades.

EJEMPLO 14.13. Como ya lo observamos, la distribución Gama tiene aplicaciones importantes para probar la duración. Supongamos, por ejemplo, que el tiempo para fallar de un generador eléctrico tiene una duración X cuya fdp está dada por

$$f(x) = \frac{\lambda^r x^{r-1}}{\Gamma(r)} e^{-\lambda x}, \quad x \geq 0,$$

en donde r y λ son dos parámetros positivos que deseamos estimar. Supóngase que es posible una muestra (X_1, \dots, X_n) de X . (Es decir, se han probado n generadores y anotado sus tiempos para fallar.) La función de verosimilitud de la muestra es

$$L(X_1, \dots, X_n; \lambda, r) = \frac{(\lambda)^nr(\prod_{i=1}^n X_i)^{r-1} \exp(-\lambda \sum_{i=1}^n X_i)}{[\Gamma(r)]^n},$$

$$\ln L = nr \ln \lambda + (r-1) \sum_{i=1}^n \ln X_i - \lambda \sum_{i=1}^n X_i - n \ln \Gamma(r).$$

Luego debemos resolver simultáneamente, $\partial \ln L / \partial \lambda = 0$ y $\partial \ln L / \partial r = 0$. Esas ecuaciones llegan a ser

$$\frac{\partial \ln L}{\partial \lambda} = \frac{nr}{\lambda} - \sum_{i=1}^n X_i = 0,$$

$$\frac{\partial \ln L}{\partial r} = n \ln \lambda + \sum_{i=1}^n \ln X_i - n \frac{\Gamma'(r)}{\Gamma(r)} = 0.$$

Luego $\partial \ln L / \partial \lambda = 0$ da directamente $\hat{\lambda} = r/\bar{X}$. Por tanto, después de sustituir λ por $\hat{\lambda}$, encontramos que $\partial \ln L / \partial r = 0$ da

$$\ln r - \frac{\Gamma'(r)}{\Gamma(r)} = \ln \bar{X} - \frac{1}{n} \sum_{i=1}^n \ln X_i.$$

Es evidente que debemos resolver la ecuación anterior para r , obtener \hat{r} , y luego $\hat{\lambda} = \hat{r}/\bar{X}$. Afortunadamente, la función $\Gamma'(r)/\Gamma(r)$ ha sido tabulada. Un mé-

todo muy rápido para obtener las soluciones pedidas ha sido presentado en un artículo por D. G. Chapman (*Annals of Mathematical Statistics*, 27, 498-506, 1956). Este ejemplo muestra que la solución de las ecuaciones de verosimilitud puede conducir a dificultades matemáticas considerables.

Como lo hemos mencionado anteriormente, los estimadores ML poseen otra propiedad que los hace muy apreciados, especialmente si las estimaciones se hacen en una muestra muy grande.

Propiedad asintótica de los estimadores de máxima verosimilitud. Si $\hat{\theta}$ es un estimador ML para el parámetro θ , definido sobre una muestra aleatoria X_1, \dots, X_n de una variable aleatoria X , entonces para n suficientemente grande, la variable aleatoria $\hat{\theta}$ tiene aproximadamente la distribución

$$N\left(\theta, \frac{1}{B}\right), \quad \text{en donde } B = nE\left[\frac{\partial}{\partial\theta} \ln f(X; \theta)\right]^2; \quad (14.4)$$

aquí f es la función de probabilidad puntual o fdp de X , dependiendo si X es discreta o continua y en donde θ se supone que es un número real.

Observaciones: (a) La propiedad anterior, por cierto, es mucho más fuerte que la propiedad de convergencia que hemos mencionado antes. La convergencia expresa que si n es suficientemente grande, $\hat{\theta}$ estará «próximo» a θ . Ahora esa propiedad nos describe cuál es la conducta probabilística de $\hat{\theta}$ para una n grande.

(b) No demostraremos esta afirmación, sino que simplemente ilustraremos su uso con un ejemplo.

EJEMPLO 14.14. Reconsideremos el ejemplo 14.7. Encontramos que $\hat{\beta} = 1/\bar{T}$ es el estimador ML de β . La fdp de T fue dada por $f(t; \beta) = \beta e^{-\beta t}$, $t > 0$. La propiedad anterior nos dice que si n es suficientemente grande, $\hat{\beta} = 1/\bar{T}$ tiene aproximadamente la distribución $N(\beta, 1/B)$, en donde B está dado por la ecuación (14.4).

Para encontrar B , consideremos $\ln f(T; \beta) = \ln \beta - \beta T$. Luego

$$(\partial/\partial\beta) \ln f(T; \beta) = (1/\beta) - T.$$

Por tanto

$$\left[\frac{\partial}{\partial\beta} \ln f(T; \beta)\right]^2 = \frac{1}{\beta^2} - \frac{2T}{\beta} + T^2.$$

Puesto que

$$E(T) = 1/\beta \quad \text{y} \quad E(T^2) = V(T) + [E(T)]^2 = 1/\beta^2 + 1/\beta^2 = 2/\beta^2,$$

tenemos

$$E\left[\frac{\partial}{\partial\beta} \ln f(T; \beta)\right]^2 = \frac{1}{\beta^2} - \frac{2}{\beta} \frac{1}{\beta} + \frac{2}{\beta^2} = \frac{1}{\beta^2}.$$

Luego encontramos que $\hat{\beta}$ tiene aproximadamente la distribución $N(\beta, \beta^2/n)$, para un n grande. (Esto verifica la propiedad de convergencia del estimador puesto que $\beta^2/n \rightarrow 0$ cuando $n \rightarrow \infty$.)

TABLA 14.4

X (altura, m)	Y (temperatura, °C)	X (altura, m)	Y (temperatura, °C)
1142	13	1008	13
1742	7	208	18
280	14	439	14
437	16	1471	14
678	13	482	18
1002	11	673	13
1543	4	407	16
1002	9	1290	7
1103	5	1609	6
475	11	910	9
1049	10	1277	1
566	15	410	14
995	10		

14.5 El método de los mínimos cuadrados

EJEMPLO 14.15. Estamos familiarizados con el hecho de que la temperatura del aire disminuye con la altura del lugar. Los datos dados en la tabla 14.4 y el *diagrama de dispersión* asociado (gráfico de puntos) (figura 14.4) lo refuerzan. El gráfico de puntos no sólo indica que la temperatura Y disminuye con la altura X , sino que una relación lineal es evidente.

Las observaciones representan la altura (metros) y la temperatura (grados centígrados) en las primeras horas de la mañana en cierto número de puestos de observación en Suiza. Los datos provienen del Observatorio Basel-St. Margarathen.

¿Cuál es un modelo razonable para los datos anteriores? Supondremos que Y es una variable aleatoria cuyo valor depende, entre otras cosas, del valor de X . Supondremos, específicamente, que

$$Y = \alpha X + \beta + \epsilon,$$

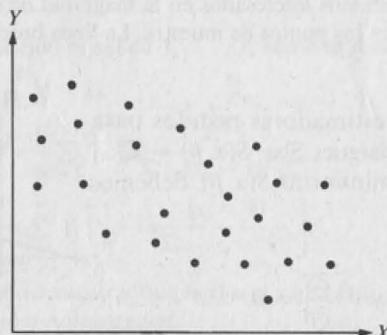


FIGURA 14.4

en donde α y β son constantes (desconocidas), X es la altura (conocida) desde la cual se mide Y , y ϵ es una variable aleatoria. El análisis de este *modelo lineal* depende de las hipótesis que hagamos acerca de la variable aleatoria ϵ . (Esencialmente decimos que la temperatura es un resultado aleatorio cuyo valor puede descomponerse estrictamente en una componente aleatoria más un término que depende de la altura X de una manera lineal.) La hipótesis que haremos acerca de ϵ es la siguiente:

$$E(\epsilon) = 0; \quad V(\epsilon) = \sigma^2 \quad \text{para todo } X.$$

Es decir, el valor esperado y la variancia de ϵ no dependen del valor de X . Luego $E(Y) = \alpha X + \beta$ y $V(Y) = \sigma^2$. Observemos que el modelo que hemos establecido depende de los tres parámetros α , β y σ^2 . No podemos usar el método de la máxima verosimilitud para estimar esos parámetros a no ser que hagamos más hipótesis acerca de la distribución de ϵ . Nada se ha supuesto acerca de la distribución de la variable aleatoria ϵ ; sólo se formuló una hipótesis acerca de su valor esperado y su varianza. (Posteriormente mencionaremos una modificación de este modelo.)

Antes de establecer cómo estimar los parámetros respectivos, digamos unas pocas palabras acerca del significado de una muestra aleatoria en el contexto presente. Supongamos que se escogen n valores de X , x_1, \dots, x_n . (Recordemos otra vez que aquí X no es una variable aleatoria.) Para cada x_i , sea Y_i una observación independiente de la variable aleatoria Y descrita anteriormente. Por tanto $(x_1, Y_1), \dots, (x_n, Y_n)$ puede ser considerada como una muestra aleatoria de la variable aleatoria Y para los valores (x_1, \dots, x_n) de X dados.)

Definición. Supóngase que tenemos $E(Y) = \alpha X + \beta$, en donde α , β y X son como se expresó anteriormente. Sea $(x_1, Y_1), \dots, (x_n, Y_n)$ una muestra aleatoria de Y . Los *estimadores de cuadrados mínimos* de α y β son los valores α y β que minimizan

$$\sum_{i=1}^n [Y_i - (\alpha x_i + \beta)]^2.$$

Observación: la interpretación del criterio anterior es muy evidente. (Ver la figura 14.5.) Para cada par (x_i, Y_i) calculamos la discrepancia entre Y_i , el valor observado, y $\alpha x_i + \beta$, el valor esperado. Puesto que sólo estamos interesados en la magnitud de esta discrepancia elevamos al cuadrado y sumamos todos los puntos de muestra. La línea buscada es aquella para la cual esta suma es más pequeña.

A fin de obtener los estimadores pedidos para α y β procedemos como sigue. Sea $S(\alpha, \beta) = \sum_{i=1}^n [Y_i - (\alpha x_i + \beta)]^2$. Para minimizar $S(\alpha, \beta)$, debemos resolver las ecuaciones

$$\frac{\partial S}{\partial \alpha} = 0 \quad \text{y} \quad \frac{\partial S}{\partial \beta} = 0.$$

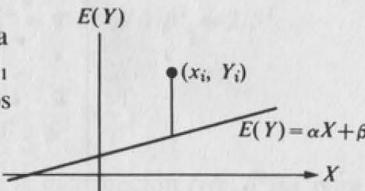


FIGURA 14.5

Derivando S respecto a α y a β , obtenemos

$$\frac{\partial S}{\partial \alpha} = \sum_{i=1}^n 2[Y_i - (\alpha x_i + \beta)](-x_i) = -2 \sum_{i=1}^n [x_i Y_i - \alpha x_i^2 - \beta x_i],$$

$$\frac{\partial S}{\partial \beta} = \sum_{i=1}^n 2[Y_i - (\alpha x_i + \beta)](-1) = -2 \sum_{i=1}^n [Y_i - \alpha x_i - \beta].$$

Luego $\partial S / \partial \alpha = 0$ y $\partial S / \partial \beta = 0$ pueden escribirse, respectivamente, como sigue:

$$\alpha \sum_{i=1}^n x_i^2 + \beta \sum_{i=1}^n x_i = \sum_{i=1}^n x_i Y_i, \quad (14.5)$$

$$\alpha \sum_{i=1}^n x_i + n\beta = \sum_{i=1}^n Y_i. \quad (14.6)$$

Tenemos así dos ecuaciones *lineales* en las incógnitas α y β . La solución puede obtenerse de la manera corriente, o por una eliminación directa o usando determinantes. Denotando las soluciones por $\hat{\alpha}$ y $\hat{\beta}$, encontramos fácilmente que

$$\hat{\alpha} = \frac{\sum_{i=1}^n Y_i(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}, \quad \text{en donde } \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \quad (14.7)$$

$$\hat{\beta} = \bar{Y} - \hat{\alpha}\bar{x}, \quad \text{en donde } \bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i. \quad (14.8)$$

Las soluciones anteriores son únicas y siempre se obtienen dado que

$$\sum_{i=1}^n (x_i - \bar{x})^2 \neq 0.$$

Esta condición, sin embargo, se satisface cada vez que todos los x_i no son iguales.

El estimador del parámetro σ^2 no puede obtenerse por los métodos anteriores. Establezcamos simplemente que el estimador corriente de σ^2 , mediante los estimadores de mínimos cuadrados $\hat{\alpha}$ y $\hat{\beta}$, es

$$\hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^n [Y_i - (\hat{\alpha}x_i + \hat{\beta})]^2.$$

Observaciones: (a) $\hat{\alpha}$ es evidentemente una función *lineal* de los valores muestrales Y_1, \dots, Y_n .

(b) $\hat{\beta}$ también es una función lineal de Y_1, \dots, Y_n como lo indica el cálculo siguiente:

$$\begin{aligned} \hat{\beta} &= \bar{Y} - \hat{\alpha}\bar{x} \\ &= \frac{1}{n} \sum_{i=1}^n Y_i - \frac{\bar{x}}{\sum_{i=1}^n (x_i - \bar{x})^2} \sum_{i=1}^n (x_i - \bar{x}) Y_i \\ &= \sum_{i=1}^n Y_i \left[\frac{1}{n} - \bar{x} \frac{(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} \right]. \end{aligned}$$

(c) Es un ejercicio sencillo demostrar que $E(\hat{\alpha}) = \alpha$ y que $E(\hat{\beta}) = \beta$. (Ver el problema 14.34.) Así, $\hat{\alpha}$ y $\hat{\beta}$ son ambos estimadores insesgados.

(d) Las varianzas de $\hat{\alpha}$ y $\hat{\beta}$ también se pueden calcular fácilmente. (Ver el problema 14.35.)

Tenemos

$$V(\hat{\alpha}) = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}; \quad V(\hat{\beta}) = \left[\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right] \sigma^2 \quad (14.9)$$

(e) Los estimadores $\hat{\alpha}$ y $\hat{\beta}$ son realmente los mejores estimadores lineales insesgados de α y β . Es decir, entre todos los estimadores insesgados lineales, éstos tienen la varianza mínima. Este es un caso especial del *teorema general de Gauss-Markoff*, que establece que bajo ciertas condiciones los estimadores de cuadrados mínimos y los mejores estimadores lineales insesgados son siempre los mismos.

(f) El método de los cuadrados mínimos puede aplicarse a modelos no lineales. Por ejemplo, si $E(Y) = \alpha X^2 + \beta X + \gamma$, podemos estimar α , β y γ de modo que

$$\sum_{i=1}^n [Y_i - (\alpha x_i^2 + \beta x_i + \gamma)]^2$$

es minimizada.

(g) Si hacemos la hipótesis adicional de que la variable aleatoria ϵ tiene distribución $N(0, \sigma^2)$, podemos aplicar el método de la máxima verosimilitud para estimar los parámetros α y β . Esos estimadores son los mismos que los estimadores de cuadrados mínimos obtenidos anteriormente. (Esto no siempre es cierto, es una consecuencia de la hipótesis de normalidad.)

EJEMPLO 14.16. Este ejemplo es presentado por Y. V. Linnik en *Method of Least Squares and Principles of the Theory of Observations*, Pergamon Press, New York, 1961. Los datos de este ejemplo fueron obtenidos por Mendeléiev y presentados en *Foundations of Chemistry*. (Ver la tabla 14.5.) Relacionan la solubilidad de nitrato de sodio (NaNO_3) con la temperatura del agua ($^{\circ}\text{C}$). A la temperatura indicada, las Y partes de NaNO_3 se disuelven en 100 partes de agua. Graficando esos datos se obtiene el diagrama de dispersión que se muestra en la figura 14.6.

TABLA 14.5

T	Y	T	Y
0	66,7	29	92,9
4	71,0	36	99,4
10	76,3	51	113,6
15	80,6	68	125,1
21	85,7		

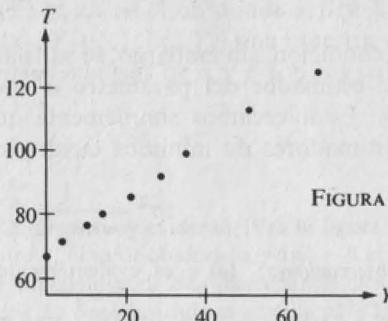


FIGURA 14.6

Este sugiere un modelo de la forma $E(Y) = bT + a$. Usando el método de los cuadrados mínimos bosquejado anteriormente, encontramos que $\hat{b} = 0,87$ y $\hat{a} = 67,5$.

14.6 El coeficiente de correlación

En la sección anterior estuvimos interesados en pares de valores (X, Y) , pero, como lo hemos señalado repetidamente, X no iba a ser considerada como una

variable aleatoria. Sin embargo, hay variables aleatorias bidimensionales (X, Y) que dan origen a una muestra aleatoria $(X_1, Y_1), \dots, (X_n, Y_n)$. Uno de los parámetros más importantes asociado con una variable aleatoria bidimensional es el coeficiente de correlación ρ_{xy} .

TABLA 14.6

X (velocidad, km/seg)	11,93	11,81	11,48	10,49	10,13	8,87
Y (altura, km)	62,56	57,78	53,10	48,61	44,38	40,57

El estimador que se acostumbra usar para ρ es el *coeficiente de correlación muestral*, definido como sigue:

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2}}.$$

Nótese que para propósitos computacionales, es más fácil calcular r como sigue:

$$r = \frac{n \sum_{i=1}^n X_i Y_i - \sum_{i=1}^n X_i \sum_{i=1}^n Y_i}{\sqrt{n \sum_{i=1}^n X_i^2 - (\sum_{i=1}^n X_i)^2} \sqrt{n \sum_{i=1}^n Y_i^2 - (\sum_{i=1}^n Y_i)^2}}.$$

EJEMPLO 14.17. Los datos anotados en la tabla 14.6 representan la velocidad (km/seg) y la altura (km) de la estrella fugaz núm. 1242 como se informó en la «Smithsonian Contributions to Astrophysics» en el *Proceedings of the Symposium on Astronomy and Physics of Meteors*. Cambridge, Mass., agosto 28-septiembre 1, 1961. Un cálculo directo da $r = 0,94$.

14.7 Intervalos de confianza

Hasta ahora sólo nos hemos interesado por la obtención de un estimador puntual para un parámetro desconocido. Como lo insinuamos al comienzo de este capítulo, existe otro enfoque que conduce a menudo a resultados muy significativos.

Supóngase que X tiene distribución $N(\mu, \sigma^2)$, donde σ^2 se supone conocido, mientras que μ es el parámetro desconocido. Sea X_1, \dots, X_n una muestra aleatoria de X y sea \bar{X} el promedio muestral.

Sabemos que \bar{X} tiene distribución $N(\mu, \sigma^2/n)$. Por tanto $Z = [(\bar{X} - \mu)/\sigma]/\sqrt{n}$ tiene una distribución $N(0, 1)$. Nótese que aunque Z depende de μ , su distribución de probabilidades no. Podemos usar este hecho a nuestra conveniencia como sigue.

Considerar

$$\begin{aligned} 2\Phi(z) - 1 &= P\left(-z \leq \frac{X - \mu}{\sigma} \sqrt{n} \leq z\right) \\ &= P\left(-\frac{z\sigma}{\sqrt{n}} - \bar{X} \leq -\mu \leq +\frac{z\sigma}{\sqrt{n}} - \bar{X}\right) \\ &= P\left(\bar{X} - \frac{z\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + \frac{z\sigma}{\sqrt{n}}\right). \end{aligned}$$

Esta última proposición probabilística debe interpretarse *muy cuidadosamente*. No significa que la probabilidad que el parámetro μ caiga en el intervalo mencionado sea igual a $2\Phi(z) - 1$; μ es un parámetro y está o no en el intervalo anterior.

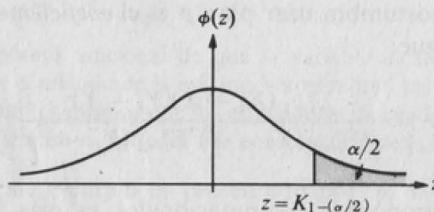


FIGURA 14.7

terior. En su lugar, lo anterior debería ser interpretado como sigue: $2\Phi(z) - 1$ es igual a la probabilidad que el intervalo aleatorio $(\bar{X} - z\sigma/\sqrt{n}, \bar{X} + z\sigma/\sqrt{n})$ contenga a μ . A tal intervalo se le llama *intervalo de confianza para el parámetro μ* . Puesto que z queda a nuestro criterio, podemos elegirlo de modo que la probabilidad anterior sea igual a $1 - \alpha$. Por tanto z está definida por la relación $\Phi(z) = 1 - \alpha/2$. Ese valor de z , denotado por $K_{1-\alpha/2}$, puede obtenerse de las tablas de la distribución normal. (Ver también la figura 14.7.) Es decir, tenemos $\Phi(K_{1-\alpha/2}) = 1 - \alpha/2$.

Para resumir: el intervalo $(\bar{X} - n^{-1/2}\sigma K_{1-\alpha/2}, \bar{X} + n^{-1/2}\sigma K_{1-\alpha/2})$ es un intervalo de confianza para el parámetro μ con *coeficiente de confianza* $(1 - \alpha)$, o un $(1 - \alpha)$ 100 por ciento de intervalo de confianza.

Supóngase que X representa la duración de una pieza de un equipo. Supóngase que se probaron 100 piezas, que dieron una duración promedio de $\bar{X} = 501,2$ horas. Se sabe que σ es 4 horas y que deseamos tener un intervalo del 95 por ciento de confianza para μ . Encontramos por tanto el siguiente intervalo de confianza para $\mu = E(X)$:

$$501,2 - \frac{4}{\sqrt{100}}(1,96), 501,2 + \frac{4}{\sqrt{100}}(1,96), \quad \text{llega a ser} \quad (500,4; 502,0)$$

Nuevamente es útil un comentario. Al establecer que $(500,4; 502,0)$ es un intervalo de 95 por ciento de confianza para μ , no estamos diciendo que el 95 por ciento de las veces el promedio muestral quedará en ese intervalo. La próxima vez que saquemos una muestra aleatoria, X será posiblemente distinto, y por tanto los extremos del intervalo de confianza serán diferentes. Estamos diciendo que el 95 por ciento de las veces μ estará contenido en el intervalo $(\bar{X} - 1,96\sigma/\sqrt{n}, \bar{X} + 1,96\sigma/\sqrt{n})$.

$1,96\sigma/\sqrt{n}$). Cuando hacemos la afirmación de que $500,4 < \mu < 502,0$ sencillamente estamos adoptando el criterio de creer que algo es así cuando sabemos que es verdadero la mayor parte del tiempo.

Observación: el intervalo de confianza construido no es único. Precisamente hay tantos como estimadores (puntuales) del parámetro, así que podemos construir muchos intervalos de confianza. Aunque no discutiremos el problema de lo que podríamos designar por un intervalo de confianza «mejor», establezcamos sin embargo un hecho obvio. Si se comparan dos intervalos de confianza que tienen el mismo coeficiente, preferiríamos el que tiene menos longitud.

La longitud L del intervalo de confianza considerado anteriormente puede escribirse como

$$L = (\bar{X} + n^{-1/2}\sigma K_{1-\alpha/2}) - (\bar{X} - n^{-1/2}\sigma K_{1-\alpha/2}) = 2\sigma n^{-1/2}K_{1-\alpha/2}.$$

Luego L es una constante. Además, resolviendo la ecuación anterior para n da

$$n = (2\sigma K_{1-\alpha/2}/L)^2.$$

Luego podemos determinar n (para α y σ dadas) de modo que el intervalo de confianza tenga una longitud prefijada. En general, (como se ilustró en el ejemplo anterior), L será una función decreciente de n : cuanto más pequeño deseemos tener a L más grande debe tomarse a n . En el caso anterior especialmente debemos cuadruplicar n a fin de dividir a L ,

14.8 La distribución *t* de Student

El análisis del ejemplo anterior dependía mucho del hecho de que la varianza σ^2 era conocida. ¿Cómo debemos modificar el procedimiento si no conocemos el valor de σ^2 ?

Supongamos que estimamos σ^2 utilizando el estimador insesgado

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2.$$

Consideremos la variable aleatoria

$$t = \frac{(\bar{X} - \mu)\sqrt{n}}{\hat{\sigma}}. \quad (14.10)$$

Debería ser intuitivamente evidente que la distribución de probabilidades de la variable aleatoria t es considerablemente más complicada que la de $Z = (\bar{X} - \mu)/\sqrt{n}/\sigma$. Ya que en la definición de t tanto el numerador como el denominador son variables aleatorias, mientras que Z sencillamente es una función lineal de X_1, \dots, X_n . Para obtener la distribución de probabilidades de t usemos los hechos siguientes:

- (a) $Z = (\bar{X} - \mu)/\sqrt{n}/\sigma$ tiene distribución $N(0, 1)$.
- (b) $V = \sum_{i=1}^n (X_i - \bar{X})^2/\sigma^2$ tiene una distribución de χ cuadrado con $(n - 1)$ grados de libertad. (Ver el teorema 13.4.)
- (c) Z y V son variables aleatorias independientes. (Esto no es muy fácil de demostrar, y no lo verificaremos aquí.) Con ayuda del teorema siguiente podemos obtener ahora la fdp de t .

Teorema 14.3. Suponiendo que las variables aleatorias Z y V son independientes y tienen respectivamente distribuciones $N(0,1)$ y χ_k^2 , Definimos

$$t = \frac{Z}{\sqrt{(V/k)}}.$$

Luego la fdp de t está dada por

$$h_k(t) = \frac{\Gamma[(k+1)/2]}{\Gamma(k/2)\sqrt{\pi k}} \left(1 + \frac{t^2}{k}\right)^{-(k+1)/2}, \quad -\infty < t < \infty. \quad (14.11)$$

Esta distribución se conoce como la *distribución t de Student* con k grados de libertad.

