

Estadística

Grados en Ingeniería de Telecomunicaciones

Versión diciembre 2012

Índice general

Índice de figuras

page VII

Parte I	Introducción	1
1.	Modelos probabilísticos en ingeniería de telecomunicaciones	3
1.1.	Los modelos matemáticos como herramientas de análisis y diseño	4
1.2.	Modelos deterministas	5
1.3.	Modelos probabilísticos	7
1.3.1.	Regularidad estadística	8
1.3.2.	Propiedades de la frecuencia relativa	8
1.3.3.	Aproximación axiomática a la teoría de la probabilidad	10
1.3.4.	Construcción de un modelo de probabilidad	11
1.4.	Un ejemplo detallado: un sistema de transmisión de paquetes de voz	12
1.5.	Otros ejemplos	15
1.5.1.	Comunicación a través de canales no fiables	15
1.5.2.	Compresión de señales	17
1.5.3.	Fiabilidad de sistemas	17
1.5.4.	Sistemas con uso compartido de recursos	18
1.5.5.	Sistemas a escala internet	19
2.	Conceptos básicos de teoría de la probabilidad	23
2.1.	Especificación de experimentos aleatorios	23
2.1.1.	El espacio muestral	25
2.1.2.	Sucesos	27
2.1.3.	Revisión de la teoría de conjuntos	28
2.1.4.	Clases de sucesos	34
2.2.	Los axiomas de la probabilidad	35
2.2.1.	Espacios muestrales discretos	39
2.2.2.	Espacios muestrales continuos	43
2.3.	Probabilidad condicionada	47
2.3.1.	Teorema de Bayes	52
2.4.	Independencia de sucesos	54
2.5.	Experimentos secuenciales	60

2.5.1.	Sucesiones de experimentos independientes	60
2.5.2.	La ley de probabilidad binomial	61
2.5.3.	La ley de probabilidad multinomial	65
2.5.4.	La ley de probabilidad geométrica	66
2.5.5.	Secuencias de experimentos dependientes	67
Parte II	Variables aleatorias	77
3.	Variables aleatorias discretas	79
3.1.	La noción de variable aleatoria	79
3.2.	Variables aleatorias discretas y la función de probabilidad	82
3.3.	Esperanza y momentos de una variable aleatoria discreta	87
3.3.1.	Esperanza de la función de una variable aleatoria	91
3.3.2.	Varianza de una variable aleatoria	93
3.4.	Función de probabilidad condicionada	95
3.4.1.	Función de probabilidad condicionada	95
3.4.2.	Esperanza condicionada	98
3.5.	Variables aleatorias importantes	100
3.5.1.	Variable aleatoria de Bernoulli	100
3.5.2.	Variable aleatoria Binomial	102
3.5.3.	Variable aleatoria Geométrica	104
3.5.4.	Variable aleatoria de Poisson	105
3.5.5.	La variable aleatoria Uniforme	110
4.	Variables aleatorias	115
4.1.	La función de distribución	115
4.1.1.	Los tres tipos de variables aleatorias	121
4.2.	La función de densidad	122
4.2.1.	pdf de variables aleatorias discretas	125
4.2.2.	cdf y pdf condicionadas	126
4.3.	La esperanza de X	129
4.3.1.	La esperanza de $Y = g(X)$	131
4.3.2.	Varianza de X	134
4.4.	Variables aleatorias continuas importantes	136
4.4.1.	Variable aleatoria Uniforme	136
4.4.2.	Variable aleatoria Exponencial	138
4.4.3.	Variable aleatoria Normal	139
4.4.4.	Variable aleatoria Gamma	141
4.4.5.	Variable aleatoria Beta	144
4.4.6.	Variable aleatoria de Cauchy	145
4.4.7.	Variable aleatoria de Pareto	145
4.5.	Funciones de una variable aleatoria	146
4.6.	Las desigualdades de Chebyshev y Markov	153

5.	Variables aleatorias bidimensionales	160
5.1.	Dos variables aleatorias	160
5.2.	Pares de variables aleatorias discretas	165
5.2.1.	Función de masa de probabilidad marginal	168
5.3.	La cdf conjunta de X e Y	170
5.3.1.	Variables aleatorias de diferente tipo	176
5.4.	La pdf conjunta de dos variables aleatorias continuas	177
5.5.	Independencia de dos variables aleatorias	183
5.6.	Momentos conjuntos y esperanza de una función de dos variables aleatorias	186
5.6.1.	Esperanza de una función de dos variables aleatorias	187
5.6.2.	Momentos conjuntos, correlación y covarianza	188
5.7.	Probabilidad condicionada y esperanza condicionada	191
5.7.1.	Probabilidad condicionada	192
5.7.2.	Esperanza condicionada	199
5.8.	Funciones de dos variables aleatorias	201
5.8.1.	Una función de dos variables aleatorias	202
5.8.2.	Transformaciones de dos variables aleatorias	205
5.8.3.	pdf de transformaciones lineales	207
5.9.	Variable aleatoria normal bivariante	210
6.	Variables aleatorias multidimensionales	219
6.1.	Variables aleatorias multidimensionales	219
6.1.1.	Sucesos y probabilidades	220
6.1.2.	Funciones de distribución conjuntas	221
6.1.3.	Independencia	225
6.2.	Funciones de varias variables aleatorias	226
6.2.1.	Una función de varias variables aleatorias	226
6.2.2.	Transformaciones de vectores aleatorios	228
6.3.	Esperanza de variables aleatorias multidimensionales	229
6.3.1.	Vector de medias y matriz de covarianza	229
6.3.2.	Transformaciones lineales de vectores aleatorios	231
6.4.	Variable aleatoria normal multivariante	233
7.	Suma de variables aleatorias	242
7.1.	Suma de variables aleatorias	242
7.1.1.	Media y varianza de la suma de variables aleatorias	243
7.1.2.	pdf de la suma de variables aleatorias independientes	244
7.2.	La media muestral y las leyes de los grandes números	247
7.3.	El teorema central del límite	251
7.3.1.	Aproximación normal de probabilidades binomiales	255
Parte III	Procesos estocásticos	259

8.	Procesos estocásticos	261
8.1.	Definición de proceso estocástico	261
8.2.	Especificación de un proceso estocástico	265
8.2.1.	Distribuciones conjuntas de las muestras en el tiempo	266
8.2.2.	Las funciones media, autocorrelación y autocovarianza	268
8.2.3.	Procesos estocásticos múltiples	270
8.3.	Proceso de suma, proceso de conteo binomial y paseo aleatorio	272
8.3.1.	Procesos estocásticos iid	273
8.3.2.	Incrementos independientes y propiedad de Markov de procesos estocásticos	275
8.3.3.	Procesos de suma: los procesos conteo binomial y paseo aleatorio	276
8.4.	Procesos estocásticos estacionarios	282
8.4.1.	Procesos estocásticos estacionarios en sentido amplio	284
8.4.2.	Procesos estocásticos gaussianos estacionarios en sentido amplio	288
8.5.	Promedios temporales de procesos estocásticos y teoremas ergódicos	288

Índice de figuras

1.1.	El proceso de modelado.	6
1.2.	Respuestas del experimento con la urna.	7
1.3.	Frecuencias relativas del experimento con la urna.	9
1.4.	Frecuencias relativas del experimento con la urna.	9
1.5.	Un sistema de transmisión de paquetes de voz.	13
1.6.	Probabilidades para el número de altavoces activos en un grupo de 48.	14
1.7.	(a) Un modelo de canal de comunicación binario. (b) Sistema de control de error.	16
1.8.	Sistemas con n componentes.	18
1.9.	Un modelo simple para un sistema servidor web.	19
1.10.	Un gran comunidad de usuarios interactuando a través de internet.	19
2.1.	Espacios muestrales de los experimentos E_7, E_9, E_{12} y E_{13} .	26
2.2.	Relaciones y operaciones entre conjuntos.	30
2.3.	Descomposición de $A \cup B$ en tres conjuntos disjuntos.	38
2.4.	Si $A \subset B$, entonces $P[A] \leq P[B]$.	39
2.5.	En n ensayos aparecen caras en el primer lanzamiento $n/2$ veces aproximadamente, en el segundo lanzamiento $n/4$ veces aproximadamente, y así sucesivamente.	43
2.6.	$(r, \infty) = \delta(r, s] \cup (s, \infty)$.	46
2.7.	Un espacio muestral bidimensional y tres sucesos.	47
2.8.	Si sabemos que ha ocurrido B , entonces A puede ocurrir solamente si $A \cap B$ ocurre.	48
2.9.	Los caminos desde el nodo superior hasta un nodo inferior se corresponden con las posibles respuestas en la extracción de dos bolas de una urna sin reemplazamiento. La probabilidad de un camino es el producto de las probabilidades de las transiciones asociadas.	50
2.10.	Probabilidades de parejas entrada-respuesta en un sistema de transmisión binario.	51
2.11.	Una partición de S en n conjuntos disjuntos.	51
2.12.	Ejemplos de sucesos independientes y dependientes.	57
2.13.	Los sucesos B, D y F son independientes dos a dos, pero el triplete B, D, F no está formado por sucesos independientes.	58
2.14.	Diagrama de enrejado de una cadena de Markov.	68
2.15.		72