Observaciones: (a) La demostración de este teorema no es dada aquí pero se sugiere en la sección de problemas. (Ver el problema 14.17.) Tenemos las herramientas disponibles con las cuales podemos encontrar $h_k(t)$ muy fácilmente. Primero necesitamos determinar la fdp de $\sqrt{V/k}$, que es obtenida fácilmente conociendo la fdp de V . Sólo necesitamos aplicar el teorema 6.5 que da la fdp del cociente de dos variables aleatorias independientes.

(b) El teorema anterior puede ser aplicado directamente para obtener la fdp de $t = (\bar{X} - \mu)/\sqrt{n}/\hat{\sigma}$, la variable aleatoria considerada anteriormente. Esta variable tiene la distribución t de Student con $(n-1)$ grados de libertad. Note que aunque el valor de t depende de μ , su distribución no.

(c) El gráfico de h_k es simétrico, como se indicó en la figura 14.8. En realidad, se asemeja al gráfico de la distribución normal, y el lector puede demostrar que

$$\lim_{k \rightarrow \infty} h_k(t) = (1/\sqrt{2\pi})e^{-t^2/2}.$$

(d) Debido a su importancia esta distribución ha sido tabulada. (Ver el apéndice.) Para un α dado, $0.5 < \alpha < 1$, los valores de $t_{k,\alpha}$, satisfacen la condición

$$\int_{-\infty}^{t_{k,\alpha}} h_k(t) dt = \alpha$$

están tabulados. (Ver la figura 14.9.) (Para los valores de α que satisfacen $0 < \alpha < 0.5$, podemos usar los valores tabulados debido a la simetría de la distribución.)

(e) Esta distribución se llama así en honor del estadístico inglés W. S. Gosset, quien publicó su trabajo bajo el pseudónimo de «Student».

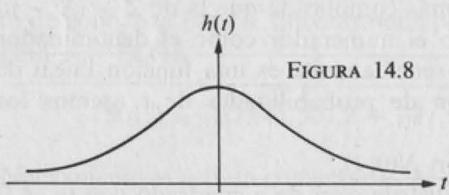


FIGURA 14.8

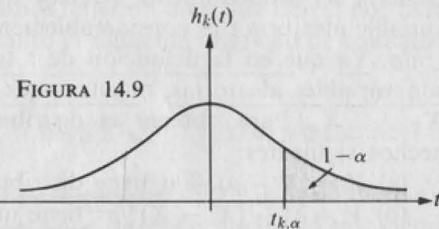


FIGURA 14.9

Volvamos ahora al problema presentado al comienzo de esta sección. ¿Cómo obtenemos un intervalo de confianza para el promedio de una variable aleatoria distribuida normalmente si la varianza es desconocida?

De una manera completamente análoga a la usada en la sección 14.7, obtenemos el siguiente intervalo de confianza para μ , con coeficiente de confianza $(1 - \alpha)$:

$$(\bar{X} - n^{-1/2}\hat{\sigma}t_{n-1,1-\alpha/2}, \bar{X} + n^{-1/2}\hat{\sigma}t_{n-1,1-\alpha/2}).$$

Luego el intervalo de confianza anterior tiene la misma estructura que el previo, con la diferencia importante de que el valor conocido de σ ha sido sustituido por el estimador $\hat{\sigma}$ y la constante $K_{1-\alpha/2}$, obtenida anteriormente de las tablas de la distribución normal, ha sido sustituida por $t_{n-1,1-\alpha/2}$, obtenida de las tablas de la distribución t .

Observación: el largo L del intervalo de confianza anterior es igual a

$$L = 2n^{-1/2}t_{n-1,1-\alpha/2}\hat{\sigma}.$$

Luego L no es una constante puesto que depende de $\hat{\sigma}$ que a su vez depende de los valores muestrales (X_1, \dots, X_n) .

EJEMPLO 14.18. Se hicieron diez mediciones sobre la resistencia de cierto tipo de alambre, que dieron los valores X_1, \dots, X_{10} . Supóngase que $\bar{X} = 10,48$ ohms y $\hat{\sigma} = \sqrt{\frac{1}{9} \sum_{i=1}^{10} (X_i - \bar{X})^2} = 1,36$ ohms. Supongamos que X tiene distribución $N(\mu, \sigma^2)$ y que deseamos obtener un intervalo de confianza para μ con coeficiente de confianza 0,90. Luego $\alpha = 0,10$. De las tablas de la distribución t encontramos que $t_{9,0.95} = 1,83$. Luego el intervalo de confianza pedido es

$$\left(10,48 - \frac{1}{\sqrt{10}}(1,36)(1,83), 10,48 + \frac{1}{\sqrt{10}}(1,36)(1,83) \right) = (9,69, 11,27).$$

14.9 Más sobre los intervalos de confianza

Aunque no intentemos dar una presentación general de este tema, deseamos continuar considerando algunos ejemplos importantes.

Algunas veces deseamos obtener un intervalo de confianza para una función particular de un parámetro desconocido, conociendo un intervalo de confianza para el parámetro mismo. Si la función es monótona esto puede satisfacerse como lo ilustra el ejemplo siguiente.

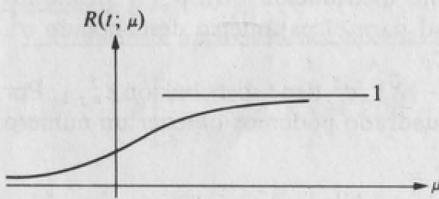


FIGURA 14.10

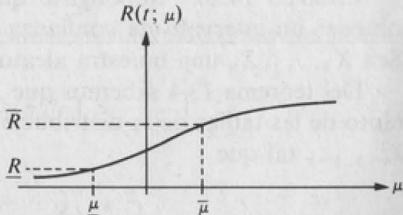


FIGURA 14.11

EJEMPLO 14.19. Suponiendo que la duración X de un artículo tiene distribución $N(\mu, \sigma^2)$. Se supone que σ^2 es conocido. La confiabilidad del artículo para un tiempo de servicio de t horas es dado por

$$R(t; \mu) = P(X > t) = 1 - \Phi\left(\frac{t - \mu}{\sigma}\right).$$

Puesto que $\partial R(t; \mu)/\partial \mu > 0$ para todo μ , tenemos que para cada t fijo, $R(t; \mu)$ es una función creciente de μ . (Ver la figura 14.10.) Luego podemos proceder como sigue para obtener un intervalo de confianza para $R(t; \mu)$. Sea $(\underline{\mu}, \bar{\mu})$ el intervalo de confianza para μ obtenido en la sección 14.7. Sean \underline{R} y \bar{R} respectivamente los extremos inferior y superior del intervalo de confianza pedido para $R(t; \mu)$. Si definimos a \underline{R} y \bar{R} por las relaciones

$$\underline{R} = 1 - \Phi\left(\frac{t - \underline{\mu}}{\sigma}\right) \quad \text{y} \quad \bar{R} = 1 - \Phi\left(\frac{t - \bar{\mu}}{\sigma}\right)$$

encontramos que $P(\underline{R} \leq R \leq \bar{R}) = P(\underline{\mu} \leq \mu \leq \bar{\mu}) = 1 - \alpha$, y que por tanto (\underline{R}, \bar{R}) no representa un intervalo de confianza para $R(t; \mu)$ con coeficiente de confianza $(1 - \alpha)$. (Ver figura 14.11.)

Usemos los valores muestrales obtenidos en la sección 14.7 para ilustrar este procedimiento. Supóngase que deseamos un intervalo de confianza para la confiabilidad de la componente cuando se usó durante $t = 500$ horas. Puesto que encontramos $\mu = 500,4$ y $\bar{\mu} = 502,0$, obtenemos

$$\underline{R} = 1 - \Phi\left(\frac{500 - 500,4}{4}\right) = 0,6554, \quad \bar{R} = 1 - \Phi\left(\frac{500 - 502}{4}\right) = 0,6915.$$

Hasta ahora sólo hemos considerado intervalos de confianza *bilaterales*. Es decir, hemos obtenido dos estadígrafos (llamados algunas veces cota superior e inferior de confianza), sean $L(X_1, \dots, X_n)$ y $(U(X_1, \dots, X_n)$, tales que $P[L \leq \theta \leq U] = 1 - \alpha$, en donde θ es el parámetro desconocido.

A menudo estamos interesados sólo en obtener intervalos de confianza *unilaterales* de la forma siguiente:

$$P[\theta \leq U] = 1 - \alpha \quad \text{o} \quad P[L \leq \theta] = 1 - \alpha.$$

Ilustremos lo anterior con ejemplos.

EJEMPLO 14.20. Supóngase que X tiene distribución $N(\mu, \sigma^2)$ y deseamos obtener un intervalo de confianza unilateral para el parámetro desconocido σ^2 . Sea X_1, \dots, X_n una muestra aleatoria de X .

Del teorema 13.4 sabemos que $\sum_{i=1}^n (X_i - \bar{X})^2 / \sigma^2$ tiene distribución χ^2_{n-1} . Por tanto de las tablas de la distribución de χ^2 cuadrado podemos obtener un número $\chi^2_{n-1, 1-\alpha}$ tal que

$$P\left[\sum_{i=1}^n \frac{(X_i - \bar{X})^2}{\sigma^2} \leq \chi^2_{n-1, 1-\alpha}\right] = 1 - \alpha.$$

(Ver la figura 14.12.) La probabilidad anterior puede escribirse como sigue:

$$P\left[\sigma^2 \geq \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\chi^2_{n-1, 1-\alpha}}\right] = 1 - \alpha.$$

Luego $(\sum_{i=1}^n (X_i - \bar{X})^2 / \chi^2_{n-1, 1-\alpha}, \infty)$ es el intervalo de confianza unilateral pedido para σ^2 con coeficiente de confianza $(1 - \alpha)$.

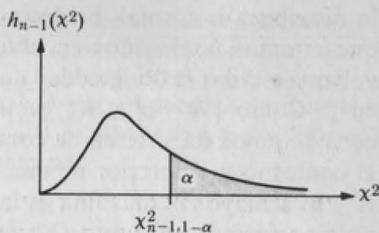


FIGURA 14.12

EJEMPLO 14.21. Supóngase que la duración X de un instrumento electrónico está distribuida exponencialmente con parámetro $1/\beta$. Luego $E(X) = \beta$. Sea X_1, \dots, X_n una muestra de X . Hemos encontrado en el ejemplo 14.7 que $\sum_{i=1}^n X_i/n$ es el estimador de β . Del colorario del teorema 10.9 encontramos que $2n\bar{X}/\beta$ tiene distribución χ^2_{2n} . Luego $P[2n\bar{X}/\beta \geq \chi^2_{2n, 1-\alpha}] = \alpha$, en donde el número $\chi^2_{2n, 1-\alpha}$ se obtiene de las tablas de distribución χ cuadrado.

Si deseamos un intervalo de confianza (inferior) para la confiabilidad $R(t; \beta) = P(X > t) = e^{-t/\beta}$, procedemos como sigue. Se multiplica la desigualdad anterior por $(-t)$ se reagrupan los términos, y se obtiene

$$P[(-t/\beta) \geq -t\chi^2_{2n, 1-\alpha}/\bar{X}2n] = 1 - \alpha.$$

Esto a su vez implica que, puesto que e^x es una función creciente de x ,

$$P\left\{R(t; \beta) = e^{-t/\beta} \geq \exp\left[-\frac{t\chi^2_{2n, 1-\alpha}}{\bar{X}2n}\right]\right\} = 1 - \alpha.$$

Luego $(\exp[-t\chi^2_{2n, 1-\alpha}/\bar{X}2n], \infty)$ es un intervalo de confianza unilateral para $R(t; \beta)$ con coeficiente de confianza $(1 - \alpha)$.

Como una ilustración final de un intervalo de confianza, encontramos un intervalo de confianza para el parámetro p asociado con una variable aleatoria X distribuida binomialmente. Sólo consideraremos el caso en donde n el número de repeticiones del experimento que da origen a X es bastante grande de modo que podamos usar la aproximación normal.

Representemos por $X/n = h$ la frecuencia relativa de un suceso A en n repeticiones de un experimento para el cual $P(A) = p$. Por tanto, $E(h) = p$ y $V(h) = pq/n$, en donde $q = 1 - p$.

Usando la aproximación normal a la distribución binomial, podemos escribir

$$\begin{aligned} P\left[\frac{|h - p|}{\sqrt{pq/n}} \leq K\right] &= P[|h - p| \leq K\sqrt{pq/n}] \simeq \frac{1}{\sqrt{2\pi}} \int_{-K}^{K} e^{-t^2/2} dt \\ &= 2\Phi(K) - 1, \end{aligned}$$

en donde, como siempre $\Phi(K) = (1/\sqrt{2\pi}) \int_{-\infty}^K e^{-t^2/2} dt$. Luego, si hacemos igual a $(1 - \alpha)$ la probabilidad anterior, podemos obtener el valor de K de la tabla de

la distribución normal. Es decir, $2\Phi(K) - 1 = 1 - \alpha$ implica $K = K_{1-\alpha/2}$. Puesto que estamos interesados en obtener un intervalo de confianza para p , debemos volver a escribir la desigualdad anterior $\{|h - p| \leq K\sqrt{pq/n}\}$ como una desigualdad en p . Como $\{|h - p| \leq K\sqrt{pq/n}\}$ es equivalente a $\{(h - p)^2 \leq K^2(1 - p)p/n\}$. Si consideramos un sistema de coordenadas (h, p) la desigualdad anterior representa el contorno y el interior de una ellipse. La forma de la ellipse es determinada por K y n ; a mayor n , más fina es la ellipse.

Considérese un punto $Q(h, p)$ en el plano hp . (Ver la figura 14.13.) Q será un punto «aleatorio» puesto que su primera coordenada h será determinada por el resultado del experimento. Puesto que Q quedará dentro de la ellipse si y sólo si $\{|h - p| \leq K\sqrt{pq/n}\}$, la probabilidad de que esto ocurra será $2\Phi(K) - 1$. Si

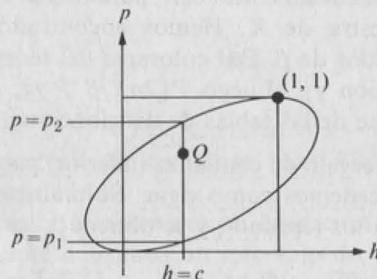


FIGURA 14.13

deseamos tener esta probabilidad igual a $(1 - \alpha)$ debemos elegir adecuadamente a K , es decir $K = K_{1-\alpha/2}$.

Actualmente p es desconocido. (Este, naturalmente, es nuestro problema.) La recta $h = c$ (constante) cortará la ellipse en dos lugares, sean $p = p_1$ y $p = p_2$. (Fácilmente puede verificarse que dados α y h , siempre habrá dos valores distintos de p .)

Los valores p_1 y p_2 pueden obtenerse como solución de la ecuación cuadrática (en p): $(h - p)^2 = K^2(1 - p)p/n$. Las soluciones son:

$$p_1 = \frac{hn + (K^2/2) - K[h(1 - h)n + (K^2/4)]^{1/2}}{n + K^2},$$

$$p_2 = \frac{hn + (K^2/2) + K[h(1 - h)n + (K^2/4)]^{1/2}}{n + K^2}. \quad (14.12)$$

Por tanto $\{|h - p| \leq K\sqrt{pq/n}\}$ es equivalente a $\{p_1 \leq p \leq p_2\}$. Luego si se escoge K de modo que el suceso anterior tenga probabilidad $(1 - \alpha)$, obtenemos un intervalo de confianza para p , sea (p_1, p_2) , con coeficiente de confianza $(1 - \alpha)$. Para resumir: con el objeto de obtener un intervalo de confianza para $p = P(A)$ con coeficiente de confianza $(1 - \alpha)$, realizar el experimento n veces y calcular la frecuencia relativa del suceso A , llamémosla h . Luego se calculan p_1 y p_2 de acuerdo con las ecuaciones (14.12), en donde K se determina de las tablas de la distribución

normal. Recordamos que este procedimiento es válido sólo si n es suficientemente grande para justificar la aproximación normal.

Observación: recuérdese que en la ecuación (14.12), $h = X/n$, en donde X = número de ocurrencias del suceso A . Luego $X = nh$ y $n - X = n(1 - h)$. Si además de n , tanto X y $n - X$ son grandes, p_1 y p_2 pueden aproximarse respectivamente, por

$$p_1 \approx h - \frac{K}{\sqrt{n}} \sqrt{h(1-h)}, \quad p_2 \approx h + \frac{K}{\sqrt{n}} \sqrt{h(1-h)}.$$

EJEMPLO 14.22. En cierto proceso de producción se fabrican 79 artículos durante cierta semana. De esos, se encontró que 3 eran defectuosos. Luego $h = 3/79 = 0,038$. Usando el procedimiento anterior, obtenemos $(0,013; 0,106)$ como un intervalo de confianza para $p = P(\text{el artículo es defectuoso})$ con un coeficiente de confianza 0,95.

EJEMPLO 14.23. Una fábrica tiene un gran almacenamiento de artículos, algunos de los cuales provienen de un método de producción considerado insatisfactorio actualmente, mientras que otros provienen de un proceso moderno. Del almacenamiento se escogen al azar 3.000 artículos. Se encuentra que de éstos 1.578 se han fabricado según el proceso insatisfactorio. Utilizando la aproximación anterior para p_1 y p_2 , el cálculo siguiente da un intervalo de confianza de 99 por ciento para p = proporción de artículos del proceso insatisfactorio.

$$h = \frac{1578}{3000} = 0,526, \quad K = 2,576,$$

$$p_1 = 0,526 - \frac{2,576}{\sqrt{3000}} \sqrt{(0,526)(0,474)} = 0,502,$$

$$p_2 = 0,526 + \frac{2,576}{\sqrt{3000}} \sqrt{(0,526)(0,474)} = 0,550.$$

PROBLEMAS

14.1. Supóngase que se mide un objeto independientemente con dos procedimientos de medida diferentes. Sean L_1 y L_2 las longitudes medidas obtenidas con el primer y segundo método respectivamente. Si ambos métodos son correctamente calibrados podemos suponer que $E(L_1) = E(L_2) = L$, la longitud verdadera. Sin embargo, la exactitud de los métodos no es necesariamente la misma. Si medimos la exactitud mediante la varianza, entonces $V(L_1) \neq V(L_2)$. Si usamos la combinación lineal $Z = aL_1 + (1 - a)L_2$ como nuestro estimador de L , inmediatamente tenemos que $E(Z) = L$. Es decir, Z es un estimador insesgado de L . ¿Para qué valor de a , $0 < a < 1$, es mínima la varianza de Z ?

14.2. Sea X una variable aleatoria con esperanza μ y varianza σ^2 . Sea (X_1, \dots, X_n) una muestra de X . Hay muchos otros estimadores de σ^2 que se sugieren además del ya propuesto. Demostrar que $C\sum_{i=1}^{n-1} (X_{i+1} - X_i)^2$ es un estimador insesgado de σ^2 para un valor apropiado de C . Encontrar el valor de C .

14.3. Supóngase que se obtienen 200 observaciones independientes X_1, \dots, X_{200} de una variable aleatoria X . Se dice que $\sum_{i=1}^{200} X_i = 300$ y que $\sum_{i=1}^{200} X_i^2 = 3754$. Usando esos valores obtener un estimador insesgado de $E(X)$ y $V(X)$.

14.4. Una variable aleatoria X tiene fdp $f(x) = (\beta + 1)x^\beta$, $0 < x < 1$.

(a) Obtener el estimador ML de β , basándose en una muestra X_1, \dots, X_n .

(b) Evaluar la estimación si los valores muestrales son 0,3; 0,8; 0,27; 0,35; 0,62 y 0,55.

14.5. Los datos de la tabla 14.4 se obtuvieron de la distribución del espesor de la madera en los postes telefónicos. (W. A. Shewhart. *Economic Control of Quality of Manufactured Products*. Macmillan and Co, New York, 1932, p. 66.) Supóngase que la variable aleatoria que se considera tiene distribución $N(\mu, \sigma^2)$ obtener el estimador ML de μ y σ^2 .

TABLA 14.4

Espesor (pulg.)	Frecuencia	Espesor (pulg.)	Frecuencia
1,0	2	3,7	123
1,3	29	4,0	82
1,6	62	4,3	48
1,9	106	4,6	27
2,2	153	4,9	14
2,5	186	5,2	5
2,8	193	5,5	1
3,1	188		Total frecuencias: 1370
3,4	151		

14.6. Supóngase que T , el tiempo para fallar (en horas) de un instrumento electrónico, tiene la siguiente fdp:

$$f(t) = \beta e^{-\beta(t-t_0)} \quad t > t_0 > 0, \\ = 0, \quad \text{para cualquier otro valor.}$$

(T tiene una distribución exponencial truncada a la izquierda en t_0). Supóngase que se prueban n artículos y que se anotan los tiempos para fallar T_1, \dots, T_n .

(a) Suponiendo que t_0 es conocido, obtener el estimador ML de β .

(b) Suponiendo que t_0 es desconocido, pero β es conocido, obtener el estimador ML de t_0 .

14.7. Considérese la misma ley de falla descrita en el problema 14.6. Esta vez se prueban N artículos durante T_0 horas ($T_0 > t_0$) y se anota el número k de artículos que fallan. Responder la pregunta (a) del problema 14.6.

14.8. Supóngase que X está distribuida uniformemente en $(-\alpha, \alpha)$. Encontrar el estimador ML de α , basándose en una muestra aleatoria de tamaño n , X_1, \dots, X_n .

14.9. (a) Se efectúa un proceso hasta que un suceso A ocurre por primera vez. En cada repetición $P(A) = p$. Se supone que son necesarias n_1 repeticiones. Luego se repite el experimento y esta vez son necesarias n_2 repeticiones para producir el suceso A . Si se hace esto k veces, obtenemos la muestra n_1, \dots, n_k . Basado en esta muestra, obtener el estimador ML de p .

(b) Supóngase que k es muy grande. Encontrar el valor aproximado de $E(\hat{p})$ y $V(\hat{p})$ en donde \hat{p} es el estimador ML obtenido en (a).

14.10. Se probó una componente que se supone que tiene una distribución exponencial de fallas y se observaron las siguientes duraciones (en horas): 108, 212, 174, 130, 198, 169, 252, 168, 143. Usando esos valores muestrales, obtener una estimación ML para la confiabilidad de la componente cuando se usa durante un período de 150 horas.

14.11. Los datos siguientes representan la duración de bombillas eléctricas (en horas):

1009, 1085, 1123, 1181, 1235, 1249, 1263, 1292, 1327, 1338, 1348, 1352, 1359, 1368, 1379, 1397, 1406, 1425, 1437, 1438, 1441, 1458, 1483, 1488, 1499, 1505, 1509, 1519, 1541, 1543, 1548, 1549, 1610, 1620, 1625, 1638, 1639, 1658, 1673, 1682, 1720, 1729, 1737, 1752, 1757, 1783, 1796, 1809, 1828, 1834, 1871, 1881, 1936, 1949, 2007.

De los valores de la muestra anterior obtener la estimación ML para la confiabilidad de tales bombillas eléctricas cuando se usan durante 1600 horas, suponiendo que la duración está distribuida normalmente.

14.12. Supóngase que se usan dos bombillas tal como se describieron en el problema 14.11 en (a) una conexión en serie y (b) en una conexión en paralelo. En cada uno de los casos encontrar la estimación ML de la confiabilidad durante una operación de 1600 horas del sistema, basándose en los valores muestrales dados en el problema 14.11.

14.13. Supóngase que una fuente radiactiva emite partículas α de acuerdo con una distribución de Poisson. Es decir, si X es el número de partículas emitidas durante un intervalo de t minutos, luego $P(X = k) = e^{-\lambda t}(\lambda t)^k/k!$ En vez de anotar el número real de partículas emitidas, supóngase que observamos el número de veces que no fue emitida ninguna partícula. Específicamente, supóngase que se observan durante 50 minutos 30 fuentes radiactivas que tiene la misma potencia y que en 25 casos al menos una partícula fue emitida. Obtener la estimación ML de λ con base en esta información.

14.14. Una variable aleatoria X tiene distribución $N(\mu, 1)$. Se toman veinte observaciones de X , pero, en vez de anotar su valor observamos sólo si X era negativa o no. Suponiendo que el suceso $\{X < 0\}$ ocurrió exactamente 14 veces, utilizar esta información para obtener la estimación ML de μ .

14.15. Suponiendo que X tiene una distribución gamma; es decir, la fdp está dada por

$$f(x) = \frac{\lambda(\lambda x)^{r-1} e^{-\lambda x}}{\Gamma(r)}, \quad x > 0.$$

Supóngase r conocido. Sea X_1, \dots, X_n una muestra de X , obtener el estimador ML de λ , basándose en esta muestra.

14.16. Supóngase que X tiene una distribución de Weisbull con fdp

$$f(x) = (\lambda x)x^{\alpha-1} e^{-\lambda x}, \quad x > 0.$$

Supóngase que α es conocido. Encontrar el estimador ML de λ basado en una muestra de tamaño n .

14.17. Demostrar el teorema 14.3. [Indicación: ver la observación (a) que sigue a este teorema.]

14.18. Comparar el valor de $P(X \geq 1)$, en donde X tiene distribución $N(0, 1)$, con $P(t \geq 1)$, en donde t tiene una distribución t de Student con:

- (a) 5 gl. (b) 10 gl. (c) 15 gl. (d) 20 gl. (e) 25 gl.

14.19. Supóngase que X tiene una distribución $N(\mu, \sigma^2)$. Una muestra de tamaño 30, sea X_1, \dots, X_{30} da los valores siguientes: $\sum_{i=1}^{30} X_i = 700,8$; $\sum_{i=1}^{30} X_i^2 = 16.395,8$. Obtener un intervalo de confianza de 95 por ciento (bilateral) para μ .

14.20. Supóngase que X tiene una distribución $N(\mu, 4)$. Una muestra de tamaño 25 produce un promedio muestral $\bar{X} = 78,3$. Obtener un intervalo de confianza de 99 por ciento (bilateral) para μ .

14.21. Supóngase que la duración de una componente tiene una distribución normal $N(\mu, 9)$. Se prueban 20 componentes y se anotan sus tiempos para fallar X_1, \dots, X_{20} . Supóngase que $\bar{X} = 100,9$ horas. Obtener un intervalo de confianza de 99 por ciento (bilateral) para la confiabilidad $R(100)$.

14.22. Obtener un intervalo de confianza unilateral de 99 por ciento (inferior) para $R(100)$ del problema 14.21.

14.23. Supóngase que X tiene distribución $N(\mu, \sigma^2)$ en donde μ y σ^2 son desconocidos. Una muestra de tamaño 15 ha producido los valores $\sum_{i=1}^{15} X_i = 8,7$ y $\sum_{i=1}^{15} X_i^2 = 27,3$. Obtener un intervalo de 95 por ciento de confianza (bilateral) para σ^2 .

14.24. Se probaron cien componentes, y 93 funcionaron más de 500 horas. Obtener un intervalo de confianza de 95 por ciento (bilateral) para $p = P$ (una componente funciona más de 500 horas). [Indicación: use la ecuación 14.12.]

14.25. Supóngase que X , la longitud de un perno, tiene distribución $N(\mu, 1)$. Se fabrica un gran número de pernos y posteriormente se separan en dos grandes lotes. El lote 1 contiene sólo aquellos pernos para los cuales $X > 5$, mientras que el lote 2 contiene el resto. Una muestra de tamaño n se saca del lote 1 y se miden las longitudes de los pernos elegidos. Así obtenemos una muestra Y_1, \dots, Y_n de la variable aleatoria Y que es una variable aleatoria distribuida normalmente y truncada en 5 a la izquierda. Escribir la ecuación que debe resolverse a fin de obtener el estimador ML de μ basado en la muestra (Y_1, \dots, Y_n) , mediante las funciones ϕ y Φ tabuladas, en donde $\phi(x) = (1/\sqrt{2\pi})e^{-x^2/2}$ y Φ es la fda de la distribución $N(0, 1)$.

14.26. (La distribución F). Sean X y Y variables aleatorias independientes con distribuciones $\chi_{n_1}^2$ y $\chi_{n_2}^2$, respectivamente. Sea definida la variable aleatoria F como sigue: $F = (X/n_1)/(Y/n_2) = n_2 X/n_1 Y$. (Esta variable aleatoria desempeña un papel importante en muchas aplicaciones estadísticas.) Demostrar que la fdp de F está dada por las expresiones siguientes:

$$h(f) = \frac{\Gamma[(n_1 + n_2)/2]}{\Gamma(n_1/2)\Gamma(n_2/2)} \left(\frac{n_1}{n_2}\right)^{n_1/2} f^{(n_1/2)-1} (1 + (n_1/n_2)f)^{-(1/2)(n_1+n_2)}, \quad f > 0.$$

[Esta se llama distribución F (Snedecor) con (n_1, n_2) grados de libertad. Debido a su importancia, se han tabulado las probabilidades asociadas con la variable aleatoria F .] [Indicación: para derivar la fdp anterior, usar el teorema 6.5.]

14.27. Dibujar el gráfico de la fdp h como se da en el problema 14.26, suponiendo que $n_1 > n_2 > 2$.

14.28. Una razón para la importancia de la distribución F es la siguiente. Suponiendo que X y Y son variables aleatorias independientes con distribuciones $N(\mu_x, \sigma_x^2)$ y $N(\mu_y, \sigma_y^2)$, respectivamente. Sean X_1, \dots, X_{n_1} e Y_1, \dots, Y_{n_2} muestras aleatorias de X y Y respectivamente. Luego el estadígrafo $C \sum_{i=1}^{n_1} (X_i - \bar{X})^2 / \sum_{i=1}^{n_2} (Y_i - \bar{Y})^2$ tiene una distribución F para una elección apropiada de C . Demostrar esto y determinar C . ¿Cuáles son los grados de libertad asociados con esta distribución?

14.29. Supóngase que la variable aleatoria t tiene una distribución t de Student con 1 grado de libertad. ¿Cuál es la distribución de t^2 ? Identifíquela.

14.30. Supóngase que X está distribuida normalmente. Se obtiene una muestra aleatoria de tamaño 4 y se calcula \bar{X} el promedio muestral. Si la suma de los cuadrados de las desviaciones de esas 4 mediciones de \bar{X} es igual a 48, obtener un intervalo de 95 por ciento de confianza (bilateral) para $E(X)$ mediante \bar{X} .

14.31. La muestra siguiente de tamaño 5 se obtuvo de la variable aleatoria bidimensional (X, Y) . Usando esos valores, calcular el coeficiente de correlación muestral.

x	1	2	3	4	5
y	4	5	3	1	2

14.32. Supóngase que $E(Y) = \alpha X + \beta$. Una muestra de tamaño 50 está disponible, sea (x_i, Y_i) , $i = 1, \dots, 50$ para la cual $\bar{x} = \bar{Y} = 0$, $\sum_{i=1}^{50} x_i^2 = 10$, $\sum_{i=1}^{50} Y_i^2 = 15$ y $\sum_{i=1}^{50} x_i Y_i = 8$.

(a) Determinar los estimadores de cuadrados mínimos de los parámetros α y β , es decir $\hat{\alpha}$ y $\hat{\beta}$.

(b) ¿Cuál es el valor mínimo de la suma de cuadrados $\sum_{i=1}^{50} [Y_i - (\hat{\alpha}x_i + \hat{\beta})]^2$.

14.33. Se podría suponer (erradamente) que siempre puede encontrarse un estimador insesgado, para un parámetro desconocido. Que esto no es así, se ilustra con el ejemplo siguiente. Supóngase que se hacen n repeticiones de un experimento y un suceso especial A se verifica exactamente k veces. Si hay una probabilidad constante $p = P(A)$ por hipótesis de que A ocurra cada vez que se hace el experimento, podríamos estar interesados en estimar la razón $r = p/(1-p)$. Para verificar que no existe un estimador insesgado de $r = p/(1-p)$ [basado en las observaciones de kA y $(n-k)\bar{A}$], suponemos que en realidad existe tal estimador. Es decir, supóngase que $\hat{r} = h(k)$ es un estadígrafo para el cual $E(\hat{r}) = p/(1-p)$. Específicamente, supóngase que $n = 2$ y por tanto $k = 0, 1$, ó 2. Designense los tres valores correspondientes de \hat{r} por a , b y c . Demostrar que $E(\hat{r}) = p/(1-p)$ conduce a una contradicción al observar lo que sucede a la izquierda y a la derecha de esta ecuación cuando $p \rightarrow 1$.

14.34. Verificar que los estimadores de cuadrados mínimos $\hat{\alpha}$ y $\hat{\beta}$ como se dan en las ecuaciones (14.7) y (14.8) son insesgados.

14.35. Verificar las expresiones para $V(\hat{\alpha})$ y $V(\hat{\beta})$ como se dan en la ecuación (14.9).

14.36. Supongamos que $E(Y) = \alpha X^2 + \beta X + \gamma$, en donde X es preasignada. Basado en una muestra (x_i, Y_i) , $i = 1, \dots, n$ determinar los estimadores de cuadrados mínimos de los parámetros α , β y γ .

14.37. Con ayuda de la tabla 7, obtener una muestra de tamaño 20 de una variable aleatoria que tenga distribución $N(2, 4)$.

(a) Supóngase que esta muestra se obtuvo de una variable aleatoria que tiene la distribución $N(\alpha, 4)$. Usar los valores muestrales para obtener un intervalo de 95 por ciento de confianza para α .

(b) Lo mismo que (a) excepto que se supone que la muestra proviene de la distribución $N(\alpha, \beta^2)$ con β^2 desconocido.

(c) Comparar las longitudes de los intervalos de confianza en (a) y (b) y comentar.

Docimasia de hipótesis

15.1 Introducción

En este capítulo presentaremos otra manera de abordar el problema para hacer una afirmación acerca de un parámetro desconocido asociado con una distribución de probabilidades, basándose en una muestra aleatoria. En vez de encontrar un estimador para el parámetro a menudo encontraremos conveniente formular una hipótesis sobre un valor para éste y luego usar la información de la muestra para confirmar o rechazar el valor de la hipótesis. Los conceptos que se van a presentar en este capítulo pueden formularse en una base teórica muy profunda. Sin embargo, no continuaremos este tema de un punto de vista formal. En su lugar, consideraremos cierto número de métodos que son intuitivamente atractivos. Estudiaremos algunas propiedades de los métodos sugeridos, pero no intentaremos indicar por qué deberían preferirse algunos métodos propuestos en vez de uno alterno. El lector interesado puede obtener un fundamento más teórico de algunos de estos procedimientos al consultar algunas de las referencias sugeridas al final del capítulo.

Consideremos el ejemplo siguiente.

EJEMPLO 15.1. Un fabricante ha estado produciendo tijeras que se usarán bajo ciertas condiciones de esfuerzo. Se ha encontrado que la duración (horas) de esos artículos está distribuida normalmente $N(100,9)$. Se ha diseñado un nuevo sistema de fabricación cuyo objetivo es alargar su duración. Es decir, se espera que la duración de vida X (correspondiente a los artículos producidos por el nuevo método) tenga una distribución $N(\mu, 9)$, en donde $\mu > 100$. (Supongamos que la varianza permanece igual. Principalmente, esto significa que la variabilidad del nuevo proceso es la misma que la del antiguo).

Luego el fabricante y el comprador en potencia están interesados en docimar la siguiente hipótesis:

$$H_0 : \mu = 100 \quad \text{contra} \quad H_1 : \mu > 100.$$

(Estamos haciendo la suposición tácita de que el nuevo proceso no puede ser peor que el antiguo.) H_0 se llama la hipótesis nula, y H_1 la hipótesis alterna.

Esencialmente estamos encarando un problema similar a uno discutido en

el capítulo 14. Estamos estudiando una variable aleatoria y no conocemos el valor de un parámetro asociado con su distribución. Podríamos resolver este problema como lo hicimos anteriormente, al estimar simplemente μ . Sin embargo, en muchas situaciones realmente estamos interesados en hacer una decisión específica: ¿deberíamos aceptar o rechazar la hipótesis H_0 ? Así, no volveremos a tratar los conceptos previos de estimación, sino que procederemos a desarrollar algunos conceptos que son especialmente apropiados para resolver el problema específico que tratamos.

Empezamos por obtener una muestra de tamaño n de la variable aleatoria X . Es decir, elegimos al azar, n artículos fabricados por el nuevo proceso y anotamos cuánto tiempo funciona cada uno, así obtenemos la muestra X_1, \dots, X_n . Luego calculamos el promedio aritmético de esos números, sea \bar{X} . Puesto que se sabe que \bar{X} es un «buen» estimador de μ , parece razonable que basemos nuestra decisión de aceptar o rechazar H_0 en el valor de \bar{X} . Puesto que estamos interesados en la discriminación entre $\mu = 100$ y valores de μ mayores que 100, parece razonable que debamos rechazar H_0 si $(\bar{X} - 100)$ es muy «grande».

Así, llegamos al siguiente procedimiento, (sobre una base intuitiva estrictamente) llamado corrientemente una *dócima* de la hipótesis: se rechaza H_0 si $\bar{X} - 100 > C'$ o, equivalentemente, si $\bar{X} > C$, (en donde C es una constante para ser determinada), y se acepta en cualquier otro caso.

Nótese que la forma particular de la dócima que estamos usando fue sugerida en parte por la hipótesis alterna H_1 . Este es un punto al cual nos referiremos nuevamente. Si en la situación anterior hubiésemos estado interesados en docimar $H_0: \mu = 100$ versus $H_1: \mu \neq 100$, habríamos usado la dócima: rechazar H_0 si $|\bar{X} - 100| > C'$. Ahora, estamos en una posición análoga a la que nos encontrábamos cuando construimos un estimador $\hat{\theta}$ para el parámetro θ . Nos preguntábamos: ¿qué tan «bueno» es el estimador? ¿Qué propiedades deseables tiene? Podemos formular preguntas similares acerca de la dócima que hemos construido. ¿Cómo es de «buena» la dócima? ¿Qué propiedades posee? ¿Cómo la podríamos comparar con otra posible dócima?

A fin de responder tales preguntas debemos verificar primeramente que no existe solución, en el sentido corriente, para el problema que estamos proponeiendo. Es decir, por una simple inspección de algunos de los artículos que están siendo fabricados nunca podemos estar seguros de que $\mu = 100$. (Nótese nuevamente la analogía al problema de estimación: no esperamos que nuestro estimador $\hat{\theta}$ sea igual a θ . Sencillamente esperamos que esté «próximo» a θ .) Lo mismo es cierto aquí: una dócima no conducirá siempre a la decisión correcta, pero una «buena» dócima conduciría «el mayor número de veces» a la decisión correcta.

Seamos más precisos. Básicamente hay dos tipos de error que podemos cometer. Podemos rechazar H_0 cuando realmente H_0 es verdadera; es decir, cuando la calidad de las tijeras *no* ha sido mejorada. Esto puede ocurrir debido a que aconteció que escogimos unas pocas tijeras más resistentes en nuestra muestra que no son típicas de la producción completa. O, alternativamente, podemos aceptar H_0 cuando realmente H_0 es falsa; es decir, cuando la calidad de las tijeras ha sido mejorada. Hagamos formalmente la siguiente definición.

Definición. *Error tipo 1:* rechazar H_0 cuando es verdadera.

Error tipo 2: aceptar H_0 cuando es falsa.

Debe ser evidente que no podemos evitar completamente el cometer esos errores. Trataremos de mantener pequeña la probabilidad de cometerlos.

Para aprehender este problema presentaremos la noción muy importante de función de operación característica de la dócima, sea L , que es la siguiente función del parámetro (desconocido) μ .

Definición. La *función de operación característica* (función OC) de la dócima anterior está definida como

$$L(\mu) = P(\text{aceptar } H_0 \mid \mu) = P(\bar{X} \leq C \mid \mu).$$

Es decir, $L(\mu)$ es la probabilidad de aceptar H_0 considerada como una función de μ .

Observación: otra función, muy relacionada con la función OC, es la *función de potencia* definida por

$$H(\mu) = P[\text{rechazar } H_0 \mid \mu].$$

Luego $H(\mu) = 1 - L(\mu)$. Usaremos la función OC para describir propiedades de la dócima aunque esto se podría hacer fácilmente mediante la función de potencia.

En el caso específico que se está considerando, podemos obtener la siguiente expresión explícita para L : si μ es el valor verdadero de $E(X)$, entonces \bar{X} tiene distribución $N(\mu, 9/n)$. Luego,

$$L(\mu) = P(\bar{X} \leq C \mid \mu) = P\left(\frac{\bar{X} - \mu}{3/\sqrt{n}} \leq \frac{C - \mu}{3/\sqrt{n}}\right) = \Phi\left[\frac{(C - \mu)\sqrt{n}}{3}\right],$$

en donde, como siempre,

$$\Phi(s) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^s e^{-x^2/2} dx.$$

Las siguientes propiedades de $L(\mu)$ son verificadas fácilmente:

(a) $L(-\infty) = 1$.

(b) $L(+\infty) = 0$.

(c) $dL/d\mu < 0$ para todo μ (Luego L es una función estrictamente decreciente de μ).

Luego el gráfico de la función L tiene en general la apariencia de la curva de la figura 15.1. (La forma específica dependerá, naturalmente, de la elección de la constante C y del tamaño de muestra n).

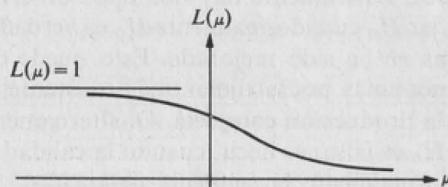


FIGURA 15.1

Considérese $1 - L(100)$. Este número representa la probabilidad de rechazar H_0 cuando H_0 es verdadero. Es decir, $1 - L(100)$ representa la probabilidad de un error del tipo 1. Si se dan n y C , entonces $1 - L(100)$ está determinado completamente. Por ejemplo, si tomamos $n = 50$ y $C = 101$, obtenemos

$$\begin{aligned} 1 - L(100) &= 1 - \Phi\left[\frac{101 - 100}{3}\sqrt{50}\right] \\ &= 1 - \Phi(2,37) = 0,009. \end{aligned}$$

Luego esta dócima particular nos conduciría a rechazar H_0 erróneamente alrededor del 0,9 por ciento de las veces.

A menudo consideramos el problema desde un punto de vista ligeramente diferente. Supóngase que se da el tamaño de muestra n y que la probabilidad de un error del tipo 1 está *especificada*, es decir, $1 - L(100) = \alpha$ o, equivalentemente, $L(100) = 1 - \alpha$. ¿Cuál sería el valor de C ?

Específicamente, si tomamos $n = 50$ y elegimos $\alpha = 0,05$, obtenemos C como solución de la siguiente ecuación:

$$0,95 = \Phi\left(\frac{C - 100}{3}\sqrt{50}\right).$$

De la tabla de la distribución normal esto da

$$1,64 = \frac{C - 100}{3}\sqrt{50}.$$

Por tanto

$$C = 100 + \frac{3(1,64)}{\sqrt{50}} = 100,69.$$

Luego si rechazamos la hipótesis cada vez que el promedio muestral es mayor que 100,69, estamos garantizando que el error del tipo 1 ocurrirá, con una probabilidad de 0,05. Puesto que ahora se conocen n y C la función OC está completamente especificada. Su gráfico se muestra en la figura 15.2. El valor 0,05 se llama *nivel de significación* de la dócima (o algunas veces *tamaño* de la dócima). En la mayor parte de los problemas se supone que es este valor menor que 0,1.

Nótese que al especificar α y el tamaño de la muestra n , sólo debe ser determinada la constante C a fin de especificar completamente la dócima. Hicimos esto insistiendo en que el gráfico de la función OC pase a través de un punto espe-

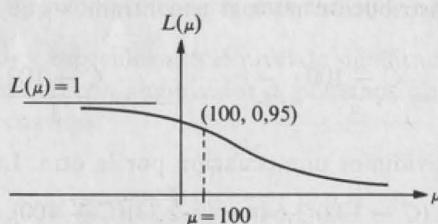


FIGURA 15.2

cificado, tal como $(100, 0,95)$. (Debe ser evidente cómo se modificaría el procedimiento anterior si hubiésemos escogido un valor distinto a 0,05 para el nivel de significación).

Ahora que la función OC está completamente especificada podemos encontrar las coordenadas de cualquier otro punto. Por ejemplo, ¿cuál es el valor de $L(102)$?

$$\begin{aligned} L(102) &= \Phi\left(\frac{100,69 - 102}{3}\sqrt{50}\right) \\ &= \Phi(-3,1) = 0,00097. \end{aligned}$$

Luego para la dómica que se considera, la probabilidad de aceptar $H_0: \mu = 100$ cuando en realidad $\mu = 102$ es igual a 0,00097. Por tanto la probabilidad de un error del tipo 2 es muy pequeña si $\mu = 102$. Puesto que L es una función decreciente de μ , observamos que $L(\mu) < 0,00097$ para todo $\mu > 102$.

Si queremos escoger a n y C , debemos especificar dos puntos a través de los cuales pase el gráfico de la función OC. De esta manera podemos controlar no sólo la probabilidad de error del tipo 1, sino también la probabilidad de error del tipo 2.

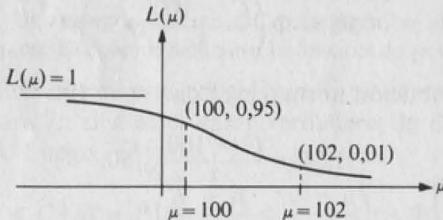


FIGURA 15.3

Supóngase que en el ejemplo que estamos considerando deseamos evitar el rechazo de H_0 cuando $\mu \geq 102$. Luego podemos hacer $L(102) = 0,01$, por ejemplo, puesto que L es una función decreciente de μ , se deduce que $L(\mu) \leq 0,01$ para $\mu > 102$. (Ver la figura 15.3). Si pedimos también un nivel de significación de 0,05, obtenemos las siguientes ecuaciones para la determinación de n y C :

$$L(100) = 0,95, \quad L(102) = 0,01.$$

Estas ecuaciones se reducen a

$$0,95 = \Phi\left(\frac{C - 100}{3}\sqrt{n}\right), \quad 0,01 = \Phi\left(\frac{C - 102}{3}\sqrt{n}\right).$$

De las tablas de la distribución normal encontramos que estas expresiones son equivalentes a

$$1,64 = \frac{C - 100}{3}\sqrt{n}, \quad -2,33 = \frac{C - 102}{3}\sqrt{n}.$$

A fin de eliminar n , dividimos una ecuación por la otra. Luego

$$(C - 102)(1,64) = (-2,33)(C - 100),$$

de la cual obtenemos

$$C = \frac{(102)(1,64) - (100)(-2,33)}{1,64 - (-2,33)} = 100,8.$$

Una vez que se conoce C , podemos obtener n elevando al cuadrado cualquiera de las ecuaciones anteriores:

Por lo tanto

$$n = \left[\frac{3(1,64)}{C - 100} \right]^2 = 34,6 \simeq 35.$$

15.2 Fórmulación general: distribución normal con varianza conocida

Hemos considerado con algún detalle un ejemplo relacionado con una hipótesis que implica el promedio de una variable aleatoria distribuida normalmente. Mientras algunos de los cálculos están aún frescos en nuestra mente, generalicemos este ejemplo como sigue.

Supóngase que X es una variable aleatoria con distribución $N(\mu, \sigma^2)$ en donde σ^2 se supone conocida. Para docirar $H_0: \mu = \mu_0$ versus $H_1: \mu > \mu_0$ proponemos lo siguiente: obtener una muestra de tamaño n , calcular el promedio muestral \bar{X} , y rechazar H_0 si $\bar{X} > C$, en donde C es una constante que se ha de determinar. La función OC de esta dómica está dada por

$$L(\mu) = P(\bar{X} \leq C) = \Phi\left(\frac{C - \mu}{\sigma} \sqrt{n}\right).$$

La forma general de la función OC es como se indica en la figura 15.4.

Las propiedades generales de $L(\mu)$ se establecen fácilmente. (Ver el problema 15.4):

- (a) $L(-\infty) = 1$.
- (b) $L(+\infty) = 0$.
- (c) $L'(\mu) < 0$, y por tanto L es una función estrictamente decreciente de μ . (15.1)
- (d) $L''(\mu) = 0$ para $\mu = C$, y por tanto el gráfico tiene aquí un punto de inflexión.
- (e) El aumento de n motiva que la curva llegue a tener mucha pendiente.

A fin de proceder, debemos considerar dos casos.

Caso 1. Si se da n y especificamos el nivel de significación (es decir, la probabilidad del error del tipo 1) con algún valor α , podemos obtener el valor de C al resolver la siguiente ecuación:

$$1 - \alpha = \Phi\left(\frac{C - \mu_0}{\sigma} \sqrt{n}\right).$$

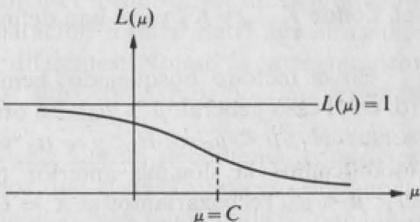


FIGURA 15.4

Definiendo K_α en la relación $1/\sqrt{2\pi} \int_{-\infty}^{K_\alpha} e^{-t^2/2} dt = \alpha$, podemos escribir lo anterior como

$$K_{1-\alpha} = \frac{C - \mu_0}{\sigma} \sqrt{n},$$

en donde $K_{1-\alpha}$ puede obtenerse de la tabla de la distribución normal. Luego rechazamos H_0 si

$$\bar{X} > \mu_0 + \frac{\sigma}{\sqrt{n}} K_{1-\alpha}.$$

Caso 2. Si vamos a determinar n y C , debemos especificar dos puntos sobre el gráfico de la curva OC: $1 - L(\mu_0) = \alpha$, el nivel de significación, y $L(\mu_1) = \beta$, la probabilidad del error del tipo 2 para $\mu = \mu_1$. Luego debemos resolver las ecuaciones siguientes para n y C :

$$1 - \alpha = \Phi\left(\frac{C - \mu_0}{\sigma} \sqrt{n}\right); \quad \beta = \Phi\left(\frac{C - \mu_1}{\sigma} \sqrt{n}\right).$$

Estas ecuaciones pueden resolverse para C y n como se indicó anteriormente. Obtenemos

$$C = \frac{\mu_1 K_{1-\alpha} - \mu_0 K_\beta}{K_{1-\alpha} - K_\beta}, \quad n = \left[\frac{\sigma K_{1-\alpha}}{C - \mu_0} \right]^2,$$

en donde $K_{1-\alpha}$ y K_β ya se han definido.

En el método bosquejado, hemos tratado la hipótesis alterna $H_1: \mu > 100$ (o, en el caso general, $\mu > \mu_0$). En otro contexto podríamos considerar $H_0: \mu = \mu_0$ versus $H'_1: \mu < \mu_0$ o $H_0: \mu = \mu_0$ versus $H''_1: \mu \neq \mu_0$. Debe ser evidente cómo modificamos la dómica anterior para tal hipótesis alterna. Si consideramos $H'_1: \mu < \mu_0$, rechazaríamos si $\bar{X} < C$ y la función OC sería definida por

$$L(\mu) = P(\bar{X} \geq C) = 1 - \Phi\left(\frac{C - \mu}{\sigma} \sqrt{n}\right).$$

Si consideramos $H''_1: \mu \neq \mu_0$, rechazaríamos H_0 cada vez que $|\bar{X} - \mu_0| > C$ y, por tanto, la función OC se definiría

$$\begin{aligned} L(\mu) &= P(|\bar{X} - \mu_0| \leq C) \\ &= \Phi\left(\frac{C + \mu_0 - \mu}{\sigma} \sqrt{n}\right) - \Phi\left(\frac{-C + \mu_0 - \mu}{\sigma} \sqrt{n}\right). \end{aligned}$$

Si escogemos el *mismo* nivel de significación α para cada una de las dómicas anteriores y dibujamos los gráficos de las respectivas funciones OC en el mismo sistema de coordenadas, obtenemos lo siguiente: (A corresponde a H_1 , B a H'_1 , y D a H''_1).