2.16.		73
3.1.	Una variable aleatoria asigna un número real $X(\zeta)$ a cada elemento ζ del espacio muestral S de un experimento aleatorio.	80
3.2.	$P[X \in B] = P[\zeta \in A]$	82
3.3.	Partición del espacio muestral S asociado a una variable aleatoria discreta.	83
3.4.	(a) Gráfico de la pmf de tres lanzamientos de una moneda; (b) gráfico de la pmf del juego de apuestas.	84
3.5.	(a) Frecuencias relativas y pmf uniforme; (b) Frecuencias relativas y pmf geométrica.	87
3.6.	Los gráficos muestran 150 repeticiones de los experimentos que generan X e Y . La variable aleatoria X se centra en el valor 5 mientras que Y varía alrededor del valor 0. Es obvio también que X es más dispersa que Y .	87
3.7.	pmf condicionada de X dado el suceso C .	96
3.8.	Función de probabilidad de la variable aleatoria Binomial (a) $p = 0.2$; (b) $p=0.5$.	102
3.9.	Función de probabilidad de la variable aleatoria de Poisson (a) $\alpha = 0.75$; (b) $\alpha = 3$, (c) $\alpha = 9$.	106
3.10.	Sucesos que ocurren en n subintervalos de $[0, T]$.	109
4.1.	cdf (a) y pdf (b) de una variable aleatoria discreta.	117
4.2.	cdf (a) y pdf (b) de una variable aleatoria continua.	118
4.3.	cdf (a) y pdf (b) de una variable aleatoria de tipo mixto.	119
4.4.	(a) La función de densidad especifica la probabilidad de intervalos de anchura infinitesimal. (b) La probabilidad de un intervalo $[a, b]$ es el área por debajo de la pdf en dicho intervalo.	123
4.5.	La pdf condicionada, dada la señal de entrada.	129
4.6.	Dos sucesos infinitesimales equivalentes.	132
4.7.	Función de densidad de una variable aleatoria Normal.	136
4.8.	Un ejemplo de variable aleatoria continua – la variable aleatoria Exponencial. La parte (a) es la cdf y la parte (b) es la pdf.	138
4.9.	Función de densidad de la variable aleatoria Gamma.	142
4.10.	El suceso equivalente para $\{Y \leq y\}$ es el suceso $\{X \leq (y - b)/a\}$, si $a > 0$.	148
4.11.	El suceso equivalente para $\{Y \leq y\}$ es el suceso $\{-\sqrt{y} \leq X \leq \sqrt{y}\}$, si $y \geq 0$.	150
4.12.	El suceso equivalente de $\{y < Y < y + dy\}$ es $\{x_1 < X < x_1 + dx_1\} \cup \{x_2 + dx_2 < X < x_2\} \cup \{x_3 < X < x_3 + dx_3\}$.	151
4.13.	$y = \cos x$ tiene dos raíces en el intervalo $(0, 2\pi)$.	152
4.14.	Límites en la función indicatriz para $A = \{t \geq a\}$.	156
4.15.		157
5.1.	(a) Una función que asigna un par de números reales a cada elemento de S . (b) Sucesos equivalentes para dos variables aleatorias.	161
5.2.	Ejemplos de sucesos bidimensionales.	162
5.3.	Diagramas de dispersión de 200 observaciones de cuatro diferentes pares de variables aleatorias.	164

5.4.	Algunos sucesos bidimensionales en forma de producto.	165
5.5.	Representaciones gráficas de la pmf: (a) en forma de tabla; (b) uso de flechas para mostrar altura; (c) puntos etiquetados con el valor de la pmf.	166
5.6.	Mostrando la pmf a través de un esquema que contiene los puntos en B .	168
5.7.	La función de distribución conjunta se define como la probabilidad del rectángulo semi-infinito definido por el punto (x_1, y_1) .	171
5.8.	Las cdf marginales son las probabilidades de estos semiplanos.	172
5.9.	La cdf conjunta se puede usar para determinar la probabilidad de varios sucesos.	173
5.10.	cdf conjunta del ejemplo del conmutador de paquetes.	173
5.11.	cdf conjunta de dos variables aleatorias uniformes.	174
5.12.	Sucesos bi-dimensionales que no atienden a la forma de producto.	177
5.13.	La probabilidad de A es la integral de $f_{X,Y}(x, y)$ en la región definida por A .	178
5.14.	Interpretación de las pdf marginales.	179
5.15.	Regiones que deben ser consideradas de forma separada en el cálculo de la cdf del ejemplo 5.15.	180
5.16.	Las variables aleatorias X e Y de los ejemplos 5.16 y 5.17 tienen una pdf que es no negativa solo en la región sombreada mostrada en la parte (a).	182
5.17.	pdf conjunta de dos variables aleatorias normales.	183
5.18.	(X, Y) es un punto seleccionado de forma aleatoria en la circunferencia de radio 1. X e Y están incorreladas pero no son independientes.	190
5.19.	Interpretación de la pdf condicionada.	196
5.20.	$P[Z \leq z] = P[X + Y \leq z]$.	202
5.21.	$\{\min(X, Y) \leq w = \{X \leq w\} \cup \{Y \leq w\}\}$ y $\{\max(X, Y) \leq z = \{X \leq z\} \cap \{Y \leq z\}\}$.	206
5.22.	Región de integración $R_{(r_0, \theta_0)}$ en el ejemplo 5.44.	207
5.23.	Imagen de un rectángulo infinitesimal bajo una transformación lineal.	208
5.24.	Mapa de isobaras de la pdf normal conjunta del ejemplo 5.45.	210
5.25.	pdf de la normal bivalente (a) $\rho = 0$ (b) $\rho = -0.9$.	211
5.26.	Orientación de las líneas de contorno de la pdf de la normal bivalente para $\rho_{X,Y} > 0$.	212
5.27.	Una rotación del sistema de coordenadas transforma un par de variables aleatorias normales dependientes en un par de variables aleatorias normales independientes.	214
7.1.	Convergencia de la secuencia de medias muestrales a $E[X]$.	250
7.2.	(a) La cdf de la suma de cinco variables aleatorias independientes de Bernoulli con $p = 1/2$ y la cdf de una variable aleatoria Normal con la misma media y varianza. (b) La cdf de la suma de 25 variables aleatorias de Bernoulli con $p = 1/2$ y la cdf de una variable aleatoria Normal con la misma media y varianza.	252

7.3.	La cdf de la suma de cinco variables aleatorias independientes discretas uniformes en el conjunto $\{0, 1, \dots, 9\}$ y la cdf de una variable aleatoria Normal con la misma media y varianza.	253
7.4.	(a) La cdf de la suma de cinco variables aleatorias independientes exponenciales con media 1 y la cdf de una variable aleatoria Normal con la misma media y varianza. (b) La cdf de la suma de 50 variables aleatorias exponenciales de media 1 y la cdf de una variable aleatoria Normal con la misma media y varianza.	253
7.5.	Aproximación normal de la pdf de S_{100} y S_{129} en los ejemplos 7.10 y 7.11.	254
7.6.	(a) Aproximación normal de probabilidades binomiales con $n = 5$ y $p = 1/2$. (b) Aproximación normal de la binomial con $n = 25$ y $p = 1/2$.	256
8.1.	Distintas realizaciones de un proceso estocástico.	262
8.2.	(a) Sinusoide amplitud aleatoria, (b) Sinusoide con fase aleatoria.	264
8.3.	(a) pdf de una sinusoide con amplitud aleatoria. (b) pdf de una sinusoide con fase aleatoria. (c) Histograma de las muestras de una sinusoide con amplitud aleatoria uniformemente distribuida en $t = 0$. (d) Histograma de las muestras de una sinusoide con fase aleatoria en $t = 0$.	265
8.4.	(a) Realización de un proceso de Bernoulli. $I_n = 1$ indica que una bombilla se estropea y se sustituye en el día n . (b) Realización de un proceso binomial. S_n denota de número de bombillas que se han estropeado hasta el día n .	274
8.5.	(a) Realización de un proceso escalón aleatorio. (b) Realización de un proceso paseo aleatorio. S_n denota la posición de una partícula en el tiempo n .	275
8.6.	El proceso de suma $S_n = X_1 + \dots + X_n$, $S_0 = 0$ puede ser generado de esta forma.	276
8.7.	(a) Proceso paseo aleatorio con $p = 1/2$. (b) Cuatro funciones muestrales de un proceso paseo aleatorio simétrico con $p = 1/2$. (c) Cuatro funciones muestrales de un proceso paseo aleatorio asimétrico con $p = 3/4$.	278
8.8.	(a) Proceso autorregresivo de primer orden. (b) Proceso de media móvil.	282
8.9.	(a) Función de autocorrelación de una señal telégrafica aleatoria. (b) Función de autocorrelación de una sinusoide con fase aleatoria. (c) función de autocorrelación de un proceso estocástico que tiene media distinta de cero, una componente periódica, y una componente "aleatoria".	287
8.10.		293

Parte I

Introducción

1 Modelos probabilísticos en ingeniería de telecomunicaciones

Los ingenieros en telecomunicaciones han jugado un papel central en el diseño de los nuevos sistemas de información y comunicación. Estos sistemas de gran éxito funcionan de forma fiable y predecible en entornos muy variables y caóticos:

- Las redes de comunicación inalámbricas proporcionan comunicaciones de voz y datos a usuarios móviles en entornos con intensas interferencias.
- La gran mayoría de las señales de los medios de comunicación, voz, audio, imágenes y vídeo, son procesadas digitalmente.
- Enormes “granjas” de servidores web ofrecen a los usuarios grandes cantidades de información muy específica.

Debido a estos éxitos, los diseñadores de hoy se enfrentan a retos aún mayores. Los sistemas que construyen no tienen precedentes en lo que se refiere a la escala y el ambiente caótico en el que deben operar es un territorio inexplorado:

- La información en la web se crea y publica a un ritmo acelerado; las futuras aplicaciones de búsqueda deberán ser más exigentes para obtener la respuesta requerida a partir de un vasto océano de información.
- Los delincuentes de la era de la información secuestran computadoras y las explotan con propósitos ilícitos, por lo que se necesitan métodos para identificar y contener estas amenazas.
- Los sistemas de aprendizaje de máquinas deben ir más allá de la navegación y la compra de aplicaciones, permitiendo el monitoreo de la salud y el medio ambiente en tiempo real.
- Los sistemas de distribución masiva de la forma “peer-to-peer” y las comunidades de computación “grid” han surgido y cambiado la naturaleza de la prestación de los medios de comunicación, los juegos y la interacción social; sin embargo, no entienden o no saben cómo controlar y administrar estos sistemas.

Los modelos de probabilidad son una de las herramientas que permiten al diseñador dar sentido al caos y construir con éxito sistemas que sean eficientes, fiables y rentables. Este manual es una introducción a la teoría subyacente a los modelos de probabilidad, así como a las técnicas básicas utilizadas en el desarrollo de dichos modelos.

En este capítulo se presentan los modelos de probabilidad y se muestra cómo se diferencian de los modelos deterministas que son dominantes en la ingeniería. Se desarrollan las propiedades clave de la noción de probabilidad, y se presentan varios ejemplos de ingeniería en telecomunicaciones, donde los modelos de probabilidad juegan un papel clave.

1.1 Los modelos matemáticos como herramientas de análisis y diseño

El diseño o modificación de cualquier sistema complejo consiste en la toma de decisiones entre diferentes alternativas viables. Las decisiones se toman sobre la base de criterios tales como coste, fiabilidad y rendimiento. La evaluación cuantitativa de estos criterios rara vez se hace a través de la aplicación real y la evaluación experimental de las configuraciones alternativas. Por el contrario, las decisiones se toman en base a estimaciones que se obtienen utilizando modelos de las alternativas.

Un modelo es una representación aproximada de una situación física. Un modelo intenta explicar el comportamiento observado utilizando una serie de reglas sencillas y comprensibles. Estas reglas pueden ser utilizadas para predecir el resultado de experimentos involucrados en dicha situación física. Un modelo útil explica todos los aspectos relevantes de una situación dada. Estos modelos pueden utilizarse en lugar de los experimentos para responder preguntas relacionadas con la situación dada. Por lo tanto, los modelos permiten al ingeniero evitar los costes de la experimentación, es decir, mano de obra, equipo y tiempo.

Los modelos matemáticos se utilizan cuando el fenómeno observado tiene propiedades medibles. Un modelo matemático consiste en un conjunto de supuestos acerca de cómo funciona un sistema o un proceso físico. Estos supuestos se presentan en forma de relaciones matemáticas que involucran los parámetros y las variables más importantes del sistema. Las condiciones bajo las cuales se realiza un experimento que involucra al sistema determinan los elementos “dados” en las relaciones matemáticas, y la solución de estas relaciones nos permite predecir las mediciones que se obtendrían si el experimento se realizara.

Los modelos matemáticos son utilizados ampliamente por los ingenieros en el diseño y modificación de sistemas de guiado. La intuición y las reglas generales no siempre son fiables para predecir el rendimiento de sistemas complejos y novedosos, y la experimentación no es posible durante las fases iniciales del diseño de sistema. Además, el coste de una experimentación extensa en los sistemas ya existentes, con frecuencia resulta ser prohibitivo. La disponibilidad de modelos adecuados para los componentes de un sistema complejo en combinación con el conocimiento de sus interacciones permiten al científico y al ingeniero desarrollar un modelo matemático completo del sistema. Es entonces posible responder de forma rápida y económica a preguntas sobre el funcionamiento de sistemas complejos. De hecho, los programas de ordenador para la obtención de la solución

de los modelos matemáticos son la base de muchos de los análisis y diseños de sistemas asistidos por ordenador.

Con el fin de ser útil, un modelo debe ajustarse a los hechos de una situación dada. Por tanto, el proceso de desarrollo y validación de un modelo necesariamente se compone de una serie de experimentos y modificaciones del modelo como se muestra en la figura 1.1. Cada experimento investiga un determinado aspecto del fenómeno bajo investigación y consiste en la toma de observaciones y mediciones bajo un conjunto específico de condiciones. El modelo se utiliza para predecir el resultado del experimento, y estas predicciones se comparan con las observaciones reales que se producen cuando el experimento se lleva a cabo. Si hay una discrepancia significativa, el modelo se modifica para dar cuenta de ello. El proceso de modelado continúa hasta que el investigador está convencido de que el comportamiento de todos los aspectos pertinentes al fenómeno se puede predecir con una precisión deseada. Hay que destacar que la decisión de cuándo parar el proceso de modelado depende de los objetivos inmediatos del investigador. Así, un modelo que sea adecuado para una aplicación puede llegar a ser completamente inadecuado en otra.

Las predicciones de un modelo matemático deben ser entendidas como hipótesis hasta que el modelo haya sido validado a través de una comparación con las mediciones experimentales. Se plantea un dilema en el caso del diseño de un sistema: el modelo no se puede validar experimentalmente porque el sistema real no existe. Modelos de simulación por ordenador desempeñan un papel útil en esta situación mediante la presentación de un medio alternativo de predicción del comportamiento del sistema y, por lo tanto, una forma de verificar las predicciones hechas por un modelo matemático. Un modelo informático de simulación consiste en un programa de ordenador que simula o imita la dinámica de un sistema. Incorporadas al programa se encuentran unas instrucciones que “miden” los parámetros de funcionamiento relevantes. En general, los modelos de simulación son capaces de representar los sistemas con mayor detalle que los modelos matemáticos. Sin embargo, tienden a ser menos flexibles y por lo general requieren más tiempo de cálculo que los modelos matemáticos.

En las dos secciones siguientes se discuten los dos tipos básicos de modelos matemáticos, los modelos deterministas y los modelos de probabilidad.

1.2 Modelos deterministas

En los modelos deterministas las condiciones bajo las cuales un experimento se lleva a cabo determinan el resultado exacto de la prueba. En los modelos matemáticos deterministas, la solución de un conjunto de ecuaciones matemáticas especifica el resultado exacto del experimento. La teoría de circuitos es un ejemplo de un modelo matemático determinista.