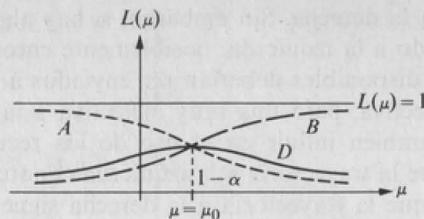


FIGURA 15.5

La figura 15.5 nos da un medio para comparar las tres dócimas que estamos considerando. Todas las dócimas tienen el mismo nivel de significación. (Debe comprenderse claramente que C es sólo un símbolo genérico de una constante y no será la misma en cada uno de los casos. Lo importante es que en cada uno de los casos se ha escogido C de modo que la dócima tendrá el nivel de significación α).

Si $\mu > \mu_0$, entonces la dócima A es mejor que las otras dos puesto que dará un valor más pequeño para la probabilidad del error del tipo 2. Sin embargo, si $\mu < \mu_0$, la dócima A es la peor de las que están siendo consideradas, mientras que la dócima B es la mejor. Finalmente, la dócima D es aceptable generalmente y aún puede ser mejorada en cualquier caso específico con la dócima A o la dócima B . Por tanto, observemos que es muy importante tener una hipótesis alterna específica en mente debido a que la dócima que escogamos puede depender de esto. (Sólo comparamos las dócimas que tienen el mismo nivel de significación; esto no es necesario en absoluto y la comparación resulta algo vaga si usamos dócimas que tienen niveles de significación diferentes. Nótese la semejanza con nuestra comparación de ciertos estimadores: sólo comparábamos las varianzas de los estimadores que eran insesgados).

En muchos casos es evidente cuál hipótesis alterna deberíamos considerar. En el ejemplo anterior, posiblemente sabemos que el nuevo proceso de fabricación produciría tijeras de la misma durabilidad o bien aumentada y, por tanto, usariamos la dócima A como sugerímos. (Si el nuevo proceso fuese a producir tijeras de calidad inferior nuestra dócima sería muy mala. Si no se garantiza ninguna de tales hipótesis sería mejor usar una dócima tal como D : rechazar H_0 si $|X\mu_0| > C$. Las dócimas tales como A y B se llaman dócimas *unilaterales* mientras que una dócima tal como D se llama dócima *bilateral*.

Al considerar hipótesis alternas podríamos encontrar la siguiente analogía útil. Supóngase que una persona se pierde de un lugar M y *sabemos* que el individuo ha ido bien a la izquierda o a la derecha de M , manteniéndose en una trayectoria rectilínea.



Si 10 personas están disponibles para la búsqueda, ¿cómo deberían dispersarse esas personas? Si nada se sabe con relación a la ubicación del individuo, podría ser razonable enviar un grupo de 5 personas en cada una de esas direcciones, luego despachar un grupo de búsqueda muy efectivo, pero no muy fuerte, tanto

a la izquierda como a la derecha. Sin embargo, si hay alguna indicación de que la persona ha caminado a la izquierda, posiblemente entonces todos o la mayor parte de los hombres disponibles deberían ser enviados a la izquierda, haciendo una búsqueda muy efectiva, pero una muy inefectiva a la derecha. Otras consideraciones podrían también influir en el uso de los recursos disponibles. Por ejemplo, supóngase que la trayectoria a la izquierda conduce a un terreno plano con bosque mientras que la trayectoria a la derecha sigue el borde de un precipicio profundo. En este caso obviamente la mayor búsqueda se concentraría a la derecha debido a que las probabilidades de estar perdido de este lado son mucho mayores que las de la izquierda.

La analogía debería ser evidente. Al docimar hipótesis, debemos estar también interesados en las consecuencias de nuestra decisión de rechazar o de aceptar H_0 . Por ejemplo, ¿es el error que cometemos al aceptar algunas tijeras que son de mala calidad (creyendo que no lo son) tan importante como no aceptar las tijeras que son de buena calidad (creyendo que no son)?

Observación: en el desarrollo anterior sugerimos que a menudo preasignamos el nivel de significación de la dócima. Esto se hace frecuentemente, sin embargo, existe otro enfoque del problema que es muy común y que deberíamos discutir.

Volvamos al ejemplo 15.1 en donde docimamos $H_0: \mu = 100$ versus $H_1: \mu > 100$. Supóngase que obtenemos sólo una muestra de tamaño 50, calculamos el promedio muestral \bar{X} y encontramos que es igual a 100,87. ¿Deberíamos aceptar o rechazar H_0 ? Podemos argumentar como sigue: si $\mu = 100$, entonces \bar{X} tiene distribución $N(100, 9/50)$. Luego podemos calcular,

$$\begin{aligned} P(\bar{X} \geq 100,87) &= P\left(\frac{\bar{X} - 100}{\sqrt{50}} \geq \frac{100,87 - 100}{\sqrt{50}}\right) \\ &= 1 - \Phi(2,06) = 0,019699. \end{aligned}$$

Puesto que $0,01 < 0,019699 < 0,05$ diremos que el valor observado \bar{X} es significativo al nivel del 5 por ciento, pero no al nivel del 1 por ciento. Es decir, si usamos $\alpha = 0,05$ rechazaríamos H_0 , mientras que al mismo tiempo si usamos $\alpha = 0,01$ no deberíamos rechazar H_0 .

Diciéndolo de un modo diferente, si $\mu = 100$ obtenemos un resultado que ocurrirá sólo alrededor del 1,9 por ciento de las veces. Si creemos que a fin de aceptar H_0 un resultado debería tener a lo menos una probabilidad de ocurrir de 0,05, luego la rechazamos. Si estamos satisfechos con una probabilidad de 0,01, la aceptamos.

Observación: la dócima anterior estipuló que H_0 debería ser rechazada siempre que $\bar{X} > C$. Suponiendo que el tamaño muestral $n = 2$, el criterio anterior se reduce a $(X_1 + X_2)/2 > C$ o, equivalentemente, $(X_1 + X_2) > k$. Luego el conjunto de valores muestrales posibles (x_1, x_2) ha sido dividido en dos regiones: $R = \{(x_1, x_2) | x_1 + x_2 > k\}$ y \bar{R} .

La región específica R depende naturalmente del valor de k , que a su vez depende del nivel de significación de la dócima R , la región de rechazo se llama algunas veces *región crítica* de la dócima. (Ver la figura 15.6.)

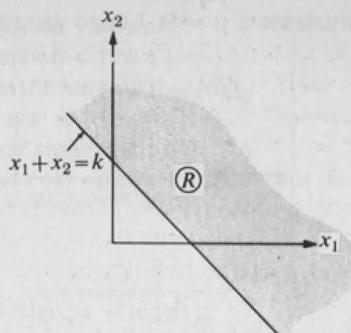


FIGURA 15.6

Generalmente, una dócima puede describirse mediante su región crítica R . Es decir, rechazamos H_0 si y sólo si $(x_1, \dots, x_n) \in R$.

15.3 Ejemplos adicionales

En vez de formular una teoría general para la docimaria de hipótesis (que existe y es muy extensa), continuaremos considerando algunos ejemplos. En cada uno de los casos, la dócima que propondremos será intuitivamente atractiva. No se hará ningún esfuerzo para indicar que una dócima especial es mejor en algún sentido.

EJEMPLO 15.2. Se comparan dos procesos de producción. El resultado del proceso A puede caracterizarse como una variable aleatoria X con distribución $N(\mu_x, \sigma_x^2)$ mientras que el resultado del proceso B puede caracterizarse como una variable aleatoria Y con distribución $N(\mu_y, \sigma_y^2)$. Supondremos se *conoce* la variabilidad intrínseca en cada uno de los procesos, medida por la varianza. Desearíamos docimar las hipótesis $H_0: \mu_x = \mu_y$ versus la hipótesis alterna $H_1: \mu_x - \mu_y > 0$.

Obtenemos una muestra de tamaño n de X , sea X_1, \dots, X_n , y una muestra de tamaño m de Y , sea Y_1, \dots, Y_m . Calculemos los respectivos promedios muestrales \bar{X} y \bar{Y} , y propongamos la dócima siguiente para docinar las hipótesis anteriores.

Rechazamos H_0 si $\bar{X} - \bar{Y} > C$ en donde C es una constante escogida de modo que la dócima tenga un nivel de significación específico igual a α .

La variable aleatoria $Z = [(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)] / \sqrt{\sigma_x^2/n + \sigma_y^2/m}$ tiene distribución $N(0, 1)$. Definiendo $\mu = \mu_x - \mu_y$, podemos expresar la función OC de μ de la manera siguiente

$$\begin{aligned} L(\mu) &= P(\bar{X} - \bar{Y} \leq C | \mu) = P\left(Z \leq \frac{C - \mu}{\sqrt{\sigma_x^2/n + \sigma_y^2/m}}\right) \\ &= \Phi\left(\frac{C - \mu}{\sqrt{\sigma_x^2/n + \sigma_y^2/m}}\right). \end{aligned}$$

Ahora $\mu_x = \mu_y$ es equivalente a $\mu = 0$. Luego para determinar C debemos resolver la ecuación

$$L(0) = 1 - \alpha$$

o

$$1 - \alpha = \Phi\left(\frac{C}{\sqrt{\sigma_x^2/n + \sigma_y^2/m}}\right).$$

Por lo tanto

$$K_{1-\alpha} = \frac{C}{\sqrt{\sigma_x^2/n + \sigma_y^2/m}}$$

en donde K_α está definida como antes, por la relación $\alpha = (1/\sqrt{2\pi}) \int_{-\infty}^{K_\alpha} e^{-t^2/2} dt$.
Luego

$$C = K_{1-\alpha} \sqrt{\sigma_x^2/n + \sigma_y^2/m}.$$

(No continuaremos el problema de determinar óptimamente a n y m . Una exposición de este problema se encuentra en Derman and Klein, *Probability and Statistical Inference for Engineers*. Oxford University Press, New York, 1959).

EJEMPLO 15.3. Un fabricante entrega fusibles de los cuales el 90 por ciento aproximadamente funcionan bien. Se inicia un nuevo proceso con el objeto de aumentar la proporción de fusibles que funcionan bien. Así, deseamos docimar la hipótesis $H_0: p = 0,90$ versus $H_1: p > 0,90$, en donde p es la proporción de fusibles que funcionan correctamente. (Es decir, estamos docimando la hipótesis de que no se ha producido ninguna mejoría sobre la hipótesis de que el nuevo proceso es superior). Obtenemos una muestra de 50 fusibles fabricados con el nuevo proceso y contamos el número de fusibles que funcionan correctamente, X . Propongamos la dócima siguiente:

Rechazar H_0 siempre que $X > 48$ y aceptarla en cualquier otro caso.

Suponiendo que la variable aleatoria X tiene una distribución binomial con parámetro p (que es una suposición realista si la muestra se toma de un lote muy grande), obtenemos la siguiente expresión para la función OC

$$\begin{aligned} L(p) &= P(X \leq 48) = 1 - P(X \geq 49) \\ &= 1 - \sum_{k=49}^{50} \binom{50}{k} p^k (1-p)^{50-k} = 1 - p^{49}(50 - 49p) \end{aligned}$$

después de algunas simplificaciones algebraicas. Por lo tanto tenemos lo siguiente

- (a) $L(0) = 1$. (c) $L'(p) < 0$ para todo $p, 0 < p < 1$.
- (b) $L(1) = 0$. (d) $L''(p) = 0$ si $p = 48/49$.

[Las propiedades (c) y (d) se verifican fácilmente con una derivación directa.] Luego el gráfico de la anterior función L tiene la forma de la curva que se muestra en la figura 15.7. El nivel de significación α de esta dócima se obtiene al calcular $1 - L(0,9)$. Obtenemos

$$\begin{aligned}\alpha &= 1 - L(0,9) \\ &= (0,9)^{49} [50 - 44,1] \\ &= 0,034.\end{aligned}$$

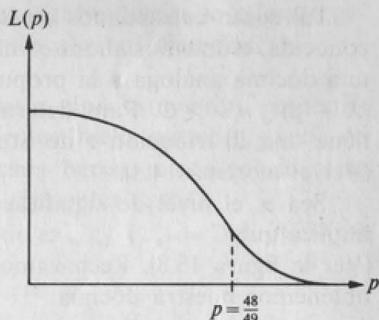


FIGURA 15.7

Observaciones: (a) El ejemplo anterior puede generalizarse como sigue. Supóngase que X es una variable aleatoria distribuida binomialmente, basada en n repeticiones de un experimento con parámetro p . Para docimar la hipótesis $H_0: p = p_0$ versus $H_1: p > p_0$, proponemos la dócima siguiente.

Rechazar H_0 siempre que $X > C$, en donde C es una constante para ser determinada. (Por tanto aceptar H_0 siempre que $X \leq C$.)

La función OC de esta dócima será de la forma

$$L(p) = \sum_{k=0}^C \binom{n}{k} p^k (1-p)^{n-k}. \quad (15.2)$$

Las siguientes propiedades de L son verificadas fácilmente. (Ver el problema 15.5).

$$(1) L(0) = 1; L(1) = 0.$$

(2) $L'(p) < 0$ para todo p , $p < 0 < 1$. Por tanto L es estrictamente decreciente.

$$(3) L'(p) = 0 \text{ si } p = C/(n-1). \text{ [Luego } L \text{ tiene un punto de inflexión en } C/(n-1).]$$

(b) Hemos dicho que en algunos casos podemos aproximar la distribución binomial con la distribución de Poisson. Es decir, si n es grande y p es pequeño, $P(X = k) \approx e^{-np}(np)^k/k!$ Usando esta fórmula de $P(X = k)$, encontramos que la función OC para la dócima propuesta anteriormente se reduce a

$$R(p) = \sum_{k=0}^C \frac{e^{-np}(np)^k}{k!} \quad (15.3)$$

Las propiedades siguientes de R también se pueden verificar fácilmente. (Ver el problema 15.6)

$$(4) R(0) = 1; R(1) = 0.$$

(5) $R'(p) < 0$ para todo p , $0 < p < 1$.

$$(6) R''(p) = 0 \text{ si } p = C/n.$$

(7) $R(C/n)$ es una función solamente de C y no de n .

Reconsideremos el problema de docimar $H_0: \mu = \mu_0$ versus $H_1: \mu > \mu_0$, en donde X tiene distribución $N(\mu, \sigma^2)$. Previamente suponíamos que σ^2 era conocido. Levantemos ahora esta restricción.

Nuestra dócima anterior rechazaba H_0 siempre que $(\bar{X} - \mu_0)\sqrt{n}/\sigma > C$; C estaba determinada considerando el hecho de que $(\bar{X} - \mu_0)\sqrt{n}/\sigma$ tiene distribución $N(0, 1)$ si $\mu = \mu_0$.

Tal como construimos un intervalo de confianza para μ cuando σ^2 era desconocida, estimemos ahora σ^2 mediante $\hat{\sigma}^2 = [1/(n-1)] \sum_{i=1}^n (X_i - \bar{X})^2$. Usemos una dómica análoga a la propuesta anteriormente: rechacemos H_0 siempre que $(\bar{X} - \mu) \sqrt{n}/\hat{\sigma} > C$. Para determinar C usemos el hecho de que $(\bar{X} - \mu_0) \sqrt{n}/\hat{\sigma}$ tiene una distribución t de Student con $(n-1)$ grados de libertad si $\mu = \mu_0$. (Ver el teorema 14.3).

Sea α el nivel de significación prefijado. Luego $\alpha = P[(\bar{X} - \mu_0) \sqrt{n}/\hat{\sigma} > C]$ implica que $C = t_{n-1, 1-\alpha}$, es obtenida de la tabla de la distribución t de Student. (Ver la figura 15.8). Rechazamos H_0 cada vez que $\bar{X} > \hat{\sigma} t_{n-1, 1-\alpha} n^{1/2} + \mu_0$, así obtenemos nuestra dómica.

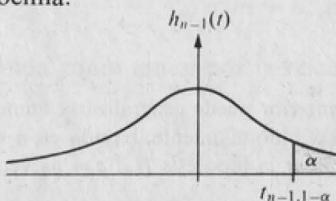


FIGURA 15.8

EJEMPLO 15.4. Supóngase que X , la lluvia caída anualmente en cierta zona, está distribuida normalmente con $E(X) = 30,0$ pulgadas. (Este valor ha sido obtenido de un gran registro histórico de datos meteorológicos.) En años recientes, ciertos cambios de clima parecen que evidentemente afectan, en otras cosas, la precipitación anual. Se establece la hipótesis de que en realidad la lluvia caída anualmente ha aumentado. Específicamente, deseamos docimar $H_0 : \mu = 30,0$ versus $H_1 : \mu > 30,0$. La varianza se supone desconocida puesto que el cambio climatológico sospechado también puede afectar la variabilidad de la lluvia caída, y, por tanto, los datos anteriores sobre la varianza no son significativos.

Supongamos que los 8 años anteriores han dado la siguiente precipitación anual (pulgadas):

$$34,1; 33,7; 27,4; 31,1; 30,9; 35,2; 28,4; 32,1.$$

Cálculos directos dan $\bar{X} = 31,6$ y $\hat{\sigma}^2 = 7,5$. De la tabla de la distribución t encontramos que

$$t_{7, 0, 95} = 1.89.$$

Luego

$$\hat{\sigma} t_{7, 0, 95} / \sqrt{8} + 30,0 = 31,8 > 31,6.$$

Por tanto no rechazamos H_0 al nivel de significación de 0,05.

15.4 Dómica para la bondad de ajuste

En la mayoría de nuestras discusiones supusimos que la variable aleatoria que se considera tiene una distribución específica. En el capítulo anterior y en las primeras secciones de este, hemos aprendido cómo resolver el problema de tener un parámetro desconocido asociado con una distribución de probabilidades.

Sin embargo, puede acontecer que aún no estemos seguros acerca de la forma general de la distribución que se tiene. Consideraremos algunos ejemplos.

EJEMPLO 15.5. Los buques mercantes de cierto tipo estuvieron expuestos durante 400 días a riesgos de accidentes de tormentas, hielo, incendio, avería de máquinas, etc. El número de accidentes X de cada barco, puede considerarse como una variable aleatoria. Se registraron los siguientes datos:

Número de accidentes (X):	0	1	2	3	4	5	6
Número de barcos con X accidentes:	1448	805	206	34	4	2	1

¿Justifican los datos anteriores que X tiene una distribución de Poisson?

EJEMPLO 15.6. Supóngase que se tienen 20 muestras de una clase especial de cables y que se miden las resistencias en ohms y se obtienen los valores siguientes:

$$\begin{aligned} & 9.8; 14.5; 13.7; 7.6; 10.5; 9.3; 11.1; 10.1; 12.7; 9.9; \\ & 10.4; 8.3; 11.5; 10.0; 9.1; 13.8; 12.9; 10.6; 8.9; 9.5. \end{aligned}$$

Si R es la variable aleatoria de la cual se obtiene la muestra anterior, ¿tenemos razón al suponer que R está distribuida normalmente?

EJEMPLO 15.7. Se probaron veinte tubos electrónicos y se anotaron las duraciones (en horas):

$$\begin{aligned} & 7.2; 37.8; 49.6; 21.4; 67.2; 41.1; 3.8; 8.1; 23.2; 72.1; \\ & 11.4; 17.5; 29.8; 57.8; 84.6; 12.8; 2.9; 42.7; 7.4; 33.4. \end{aligned}$$

¿Son los datos anteriores consistentes con la hipótesis de que la variable aleatoria T , que está siendo muestreada, está distribuida exponencialmente?

Los ejemplos anteriores son típicos de una gran clase de problemas que aparecen frecuentemente en aplicaciones. Hay varias técnicas estadísticas disponibles con las cuales se pueden analizar tales problemas; consideraremos a continuación algunas de ellas.

El problema de docirar la hipótesis de que una variable aleatoria tiene cierta distribución específica puede considerarse como un caso especial del siguiente problema general.

Considérense nuevamente las condiciones que dan origen a la distribución multinomial (Ver la sección 8.7). Se efectúa un experimento εn veces. Cada una de las repeticiones de ε da origen a uno y sólo uno de los sucesos A_i , $i = 1, 2, \dots, k$. Supóngase que $P(A_i) = p_i$. Sea n_i el número de veces que ocurre A_i entre las n repeticiones de ε , $n_1 + \dots + n_k = n$.

Deseamos docirar la hipótesis $H_0: p_i = p_{io}$, $i = 1, \dots, k$, en donde p_{io} es

un valor específico. Karl Pearson (1900) introdujo la siguiente dócima de «bondad de ajuste» para docimar la hipótesis anterior:

$$\text{Rechazar } H_0 \text{ siempre que } D^2 = \sum_{i=1}^k \frac{(n_i - np_{io})^2}{np_{io}} > C, \quad (15.4)$$

en donde C es una constante que se va a determinar.

Observaciones: (a) Puesto que $E(n_i) = np_{io}$ si $p_i = p_{io}$, este criterio para docimar tiene mucho atractivo intuitivo. Exige que rechacemos H_0 siempre que la discrepancia entre los valores observados n_i y los valores esperados np_{io} sea «muy grande». Algunas veces el estadígrafo D^2 anterior se escribe de una manera muy sugerente como $\sum_{i=1}^k (o_i - e_i)^2/e_i$, en donde o_i y e_i representan respectivamente el valor observado y esperado de n_i .

(b) Es importante establecer que D^2 es un estadígrafo (es decir, una función de los valores observados n_1, \dots, n_k) y es, por tanto, una variable aleatoria. En realidad, D^2 es una variable aleatoria discreta que toma un gran número finito de valores. La distribución actual de D^2 es muy complicada. Afortunadamente, hay una aproximación disponible para la distribución de D^2 , válida si n es grande, y que hace muy útil el procedimiento sugerido.

Teorema 15.1. Si n es suficientemente grande, y si $p_i = p_{io}$, la distribución de D^2 tiene aproximadamente la distribución de χ^2 cuadrado con $(k-1)$ grados de libertad.

Demostración: el argumento siguiente no es una demostración rigurosa. Sólo es un intento de hacer aceptable el resultado.

Consideremos un caso especial, es decir, $k = 2$. Luego

$$D^2 = \frac{(n_1 - np_{1o})^2}{np_{1o}} + \frac{(n_2 - np_{2o})^2}{np_{2o}}.$$

Usando el hecho que $n_1 + n_2 = n$ y que $p_{1o} + p_{2o} = 1$, podemos escribir

$$\begin{aligned} D^2 &= \frac{(n_1 - np_{1o})^2}{np_{1o}} + \frac{(n - n_1 - n(1 - p_{1o}))^2}{np_{2o}} \\ &= \frac{(n_1 - np_{1o})^2}{np_{1o}} + \frac{(n_1 - np_{1o})^2}{np_{2o}} = (n_1 - np_{1o})^2 \left[\frac{1}{np_{1o}} + \frac{1}{np_{2o}} \right] \\ &= (n_1 - np_{1o})^2 \left[\frac{np_{2o} + np_{1o}}{n^2 p_{1o} p_{2o}} \right] = \frac{(n_1 - n_1 p_{1o})^2}{np_{1o}(1 - p_{1o})}. \end{aligned}$$

Como $n_1 = \sum_{j=1}^n Y_{1j}$, en donde

$$\begin{aligned} Y_{1j} &= 1 && \text{si } A_1 \text{ ocurre en la } j\text{-ésima repetición,} \\ &= 0, && \text{en cualquier otro caso.} \end{aligned}$$

Así n_1 puede expresarse como la suma de n variables aleatorias independientes y de acuerdo con el teorema del límite central, tiene aproximadamente una distribución normal si n es grande. Además, $E(n_1) = np_{1o}$ y $V(n_1) = np_{1o}(1 - p_{1o})$ si p_{1o} es el valor verdadero de p_1 . Por tanto si $p_1 = p_{1o}$, para un gran valor de n ,

entonces, la variable aleatoria $(n_1 - np_{1o})/\sqrt{np_{1o}(1 - p_{1o})}$ tiene aproximadamente la distribución $N(0, 1)$. Luego, de acuerdo con el teorema 10.8 para un n grande, la variable aleatoria

$$\left[\frac{n_1 - np_{1o}}{\sqrt{np_{1o}(1 - p_{1o})}} \right]^2$$

tiene aproximadamente la distribución χ^2_1 .

Hemos demostrado que si n es suficientemente grande, D^2 (con $k = 2$) tiene aproximadamente la distribución χ^2_1 . Pero precisamente es esto lo que sostienemos en el teorema. La demostración para un k en general sigue la misma línea: debemos demostrar que D^2 puede expresarse como la suma de cuadrados de $(k - 1)$ variables aleatorias independientes cada una con distribución $N(0, 1)$ si n es grande y si $p_i = p_{io}$ y, recurriendo al teorema 10.8, encontramos que se deduce el resultado anterior.

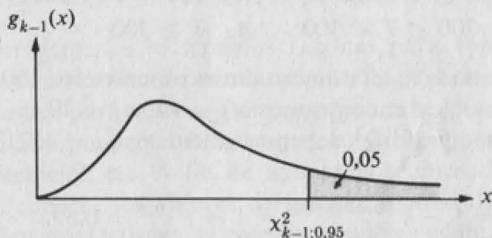


FIGURA 15.9

Podemos usar el resultado ya establecido para responder a la pregunta de «qué tan grande» debería ser D^2 a fin de rechazar la hipótesis $H_0: p_i = p_{io}$.