La teoría de circuitos modeliza la interconexión de dispositivos electrónicos mediante circuitos ideales formados por componentes discretos con características

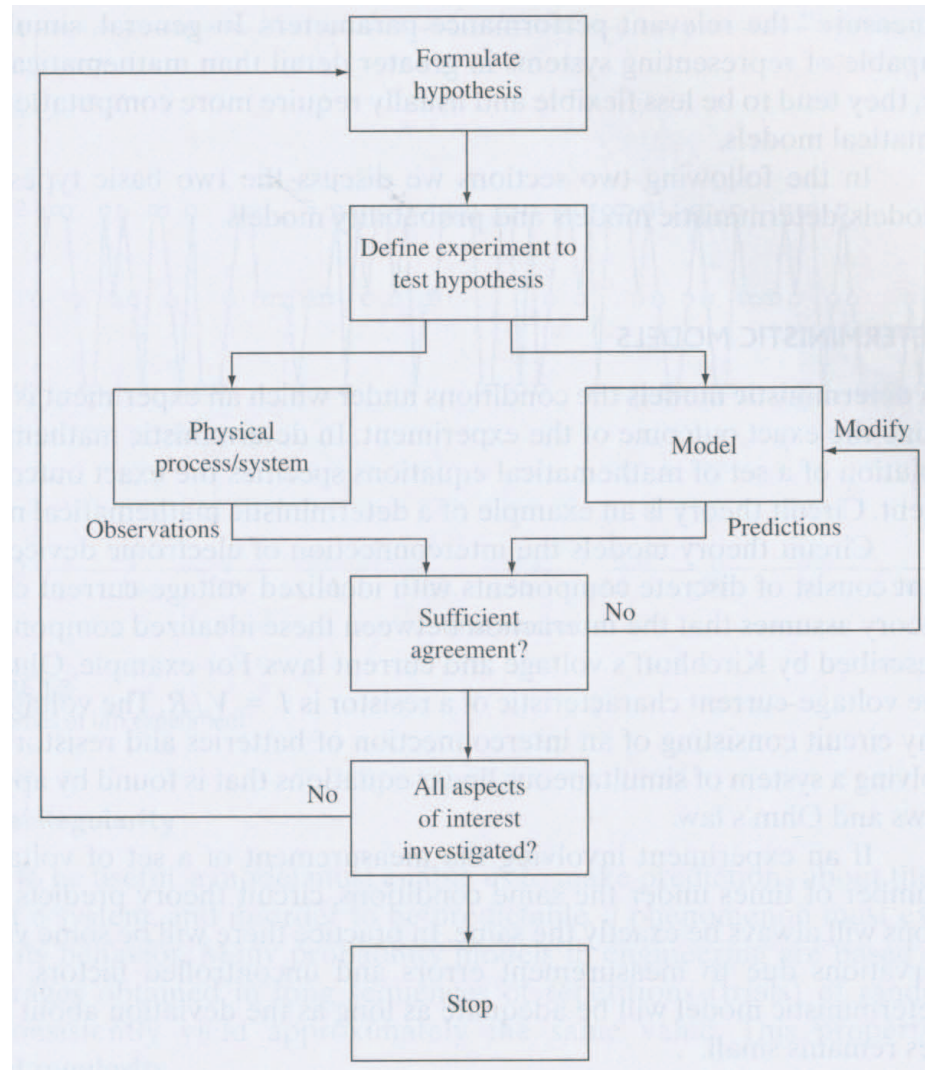


Figura 1.1 El proceso de modelado.

de tensión-corriente idealizadas. La teoría asume que la interacción entre estos componentes idealizados se describe completamente mediante la ley de tensiones de Kirchhoff y las leyes de corriente. Por ejemplo, la ley de Ohm establece que la característica tensión-corriente de una resistencia es $I = V/R$. Las tensiones y corrientes en cualquier circuito que consista en una interconexión de baterías y resistencias se pueden calcular resolviendo el sistema de ecuaciones lineales simultáneas que se plantea al aplicar las leyes de Kirchhoff y la ley de Ohm.

Si un experimento relacionado con la medición de un conjunto de tensiones se repite un número de veces bajo las mismas condiciones, la teoría de circui-

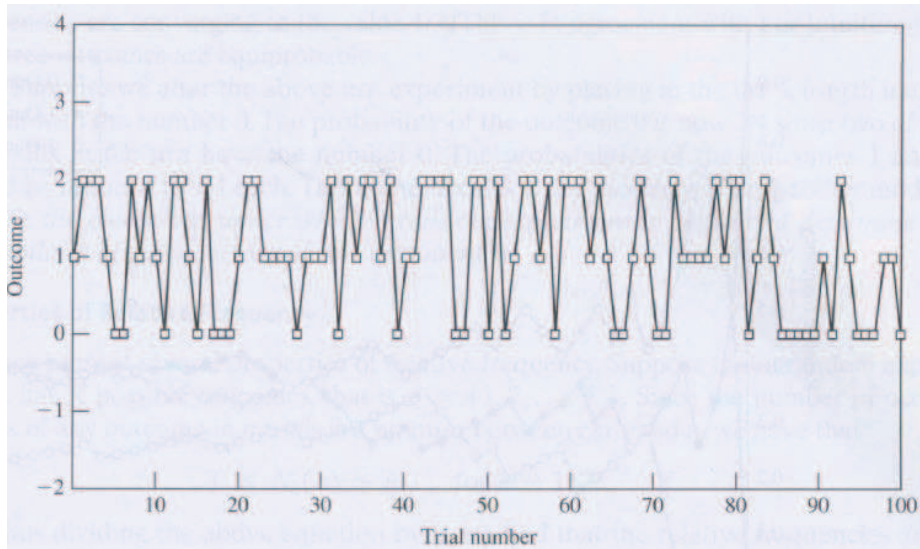


Figura 1.2 Respuestas del experimento con la urna.

tos predice que las observaciones siempre serán exactamente las mismas. En la práctica habrá variaciones en las observaciones debidas a errores de medición y factores no controlados. Sin embargo, este modelo determinista será adecuado, siempre y cuando la desviación de los valores previstos siga siendo pequeña.

1.3 Modelos probabilísticos

Muchos sistemas de interés involucran fenómenos que presentan variaciones impredecibles y aleatoriedad. Se define un experimento aleatorio como un experimento en el que el resultado varía de forma impredecible cuando el experimento se repite bajo las mismas condiciones. Los modelos deterministas no son apropiados para experimentos aleatorios ya que predicen el mismo resultado en cada repetición de un experimento. En esta sección se presentan los modelos de probabilidad orientados a experimentos aleatorios.

Como ejemplo de un experimento aleatorio, supongamos que se selecciona una bola de una urna que contiene tres bolas idénticas, numeradas 0, 1 y 2. Se sacude la urna en primer lugar para aleatorizar la posición de las bolas, y entonces se selecciona una bola. Se anota el número de la bola y se devuelve a la urna. El resultado de este experimento es un número perteneciente al conjunto $S = \{0, 1, 2\}$. Llamamos espacio muestral al conjunto S de todos los posibles resultados. La figura 1.2 muestra los resultados de 100 repeticiones (ensayos) de una simulación por ordenador de este experimento con la urna. Está claro que el resultado de este experimento no siempre puede predecirse de forma correcta.

1.3.1 Regularidad estadística

Con el fin de ser útil, un modelo debe permitir hacer predicciones sobre el comportamiento futuro de un sistema mientras que, con el fin de ser predecible, un fenómeno debe exhibir regularidad en su comportamiento. Muchos de los modelos de probabilidad en ingeniería se basan en el hecho de que los promedios obtenidos a partir de largas secuencias de repeticiones (ensayos) de experimentos aleatorios, consecuentemente den de forma aproximada el mismo valor. Esta propiedad se llama regularidad estadística.

Supongamos que el experimento anterior con la urna se repite n veces en idénticas condiciones. Sean $N_0(n)$, $N_1(n)$ y $N_2(n)$ el número de veces en las que el resultado del experimento es bola 0, bola 1 y bola 2, respectivamente, y definamos la frecuencia relativa del resultado k como

$$f_k(n) = \frac{N_k(n)}{n}. \quad (1.1)$$

Con el término regularidad estadística nos referimos a que $f_k(n)$ difiere cada vez menos de un valor constante a medida que n se hace grande, es decir,

$$\lim_{n \rightarrow \infty} f_k(n) = p_k. \quad (1.2)$$

La constante p_k se denomina probabilidad del resultado k . La ecuación (1.2) establece que la probabilidad de un resultado es la proporción de veces a largo plazo que se obtiene dicho resultado en una secuencia grande de ensayos. Veremos a lo largo del curso que la ecuación (1.2) proporciona la conexión clave en la transición entre la medición de cantidades físicas y los modelos de probabilidad analizados en este curso.

Las figuras 1.3 y 1.4 muestran las frecuencias relativas de los tres posibles resultados para el experimento anterior de la urna a medida que el número de ensayos n es mayor. Está claro que todas las frecuencias relativas convergen al valor $1/3$. Esto concuerda con nuestra intuición de que los tres resultados son equiprobables.

Supongamos que se modifica el experimento anterior mediante la colocación en la urna de una cuarta bola idéntica con el número 0. La probabilidad del resultado 0 es ahora $2/4$, ya que dos de las cuatro bolas en la urna tienen el número 0. Las probabilidades de los resultados 1 y 2 se reducirían a $1/4$ cada una. Esto demuestra una propiedad clave de los modelos de probabilidad, a saber, las condiciones en que se lleva a cabo un experimento aleatorio determinan las probabilidades de los resultados de un experimento.

1.3.2 Propiedades de la frecuencia relativa

A continuación se presentan varias propiedades de la frecuencia relativa. Supongamos que un experimento aleatorio tiene K resultados posibles, es decir, $S = \{1, 2, \dots, K\}$. Dado que el número de ocurrencias de cualquier resultado en

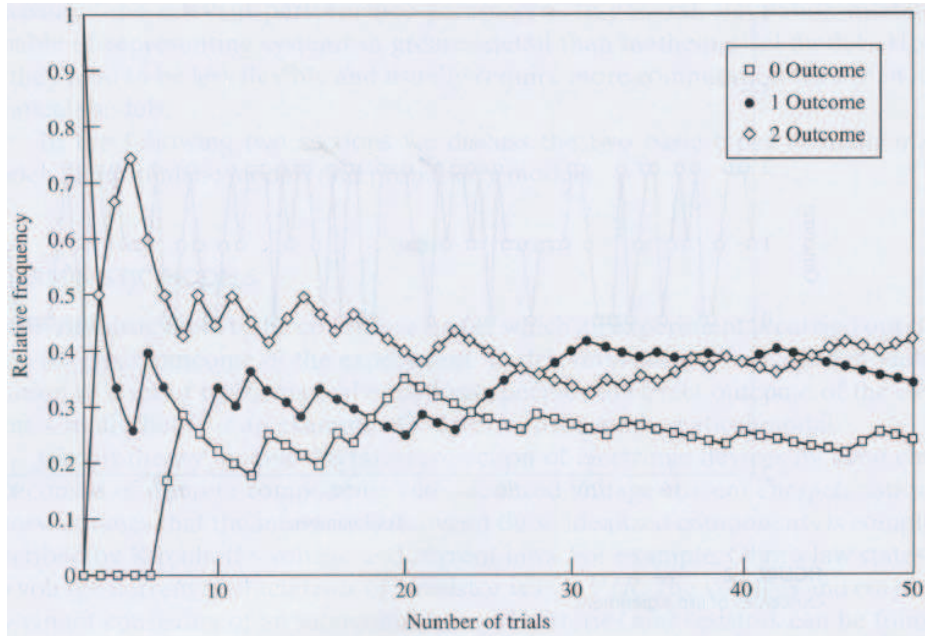


Figura 1.3 Frecuencias relativas del experimento con la urna.

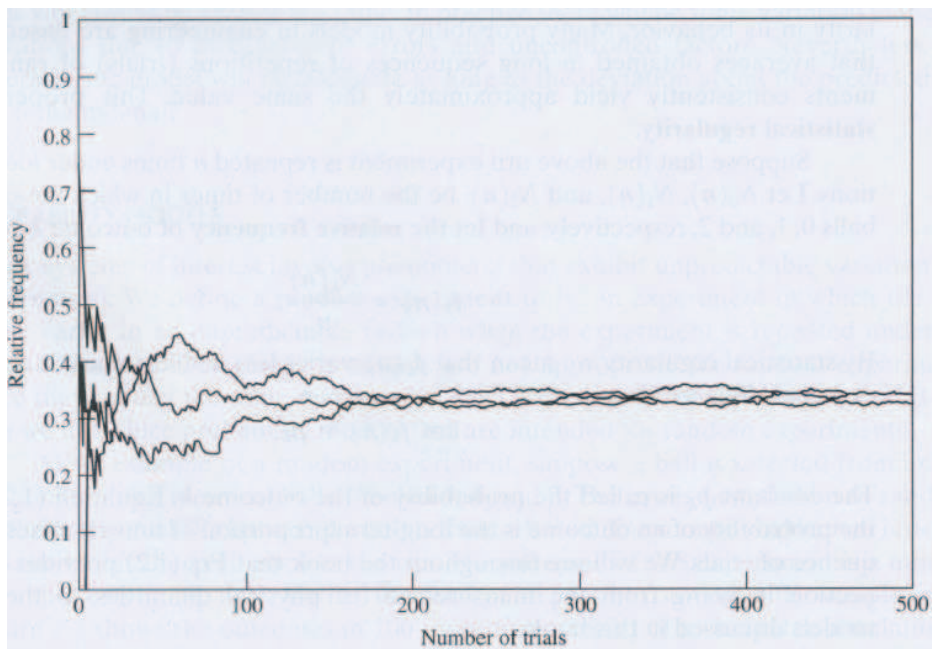


Figura 1.4 Frecuencias relativas del experimento con la urna.

n ensayos es un número entre cero y n , tenemos que

$$0 \leq N_k(n) \leq n, \text{ for } k = 1, 2, \dots, K,$$

y por tanto dividiendo la ecuación anterior por n , nos encontramos con que las frecuencias relativas son un número entre cero y uno:

$$0 \leq f_k(n) \leq 1, \text{ for } k = 1, 2, \dots, K. \quad (1.3)$$

La suma del número de ocurrencias de todos los resultados posibles debe ser n :

$$\sum_{k=1}^K N_k(n) = n.$$

Si dividimos por n ambos lados de la ecuación anterior, nos encontramos con que la suma de todas las frecuencias relativas es igual a uno:

$$\sum_{k=1}^K f_k(n) = 1. \quad (1.4)$$

A veces estamos interesados en la ocurrencia de sucesos asociados con los resultados de un experimento. Por ejemplo, consideremos el suceso “seleccionar una bola con número par” en el experimento anterior con la urna. ¿Cuál es la frecuencia relativa de este suceso? Este suceso tendrá lugar siempre que el número de la bola sea 0 ó 2. El número de experimentos en el que el resultado es una bola con número par es por lo tanto, $N_E(n) = N_0(n) + N_2(n)$. La frecuencia relativa de este suceso es entonces

$$f_E(n) = \frac{N_E(n)}{n} = \frac{N_0(n) + N_2(n)}{n} = f_0(n) + f_2(n).$$

Este ejemplo muestra que la frecuencia relativa de un suceso es la suma de las frecuencias relativas de los resultados asociados. Más en general, sea C el suceso “ocurre A u ocurre B ”, donde A y B son dos sucesos que no pueden ocurrir simultáneamente, entonces el número de veces que ocurre C es $N_C(n) = N_A(n) + N_B(n)$, por lo que

$$f_C(n) = f_A(n) + f_B(n). \quad (1.5)$$

Las ecuaciones (1.3), (1.4) y (1.5) son las tres propiedades básicas de la frecuencia relativa a partir de las cuales podemos obtener muchos otros resultados útiles.

1.3.3 Aproximación axiomática a la teoría de la probabilidad

La ecuación (1.2) sugiere que definamos la probabilidad de un suceso a través de su frecuencia relativa a largo plazo. Existen problemas para desarrollar una teoría matemática de la probabilidad usando esta definición de probabilidad. En primer lugar, no está claro cuándo y en qué sentido matemático existe el límite de la ecuación (1.2). En segundo lugar, no podemos realizar un experimento un número infinito de veces, por lo que nunca se podrán conocer con exactitud las

probabilidades p_k . Por último, el uso de la frecuencia relativa para definir la probabilidad excluiría la posibilidad de aplicar la teoría de la probabilidad en situaciones en las que un experimento no puede repetirse. Por lo tanto, tiene sentido práctico desarrollar una teoría matemática de la probabilidad que no esté vinculada a ninguna aplicación particular o a ninguna noción particular de lo que la probabilidad significa. Por otro lado, debemos insistir en que, cuando sea apropiado, la teoría debe permitirnos usar nuestra intuición e interpretar la probabilidad como una frecuencia relativa.

Con el fin de ser coherente con la interpretación de frecuencia relativa, cualquier definición de la “probabilidad de un suceso” debe satisfacer las propiedades de las ecuaciones (1.3) – (1.5). La teoría moderna de la probabilidad se inicia con la construcción de un conjunto de axiomas que especifican que las asignaciones de probabilidad deben satisfacer estas propiedades. Se supone que: (1) se ha definido un experimento de forma aleatoria, y se ha identificado un conjunto S de todos los resultados posibles; (2) se ha especificado una clase de subconjuntos de S denominados sucesos, y (3) a cada suceso A se le ha asignado un número, $P[A]$, de tal manera que se satisfagan los siguientes axiomas:

1. $0 \leq P[A] \leq 1$
2. $P[S] = 1$
3. Si A y B son sucesos que no pueden ocurrir simultáneamente, entonces $P[A \text{ o } B] = P[A] + P[B]$.