Supóngase que queremos obtener una probabilidad del error del tipo 1 (es decir, el nivel de significación) igual a 0,05. Esto significa que esperamos rechazar H_0 alrededor del 5 por ciento de las veces, cuando en realidad H_0 es verdadero. Luego elegimos C para satisfacer

$$P(D^2 > C | p_i = p_{io}) = 0,05.$$

Puesto que D^2 tiene distribución χ^2_{k-1} si $p_i = p_{io}$ podemos obtener el valor de C de la tabla de la distribución χ cuadrado; es decir, $C = \chi^2_{k-1, 0.95}$; en donde $\chi^2_{k-1, 0.95}$ está definida por la relación

$$\int_{\chi^2_{k-1, 0.95}}^{\infty} g_{k-1}(x) dx = 0,05,$$

en donde $g_{k-1}(x)$ es la fdp de una variable aleatoria con distribución χ^2_{k-1} . (Ver la figura 15.9.)

Usemos las ideas que desarrollamos para responder algunas preguntas surgidas al comienzo de esta sección: ¿podemos encontrar una décima para decidir si se acepta o se rechaza la hipótesis de que se tomó una muestra particular de una variable aleatoria con una distribución específica?

En este punto debemos hacer una distinción entre los dos tipos de problemas. Simplemente podríamos formular la hipótesis de que la variable aleatoria que se muestrea tiene una distribución normal sin especificar los parámetros implicados.

O podríamos ser más explícitos y formular la hipótesis de que la variable aleatoria que se considera tiene una distribución normal con promedio y varianza específicos. Los dos problemas pueden ser tratados de una manera semejante, pero la última (cuando especificamos completamente la distribución hipotética) es un poco más simple y la consideraremos primero.

Caso 1. Docimación para una distribución completamente especificada.

EJEMPLO 15.8. Supóngase que creemos que la duración T de bombillos eléctricos está distribuida exponencialmente con parámetro $\beta = 0,005$. (Es decir, el tiempo esperado para fallar es de 200 horas.) Saquemos una muestra de 150 bombillos, probémoslos, y anotemos sus tiempos para quemarse, T_1, \dots, T_{150} . Consideremos los cuatro sucesos mutuamente excluyentes:

$$\begin{aligned} A_1: 0 \leq T < 100; \quad A_2: 100 \leq T < 200; \\ A_3: 200 \leq T < 300; \quad A_4: T \geq 300. \end{aligned}$$

Supóngase que registramos n_i , el número de veces (entre los 150 tiempos de falla) que ocurrió el suceso A_i , y encontramos $n_1 = 47$, $n_2 = 40$, $n_3 = 35$, y $n_4 = 28$. A fin de evaluar el estadígrafo D^2 debemos calcular p_i , $i = 1, 2, 3, 4$.

Ahora

$$p_1 = P(T \leq 100) = 1 - e^{-0,005(100)} = 1 - e^{-0,5} = 0,39,$$

$$p_2 = P(100 \leq T < 200) = 1 - e^{-0,005(200)} - 0,39 = 0,24,$$

$$p_3 = P(200 \leq T < 300) = 1 - e^{-0,005(300)} - (1 - e^{-0,005(200)}) = 0,15,$$

$$p_4 = P(T > 300) = e^{-0,005(300)} = 0,22.$$

Ahora podemos calcular

$$\begin{aligned} D^2 &= \sum_{i=1}^4 \frac{(n_i - np_{io})^2}{np_{io}} \\ &= \frac{(47 - 58,5)^2}{58,5} + \frac{(40 - 36)^2}{36} + \frac{(35 - 22,5)^2}{22,5} + \frac{(28 - 33)^2}{33} = 11,56. \end{aligned}$$

En las tablas de la distribución χ^2 cuadrado encontramos que $P(D^2 > 11,56) < 0,01$, en donde D^2 tiene aproximadamente la distribución χ^2 cuadrado con $4 - 1 = 3$ grados de libertad. Por lo tanto rechazaríamos (al nivel del 1 por ciento) la hipótesis de que los datos representan una muestra de una distribución exponencial con parámetro $\beta = 0,005$.

El ejemplo anterior ilustra el procedimiento general que usamos para docimar la hipótesis de que X_1, \dots, X_n representa una muestra de una variable aleatoria con una distribución completamente especificada:

- Dividir la recta real en k intervalos mutuamente excluyentes, A_1, \dots, A_k .
- Sea N_i el número de valores muestrales que caen en A_i , $i = 1, 2, \dots, k$.
- Sea $p_{io} = P(A_i)$. Esos valores se pueden calcular puesto que la hipótesis especifica completamente la distribución.

- (d) Calcular D^2 y rechazar la hipótesis si $D^2 > C$ en donde C se obtiene de la tabla de la distribución χ^2 cuadrado. Si se requiere un nivel de significación α , $C = \chi_{k-1, 1-\alpha}^2$.

Observación: no discutiremos cómo se deben escoger los intervalos A_i o cuántos deberían escogerse. Establezcamos sólo la regla siguiente: si $np_{io} < 5$ para cualquier A_i combinar los datos con A_{i+1} o A_{i-1} . Es decir, no deseamos subdividir el espacio muestral de la variable aleatoria en partes tales que el número esperado de ocurrencias en cualquier subdivisión particular sea menor que 5. [Una exposición amena de este problema puede encontrarse en un artículo de W. G. Cochran, titulado «The χ^2 -Test of Goodness of Fit» que apareció en *Ann. Math. Stat.* 23, 315-345 (1952).]

Caso 2. Docimación para una distribución si deben estimarse los parámetros.

En muchos problemas sólo tenemos razones para suponer que la variable aleatoria que se está muestreando tiene una distribución de cierto *tipo* sin que podamos especificar los parámetros. Por ejemplo, sabemos que ciertas hipótesis que hemos formulado pueden conducirnos a una distribución de Poisson, a una distribución exponencial, etc. A fin de aplicar la técnica sugerida en la sección anterior debemos *conocer* los valores de los parámetros de la distribución.

Si no conocemos esos valores, el enfoque obvio es estimarlos primero, y luego usar esas estimaciones a fin de evaluar las probabilidades p_i . Aparecen dos interrogantes.

(a) ¿Cómo deberían estimarse los parámetros?

(b) Si se usa \hat{p}_i (es decir, el parámetro estimado) en vez de p_{io} en la expresión de D^2 , ¿cómo afecta esto la distribución de D^2 ? (El hecho de que afectará la distribución debe ser evidente si comprobamos que originalmente los p_{io} eran constantes, mientras que ahora los p_i son ellas mismas variables aleatorias, dando así una estructura mucho más complicada a la variable aleatoria D^2 .)

Daremos (sin demostración) algunas respuestas a las interrogantes anteriores.

(a) Los estimadores usados corrientemente para los parámetros son los obtenidos por el método de la máxima verosimilitud.

(b) Si el número de parámetros estimados es igual a $r < k$, entonces para un gran n , la variable aleatoria D^2 tiene nuevamente una distribución χ^2 cuadrado, esta vez con $n - k - r$ grados de libertad.

Observación: este último hecho es muy notable. Significa que D^2 tiene la misma distribución fundamental χ^2 que antes, la única diferencia es que se pierde un grado de libertad por cada parámetro que debe estimarse.

EJEMPLO 15.9. Se consideran los datos del contenido de ceniza en el carbón tal como aparecen en el ejemplo 14.6. Supóngase que deseamos docimar la hipótesis de que esos datos fueron obtenidos de una variable aleatoria distribuida normalmente. Primero debemos estimar los parámetros correspondientes μ y σ^2 .

Previamente obtuvimos las estimaciones M.L. $\hat{\mu} = 17,0$ y $\hat{\sigma}^2 = 7,1$. Dividamos los valores posibles de X en cinco categorías:

$$\begin{aligned} A_1: X < 12; \quad A_2: 12 \leq X < 15; \quad A_3: 15 \leq X < 18; \\ A_4: 18 \leq X < 21; \quad A_5: X \geq 21. \end{aligned}$$

Sea n_i el número de veces que ocurre A_i . Encontramos que

$$n_1 = 7; \quad n_2 = 49; \quad n_3 = 109; \quad n_4 = 67; \quad n_5 = 18.$$

Debemos calcular a continuación $p_i = P(A_i)$, usando los valores estimados de $\hat{\mu}$ y $\hat{\sigma}^2$ ya obtenidos. Tenemos

$$\begin{aligned} p_1 &= P(X < 12) = P\left(\frac{X - 17}{2,7} < \frac{12 - 17}{2,7}\right] = \Phi(-1,85) = 0,03, \\ p_2 &= P(12 \leq X < 15) = \Phi(-0,74) - \Phi(-1,85) = 0,20, \\ p_3 &= P(15 \leq X < 18) = \Phi(0,37) - \Phi(-0,74) = 0,41, \\ p_4 &= P(18 \leq X < 21) = \Phi(1,48) - \Phi(0,37) = 0,29, \\ \hat{p}_5 &= P(X \geq 21) = 1 - \Phi(1,48) = 0,07. \end{aligned}$$

Ahora podemos calcular

$$\begin{aligned} D^2 &= \sum_{i=1}^5 \frac{(n_i - 250\hat{p}_i)^2}{250\hat{p}_i} \\ &= \frac{(7 - 7,5)^2}{7,5} + \frac{(49 - 50)^2}{50} + \frac{(109 - 102,5)^2}{102,5} \\ &\quad + \frac{(67 - 72,5)^2}{72,5} + \frac{(18 - 17,5)^2}{17,5} \\ &= 0,82. \end{aligned}$$

Puesto que D^2 tiene $5 - 1 - 2 = 2$ grados de libertad, encontramos en las tablas de la distribución de χ^2 cuadrado que $P(D^2 \geq 0,82) \approx 0,65$, y que, por tanto, deberíamos aceptar la hipótesis de normalidad.

EJEMPLO 15.10. Consideremos los datos presentados en el ejemplo 15.5. ¿Tiene la variable aleatoria X = número de accidentes durante el período especificado de 400 días, una distribución de Poisson?

Estimemos primero el parámetro λ de la distribución. En el capítulo 14 encontramos que el estimador ML de λ está dado por el promedio muestral. Luego

$$\begin{aligned} \hat{\lambda} &= \frac{0(1448) + 1(805) + 2(206) + 3(34) + 4(4) + 5(2) + 6(1)}{1448 + 805 + 206 + 34 + 4 + 2 + 1} \\ &= \frac{1351}{2500} \\ &= 0,54. \end{aligned}$$

Sea $A_1: X = 0$; $A_2: X = 1$; $A_3: X = 2$; $A_4: X = 3$; $A_5: X \geq 4$. Luego $n_1 = 1448$; $n_2 = 805$; $n_3 = 206$; $n_4 = 34$; $n_5 = 7$, y obtenemos

$$\begin{aligned}\hat{p}_1 &= P(X = 0) = e^{-(0.54)} = 0,58 && y \quad n\hat{p}_1 = 1450 \\ \hat{p}_2 &= P(X = 1) = e^{-0.54}(0.543) = 0,31 && y \quad n\hat{p}_2 = 775 \\ \hat{p}_3 &= P(X = 2) = e^{-0.54} \frac{(0.543)^2}{2} = 0,08 && y \quad n\hat{p}_3 = 200 \\ \hat{p}_4 &= P(X = 3) = e^{-0.54} \frac{(0.543)^3}{6} = 0,01 && y \quad n\hat{p}_4 = 25 \\ \hat{p}_5 &= P(X \geq 4) = 1 - P(X < 4) = 0,02 && y \quad n\hat{p}_5 = 50\end{aligned}$$

Ahora podemos calcular

$$\begin{aligned}D^2 + \sum_{i=1}^5 \frac{(n_i - np_i)^2}{np_i} &= \frac{(1448 - 1450)^2}{1450} + \frac{(805 - 775)^2}{775} \\ &\quad + \frac{(206 - 200)^2}{200} + \frac{(34 - 25)^2}{25} + \frac{(7 - 50)^2}{50} = 42,2.\end{aligned}$$

Puesto que hay cinco categorías y estimábamos un parámetro, la variable aleatoria D^2 tiene la distribución aproximada χ^2_3 . Encontramos en las tablas de la distribución χ cuadrado que $P(D^2 \geq 42,2) \approx 0$ y, por tanto, deberíamos rechazar la hipótesis.

PROBLEMAS

15.1. Se supone que X tiene distribución $N(\mu, \sigma^2)$ con σ^2 conocida. Para docimar $H_0: \mu = \mu_0$ versus $H_1: \mu < \mu_0$ se propone el método siguiente: una muestra de tamaño n y rechazar H_0 siempre que el promedio muestral $\bar{X} < C$, en donde C es una constante que debe determinarse.

(a) Obtener una expresión para la función OC, $L(\mu)$ mediante la distribución normal tabulada.

(b) Si el nivel de significación de la dócima es $\alpha = 0,01$, obtener una expresión para C .

(c) Supóngase que $\sigma^2 = 4$ y supóngase que se está docimando $H_0: \mu = 30$ versus $H_1: \mu < 30$. Determinar el tamaño muestral n y la constante C a fin de satisfacer las condiciones $L(30) = 0,98$ y $L(27) = 0,01$.

(d) Supóngase que se obtienen los siguientes valores de muestra de X :

$$27,1; 29,3; 31,5; 33,0; 30,1; 30,9; 28,4; 32,4; 31,6; 28,9; 27,3; 29,1.$$

¿Rechazaría H_0 (versus H_1) como se estableció en (c) al nivel de significación del 5 por ciento?

15.2. Considerar la situación descrita en el problema 15.1 exceptuando que la hipótesis alterna sea de la forma $H_1: \mu \neq \mu_0$. Por tanto rechazamos H_0 siempre que $|\bar{X} - \mu_0| > C$. Responder las preguntas (a) y (b) anteriores.

15.3. Supóngase que X tiene una distribución de Poisson con parámetro λ . Para docimar $H_0: \lambda = \lambda_0$ versus $H_1: \lambda > \lambda_0$ se propone la siguiente dócima. Sacar una muestra de

tamaño n , calcular el promedio muestral \bar{X} , y rechazar H_0 siempre que $\bar{X} > C$, en donde C es una constante para ser determinada.

(a) Obtener una expresión para la función OC de la dócima anterior, $L(\lambda)$. [Indicación: use la propiedad reproductiva de la distribución de Poisson.]

(b) Graficar la función OC.

(c) Supóngase que se está docimando $H_0: \lambda = 0,2$ versus $H_1: \lambda > 0,2$. Se obtiene una muestra de tamaño $n = 10$ y rechazamos H_0 si $\bar{X} > 0,25$. ¿Cuál es el nivel de significación de esta dócima?

15.4. Establecer las propiedades de la ecuación (15.1) para la función OC, $L(\mu) = \Phi[(C - \mu)\sqrt{n/\sigma}]$.

15.5. Verificar las propiedades para la función OC, $L(p)$, como se definieron en la ecuación (15.2).

15.6. Verificar las propiedades para la función OC, $R(p)$, como se definieron en la ecuación (15.3).

15.7. Se sabe que una gran remesa de voltímetros contiene cierta proporción p de defectuosos. Para docimar $H_0: p = 0,2$ versus $H_1: p > 0,2$ se usa el siguiente método. Se saca una muestra de tamaño 5 y se cuenta X , el número de voltímetros defectuosos. Si $X \leq 1$, se acepta H_0 ; si $X > 4$, se rechaza H_0 ; y si $\bar{X} = 2, 3$, ó 4 se saca una segunda muestra de tamaño 5. Sea Y el número de instrumentos defectuosos en la segunda muestra. Se rechaza H_0 si $Y \geq 2$ y se acepta en otro caso. (Se supone que el lote muestreado es suficientemente grande de modo que pueda suponerse que X y Y son variables aleatorias independientes distribuidas binomialmente.)

(a) Obtener una expresión para $L(p)$, la función OC de la dócima anterior, y dibujar su gráfico.

(b) Encontrar el tamaño de la dócima anterior.

(c) ¿Cuál es la probabilidad del error del tipo 2 si $p = 0,5$?

15.8. Si $n = 4$ y $k = 3$, ¿cuántos valores posibles puede tomar la variable aleatoria D^2 , como se definió en la ecuación (15.4)?

15.9. (a) Calcular el valor esperado de la variable aleatoria D^2 como se definió en la ecuación (15.4).

(b) ¿Cómo se compara este valor con el valor esperado (asintótico) de D^2 obtenido con la distribución χ^2 cuadrado que puede usarse para aproximar la distribución de D^2 cuando n es grande?

15.10. Mediante un nuevo proceso se preparan tres clases de lubricantes. Cada uno de los lubricantes se prueba con cierto número de máquinas, y el resultado es luego clasificado como aceptable. Los datos de la tabla 15.1 representan los resultados de este experimento. Docimar la hipótesis de que la probabilidad p de que un lubricante tenga un resultado aceptable es la misma para los tres lubricantes. [Indicación: estime primero p de la muestra.]

TABLA 15.1

	Lubricante 1	Lubricante 2	Lubricante 3
Aceptable	144	152	140
Inaceptable	56	48	60
Total	200	200	200

15.11. Al usar varias leyes de falla hemos encontrado que la distribución exponencial desempeña un papel muy importante y que por tanto interesa poder decidir si una muestra particular de tiempos de falla proviene de una distribución exponencial básica. Supóngase que se han probado 335 bombillas y el resumen siguiente de su duración T (en horas) está disponible:

Duración (en horas)	$0 \leq T < 100$	$100 \leq T < 200$	$200 \leq T < 300$	$300 \leq T < 400$	$T \geq 400$
Número	82	71	68	62	52

De los tiempos de falla que se registraron, se encontró que \bar{T} el promedio muestral era igual a 123,5 horas. Usando esta información, docimar la hipótesis de que T , el tiempo para fallar está distribuido exponencialmente.

15.12. Supóngase que la variable aleatoria X tiene la siguiente fdp:

$$f(x) = \frac{\alpha \cos \alpha x}{\sin(\pi/2)\alpha}, \quad 0 < x < \pi/2, \quad \text{en donde } 0 < \alpha < 1.$$

Para docimar $H_0: \alpha = \alpha_0$ se propone la siguiente dócima. Sacar una observación de X , sea X_1 y rechazar H_0 si $X_1 > 1$.

- (a) Obtener una expresión para la función OC, $L(\alpha)$, de esta dócima y graficarla.
- (b) ¿Cuál es el tamaño de esta dócima si $\alpha_0 = \pi/4$?

15.13. En una malla de 165 celdas, se contó el número de granos de grafito en cada celda. Así se obtuvieron los datos de la tabla 15.2. Docimar la hipótesis de que el número de granos en cada una de las celdas es una variable aleatoria con una distribución de Poisson. [Indicación: junte las observaciones ≤ 2 y también aquellas ≥ 10 .]

TABLA 15.2

Número de granos de grafito por celda	Observados	Número de granos de grafito por celda	Observados
0	1	7	17
1	1	8	22
2	5	9	21
3	7	10	4
4	20	11	2
5	34	12	1
6	30		

Referencias

La siguiente lista de referencias, aunque no es, ni mucho menos, exhaustiva, deberá dar al lector interesado la oportunidad de encontrar mucha lectura suplementaria y complementaria sobre diversos temas presentados en el texto. Además de encontrar varios temas que no están incluidos en esta presentación, el lector encontrará ciertos otros considerados o con mucho mayor detalle o bajo un punto de vista algo diferente.

Además de tomar nota de los textos que figuran a continuación, el lector también deberá estar consciente de ciertas revistas profesionales en las cuales se hacen avances importantes. Muchas de esas revistas, naturalmente, son escritas por personas con una experiencia y comprensión de este tema considerablemente mayores que las que pueden obtenerse estudiando un semestre. Sin embargo, hay varias de ellas que a veces contienen artículos muy comprensibles y al alcance de un estudiante que ha madurado el tema de este texto; y entre éstas el *Journal of the American Statistical Association* y *Technometrics*. La última es en realidad subtitulada «*A Journal of Statistics for the Physical, Chemical, and Engineering Sciences*», y, por lo tanto, puede ser de interés particular para aquellos estudiantes a quienes está especialmente dedicado este texto.

Muchos de los libros en esta lista contienen exposiciones de la mayoría de los temas incluidos en este texto. Sin embargo, algunas de las referencias son algo más especializadas y especialmente adecuados sólo algunos capítulos. En tales casos, los capítulos específicos figuran entre paréntesis después de las referencias.

En inglés:

- BAZOVSKY, I., *Reliability Theory and Practice*. Englewood Cliffs, N. J.: Prentice-Hall, Inc., 1961 (11).
- BERMAN, SIMEON M., *The Elements of Probability*. Reading, Mass.: Addison-Wesley Publishing Co., 1969.
- BOWKER, A. H., y G. J. LIEBERMAN, *Engineering Statistics*. Englewood Cliffs, N. J.: Prentice-Hall, Inc., 1959.
- DERMAN, C., y M. KLEIN, *Probability and Statistical Inference for Engineers*. New York: Oxford University Press, 1959.
- EHRENFELD, S., y S. B. LITTAUER, *Introduction to Statistical Method*. New York: McGraw-Hill Book Co., 1964.
- FELLER, W., *An Introduction to Probability Theory and Its Applications*, vol. 1, 3.^a edición. New York: John Wiley and Sons, Inc., 1968 (1, 2, 3).
- FREUND, J. E., *Mathematical Statistics*. Englewood Cliffs, N. J.: Prentice-Hall, Inc., 1962.
- FRY, T. C., *Probability and Its Engineering Uses*, 2.^a edición. Princeton, N. J.: D. Van Nostrand, 1964.

- GNEDENKO, B. V., *The Theory of Probability* (traducción del ruso). New York: Chelsea Publishing Co., Inc., 1962.
- GUTTMAN, I., y S. S. WILKS, *Introductory Engineering Statistics*. New York: John Wiley and Sons, Inc., 1965.
- HABER, A., R. P. RUNYON y P. BADIA, *Readings in Statistics*. Reading, Mass.: Addison-Wesley Publishing Co., 1970.
- LINDGREN, B. W., y G. W. McELRATH, *Introduction to Probability and Statistics*, 3.^a edición. New York: The Macmillan Co., 1969 (13, 14, 15).
- LLOYD, D. K., y M. LIPOW, *Reliability: Management, Methods, and Mathematics*. Englewood Cliffs, N. J.: Prentice-Hall, Inc., 1962 (11).
- MCCORD, J. R., III, y R. M. MORONEY, JR., *Introduction to Probability Theory*. New York: The Macmillan Co., 1964.
- MILLER, I., y J. E. FREUND, *Probability and Statistics for Engineers*. Englewood Cliffs, N. J.: Prentice-Hall, Inc., 1965.
- MOOD, A. M., y F. A. GRAYBILL, *Introduction to the Theory of Statistics*, 2.^a edición. New York: McGraw-Hill Book Co., 1963.
- PARZEN, E., *Modern Probability Theory and Its Applications*. New York: John Wiley and Sons, Inc., 1960.
- WADSWORTH, B. P., y J. G. BRYAN, *Introduction to Probability and Random Variables*. New York: McGraw-Hill Book Co., 1960.

En español:

Damos a continuación una lista de libros de los cuales existe versión española, para beneficio del estudiante y del lector en general que quiera investigar algunos de los temas tratados en este libro y, en especial, en relación con algunas áreas particulares. (N. del T.)

- ARLEY, NIELS, y RANDER BUCH, *Introducción a la teoría de la probabilidad y de la Estadística*. Madrid: Alhambra, 1968.
- CARRANZA, ROQUE, *Probabilidad y Estadística*. Buenos Aires: Universidad de Buenos Aires, 1965.
- CRAMER, HARALD, *Métodos matemáticos de Estadística*. Madrid: Aguilar, 1968.
- DIXON, JOHN, *Introducción a la probabilidad*. México: Limusa-Wiley, 1970.
- DIXON, WILFRID, y FRANK MASSEY, *Introducción al análisis estadístico*. Madrid: Ediciones del Castillo, 1965.
- FREEMAN, HAROLD A., *Introducción a la inferencia estadística*. México: Trillas, 1970.
- GNEDENKO, B. V., y A. I. JINCHIN, *Introducción al cálculo de probabilidades* (traducción del ruso). Buenos Aires: Eudeba, 1971.
- GUENTHER, WILLIAM, *Introducción a la inferencia estadística*. New York: McGraw-Hill, 1965.
- HOEL, PAUL, *Introducción a la estadística matemática*. Barcelona: Ariel, 1968.
- LIPSCHUTZ, SEYMOUR, *Probabilidad*. México: McGraw-Hill, 1970.
- MAISEL, LOUIS, *Probabilidad y Estadística*. Bogotá: Fondo Educativo Interamericano, S. A., 1973.
- RÍOS, SIXTO, *Métodos estadísticos*, 5.^a edición. New York: McGraw-Hill, 1967.