La correspondencia entre los tres axiomas y las propiedades de la frecuencia relativa formuladas en las ecuaciones (1.3) – (1.5), es evidente. Estos tres axiomas conducen a muchos resultados útiles y de gran alcance. De hecho, el resto de este curso desarrolla muchos de estos resultados.

Observemos que la teoría de la probabilidad no se ocupa de cómo se obtienen las probabilidades o lo que significan. Cualquier asignación de probabilidades a sucesos que satisfaga los axiomas anteriores es legítima. Es responsabilidad del usuario de la teoría, el constructor del modelo, el determinar cuál debería ser la asignación de probabilidad y que interpretación de la probabilidad tiene sentido en cualquier aplicación dada.

1.3.4 Construcción de un modelo de probabilidad

Veamos cómo pasar de un problema del mundo real que conlleve aleatoriedad a un modelo de probabilidad para dicho problema. La teoría requiere que identifiquemos los elementos que intervienen en los axiomas anteriores. Esto implica (1) la definición del experimento aleatorio inherente a la aplicación, (2) la especificación del conjunto S de todos los posibles resultados y los sucesos de interés, y (3) la especificación de una asignación de probabilidad a partir de la cual se puedan calcular todas las probabilidades de los sucesos de interés. El desafío es desarrollar el modelo más simple que explique todos los aspectos relevantes del problema del mundo real.

A modo de ejemplo, supongamos que ponemos a prueba una conversación telefónica para determinar si la persona que habla está efectivamente hablando o en silencio. Sabemos que en promedio un interlocutor está activo solamente $1/3$ del tiempo, el resto del tiempo está escuchando al otro interlocutor o haciendo una pausa entre las palabras y frases que utiliza. Podemos modelar esta situación física como un experimento de una urna en la que se selecciona una bola de una urna que contiene dos bolas blancas (silencio) y una bola de color negro (voz activa). Estamos haciendo una gran simplificación, no todos los interlocutores son iguales, no todas las lenguas tienen el mismo comportamiento silencio - actividad, etcétera. La utilidad y el potencial de esta simplificación se hace evidente cuando empezamos a hacernos preguntas que surgen durante el diseño del sistema, tales como: ¿Cuál es la probabilidad de que más de 24 interlocutores de entre 48 interlocutores independientes estén activos al mismo tiempo? Esta pregunta es equivalente a: ¿Cuál es la probabilidad de que más de 24 bolas de color negro negro sean seleccionados en 48 repeticiones independientes del experimento de la urna? Al final del capítulo 2 seremos capaces de responder a esta última pregunta, y a todos los problemas del mundo real que se pueden reducir a ella.

1.4 Un ejemplo detallado: un sistema de transmisión de paquetes de voz

En el comienzo de este capítulo hemos afirmado que los modelos de probabilidad proporcionan una herramienta que permite al diseñador modelar con éxito sistemas que deben operar en un entorno estocástico, mientras que, por otra parte, sean eficientes, fiables y rentables. En esta sección, se presenta un ejemplo detallado de tal sistema. Nuestro objetivo es convencernos del poder y la utilidad de la teoría de la probabilidad. La presentación intencionalmente hace uso de la intuición. Muchos de los pasos que ahora pueden parecer poco rigurosos se precisarán más adelante.

Supongamos que a un sistema de comunicación se le exige transmitir simultáneamente 48 conversaciones desde el lugar A al lugar B usando “paquetes” de información de voz. Las palabras de cada interlocutor se convierten en formas de onda de voltaje que primero se digitalizan (es decir, se convierten en una secuencia de números binarios), y luego se agrupan en paquetes de información que corresponden a segmentos de habla de 10 milisegundos (ms). Una dirección fuente y una dirección de destino se añaden a cada paquete de voz antes de su transmisión (ver figura 1.5).

El diseño más simple para el sistema de comunicación transmitiría 48 paquetes cada 10 ms en cada dirección. Sin embargo, se trata de un diseño ineficiente, ya que se sabe que en promedio aproximadamente $2/3$ de todos los paquetes contienen silencio y, por lo tanto, no hay información de voz. Dicho de otra forma, en promedio, los 48 interlocutores sólo producen alrededor de $48/3 = 16$

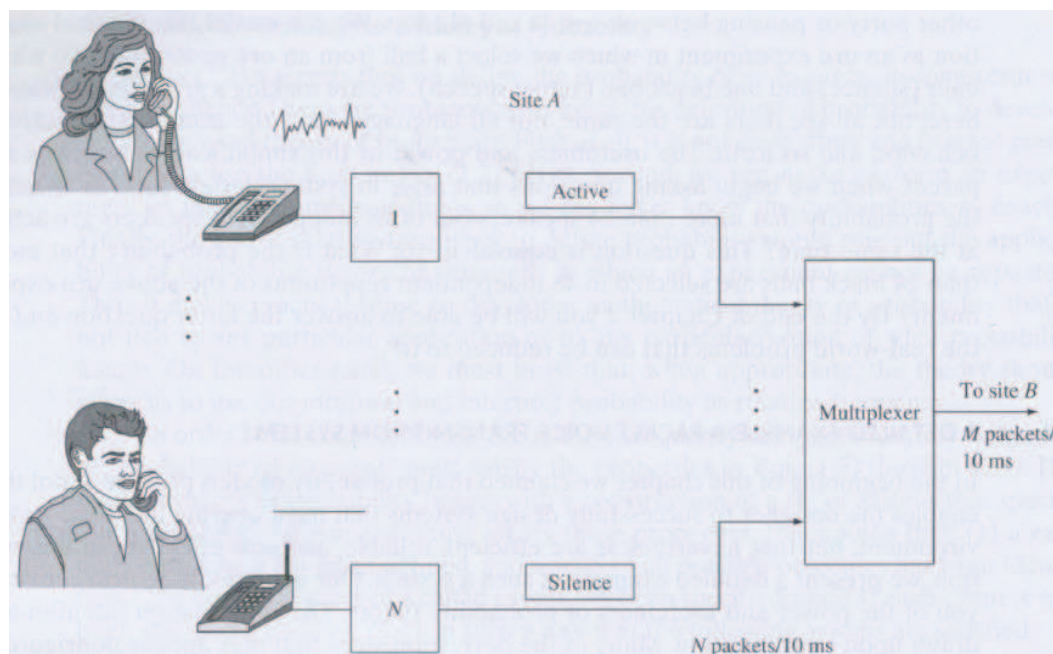


Figura 1.5 Un sistema de transmisión de paquetes de voz.

paquetes activos (no silenciosos) por cada período de 10 ms. Por consiguiente, consideraremos otro sistema que sólo transmita $M < 48$ paquetes cada 10 ms.

Cada 10 ms, el nuevo sistema determina qué interlocutores han producido paquetes con voz activa. Supongamos que A es el resultado de este experimento aleatorio, el número de paquetes activos producidos en un determinado segmento de 10 ms. La cantidad A toma valores que oscilan entre 0 (todos los interlocutores en silencio) a 48 (todos los interlocutores activos). Si $A \leq M$, entonces se transmiten todos los paquetes activos. Sin embargo, si $A > M$, entonces el sistema es incapaz de transmitir todos los paquetes activos, por lo que $A - M$ de los paquetes activos son seleccionados al azar y se desechan. El descarte de los paquetes activos da como resultado la pérdida de voz, por lo que nos gustaría mantener la fracción de los paquetes activos descartados a un nivel que los interlocutores no encuentren objetable.

En primer lugar consideremos la frecuencia relativa de A . Supongamos que el experimento anterior se repite n veces. Sea $A(j)$ el resultado en el j -ésimo ensayo. Sea $N_k(n)$ el número de ensayos para el que el número de paquetes activos es k . La frecuencia relativa del resultado k en los primeros n ensayos es entonces $f_k(n) = N_k(n)/n$, que suponemos converge a una probabilidad p_k :

$$\lim_{n \rightarrow \infty} f_k(n) = p_k \quad 0 \leq k \leq 48. \quad (1.6)$$

En el capítulo 2 se obtendrá la probabilidad p_k de que k interlocutores estén

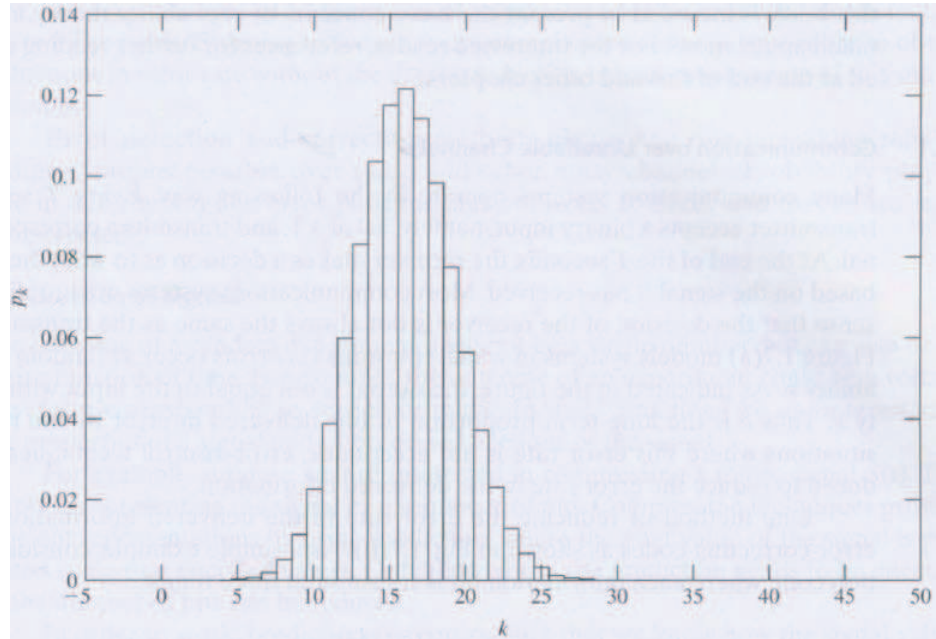


Figura 1.6 Probabilidades para el número de altavoces activos en un grupo de 48.

activos. La figura 1.6 muestra p_k frente a k . Se puede observar que el número más frecuente de interlocutores activos es de 16 y que el número de interlocutores activos es rara vez superior a 24 o más.

Consideremos ahora el ratio de producción de los paquetes activos. El número medio de paquetes activos producidos en cada intervalo de 10 ms está dado por la media muestral del número de paquetes activos:

$$\langle A \rangle_n = \frac{1}{n} \sum_{j=1}^n A(j) \quad (1.7)$$

$$= \sum_{j=1}^n k N_k(n). \quad (1.8)$$

La primera expresión agrega el número de paquetes activos producidos en los primeros n ensayos en el orden en que las observaciones fueron registradas. La segunda expresión contabiliza cuantas de estas observaciones tienen k paquetes activos para cada valor posible de k , y luego calcula el total. A medida que n crece, el ratio $N_k(n)/n$ en la segunda expresión se aproxima a p_k . Así, el número medio de paquetes activos producidos en cada segmento de 10 ms se aproxima a

$$\langle A \rangle_n \rightarrow \sum_{k=0}^{48} k p_k \triangleq E[A]. \quad (1.9)$$

La expresión en el lado derecho se definirá como el valor esperado de A en la

sección 3.3. $E[A]$ está completamente determinada por las probabilidades p_k y en el capítulo 3 se mostrará que $E[A] = 48 \times 1/3 = 16$. La ecuación (1.9) establece que el promedio a largo plazo del número de paquetes activos producidos por periodo de 10 ms es de $E[A] = 16$ interlocutores cada 10 ms.

La información proporcionada por las probabilidades p_k nos permite diseñar sistemas que son eficientes y que ofrecen una buena calidad de voz. Por ejemplo, podemos reducir la capacidad de transmisión a la mitad, a 24 paquetes por período de 10 ms, mientras descartamos un número imperceptible de paquetes activos.

Vamos a resumir lo que hemos hecho en esta sección. Hemos presentado un ejemplo en el que el comportamiento del sistema es intrínsecamente aleatorio, y en el que las medidas de rendimiento del sistema se expresan en términos de promedios a largo plazo. Hemos mostrado cómo estas medidas a largo plazo conducen a expresiones que involucran las probabilidades de los distintos resultados. Finalmente, hemos indicado que, en algunos casos, la teoría de probabilidad nos permite obtener estas probabilidades. Por tanto, somos capaces de predecir los promedios a largo plazo de varias cantidades de interés y proceder con el diseño del sistema.

1.5 Otros ejemplos

En esta sección se presentan ejemplos adicionales de las ingenierías en informática y de telecomunicaciones, donde los modelos de probabilidad se utilizan para diseñar sistemas que funcionan en un entorno aleatorio. Nuestra intención aquí es mostrar cómo las probabilidades y los promedios a largo plazo surgen de forma natural como medidas de rendimiento en muchos sistemas. No obstante, este curso tiene como objetivo presentar los conceptos básicos de teoría de la probabilidad y no sus aplicaciones con un elevado grado de detalle.

1.5.1 Comunicación a través de canales no fiables

Muchos de los sistemas de comunicación funcionan de la siguiente manera. Cada T segundos, el transmisor acepta una entrada binaria, es decir, un 0 o un 1, y transmite una señal correspondiente. Al final de los T segundos, el receptor toma una decisión acerca de cuál es la entrada, basada en la señal que ha recibido. La mayoría de los sistemas de comunicación no son fiables en el sentido de que la decisión del receptor no siempre coincide con la entrada del transmisor. La figura 1.7(a) modeliza sistemas en los que los errores de transmisión ocurren al azar con una probabilidad ε . Como se indica en la figura, la salida no es igual a la entrada con una probabilidad ε . Por tanto, ε es la proporción a largo plazo de bits erróneamente enviados por el receptor. En aquellas situaciones en las que este porcentaje de error no es aceptable, se introducen técnicas de control de error para reducir la tasa de error en la información suministrada.

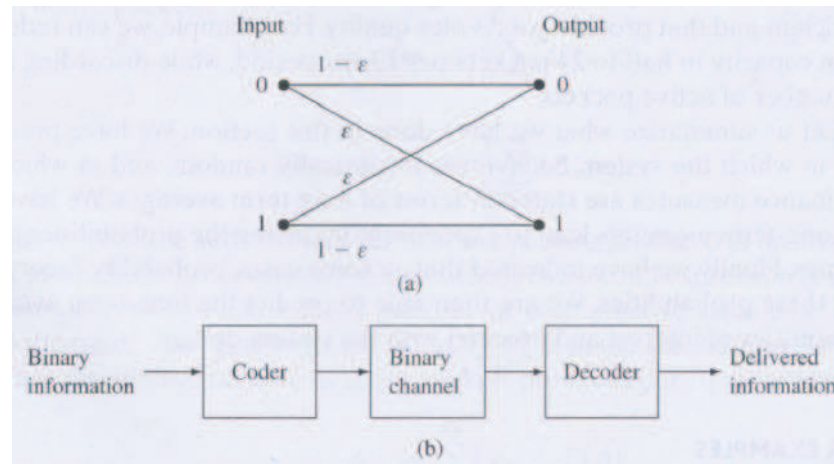


Figura 1.7 (a) Un modelo de canal de comunicación binario. (b) Sistema de control de error.

Un método para reducir la tasa de error en la información enviada es utilizar códigos correctores de errores, como se muestra en la figura 1.7(b). Como un simple ejemplo, consideremos un código de repetición en el que cada bit de información se transmite tres veces:

$$0 \rightarrow 000$$

$$1 \rightarrow 111$$

Si suponemos que el decodificador toma una decisión sobre el bit de información mediante la adopción de un voto por mayoría de entre los tres bits de respuesta del receptor, el decodificador tomará una decisión equivocada sólo si dos o tres de los bits son erróneos. En el ejemplo 2.28, se muestra que esto ocurre con probabilidad $3\epsilon^2 - 2\epsilon^3$. Por tanto, si el BER (Bit Error Ratio, la tasa de error de recepción de bits) del canal sin codificación es de 10^{-3} , entonces el error de envío de bits con el sencillo código anterior será 3×10^{-3} , una reducción de tres órdenes de magnitud. Esta mejora, sin embargo, tiene un coste: La tasa de transmisión de información se ha ralentizado a 1 bit cada $3T$ segundos. Usando códigos más largos y complicados, es posible obtener reducciones en la tasa de error sin esta drástica reducción en la tasa de transmisión en este sencillo ejemplo.

La detección de errores y los métodos de corrección juegan un papel clave a la hora de hacer posible comunicaciones fiables a través de la radio y de otros canales ruidosos. La probabilidad juega un papel en la determinación de los patrones de error que probablemente ocurran y que, por lo tanto, deben ser corregidos.

1.5.2 Compresión de señales

El resultado de un experimento aleatorio no tiene por qué ser un número único, sino que puede ser también una función dependiente del tiempo. Por ejemplo, el resultado de un experimento puede ser una forma de onda de tensión que corresponde a una conversación o a música. En estas situaciones nos interesan las propiedades de una señal y de versiones procesadas de dicha señal.