Apéndice

TABLA 1. Valores de la función distribución normal estándar*

$$\Phi(z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du = P(Z \leq z)$$

<i>z</i>	0	1	2	3	4	5	6	7	8	9
-3,0	0,0013	0,0010	0,0007	0,0005	0,0003	0,0002	0,0002	0,0001	0,0001	0,0000
-2,9	0,0019	0,0018	0,0017	0,0017	0,0016	0,0016	0,0015	0,0015	0,0014	0,0014
-2,8	0,0026	0,0025	0,0024	0,0023	0,0023	0,0022	0,0021	0,0021	0,0020	0,0019
-2,7	0,0035	0,0034	0,0033	0,0032	0,0031	0,0030	0,0029	0,0028	0,0027	0,0026
-2,6	0,0047	0,0045	0,0044	0,0043	0,0041	0,0040	0,0039	0,0038	0,0037	0,0036
-2,5	0,0062	0,0060	0,0059	0,0057	0,0055	0,0054	0,0052	0,0051	0,0049	0,0048
-2,4	0,0082	0,0080	0,0078	0,0075	0,0073	0,0071	0,0069	0,0068	0,0066	0,0064
-2,3	0,0107	0,0104	0,0102	0,0099	0,0096	0,0094	0,0091	0,0089	0,0087	0,0084
-2,2	0,0139	0,0136	0,0132	0,0129	0,0126	0,0122	0,0119	0,0116	0,0113	0,0110
-2,1	0,0179	0,0174	0,0170	0,0166	0,0162	0,0158	0,0154	0,0150	0,0146	0,0143
-2,0	0,0228	0,0222	0,0217	0,0212	0,0207	0,0202	0,0197	0,0192	0,0188	0,0183
-1,9	0,0287	0,0281	0,0274	0,0268	0,0262	0,0256	0,0250	0,0244	0,0238	0,0233
-1,8	0,0359	0,0352	0,0344	0,0336	0,0329	0,0322	0,0314	0,0307	0,0300	0,0294
-1,7	0,0446	0,0436	0,0427	0,0418	0,0409	0,0401	0,0392	0,0384	0,0375	0,0367
-1,6	0,0548	0,0537	0,0526	0,0516	0,0505	0,0495	0,0485	0,0475	0,0465	0,0455
-1,5	0,0668	0,0655	0,0643	0,0630	0,0618	0,0606	0,0594	0,0582	0,0570	0,0559
-1,4	0,0808	0,0793	0,0778	0,0764	0,0749	0,0735	0,0722	0,0708	0,0694	0,0681
-1,3	0,0968	0,0951	0,0934	0,0918	0,0901	0,0885	0,0869	0,0853	0,0838	0,0823
-1,2	0,1151	0,1131	0,1112	0,1093	0,1075	0,1056	0,1038	0,1020	0,1003	0,0985
-1,1	0,1357	0,1335	0,1314	0,1292	0,1271	0,1251	0,1230	0,1210	0,1190	0,1170
-1,0	0,1587	0,1562	0,1539	0,1515	0,1492	0,1469	0,1446	0,1423	0,1401	0,1379
-0,9	0,1841	0,1814	0,1788	0,1762	0,1736	0,1711	0,1685	0,1660	0,1635	0,1611
-0,8	0,2119	0,2090	0,2061	0,2033	0,2005	0,1977	0,1949	0,1922	0,1894	0,1867
-0,7	0,2420	0,2389	0,2358	0,2327	0,2297	0,2266	0,2236	0,2206	0,2177	0,2148
-0,6	0,2743	0,2709	0,2676	0,2643	0,2611	0,2578	0,2546	0,2514	0,2483	0,2451
-0,5	0,3085	0,3050	0,3015	0,2981	0,2946	0,2912	0,2877	0,2843	0,2810	0,2776
-0,4	0,3446	0,3409	0,3372	0,3336	0,3300	0,3264	0,3228	0,3192	0,3156	0,3121
-0,3	0,3821	0,3783	0,3745	0,3707	0,3669	0,3632	0,3594	0,3557	0,3520	0,3483
-0,2	0,4207	0,4168	0,4129	0,4090	0,4052	0,4013	0,3974	0,3936	0,3897	0,3859
-0,1	0,4602	0,4562	0,4522	0,4483	0,4443	0,4404	0,4364	0,4325	0,4286	0,4247
-0,0	0,5000	0,4960	0,4920	0,4880	0,4840	0,4801	0,4761	0,4721	0,4681	0,4641

* B. W. Lindgren, *Statistical Theory*. New York: The Macmillan Co., 1960.

TABLA 1. (*Continuación*)

$$\Phi(z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du = P(Z \leq z)$$

<i>z</i>	0	1	2	3	4	5	6	7	8	9
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	0,7703	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	0,9278	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9430	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	0,9648	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9700	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9762	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	0,9874	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	0,9990	0,9993	0,9995	0,9997	0,9998	0,9998	0,9999	0,9999	1,0000

TABLA 2. Función de la distribución binomial

$$1 - F(x - 1) = \sum_{r=x}^{r=n} \binom{n}{r} p^r q^{n-r}$$

$n = 10$ $x = 10$	$n = 10$ $x = 9$	$n = 10$ $x = 8$	$n = 10$ $x = 7$	p
0,0000000	0,0000000	0,0000000	0,0000000	0,01
,0000000	,0000000	,0000000	,0000000	,02
,0000000	,0000000	,0000000	,0000000	,03
,0000000	,0000000	,0000000	,0000000	,04
,0000000	,0000000	,0000000	,0000001	,05
,0000000	,0000000	,0000000	,0000003	,06
,0000000	,0000000	,0000000	,0000008	,07
,0000000	,0000000	,0000001	,0000020	,08
,0000000	,0000000	,0000002	,0000045	,09
,0000000	,0000000	,0000004	,0000091	,10
,0000000	,0000000	,0000008	,0000173	,11
,0000000	,0000000	,0000015	,0000308	,12
,0000000	,0000001	,0000029	,0000525	,13
,0000000	,0000002	,0000051	,0000856	,14
,0000000	,0000003	,0000087	,0001346	,15
,0000000	,0000006	,0000142	,0002051	,16
,0000000	,0000010	,0000226	,0003042	,17
,0000000	,0000017	,0000350	,0004401	,18
,0000001	,0000027	,0000528	,0006229	,19
,0000001	,0000042	,0000779	,0008644	,20
,0000002	,0000064	,0001127	,0011783	,21
,0000003	,0000097	,0001599	,0015804	,22
,0000004	,0000143	,0002232	,0020885	,23
,0000006	,0000207	,0003068	,0027228	,24
,0000010	,0000296	,0004158	,0035057	,25
,0000014	,0000416	,0005362	,0044618	,26
,0000021	,0000577	,0007350	,0056181	,27
,0000030	,0000791	,0009605	,0070039	,28
,0000042	,0001072	,0012420	,0086507	,29
,0000059	,0001437	,0015904	,0105921	,30
,0000082	,0001906	,0020179	,0128637	,31
,0000113	,0002505	,0025384	,0155029	,32
,0000153	,0003263	,0031673	,0185489	,33
,0000206	,0004214	,0039219	,0220422	,34
,0000276	,0005399	,0048213	,0260243	,35
,0000366	,0006865	,0058864	,0305376	,36
,0000481	,0008668	,0071403	,0356252	,37
,0000628	,0010871	,0086079	,0413301	,38
,0000814	,0013546	,0103163	,0476949	,39
,0001049	,0016777	,0122946	,0547619	,40
,0001342	,0020658	,0145738	,0625719	,41
,0001708	,0025295	,0171871	,0711643	,42
,0002161	,0030809	,0201696	,0805763	,43
,0002720	,0037335	,0235583	,0908427	,44
,0003405	,0045022	,0273918	,1019949	,45
,0004242	,0054040	,0317105	,1140612	,46
,0005260	,0064574	,0365560	,1270655	,47
,0006493	,0076828	,0419713	,1410272	,48
,0007979	,0091028	,0480003	,1559607	,49
,0009766	,0107422	,0546875	,1718750	,50

TABLA 2 (*Continuación*)

$$1 - F(x - 1) = \sum_{r=x}^{r=n} \binom{n}{r} p^r q^{n-r}$$

$n = 10$	p					
$x = 6$	$x = 5$	$x = 4$	$x = 3$	$x = 2$	$x = 1$	
0,0000000	0,0000000	0,0000020	0,0001138	0,0042662	0,0956179	0,01
,0000000	,0000007	,0000305	,0008639	,0161776	,1829272	,02
,0000001	,0000054	,0001471	,0027650	,0345066	,2625759	,03
,0000007	,0000218	,0004426	,0062137	,0581538	,3351674	,04
,0000028	,0000637	,0010285	,0115036	,0861384	,4012631	,05
,0000079	,0001517	,0020293	,0188378	,1175880	,4613849	,06
,0000193	,0003139	,0035761	,0283421	,1517299	,5160177	,07
,0000415	,0005857	,0058013	,0400754	,1878825	,5656115	,08
,0000810	,0010096	,0088338	,0540400	,2254471	,6105839	,09
,0001469	,0016349	,0127952	,0701908	,2639011	,6513216	,10
,0002507	,0025170	,0177972	,0884435	,3027908	,6881828	,11
,0004069	,0037161	,0239388	,1086818	,3417250	,7214990	,12
,0006332	,0052967	,0313048	,1307642	,3803692	,7515766	,13
,0009505	,0073263	,0399642	,1545298	,4184400	,7786984	,14
,0013832	,0098741	,0499698	,1798035	,4557002	,8031256	,15
,0019593	,0130101	,0613577	,2064005	,4919536	,8250988	,16
,0027098	,0168038	,0741472	,2341305	,5270412	,8448396	,17
,0036694	,0213229	,0883411	,2628010	,5608368	,8625520	,18
,0048757	,0266325	,1039261	,2922204	,5932435	,8784233	,19
,0063694	,0327935	,1208739	,3222005	,6241904	,8926258	,20
,0081935	,0398624	,1391418	,3525586	,6536289	,9053172	,21
,0103936	,0478897	,1586739	,3831197	,6815306	,9166422	,22
,0130167	,0569196	,1794024	,4137173	,7078843	,9267332	,23
,0161116	,0669890	,2012487	,4441949	,7326936	,9357111	,24
,0197277	,0781269	,2241249	,4744072	,7559748	,9436865	,25
,0239148	,0903542	,2479349	,5042200	,7777550	,9507601	,26
,0287224	,1036831	,2725761	,5335112	,7980705	,9570237	,27
,0341994	,1181171	,2979405	,5621710	,8169646	,9625609	,28
,0403932	,1336503	,3239164	,5901015	,8344869	,9674476	,29
,0473490	,1502683	,3503893	,6172172	,8506917	,9717525	,30
,0551097	,1679475	,3772433	,6434445	,8656366	,9755381	,31
,0637149	,1866554	,4043626	,6687212	,8793821	,9788608	,32
,0732005	,2063514	,4316320	,6929966	,8919901	,9817716	,33
,0835979	,2269866	,4589388	,7162304	,9035235	,9843166	,34
,0949341	,2485045	,4861730	,7383926	,9140456	,9865373	,35
,1072304	,2708415	,5132284	,7594627	,9236190	,9884708	,36
,1205026	,2939277	,5400038	,7794292	,9323056	,9901507	,37
,1347603	,3176870	,5664030	,7982887	,9401661	,9916070	,38
,1500068	,3420385	,5923361	,8160453	,9472594	,9928666	,39
,1662386	,3668967	,6177194	,8327102	,9536426	,9939534	,40
,1834452	,3921728	,6424762	,8483007	,9593705	,9948888	,41
,2016092	,4177749	,6665372	,8628393	,9644958	,9956920	,42
,2207058	,4436094	,6898401	,8763538	,9690684	,9963797	,43
,2407033	,4695813	,7123307	,8888757	,9731358	,9969669	,44
,2615627	,4955954	,7339621	,9004403	,9767429	,9974670	,45
,2832382	,5215571	,7546952	,9110859	,9799319	,9978917	,46
,3056772	,5473730	,7744985	,9208530	,9827422	,9982511	,47
,3288205	,5729517	,7933480	,9297839	,9852109	,9985544	,48
,3526028	,5982047	,8112268	,9379222	,9873722	,9988096	,49
,3769531	,6230469	,8281250	,9453125	,9892578	,9990234	,50

TABLA 3. Función de la distribución de Poisson *

$$1 - F(x - 1) = \sum_{r=x}^{r=\infty} \frac{e^{-a} a^r}{r!}$$

<i>x</i>	<i>a</i> = 0,2	<i>a</i> = 0,3	<i>a</i> = 0,4	<i>a</i> = 0,5	<i>a</i> = 0,6
0	1,0000000	1,0000000	1,0000000	1,0000000	1,0000000
1	,1812692	,2591818	,3296800	,393469	,451188
2	,0175231	,0369363	,0615519	,090204	,121901
3	,0011485	,0035995	,0079263	,014388	,023115
4	,0000568	,0002658	,0007763	,001752	,003358
5	,0000023	,0000158	,0000612	,000172	,000394
6	,0000001	,0000008	,0000040	,000014	,000039
7			,0000002	,000001	,000003
<i>x</i>	<i>a</i> = 0,7	<i>a</i> = 0,8	<i>a</i> = 0,9	<i>a</i> = 1,0	<i>a</i> = 1,2
0	1,0000000	1,0000000	1,0000000	1,0000000	1,0000000
1	,503415	,550671	,593430	,632121	,698806
2	,155805	,191208	,227518	,264241	,337373
3	,034142	,047423	,062857	,080301	,120513
4	,005753	,009080	,013459	,018988	,033769
5	,000786	,001411	,002344	,003660	,007746
6	,000090	,000184	,000343	,000594	,001500
7	,000009	,000021	,000043	,000083	,000251
8	,000001	,000002	,000005	,000010	,000037
9				,000001	,000005
10					,000001
<i>x</i>	<i>a</i> = 1,4	<i>a</i> = 1,6	<i>a</i> = 1,8	<i>a</i> = 1,9	<i>a</i> = 2,0
0	1,000000	1,000000	1,000000	1,000000	1,000000
1	,753403	,798103	,834701	,850431	,864665
2	,408167	,475069	,537163	,566251	,593994
3	,166502	,216642	,269379	,296280	,323324
4	,053725	,078813	,108708	,125298	,142877
5	,014253	,023682	,036407	,044081	,052653
6	,003201	,006040	,010378	,013219	,016564
7	,000622	,001336	,002569	,003446	,004534
8	,000107	,000260	,000562	,000793	,001097
9	,000016	,000045	,000110	,000163	,000237
10	,000002	,000007	,000019	,000030	,000046
11		,000001	,000003	,000005	,000008

* E. C. Molina, *Poisson's Exponential Binomial Limit*. Princeton, N. J.: D. Van Nostrand, Inc., 1947.

TABLA 3 (*Continuación*)

$$1 - F(x - 1) = \sum_{r=x}^{r=\infty} \frac{e^{-a} a^r}{r!}$$

<i>x</i>	<i>a</i> = 2,5	<i>a</i> = 3,0	<i>a</i> = 3,5	<i>a</i> = 4,0	<i>a</i> = 4,5	<i>a</i> = 5,0
0	1,000000	1,000000	1,000000	1,000000	1,000000	1,000000
1	,917915	,950213	,969803	,981684	,988891	,993262
2	,712703	,800852	,864112	,908422	,938901	,959572
3	,456187	,576810	,679153	,761897	,826422	,875348
4	,242424	,352768	,463367	,566530	,657704	,734974
5	,108822	,184737	,274555	,371163	,467896	,559507
6	,042021	,083918	,142386	,214870	,297070	,384039
7	,014187	,033509	,065288	,110674	,168949	,237817
8	,004247	,011905	,026739	,051134	,086586	,133372
9	,001140	,003803	,009874	,021363	,040257	,068094
10	,000277	,001102	,003315	,008132	,017093	,031828
11	,000062	,000292	,001019	,002840	,006669	,013695
12	,000013	,000071	,000289	,000915	,002404	,005453
13	,000002	,000016	,000076	,000274	,000805	,002019
14		,000003	,000019	,000076	,000252	,000698
15		,000001	,000004	,000020	,000074	,000226
16			,000001	,000005	,000020	,000069
17				,000001	,000005	,000020
18					,000001	,000005
19						,000001

TABLA 4. Valores críticos para la distribución t de Student*
 $\Pr \{t \text{ de Student} \leq \text{valor tabulado}\} = \gamma$

f	0,75	0,90	0,95	0,975	0,99	0,995
1	1,0000	3,0777	6,3138	12,7062	31,8207	63,6574
2	0,8165	1,8856	2,9200	4,3027	6,9646	9,9248
3	0,7649	1,6377	2,3534	3,1824	4,5407	5,8409
4	0,7407	1,5332	2,1318	2,7764	3,7469	4,6041
5	0,7267	1,4759	2,0150	2,5706	3,3649	4,0322
6	0,7176	1,4398	1,9432	2,4469	3,1427	3,7074
7	0,7111	1,4149	1,8946	2,3646	2,9980	3,4995
8	0,7064	1,3968	1,8595	2,3060	2,8965	3,3554
9	0,7027	1,3830	1,8331	2,2622	2,8214	3,2498
10	0,6998	1,3722	1,8125	2,2281	2,7638	3,1693
11	0,6974	1,3634	1,7959	2,2010	2,7181	3,1058
12	0,6955	1,3562	1,7823	2,1788	2,6810	3,0545
13	0,6938	1,3502	1,7709	2,1604	2,6503	3,0123
14	0,6924	1,3450	1,7613	2,1448	2,6245	2,9768
15	0,6912	1,3406	1,7531	2,1315	2,6025	2,9467
16	0,6901	1,3368	1,7459	2,1199	2,5835	2,9208
17	0,6892	1,3334	1,7396	2,1098	2,5669	2,8982
18	0,6884	1,3304	1,7341	2,1009	2,5524	2,8784
19	0,6876	1,3277	1,7291	2,0930	2,5395	2,8609
20	0,6870	1,3253	1,7247	2,0860	2,5280	2,8453
21	0,6864	1,3232	1,7207	2,0796	2,5177	2,8314
22	0,6858	1,3212	1,7171	2,0739	2,5083	2,8188
23	0,6853	1,3195	1,7139	2,0687	2,4999	2,8073
24	0,6848	1,3178	1,7109	2,0639	2,4922	2,7969
25	0,6844	1,3163	1,7081	2,0595	2,4851	2,7874
26	0,6840	1,3150	1,7056	2,0555	2,4786	2,7787
27	0,6837	1,3137	1,7033	2,0518	2,4727	2,7707
28	0,6834	1,3125	1,7011	2,0484	2,4671	2,7633
29	0,6830	1,3114	1,6991	2,0452	2,4620	2,7564
30	0,6828	1,3104	1,6973	2,0423	2,4573	2,7500
31	0,6825	1,3095	1,6955	2,0395	2,4528	2,7440
32	0,6822	1,3086	1,6939	2,0369	2,4487	2,7385
33	0,6820	1,3077	1,6924	2,0345	2,4448	2,7333
34	0,6818	1,3070	1,6909	2,0322	2,4411	2,7284
35	0,6816	1,3062	1,6896	2,0301	2,4377	2,7238
36	0,6814	1,3055	1,6883	2,0281	2,4345	2,7195
37	0,6812	1,3049	1,6871	2,0262	2,4314	2,7154
38	0,6810	1,3042	1,6860	2,0244	2,4286	2,7116
39	0,6808	1,3036	1,6849	2,0227	2,4258	2,7079
40	0,6807	1,3031	1,6839	2,0211	2,4233	2,7045
41	0,6805	1,3025	1,6829	2,0195	2,4208	2,7012
42	0,6804	1,3020	1,6820	2,0181	2,4185	2,6981
43	0,6802	1,3016	1,6811	2,0167	2,4163	2,6951
44	0,6801	1,3011	1,6802	2,0154	2,4141	2,6923
45	0,6800	1,3006	1,6794	2,0141	2,4121	2,6896

* D. B. Owen, *Handbook of Statistical Tables*. Reading, Mass.: Addison-Wesley Publishing Co., 1962. (Cortesía de la Atomic Energy Commission, Washington, D.C.)

TABLA 4 (*Continuación*) $\Pr \{t \text{ de Student} \leq \text{valores tabulados}\} \gamma$

<i>f</i>	0,75	0,90	0,95	0,975	0,99	0,995
46	0,6799	1,3002	1,6787	2,0129	2,4102	2,6870
47	0,6797	1,2998	1,6779	2,0117	2,4083	2,6846
48	0,6796	1,2994	1,6772	2,0106	2,4066	2,6822
49	0,6795	1,2991	1,6766	2,0096	2,4049	2,6800
50	0,6794	1,2987	1,6759	2,0086	2,4033	2,6778
51	0,6793	1,2984	1,6753	2,0076	2,4017	2,6757
52	0,6792	1,2980	1,6747	2,0066	2,4002	2,6737
53	0,6791	1,2977	1,6741	2,0057	2,3988	2,6718
54	0,6791	1,2974	1,6736	2,0049	2,3974	2,6700
55	0,6790	1,2971	1,6730	2,0040	2,3961	2,6682
56	0,6789	1,2969	1,6725	2,0032	2,3948	2,6665
57	0,6788	1,2966	1,6720	2,0025	2,3936	2,6649
58	0,6787	1,2963	1,6716	2,0017	2,3924	2,6633
59	0,6787	1,2961	1,6711	2,0010	2,3912	2,6618
60	0,6786	1,2958	1,6706	2,0003	2,3901	2,6603
61	0,6785	1,2956	1,6702	1,9996	2,3890	2,6589
62	0,6785	1,2954	1,6698	1,9990	2,3880	2,6575
63	0,6784	1,2951	1,6694	1,9983	2,3870	2,6561
64	0,6783	1,2949	1,6690	1,9977	2,3860	2,6549
65	0,6783	1,2947	1,6686	1,9971	2,3851	2,6536
66	0,6782	1,2945	1,6683	1,9966	2,3842	2,6524
67	0,6782	1,2943	1,6679	1,9960	2,3833	2,6512
68	0,6781	1,2941	1,6676	1,9955	2,3824	2,6501
69	0,6781	1,2939	1,6672	1,9949	2,3816	2,6490
70	0,6780	1,2938	1,6669	1,9944	2,3808	2,6479
71	0,6780	1,2936	1,6666	1,9939	2,3800	2,6469
72	0,6779	1,2934	1,6663	1,9935	2,3793	2,6459
73	0,6779	1,2933	1,6660	1,9930	2,3785	2,6449
74	0,6778	1,2931	1,6657	1,9925	2,3778	2,6439
75	0,6778	1,2929	1,6654	1,9921	2,3771	2,6430
76	0,6777	1,2928	1,6652	1,9917	2,3764	2,6421
77	0,6777	1,2926	1,6649	1,9913	2,3758	2,6412
78	0,6776	1,2925	1,6646	1,9908	2,3751	2,6403
79	0,6776	1,2924	1,6644	1,9905	2,3745	2,6395
80	0,6776	1,2922	1,6641	1,9901	2,3739	2,6387
81	0,6775	1,2921	1,6639	1,9897	2,3733	2,6379
82	0,6775	1,2920	1,6636	1,9893	2,3727	2,6371
83	0,6775	1,2918	1,6634	1,9890	2,3721	2,6364
84	0,6774	1,2917	1,6632	1,9886	2,3716	2,6356
85	0,6774	1,2916	1,6630	1,9883	2,3710	2,6349
86	0,6774	1,2915	1,6628	1,9879	2,3705	2,6342
87	0,6773	1,2914	1,6626	1,9876	2,3700	2,6335
88	0,6773	1,2912	1,6624	1,9873	2,3695	2,6329
89	0,6773	1,2911	1,6622	1,9870	2,3690	2,6322
90	0,6772	1,2910	1,6620	1,9867	2,3685	2,6316

TABLA 5. Valores críticos para la distribución de chi cuadrado *

$$\Pr \{ \chi^2 \text{ r.v. con } f \text{ grados de libertad} \leq \text{valores tabulados} \} = \gamma$$

<i>f</i>	0,005	0,01	0,025	0,05	0,10	0,25
1	—	—	0,001	0,004	0,016	0,102
2	0,010	0,020	0,051	0,103	0,211	0,575
3	0,072	0,115	0,216	0,352	0,584	1,213
4	0,207	0,297	0,484	0,711	1,064	1,923
5	0,412	0,554	0,831	1,145	1,610	2,675
6	0,676	0,872	1,237	1,635	2,204	3,455
7	0,989	1,239	1,690	2,167	2,833	4,255
8	1,344	1,646	2,180	2,733	3,490	5,071
9	1,735	2,088	2,700	3,325	4,168	5,899
10	2,156	2,558	3,247	3,940	4,865	6,737
11	2,603	3,053	3,816	4,575	5,578	7,584
12	3,074	3,571	4,404	5,226	6,304	8,438
13	3,565	4,107	5,009	5,892	7,042	9,299
14	4,075	4,660	5,629	6,571	7,790	10,165
15	4,601	5,229	6,262	7,261	8,547	11,037
16	5,142	5,812	6,908	7,962	9,312	11,912
17	5,697	6,408	7,564	8,672	10,085	12,792
18	6,265	7,015	8,231	9,390	10,865	13,675
19	6,844	7,633	8,907	10,117	11,651	14,562
20	7,434	8,260	9,591	10,851	12,443	15,452
21	8,034	8,897	10,283	11,591	13,240	16,344
22	8,643	9,542	10,982	12,338	14,042	17,240
23	9,260	10,196	11,689	13,091	14,848	18,137
24	9,886	10,856	12,401	13,848	15,659	19,037
25	10,520	11,524	13,120	14,611	16,473	19,939
26	11,160	12,198	13,844	15,379	17,292	20,843
27	11,808	12,879	14,573	16,151	18,114	21,749
28	12,461	13,565	15,308	16,928	18,939	22,657
29	13,121	14,257	16,047	17,708	19,768	23,567
30	13,787	14,954	16,791	18,493	20,599	24,478
31	14,458	15,655	17,539	19,281	21,434	25,390
32	15,134	16,362	18,291	20,072	22,271	26,304
33	15,815	17,074	19,047	20,867	23,110	27,219
34	16,501	17,789	19,806	21,664	23,952	28,136
35	17,192	18,509	20,569	22,465	24,797	29,054
36	17,887	19,233	21,336	23,269	25,643	29,973
37	18,586	19,960	22,106	24,075	26,492	30,893
38	19,289	20,691	22,878	24,884	27,343	31,815
39	19,996	21,426	23,654	25,695	28,196	32,737
40	20,707	22,164	24,433	26,509	29,051	33,660
41	21,421	22,906	25,215	27,326	29,907	34,585
42	22,138	23,650	25,999	28,144	30,765	35,510
43	22,859	24,398	26,785	28,965	31,625	36,436
44	23,584	25,148	27,575	29,787	32,487	37,363
45	24,311	25,901	28,366	30,612	33,350	38,291

* D. B. Owen, *Handbook of Statistical Tables*. Reading, Mass.: Addison-Wesley Publishing Co., 1962. (Cortesía de la Atomic Energy Commission, Washington, D.C.)

TABLA 5 (*Continuación*)
 $\Pr \{ \chi^2 \text{ r.v. con } f \text{ grados de libertad} \leq \text{valores tabulados} \} = \gamma$

<i>f</i>	0,75	0,90	0,95	0,975	0,99	0,995
1	1,323	2,706	3,841	5,024	6,635	7,879
2	2,773	4,605	5,991	7,378	9,210	10,597
3	4,108	6,251	7,815	9,348	11,345	12,838
4	5,385	7,779	9,488	11,143	13,277	14,860
5	6,626	9,236	11,071	12,833	15,086	16,750
6	7,841	10,645	12,592	14,449	16,812	18,548
7	9,037	12,017	14,067	16,013	18,475	20,278
8	10,219	13,362	15,507	17,535	20,090	21,955
9	11,389	14,684	16,919	19,023	21,666	23,589
10	12,549	15,987	18,307	20,483	23,209	25,188
11	13,701	17,275	19,675	21,920	24,725	26,757
12	14,845	18,549	21,026	23,337	26,217	28,299
13	15,984	19,812	22,362	24,736	27,688	29,819
14	17,117	21,064	23,685	26,119	29,141	31,319
15	18,245	22,307	24,996	27,488	30,578	32,801
16	19,369	23,542	26,296	28,845	32,000	34,267
17	20,489	24,769	27,587	30,191	33,409	35,718
18	21,605	25,989	28,869	31,526	34,805	37,156
19	22,718	27,204	30,144	32,852	36,191	38,582
20	23,828	28,412	31,410	34,170	37,566	39,997
21	24,935	29,615	32,671	35,479	38,932	41,401
22	26,039	30,813	33,924	36,781	40,289	42,796
23	27,141	32,007	35,172	38,076	41,638	44,181
24	28,241	33,196	36,415	39,364	42,980	45,559
25	29,339	34,382	37,652	40,646	44,314	46,928
26	30,435	35,563	38,885	41,923	45,642	48,290
27	31,528	36,741	40,113	43,194	46,963	49,645
28	32,620	37,916	41,337	44,461	48,278	50,993
29	33,711	39,087	42,557	45,722	49,588	52,336
30	34,800	40,256	43,773	46,979	50,892	53,672
31	35,887	41,422	44,985	48,232	52,191	55,003
32	36,973	42,585	46,194	49,480	53,486	56,328
33	38,058	43,745	47,400	50,725	54,776	57,648
34	39,141	44,903	48,602	51,966	56,061	58,964
35	40,223	46,059	49,802	53,203	57,342	60,275
36	41,304	47,212	50,998	54,437	58,619	61,581
37	42,383	48,363	52,192	55,668	59,892	62,883
38	43,462	49,513	53,384	56,896	61,162	64,181
39	44,539	50,660	54,572	58,120	62,428	65,476
40	45,616	51,805	55,758	59,342	63,691	66,766
41	46,692	52,949	56,942	60,561	64,950	68,053
42	47,766	54,090	58,124	61,777	66,206	69,336
43	48,840	55,230	59,304	62,990	67,459	70,616
44	49,913	56,369	60,481	64,201	68,710	71,893
45	50,985	57,505	61,656	65,410	69,957	73,166

TABLA 6. Números aleatorios*

07018	31172	12572	23968	55216	85366	56223	09300	94564	18172
52444	65625	97918	46794	62370	59344	20149	17596	51669	47429
72161	57299	87521	44351	99981	55008	93371	60620	66662	27036
17918	75071	91057	46829	47992	26797	64423	42379	91676	75127
13623	76165	43195	50205	75736	77473	07268	31330	07337	55901
27426	97534	89707	97453	90836	78967	00704	85734	21776	85764
96039	21338	88169	69530	53300	29895	71507	28517	77761	17244
68282	98888	25545	69406	29470	46476	54562	79373	72993	98998
54262	21477	33097	48125	92982	98382	11265	25366	06636	25349
66290	27544	72780	91384	47296	54892	59168	83951	91075	04724
53348	39044	04072	62210	01209	43999	54952	68699	31912	09317
34482	42758	40128	48436	30254	50029	19016	56837	05206	33851
99268	98715	07545	27317	52459	75366	43688	27460	65145	65429
95342	97178	10401	31615	95784	77026	33087	65961	10056	72834
38556	60373	77935	64608	28949	94764	45312	71171	15400	72182
39159	04795	51163	84475	60722	35268	05044	56420	39214	89822
41786	18169	96649	92406	42773	23672	37333	85734	99886	81200
95627	30768	30607	89023	60730	31519	53462	90489	81693	17849
98738	15548	42263	79489	85118	97073	01574	57310	59375	54417
75214	61575	27805	21930	94726	39454	19616	72239	93791	22610
73904	89123	19271	15792	72675	62175	48746	56084	54029	22296
33329	08896	94662	05781	59187	53284	28024	45421	37956	14252
66364	94799	62211	37539	80172	43269	91133	05562	82385	91760
68349	16984	86532	96186	53891	48268	82821	19526	63257	14288
19193	99621	66899	12351	72438	99839	24228	32079	53517	18558
49017	23489	19172	80439	76263	98918	59330	20121	89779	58862
76941	77008	27646	82072	28048	41589	70883	72035	81800	50296
55430	25875	26446	25738	32962	24266	26814	01194	48587	93319
33023	26895	65304	34978	43053	28951	22676	05303	39725	60054
87337	74487	83196	61939	05045	20405	69324	80823	20905	68727
81773	36773	21247	54735	68996	16937	18134	51873	10973	77090
74279	85087	94186	67793	18178	82224	17069	87880	54945	73489
34968	76028	54285	90845	35464	68076	15868	70063	26794	81386
99696	78454	21700	12301	88832	96796	59341	16136	01803	17537
55282	61051	97260	89829	69121	86547	62195	72492	33536	60137
31337	83886	72886	42598	05464	88071	92209	50728	67442	47529
94128	97990	58609	20002	76530	81981	30999	50147	93941	80754
06511	48241	49521	64568	69459	95079	42588	98590	12829	64366
69981	03469	56128	80405	97485	88251	76708	09558	86759	15065
23701	56612	86307	02364	88677	17192	23082	00728	78660	74196
09237	24607	12817	98120	30937	70666	76059	44446	94188	14060
11007	45461	24725	02877	74667	18427	45658	40044	59484	59966
60622	78444	39582	91930	97948	13221	99234	99629	22430	49247
79973	43668	19599	30021	68572	31816	63033	14597	28953	21162
71080	71367	23485	82364	30321	42982	74427	25625	74309	15855
09923	26729	74573	16583	37689	06703	21846	78329	98578	25447
63094	72826	65558	22616	33472	67515	75585	90005	19747	08865
19806	42212	41268	84923	21002	30588	40676	94961	31154	83133
17295	74244	43088	27056	86338	47331	09737	83735	84058	12382
59338	27190	99302	84020	15425	14748	42380	99376	30496	84523

* The Rand Corporation, *A Million Random Digits with 100,000 Normal Deviates*. New York: The Free Press, 1955.

TABLA 6 (*Continuación*)

96124	73355	01925	17210	81719	74603	30305	29383	69753	61156
31283	54371	20985	00299	71681	22496	71241	35347	37285	02028
49988	48558	20397	60384	24574	14852	26414	10767	60334	36911
82790	45529	48792	31384	55649	08779	94194	62843	11182	49766
51473	13821	75776	24401	00445	61570	80687	39454	07628	94806
07785	02854	91971	63537	84671	03517	28914	48762	76952	96837
16624	68335	46052	07442	41667	62897	40326	75187	36639	21396
28718	92405	07123	22008	83082	28526	49117	96627	38470	78905
33373	90330	67545	74667	20398	58239	22772	34500	34392	92989
36535	48606	11139	82646	18600	53898	70267	74970	35100	01291
47408	62155	47467	14813	56684	56681	31779	30441	19883	17044
56129	36513	41292	82142	13717	49966	35367	43255	06993	17418
35459	10460	33925	75946	26708	63004	89286	24880	38838	76022
61955	55992	36520	08005	48783	08773	45424	44359	25248	75881
85374	69791	18857	92948	90933	90290	97232	61348	22204	43440
15556	39555	09325	16717	74724	79343	26313	39585	56285	22525
75454	90681	73339	08810	89616	99234	36613	43440	60269	90899
27582	90856	04254	23715	00086	12164	16943	62099	32132	93031
89658	47708	01691	22284	50446	05451	68947	34932	81628	22716
57194	77203	26072	92538	85097	58178	46391	58980	12207	94901
64219	53416	03811	11439	80876	38314	77078	85171	06316	29523
53166	78592	80640	58248	68818	78915	57288	85310	43287	89223
58112	88451	22892	29765	20908	49267	18968	39165	03332	94932
14548	36314	05831	01921	97159	55540	00867	84293	54653	81281
21251	15618	40764	99303	38995	97879	98178	03701	70069	80463
30953	63369	05445	20240	35362	82072	29280	72468	94845	97004
12764	79194	36992	74905	85867	18672	28716	17995	63510	67901
72393	71563	42596	87316	80039	75647	66121	17083	07327	39209
11031	40757	10904	22385	39813	63111	33237	95008	09057	50820
91948	69586	45045	67557	86629	67943	23405	86552	17393	24221
18537	07384	13059	47389	97265	11379	24426	09528	36035	02501
66885	11985	38553	97029	88433	78988	88864	03876	48791	72613
96177	71237	08744	38483	16602	94343	18593	84747	57469	08334
37321	96867	64979	89159	33269	06367	09234	77201	92195	89547
77905	69703	77702	90176	04883	84487	88688	09360	42803	88379
53814	14560	43698	86631	87561	90731	59632	52672	24519	10966
16963	37320	40740	79330	04318	56078	23196	49668	80418	73842
87558	58885	65475	25295	59946	47877	81764	85986	61687	04373
84269	55068	10532	43324	39407	65004	35091	20714	20880	19385
94907	08019	05159	64613	26962	30688	51677	05111	51215	53285
45735	14319	78439	18033	72250	87674	67405	94163	16622	54994
11755	40589	83489	95820	70913	87328	04636	42466	68427	79135
51242	05075	80028	35144	70599	92270	62912	08859	87405	08266
00281	25893	94848	74342	45848	10404	28635	92136	42852	40812
12233	65661	10625	93343	21834	95563	15070	99901	09382	01498
88817	57827	02940	66788	76246	85094	44885	72542	31695	83843
75548	53699	90888	94921	04949	80725	72120	80838	38409	72270
42860	40656	33282	45677	05003	46597	67666	70858	41314	71100
71208	72822	17662	50330	32576	95030	87874	25965	05261	95727
44319	22313	89649	47415	21065	42846	78055	64776	64993	48051

TABLA 7. Desviaciones normales aleatorias

	00	01	02	03	04	05	06	07	08	09
00	,31	-,51	-1,45	-,35	,18	,09	,00	,11	-1,91	-1,07
01	,90	-,36	,33	-,28	,30	-2,62	-1,43	-1,79	-,99	-,35
02	,22	,58	,87	-,02	,04	,12	-,17	,78	-1,31	,95
03	-1,00	,53	-1,90	-,77	,67	,56	-,94	,16	2,22	-,08
04	-,12	-,43	,69	,75	-,32	,71	-1,13	-,79	-,26	-,86
05	,01	,37	-,36	,68	,44	,43	1,18	-,68	-,13	-,41
06	,16	-,83	-1,88	,89	-,39	,93	-,76	-,12	,66	2,06
07	1,31	-,82	-,36	,36	,24	-,95	,41	-,77	,78	-,27
08	-,38	-,26	-1,73	,06	-,14	1,59	,96	-1,39	,51	-,50
09	,38	,42	-1,39	-,22	-,28	-,03	2,48	1,11	1,10	,40
10	1,07	2,26	-1,68	-,04	,19	1,38	-1,53	-1,41	,09	1,91
11	-1,65	-1,29	-1,03	,06	2,18	-,55	-,34	-1,07	,80	1,77
12	1,02	-,67	-1,11	,08	-1,92	-,97	-,70	-,40	-,72	-,47
13	,06	1,43	-,46	-,62	-,11	,36	,64	-,27	,72	,68
14	,47	-1,84	,69	-1,07	,83	-,25	-,91	-1,94	,96	,75
15	,10	1,00	-,54	,61	-1,04	-,33	,94	,56	,62	,07
16	-,71	,04	,63	-,26	-1,35	-1,20	1,52	,63	-1,29	1,16
17	-,94	-,94	,56	-,09	,63	-,36	,20	-,60	-,29	,94
18	,29	,62	-1,09	1,84	-,11	,19	-,45	,23	-,63	-,06
19	,57	,54	-,21	,09	-,57	-,10	-1,25	-,26	,88	-,26
20	,24	,19	-,67	3,04	1,26	-1,21	,52	-,05	,76	-,09
21	-1,47	1,20	,70	-1,80	-1,07	,29	1,18	,34	-,74	1,75
22	-,01	,49	1,16	,17	-,48	,81	1,40	,17	,57	,64
23	-,63	-,26	,55	-,21	-,07	-,37	,47	-1,69	,05	-,96
24	,85	-,65	-,94	,12	-1,67	,28	-,42	,14	-1,15	-,41
25	1,07	-,36	1,10	,83	,37	-,20	-,75	-,50	,18	1,31
26	1,18	2,09	-,61	,44	,40	,42	-,61	-2,55	-,09	-1,33
27	,47	,88	,71	,31	,41	-1,96	,34	-,17	1,73	-,33
28	,26	,90	,11	,28	,76	-,12	-1,01	1,29	-,71	2,15
29	,39	-,88	-,15	-,38	,55	-,41	-,02	-,74	-,48	,46
30	-1,01	-,89	-1,23	,07	-,07	-,08	-,08	1,95	,34	-,29
31	1,36	,18	,85	,55	,00	-,43	,27	-,39	,25	,69
32	1,02	-2,49	1,79	,04	-,03	,85	-,29	-,77	,28	-,33
33	-,53	-1,13	,75	-,39	,43	,10	-2,17	,37	-1,85	,96
34	,76	1,21	-,68	,26	,93	,99	,1,12	-1,72	-,04	-,73
35	,07	-,23	-,88	-,23	,68	,24	1,38	-2,10	-,79	-,27
36	,27	,61	,43	-,38	,68	-,72	,90	-,14	-1,61	-,88
37	,93	,72	-,45	2,80	-,12	,74	-1,47	,39	-,61	-2,77
38	1,03	-,43	,95	-1,49	-,63	,22	,79	-2,80	-,41	,61
39	-,32	1,41	-,23	-,36	,60	-,59	,36	,63	,73	,81
40	1,41	,64	,06	,25	-1,75	,39	1,84	1,23	-1,27	-,75
41	,25	-,70	,33	,12	,04	1,03	-,64	,08	1,63	,34
42	-1,15	,57	,34	-,32	2,31	,74	,85	-1,25	-,17	,14
43	,72	,01	,50	-1,42	,26	-,74	-,55	1,86	-,17	-,10
44	-,92	,15	-,66	,83	,50	,24	-,40	1,90	,35	,69
45	-,42	,62	,24	,55	-,06	,14	-1,09	-1,53	,30	-1,56
46	-,54	1,21	-,53	,29	1,04	-,32	-1,20	,01	,05	,20
47	-,13	-,70	,07	,69	,88	1,18	,61	-,46	-1,54	,50
48	-,29	,36	1,44	-,44	,53	-,14	,66	,00	,33	-,36
49	1,90	-1,21	-1,87	-,27	-1,86	-,49	,25	,25	,14	1,73

TABLA 7 (*Continuación*)

	10	11	12	13	14	15	16	17	18	19
00	-,73	,25	-2,08	,17	-1,04	-,23	,74	,23	,70	-,79
01	-,87	-,74	1,44	-,79	-,76	-,42	1,93	,88	,80	-,53
02	1,18	,05	,10	-1,15	,05	1,06	,82	,90	-1,38	,51
03	-2,09	1,13	-,50	,37	-1,18	-,16	-1,85	-,90	1,32	-,83
04	-,32	1,06	1,14	-,23	,49	1,10	-,27	-,64	,47	-,05
05	,90	-,86	,63	-1,62	-,52	-1,55	,78	-,54	-,29	,19
06	-,16	-,22	-,17	-,81	,49	,96	,53	1,73	,14	1,21
07	,15	-1,12	,80	-,30	-,77	-,91	,00	,94	-1,16	,44
08	-1,87	,72	-1,17	-,36	-1,42	-,46	-,58	,03	2,08	1,11
09	,87	,95	,05	,46	-,01	,85	1,19	-1,61	-,10	,87
10	,52	,12	-1,04	-,56	-,91	-,13	,17	1,17	-1,24	,84
11	-1,39	-1,18	1,67	2,88	-2,06	,10	,05	-,55	,74	,33
12	-,94	-,46	-,85	-,29	,54	,71	,90	-,42	-1,30	,50
13	-,51	,04	-,44	-1,87	-1,06	1,18	-,39	,22	-,55	-,54
14	-1,50	-,21	-,89	,43	-1,81	-,07	-,66	-,02	1,77	-1,54
15	-,48	1,54	1,88	,66	-,62	,28	-,34	2,42	-1,65	2,06
16	,89	-,23	,57	,23	1,81	1,02	,33	1,23	1,31	,06
17	,38	1,52	-1,32	2,13	-,14	,28	-,46	,25	,65	1,18
18	-,53	,37	,19	-2,41	,16	,36	-,15	,14	-,15	-,73
19	,15	,62	-1,29	1,84	,80	-,65	1,72	-1,77	,07	,46
20	-,81	-,22	1,16	1,09	-,73	-,15	,87	-,88	,92	-,04
21	-1,61	2,51	-2,17	,49	-1,24	1,16	,97	,15	,37	,18
22	,26	-,48	-,43	-2,08	,75	1,59	1,78	-,55	,85	-1,87
23	-,32	,75	-,35	2,10	-,70	1,29	,94	,20	-1,16	,89
24	-1,00	1,37	,68	,00	1,87	-,14	,77	-,12	,89	-,73
25	,66	,04	-1,73	,25	,26	1,46	-,77	-1,67	,18	-,92
26	-,20	-1,53	,59	-,15	-,15	-,11	,68	-,14	-42	-1,51
27	1,01	-,44	-,20	-2,05	-,27	-,50	-,27	-,45	,83	,49
28	-1,81	,45	,27	,67	-,74	-,17	-1,11	,13	-1,18	-1,41
29	-,40	1,34	1,50	,57	-1,78	,08	,95	,69	,38	,71
30	-,01	,15	-1,83	1,18	,11	,62	1,86	,42	,03	-,14
31	-,23	-,19	-1,08	,44	-,41	-1,32	,14	,65	-,76	,76
32	-1,27	,13	-,17	-,74	-,44	1,67	-,07	-,99	,51	,76
33	-1,72	1,70	-,61	,18	,48	-,26	-,12	-2,83	2,35	1,25
34	,78	1,55	-,19	,43	-1,53	-,76	,83	-,46	,48	-,43
35	1,86	1,12	-2,09	1,82	-,71	-1,76	-,20	-,38	,82	-1,08
36	-,50	-,93	-,68	-1,62	-,88	,05	-,27	,23	-,58	-,24
37	1,02	-,81	,62	1,46	-,31	-,37	,08	,59	-,27	,37
38	-1,57	,10	,11	-1,48	1,02	2,35	,27	-1,22	-1,26	2,22
39	2,27	-,61	,61	-,28	-,39	-,45	,89	1,43	1,03	,01
40	-2,17	-,69	1,33	-,26	,15	-,10	-,78	,64	-,70	,14
41	,05	-1,71	,21	,55	-,60	-,74	-,90	2,52	-,07	-1,11
42	-,38	1,75	,93	-1,36	-,60	-1,76	-1,10	,42	1,44	-,58
43	,40	-1,50	,24	-,66	,83	,37	-,35	,16	,96	,79
44	,39	,66	,19	-2,08	,32	-,42	-,53	,92	,69	-,03
45	-,12	1,18	-,08	,30	-,21	,45	-1,84	,26	,90	,85
46	1,20	-,91	-1,08	-,99	1,76	-,80	,51	,25	-,11	-,58
47	-1,04	1,28	2,50	1,56	-,95	-1,02	,45	-1,90	-,02	-,73
48	-,32	,56	-1,03	,11	-,72	,53	-,27	-,17	1,40	1,61
49	1,08	,56	,34	-,28	-,37	,46	,03	-1,13	,34	-1,08

Respuestas a problemas seleccionados

CAPITULO 1

- 1.1.** (a) $\{5\}$ (b) $\{1, 3, 4, 5, 6, 7, 8, 9, 10\}$ (c) $\{2, 3, 4, 5\}$
 (d) $\{1, 5, 6, 7, 8, 9, 10\}$ (e) $\{1, 2, 5, 6, 7, 8, 9, 10\}$
- 1.2.** (a) $\{x \mid 0 \leq x < \frac{1}{2}\} \cup \{x \mid \frac{3}{2} \leq x \leq 2\}$
 (b) $\{x \mid 0 \leq x < \frac{1}{2}\} \cup \{x \mid \frac{1}{2} < x \leq 1\} \cup \{x \mid \frac{3}{2} \leq x \leq 2\}$
 (c) $\{x \mid 0 \leq x \leq \frac{1}{2}\} \cup \{x \mid 1 < x \leq 2\}$
 (d) $\{x \mid \frac{1}{4} \leq x \leq \frac{1}{2}\} \cup \{x \mid 1 < x < \frac{3}{2}\}$
- 1.3.** (a) Verdadero (b) Verdadero (c) Falso (d) Falso (e) Verdadero
- 1.4.** (a) $A = \{(0, 0), (1, 0), (2, 0), (0, 1), (1, 1), (2, 1), (0, 2), (1, 2)\}$
 (b) $B = \{(0, 0), (1, 0), (2, 0), (3, 0), (4, 0), (5, 0), (6, 0), (1, 1), (2, 1), (3, 1), (4, 1), (5, 1), (6, 1), (2, 2), (3, 2), (4, 2), (5, 2), (6, 2), (2, 3), (3, 3), (4, 3), (5, 3), (6, 3), (2, 4), (3, 4), (4, 4), (5, 4), (6, 4), (3, 5), (4, 5), (5, 5), (6, 5), (3, 6), (4, 6), (5, 6), (6, 6)\}$
- 1.6.** $\{DD, NDD, DNDD, DNDN, DNND, DNNN, NDND, NDNN, NNDD,$
 $NNDN, NNND, NNNN\}$
- 1.10.** (a) $\{(x, y) \mid 0 \leq x < y \leq 24\}$
- 1.11.** (a) $\frac{A \cup B \cup C}{A \cup B \cap C}$ (b) $[A \cap \bar{B} \cap \bar{C}] \cup [\bar{A} \cap B \cap \bar{C}] \cup [\bar{A} \cap \bar{B} \cap C]$
- 1.15.** $\frac{1}{13}, \frac{4}{13}, \frac{8}{13}$
- 1.16.** (a) $1 - z$ (b) $y - z$
- 1.17.** $\frac{5}{8}$

CAPITULO 2

- 2.1.** (a) $\frac{3}{48}$ (b) $\frac{1}{6}$
- 2.2.** (a) $\frac{1}{12}$ (b) $\frac{1}{20}$
- 2.3.** (a) $\frac{2}{3}$ (b) $\frac{5}{8}$
- 2.4.** (a) $\frac{\binom{400}{90} \binom{1100}{110}}{\binom{1500}{200}}$ (b) $1 - \left[\frac{\binom{400}{0} \binom{1100}{200} + \binom{400}{1} \binom{1100}{199}}{\binom{1500}{200}} \right]$
- 2.5.** $\frac{4}{45}$
- 2.6.** (a) $\frac{5}{8}$ (b) $\frac{7}{8}$ (c) $\frac{3}{4}$
- 2.7.** (a) $\frac{3}{8}$ (b) $\frac{1}{120}$ (c) $\frac{7}{8}$ (d) $\frac{5}{8}$ (e) $\frac{1}{2}$ (f) $\frac{91}{120}$ (g) $\frac{1}{8}$
- 2.8.** 120
- 2.9.** 720
- 2.10.** 455
- 2.11.** (a) 120 (b) 2970
- 2.12.** (a) 4^8 (b) $4 \cdot 3^7$ (c) 70 (d) 336
- 2.13.** $(N - 1)!/(N - n)!N^{n-1}$
- 2.14.** (a) 360 (b) 1296
- 2.15.** $a + b$

2.16. (a) $2/n$ (b) $2(n-1)/n^2$

2.20. 120

2.22. $\frac{10!}{10^r(10-r)!}$

2.18. 0,24

2.21.
$$\frac{\binom{r}{r-1} \binom{n-r}{k-r}}{\binom{n}{k-1}} \cdot \frac{1}{n-k+1}$$

CAPITULO 3

3.1. $\left(\frac{x}{x+y}\right)\left(\frac{z+1}{z+v+1}\right) + \left(\frac{y}{x+y}\right)\left(\frac{z}{z+v+1}\right)$

3.2. (a) $\frac{1}{6}$ (b) $\frac{1}{3}$ (c) $\frac{1}{2}$

3.4. (a) $\frac{2}{105}$ (b) $\frac{2}{5}$

3.7. $\frac{2}{3}$

3.12. (a) $\frac{1}{4}$ (b) $\frac{3}{4}$

3.15. $\frac{5}{16}$

3.20. (a) $2p^2 + 2p^3 - 5p^4 + 2p^5$

3.23. $\frac{3}{16}$

3.34. $\beta_n = \frac{1}{2} + (2p-1)^n(\beta - \frac{1}{2})$

3.37. $p_n = \frac{\alpha}{\alpha+\beta} + \frac{\beta}{(\alpha+\beta)(\alpha+\beta+1)^{n-1}}$

3.39. $(n-1)p^2/(1-2p+np^2)$

3.3. $\frac{5}{9}$

3.6. (a) $\frac{3}{5}\frac{3}{5}$ (b) $\frac{1}{2}\frac{4}{5}$ (c) $\frac{4}{5}\frac{8}{5}$

3.9. 0,362, 0,406, 0,232

3.13. $1 - (1-p)^n$

3.17. (a) 0,995 (b) 0,145

(b) $p + 3p^2 - 4p^3 - p^4 + 3p^5 - p^6$

3.25. (a) 0,50 (b) 0,05

3.35. (a) 0,65 (b) 0,22 (c) $8/35$

CAPITULO 4

4.1. $P(X=0) = \frac{1}{64}, P(X=1) = \frac{9}{64}, P(X=2) = \frac{27}{64}, P(X=3) = \frac{27}{64}$

4.2. (a) $\binom{4}{x} \left(\frac{1}{5}\right)^x \left(\frac{4}{5}\right)^{4-x}$ (b) $\binom{5}{x} \binom{20}{4-x} / \binom{25}{4}$

4.3. (a) $\frac{1}{3}$ (b) $\frac{1}{16}$ (c) $\frac{1}{4}$

4.9. $\frac{27}{64}$

4.10. $a = e^{-b}$

4.11. $P(X > b | X < b/2) = -7b^3/(b^3 + 8)$

4.13. (a) $a = (2/b) - b$

4.14. (a) $F(t) = 5t^4 - 4t^5$

4.15. (a) $a = \frac{1}{2}$ (c) $\frac{3}{8}$

4.16. (b) $F(x) = 3x^2 - 2x^3$

4.17. (a) $f(x) = \frac{1}{5}, 0 < x < 5$

(b) $f(x) = (1/\pi)(x - x^2)^{-1/2}, 0 < x < 1$

(c) $f(x) = 3e^{3x}, x < 0$

4.20. (a) $\alpha = 3$ (c) $\alpha = \frac{5}{4}$

4.23. (b) $P(X=k) = \binom{10}{k} (0,09)^k (0,91)^{10-k}$

4.25. (a) $\frac{1}{4}$

4.28. $\frac{3}{5}$

4.29. (a) $k = \beta$ (b) $r = 1$

4.30. $-\frac{1}{3} \leq x \leq \frac{1}{4}$

CAPITULO 5

5.1. $g(y) = \frac{1}{2}(4-y)^{-1/2}, 3 < y < 4$

5.2. (a) $g(y) = \frac{1}{6}, 7 < y < 13$ (b) $h(z) = 1/2z, e < z < e^3$

- 5.3.** (a) $g(y) = \frac{1}{3}y^{-2/3}e^{-y^{1/3}}, y > 0$
5.6. (a) $g(y) = (1/\pi)(1 - y^2)^{-1/2}, -1 < y < 1$
(b) $h(z) = (2/\pi)(1 - z^2)^{-1/2}, 0 < z < 1$ (c) $f(w) = 1, 0 < w < 1$
5.7. (a) $g(v) = (3/2\pi)[(3v/4\pi)^{-1/3} - 1], 0 < v < 4\pi/3$
(b) $h(s) = (3/4\pi)[1 - (s/4\pi)^{1/2}], 0 < s < 4\pi$
5.8. $g(p) = \frac{1}{8}(2/p)^{1/2}, 162 < p < 242$
5.10. (a) $g(0) = 1; g(y) = 0, y \neq 0$
(b) $g(0) = a/k; g(y) = (x_0 - a)/ky_0, 0 < y < y_0[(k - a)/(x_0 - a)]$
(c) $g(0) = a/k; g(y) = (x_0 - a)/ky_0, 0 < y < y_0; g(y_0) = 1 - x_0/k$
5.13. 0,71

CAPITULO 6

- 6.2.** (a) $k = \frac{1}{8}$ (b) $h(x) = x^3/4, 0 < x < 2$
(c) $g(y) = \begin{cases} \frac{1}{3} - y/4 + y^3/48, & 0 \leq y \leq 2 \\ \frac{1}{3} - y/4 + (5/48)y^3, & -2 \leq y \leq 0 \end{cases}$
6.3. (a) $\frac{5}{6}$ (b) $\frac{7}{24}$ (c) $\frac{5}{32}$ **6.5.** $k = 1/(1 - e^{-1})^2$
6.6. (a) $k = \frac{1}{2}$ (b) $h(x) = 1 - |x|, -1 < x < 1$
(c) $g(y) = 1 - |y|, -1 < y < 1$
6.8. $h(z) = 1/2z^2, z \geq 1$
 $= 1/2, 0 < z < 1$
6.11. $g(h) = (1600 - 9h^2)/80h^2, 8 < h < \frac{40}{3}$
 $= \frac{1}{5}, \frac{20}{3} < h \leq 8$
 $= (5h^2 - 80)/16h^2, 4 \leq h \leq \frac{20}{3}$
6.12. $h(i) = e^{-(2/i)^{1/2}}[-(2/i) - 2(2/i)^{1/2} - 2]$
 $+ e^{-(1/i)^{1/2}}[(1/i) + 2(1/i)^{1/2} + 2], i > 0$
6.13. $h(w) = 6 + 6w - 12w^{1/2}, 0 < w < 1$
6.14. (a) $g(x) = e^{-x}, x > 0$ (b) $h(y) = ye^{-y}, y > 0$

CAPITULO 7

- 7.3.** 3,4 **7.4.** $\frac{1}{3}(2C_3 + C_2 - 3C_1)$
7.6. \$0,03 **7.8.** $7T_5^2$
7.9. \$50 **7.10.** (a) $C = \frac{1}{6}$ (b) $E(D) = \frac{19}{9}$
7.12. (b) $E(Z) = \frac{8}{3}$ **7.13.** (b) $E(W) = \frac{4}{3}$
7.14. 154
7.15. $E(Y) = 10, V(Y) = 3, E(Z) = (e/2)(e^2 - 1), V(Z) = (e^2/2)(e^2 - 1)$
7.18. $E(Y) = 0, V(Y) = \frac{1}{2}, E(Z) = 2/\pi, V(Z) = (\pi^2 - 8)/2\pi^2, E(W) = \frac{1}{2}, V(W) = \frac{1}{12}$
7.20. Caso 2: $E(Y) = (y_0/2k)(x_0 - a), V(y) = \frac{(x_0 - a)y_0^2}{k} \left[\frac{1}{3} - \frac{x_0 - a}{4k} \right]$
7.24. $a = \frac{1}{3}, b = 2$ **7.25.** (a) $E(V) = \frac{3}{2}$ (b) $E(P) = \frac{4}{3}$
7.26. $V(X) = \frac{4}{3}a^2$ **7.27.** $E(S) = \frac{65}{6}$

7.30. (a) $g(x) = x/2, 0 < x < 2; h(y) = 1/2 - y/8, 0 < y < 4$
 (b) $V(X) = \frac{7}{9}$

7.31. $E(Z) \simeq \mu_x/\mu_y + 2(\mu_x/\mu_y^3)\sigma_y^2; V(Z) \simeq (1/\mu_y^2)\sigma_x^2 + (\mu_x^2/\mu_y^4)\sigma_y^2$

7.32. $E(Z) \simeq \frac{29}{27} \quad V(Z) \simeq \frac{3}{27}$

7.35. $\frac{1}{4}; 0$

7.46. $\frac{48}{65}$

7.48. $P(X_i = 1) = p(r/n) + q[(n - r)/n]$

CAPITULO 8

8.1. 0,219

8.3. (a) 0,145 (b) 4 (c) 2 (d) 1 ó 2 (e) 1,785 (f) 0,215

8.4. 0,3758 (binomial), 0,4060 (Poisson)

8.5. 0,067

8.6. $P(X = 0) = 0,264$

8.7. $E(P) = \$32,64$

8.9. (b) 0,027

8.10. 0,215

8.12. (a) $(0,735)^7$ (b) $1 - (0,265)^7$

8.16. (a) $(1 - p_1)(1 - p_2)(1 - p_3)(1 - p_4)$ (c) 0,0964

8.17. (a) 0,064

8.20. $(2 - \ln 3)/3$

8.24. \$19,125

CAPITULO 9

9.1. (a) 0,2266 (b) 0,2902

9.2. 0,3085

9.3. 0,21

9.5. (a) D_2 (b) D_2

9.6. $E(Y) = (2/\pi)^{1/2}, V(Y) = (\pi - 2)/\pi$

9.10. $C = 0,433\sigma + \mu$

9.11. 0,5090

9.12. $E(L) = \$0,528$

9.13. (a) $\frac{1}{4}; 0,0456$ (b) $\frac{1}{4} 0,069$ (c) $\frac{1}{4} 0,049$

9.15. \$23,40

9.17. (a) \$0,077

9.24. 0,10, 0,80

9.25. $E(Y) \simeq \ln \mu - (1/2\mu^2)\sigma^2; V(Y) \simeq (1/\mu^2)\sigma^2$

9.28. (b) $E(X) = np[(1 - p^{n-1})/(1 - p^n)]$

9.32. 0,15

CAPITULO 10

10.1. (a) $M_X(t) = (2/t^2)[e^t(t - 1) + 1]$ (b) $E(X) = \frac{2}{3}, V(X) = \frac{1}{18}$

10.3. (a) $M_X(t) = \lambda e^{\lambda t}/(\lambda - t)$ (b) $E(X) = (a\lambda + 1)/\lambda, V(X) = 1/\lambda^2$

10.4. (a) $M_X(t) = \frac{1}{6}(e^t + e^{2t} + e^{3t} + e^{4t} + e^{5t} + e^{6t})$

10.6. (a) $(1 - t^2)^{-1}$

10.8. (b) $E(X) = 3,2$

10.9. (a) 0,8686

10.12. 0,30

10.13. 0,75

10.14. 0,579

10.18. $\frac{1}{3}$

10.19. $(e^{3t} - 1)/3te^t$

CAPITULO 11

- 11.1.** 83,55 horas, 81,77 horas, 78,35 horas
11.2. 48,6 horas
11.3. $f(t) = C_0 \exp[-C_0 t], 0 \leq t \leq t_0$
 $= C_1 \exp[-C_0 t_0 + C_1(t_0 - t)], t > t_0$
11.4. (a) $f(t) = Ce^{-C(t-A)}, t \geq A$ (b) $R(t) = [A/(A+t)]^{r+1}$
11.7. (a) Aproximadamente (0,5)⁶ (b) 0,007
11.10. $R(t) = 2e^{-0.06t} - e^{-0.09t}$ (c) 0,014
11.12. (a) $m = \ln(\sqrt{2})$ (b) $m = 0,01$ (c) $+\infty$
11.13. $R(t) = \exp(-C_0 t), 0 < t < t_0$
 $= \exp[t(C_1 t_0 - C_0) - (C_1/2)(t^2 + t_0^2)], t > t_0$
11.14. (a) $R(t) = e^{-\beta_1 t} + e^{-\beta_2 t} + e^{-\beta_3 t} - e^{-(\beta_1 + \beta_2)t} - e^{-(\beta_1 + \beta_3)t}$
 $- e^{-(\beta_2 + \beta_3)t} + e^{-(\beta_1 + \beta_2 + \beta_3)t}$
11.15. (a) $R_S = 1 - (1 - R)^k$ (b) $R_S = [1 - (1 - R)^k]^n$
11.18. (a) 0,999926 (b) 0,99 (c) 0,68
11.19. (a) $R_S = [1 - (1 - R_A)(1 - R_B)(1 - R_C)][1 - (1 - R_{A'})(1 - R_{B'})]$
(b) $R_S = 1 - R_C(1 - R_{A'})(1 - R_{B'})$
 $- (1 - R_C)(1 - R_A R_{A'})(1 - R_B R_{B'})$
11.22. $M_X(t) = -2\lambda[1/(t-\lambda) - 1/(t-2\lambda)]$

CAPITULO 12

- 12.1.** (a) $n = 392$ (b) $n = 5000$ (c) 0,083
12.3. (a) 0,9662 (b) $n = 24$ (d) 0,1814
12.5. (a) 0,1802 (b) $n = 374$ (c) 0,1112 (d) 0,1915
12.7. $g(r) = (15.000)^{-1}(r^3 - 60r^2 + 600r), 0 \leq r \leq 10$
 $= (15.000)^{-1}(-r^3 + 60r^2 - 1200r + 8000), 10 \leq r \leq 20$
12.9. (a) $g(s) = 50[1 - e^{-0.2(s+0.01)}], \text{ si } -0,01 \leq s \leq 0,01$
 $= 50[e^{-0.2(s-0.01)} - e^{-0.2(s+0.01)}], \text{ si } s > 0,01$
12.12. $f(s) = \frac{1}{2} \left[\Phi\left(\frac{s-99}{2}\right) - \Phi\left(\frac{s-101}{2}\right) \right]$
12.13. (a) $\frac{p_1 p_2 (1-p_1)^{k-1}}{p_2 - p_1} \left[1 - \left(\frac{1-p_2}{1-p_1} \right)^{k-1} \right]$ (b) 0,055

CAPITULO 13

- 13.3.** $P(M = m) = [1 - (1 - p)^m]^n - [1 - (1 - p)^{m-1}]^n$
13.4. (a) 0,13 (b) 0,58 (c) 0,018 (d) 0,77
13.6. 0,89 (e) $(1 - 2t)^{-1}$

CAPITULO 14

14.1. $\frac{V(L_2)}{V(L_1) + V(L_2)}$

14.3. 16,1

14.6. (a) $1/(\bar{T} - t_0)$

14.9. (a) $k/\sum_{i=1}^k n_i$

14.14. -0,52

14.16. $n/\sum_{i=1}^n X_i^2$

14.29. Distribución F con (1, 1) grados de libertad

14.31. (-4/5)

14.2. $C = 1/2(n - 1)$

14.4. $-1 - n/\sum_{i=1}^n \ln X_i$

14.7. (a) $\frac{1}{T_0 - t_0} \ln \left(\frac{n}{n - k} \right)$

14.13. 0,034

14.15. r/\bar{X}

14.32. 8,6

CAPITULO 15

15.1. (a) $1 - \Phi\left(\frac{C - \mu}{\sigma}\sqrt{n}\right)$ (b) $C = -2,33 \frac{\sigma}{\sqrt{n}} + \mu_0$

15.2. (a) $\Phi\left[\frac{C + \mu_0 - \mu}{\sigma}\sqrt{n}\right] - \Phi\left[\frac{-C + \mu_0 - \mu}{\sigma}\sqrt{n}\right]$

(b) $C = 2,575\sigma n^{-1/2}$

15.3. (a) $\sum_{k=0}^{[nC]} \frac{e^{-n\lambda}(n\lambda)^k}{k!}$, donde $[nC] =$ mayor entero $\leq nC$. (c) 0,3233

15.7. (a) $(1 - p)^4[5p + (1 + 25p^5)(1 - p) + 55p^4(1 - p)^2$

15.8. 15 $+ 60p^3(1 - p)^3 + 10p^2(1 - p)^3]$

15.9. (a) $(k - 1)$ (b) Igual

15.12. (a) $\sin \alpha / \sin\left(\frac{\pi}{2}\alpha\right)$

Indice de materias

Aproximación de DeMoivre-Laplace para la distribución binomial, 257

Arbol, diagrama de, 41

Bondad de ajuste de pruebas, 336
prueba de ajuste de distribuciones asintóticas, 338
prueba de una distribución específica, 340, 341

Coeficiente binomial, 27

Coeficiente de confianza, 312

Coeficiente de correlación, 148
ejemplo de, 310
evaluación de, 148
interpretación de, 149, 150, 151
propiedades de, 148

Comparación entre varias distribuciones, 205

Confiabilidad, 233
de los sistemas, 243
y la función de distribución acumulativa, 234
y la longitud de vida esperada, 237
y la tasa de falla, 234

Conjunto, 3
complemento de, 4
identidad de, 5
intersección de, 4
número de elementos, 6
subconjunto, 4
unión de, 4
universal, 4
vacío, 4

Convergencia en probabilidad, 253

Covarianza, 148

Desigualdad de Boole, 20

Desigualdad de Chebyshev, 146, 147

Desviación estándar de una variable aleatoria, 138

Desviaciones normales, 287

Distribución binomial, 64

aproximación normal a la, 255
distribución de Pascal, y, 179
propiedades de, 79
valor esperado de, 122, 136
varianza de, 141

Distribución binomial negativa, 178

Distribución de Cauchy, 213

Distribución de Pascal, 178

esperanza de, 179
varianza de, 179
y la distribución binomial, 179

Distribución de Poisson, 164
propiedad reproductiva de, 226
valor esperado de, 164
varianza de, 165
y la distribución binomial, 165
y la distribución multinomial, 232

Distribución de Rayleigh, 228

Distribución de Weisbull, 242

aplicación de, 242
esperanza de, 243
varianza de, 243

Distribución exponencial, 195
propiedades de, 196
valor esperado de, 197

y la distribución exponencial con dos parámetros, 230
y la distribución Gama, 229

y varianza de, 197

Distribución F de Snedecor, 322

- Distribución Gama, 201, 229
 valor esperado de, 202, 203
 varianza de, 202, 203
 y la distribución de Poisson, 202
 y la distribución exponencial, 201, 229
- Distribución geométrica, 175
 -propiedades de, 177
 valor esperado de, 176
 varianza de, 176
- Distribución hipergeométrica, 29, 180
 valor esperado de, 181
 varianza de, 181
 y la distribución binomial, 181
- Distribución log-normal, 213
- Distribución de Maxwell, 228
- Distribución multinomial, 181
 valor esperado de, 182
 varianza de, 182
 y la distribución de Poisson, 232
- Distribución normal, 187
 aproximación a la distribución binomial, 255, 257
 bivariada, 206
 estandarizada, 190
 función lineal de, 190
 propiedad reproductiva de, 226
 suma de variables aleatorias independientes, 225
 tabulación, 191
 truncada, 208
 valor esperado de, 190
 varianza de, 190
- Distribución normal bivariada, 206, 207
- Distribución *t* de Student, 313
 propiedades de, 313
- Distribución truncada, 208
- Distribución uniforme, 76
 bidimensional, 102
 unidimensional, 76, 77
 valor esperado de, 122
 varianza de, 143
- Distribución χ^2 , 203
 distribución Gama, y la, 203
 distribución normal, y la, 206, 227
 grandes grados de libertad, para, 207
 propiedad reproductiva, de, 227
 valor esperado de, 204
 varianza de, 205
- Distribuciones binomial, 64
 bivariada normal, 206, 207
 exponencial, 195
F de Snedecor, 322
 gama, 201
 geométrica, 175
 hipergeométrica, 29, 180
 log-normal, 213
 Maxwell, 228
 multinomial, 181
 normal, 187
 Pascal, 178
 Poisson, 164
 Rayleigh, 228
t de Student, 313
 uniforme, 76, 102
 Weisbull, 242
- Distribuciones mixtas, 75
- Ensayos de Bernoulli, 64
- Enumeración, métodos de, 24
 adición, principio de, 25
 combinaciones, 26
 multiplicación, principio de, 24
 permutaciones, 25, 30
- Errores tipo 1 y tipo 2, 326
- Escogencia al azar de un objeto, 23, 29, 39
- Escogencia de un punto al azar en un intervalo, 76
- Espacio muestral, 56
- Espacio muestral, el, 8
 ejemplos de, 9
 finito, 21
 partición de, 38
- Espacio muestral finito, el, 21
 y resultados igualmente probables, 22
- Esperanza condicional, 152
- Estadígrafo, 277
- Estimación, 291
 Estimación de máxima verosimilitud, 298, 300
 por el parámetro de una distribución exponencial, 300, 304
 por el parámetro de una distribución gama, 305
 por el parámetro de una distribución normal, 302
 por el parámetro de una distribución uniforme, 302, 303
 por la ecuación de probabilidad, 301

- propiedades, 300
- propiedades asintóticas, 306
- Estimación de parámetros, 291
 - estimador consistente, 292
 - estimador insesgado, 291
 - estimador de máxima probabilidad, 298, 299
 - estimador de cuadros mínimos, 308
 - estimador inexistente o insesgado, 323
 - estimador insesgado de varianza mínima, 292
 - mejor estimador lineal, 293
- Estimador, 291
 - insesgado, 291
- Experimento aleatorio, 7
- Experimento no determinístico, 7

- Factorial, 26
- Falla, tasa de, 234
- Fallas, ley exponencial de, 237
 - y la distribución de Poisson, 240
- Fallas, ley Gama de, 241
- Fallas, ley normal de, 236
- Fórmula de Stirling, 256
- Frecuencia relativa, 12
- Función de densidad de probabilidades, 68
 - del cociente de variables aleatorias independientes, 113
 - conjuntas, 97
 - marginal, 101
 - de la suma de variables aleatorias independientes, 266
 - del máximo de la muestra, 280
 - del mínimo de la muestra, 280
 - del producto de variables aleatorias independientes, 112, 112
 - del recorrido de la muestra, 283
- Función de densidad de probabilidad condicional, 104
- Función de densidad de probabilidad conjunta, 97
- Función de densidad de probabilidad marginal, 101
- Función de distribución, 75
- Función de distribución acumulativa, 71, 89
 - características de, 72
 - función de densidad de probabilidad, y la, 73, 74, 97
 - gráfica de, 72, 73
- propiedades de, 73, 74
- Función de distribución acumulativa conjunta, 100
- Función de operación característica, 326
 - para probar el parámetro P de una variable aleatoria con distribución binomial, 334
 - para probar la media de una distribución normal con varianza conocida, 326, 329
 - para probar la media de una distribución normal con varianza desconocida, 336
 - y escogencia del tamaño de la muestra, 331
- Función de regresión, 154
 - aproximación de los mínimos cuadrados, 158
 - lineal, 156
- Función de riesgo, 234
- Función gama, 199
- Función generadora de momentos, 218
 - de una distribución binomial, 219
 - de una distribución χ^2 , 221
 - de una distribución de Poisson, 219
 - de una distribución exponencial, 220
 - de una distribución gama, 221
 - de una distribución geométrica, 223
 - de una distribución normal, 220
 - de una distribución uniforme, 219
 - de una función lineal de una variable aleatoria, 224
- de series de variables aleatorias, 229
- de sumas de variables aleatorias independientes, 225
- de propiedad de unicidad, 224
- y momento, 222
- Funciones de variables aleatorias, 83
 - caso bidimensional, 108, 109
 - caso continuo, 88, 89
 - caso discreto, 86
 - función monótona, 90
 - valor esperado de, 126, 127, 132

- Gráfico de Venn, 5
- Grandes números, ley de, 252

- Hipótesis alterna, 324
- Hipótesis, docimasia de, 324
 - para la media de una variable aleatoria con distribución normal, 329

- nivel de significación, 329
- región crítica, 332
- Hipótesis nula, 324

- Integral de convolución, 265
- Intervalo de confianza, 311
 - para el parámetro de una distribución binomial, 317
 - para la media de una distribución normal (varianza conocida), 312
 - para la media de una distribución normal (varianza desconocida), 314
 - para la varianza de una distribución normal, 317

- Jacobiano de una transformación, 109

- Máximo de la muestra, 304
- Mecánica, analogía con, 61, 70, 124, 139, 141
- Mínimo de la muestra, 278, 280
 - y la distribución exponencial, 281
- Modelos matemáticos, 1
- Momento de una variable aleatoria respecto de su esperanza, 140
- Muestra, tamaño de, 327
- Muestra aleatoria de una variable aleatoria, 275
 - generación de una muestra aleatoria, 285, 287
- Muestras, 273
 - coeficiente de correlación, 311
 - máximo de, 278
 - media de, 277
 - mínimo de, 278
 - recorrido de, 278
 - varianza de, 278, 283

- Números aleatorios, 284

- Orden estadístico, 278

- Parámetro de una distribución, 120
- Probabilidad, 13
 - axiomas para, 14
 - condicional, 36
 - inducida, 59, 97
 - teoremas básicos, 15, 16
- Probabilidad condicional, 34
- Probabilidad inducida, 59, 86

- Probabilidad no condicional, 36
- Proceso de Poisson, 170, 174
 - aplicaciones de, 173, 174
 - supuestos de, 171
- Producto cartesiano, 6
- Promedio muestral, 278
- Propiedades reproductivas, 225
 - de la distribución χ^2 , 227
 - de la distribución normal, 226
 - de la distribución de Poisson, 226
- Punto muestral, 278

- Recorrido de la muestra, 283
- Regularidad estadística, 13
- Resultados igualmente probables, 22

- Serie geométricas, 63
- Sucesos, 10
 - complementarios, 11
 - independientes, 42
 - mutuamente excluyentes, 11
- Sucesos equivalentes, 58, 84
- Sucesos independientes, 42, 44
- Sucesos mutuamente excluyentes, 11
- Sucesos mutuamente independientes, 46
- Suma de variables aleatorias independientes, 265, 269

- Teorema de Bayes, 40
- Teorema de la multiplicación de probabilidades, 37
 - generalización de, 52
- Teorema del binomio, 27
 - forma generalizada, 28
- Teorema del límite central, 259, 260
- Teorema Gauss-Markoff, 310
- Transformación integral, 285

- Valor esperado de una variable aleatoria, 120, 122, 124
 - aproximación para, 143
 - de la suma de variables aleatorias, 132
 - de una función de una variable aleatoria, 126, 132
 - de una función lineal, 133
 - del producto de variables aleatorias independientes, 134
 - del valor condicional esperado, 152
 - propiedades del valor esperado, 132

Valor^a medio de una variable aleatoria, 120

Variable aleatoria, 55

continua, 67

discreta, 60

distribución de probabilidad, 61

espacio muestral de, 56

no correlacionada, 149

sucesiones de, 229

Variable aleatoria bidimensional, 95

continua, 96

discreta, 96

normal, 206

Variable aleatoria continua, 67

Variables aleatorias independientes, 105, 106
criterio de, 106

Variables aleatorias *n*-dimensionales, 114

Varianza de una variable aleatoria, 138

aproximación a, 143

evaluación de, 139

propiedades de, 140

de la suma de variables aleatorias independientes, 141

Varianza muestral, 273