Por ejemplo, supongamos que estamos interesados en comprimir una señal musical $S(t)$. Ello implica representar la señal mediante una secuencia de bits. Las técnicas de compresión proporcionan representaciones eficientes mediante el uso de predicción, donde se predice el siguiente valor de la señal usando valores pasados codificados. Sólo se necesita codificar el error en la predicción de manera que se pueda reducir el número de bits.

Para que funcionen, los sistemas de predicción requieren que sepamos cómo están correlados entre sí los valores de la señal. Dada esta estructura de correlación se pueden diseñar sistemas de predicción óptimos. La probabilidad juega un papel clave en la solución de estos problemas. Los sistemas de compresión han tenido un gran éxito y se encuentran en los teléfonos móviles, cámaras digitales y videocámaras.

1.5.3 Fiabilidad de sistemas

La fiabilidad es un cometido principal en el diseño de los sistemas modernos. Un buen ejemplo es el sistema de computadores y las redes de comunicación que permiten la transferencia electrónica de fondos entre bancos. Es de suma importancia que este sistema siga funcionando incluso en caso de fallos en el subsistema. La pregunta clave es, ¿cómo se construyen sistemas fiables a partir de componentes no fiables? Los modelos de probabilidad nos proporcionan las herramientas para abordar esta cuestión de forma cuantitativa.

El funcionamiento de un sistema requiere el funcionamiento de algunas o todas sus componentes. Por ejemplo, la figura 1.8(a) muestra un sistema que funciona sólo cuando todas sus componentes están funcionando, y la figura 1.8(b) muestra un sistema que funciona siempre y cuando al menos una de sus componentes está funcionando. Los sistemas más complejos se pueden obtener como combinaciones de estas dos configuraciones básicas.

Todos sabemos por experiencia que no es posible predecir con exactitud cuándo una componente fallará. La teoría de la probabilidad nos permite evaluar medidas de seguridad tales como el tiempo medio transcurrido hasta la ocurrencia del fallo y la probabilidad de que una componente siga funcionando después de que cierto tiempo haya transcurrido. Además, veremos en los capítulos 2 y 4 que la teoría de probabilidad nos permite determinar estos promedios y las probabilidades para todo un sistema en términos de los promedios y las probabilidades de sus componentes. Esto nos permite evaluar las configuraciones del sistema en

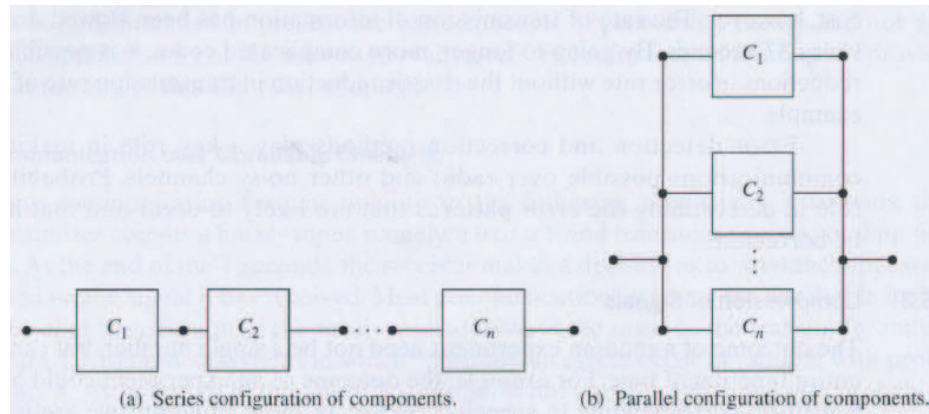


Figura 1.8 Sistemas con n componentes.

términos de su fiabilidad y, por lo tanto, seleccionar los diseños del sistema que son fiables.

1.5.4 Sistemas con uso compartido de recursos

Muchas aplicaciones implican compartir recursos que están sujetos a una demanda inestable y aleatoria. Los clientes intercalan sus demandas durante períodos cortos de servicio entre períodos relativamente largos de inactividad. Las demandas de los clientes pueden satisfacerse mediante la dedicación de recursos suficientes para cada cliente, pero este enfoque puede resultar un derroche porque los recursos no se utilizan cuando un cliente está inactivo. Un mejor enfoque consiste en configurar sistemas en donde las exigencias de los clientes se cumplen a través del intercambio dinámico de los recursos.

Por ejemplo, muchos sistemas servidores web funcionan como se muestra en la figura 1.9. Estos sistemas permiten que c clientes se conecten a un servidor en un momento dado. Los clientes envían consultas al servidor. La consulta se coloca en una línea de espera y luego es procesada por el servidor. Después de recibir la respuesta del servidor, cada cliente invierte un tiempo pensando antes de hacer la siguiente consulta. El sistema cierra una conexión de un cliente existente después de un período de tiempo de espera, y lo reemplaza con un nuevo cliente.

El sistema debe estar configurado para proporcionar respuestas rápidas a los clientes, con el fin de evitar el cierre prematuro de las conexiones y de utilizar los recursos informáticos de manera efectiva. Esto requiere la caracterización probabilística del tiempo de procesamiento de las consultas, del número de clics por cada conexión, y el tiempo entre clics (el tiempo para pensar). Estos parámetros se utilizan entonces para determinar el valor óptimo de c , así como el valor de tiempo de espera.

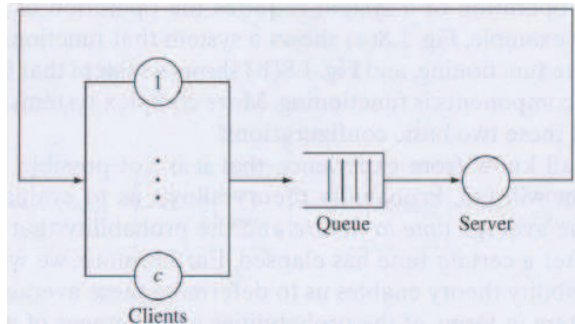


Figura 1.9 Un modelo simple para un sistema servidor web.

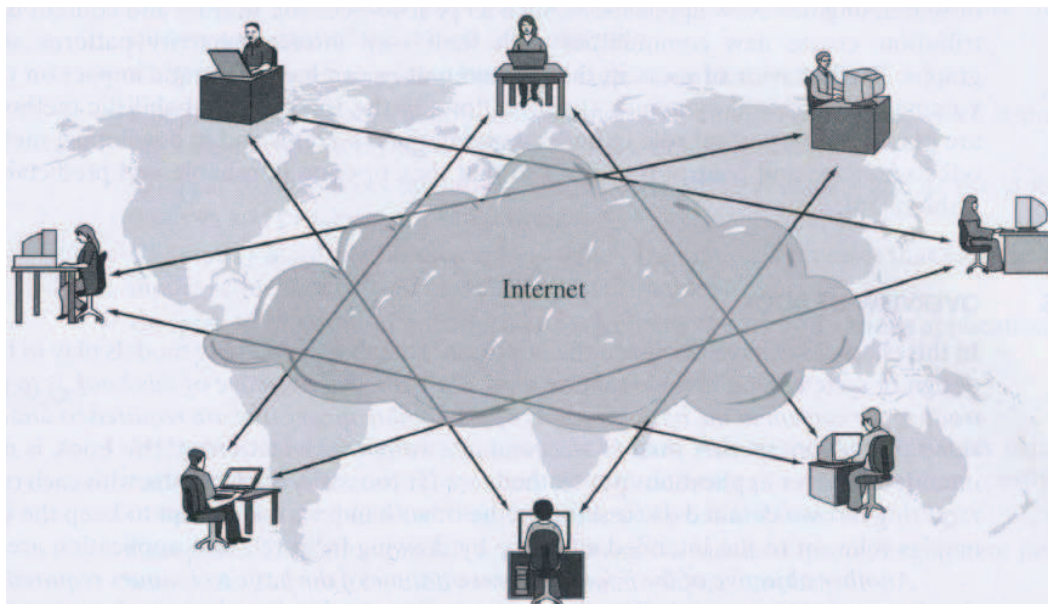


Figura 1.10 Un gran comunidad de usuarios interactuando a través de internet.

1.5.5 Sistemas a escala internet

Uno de los desafíos más importantes en la actualidad es el diseño de sistemas a escala internet dado que los sistemas cliente-servidor de la figura 1.9 evolucionan hacia sistemas masivamente distribuidos, como en la figura 1.10. En estos nuevos sistemas el número de usuarios que están conectados al mismo tiempo puede estar en las decenas de miles de personas y en el caso de los sistemas peer-to-peer (P2P, red de pares o red punto a punto) en millones.

Las interacciones entre los usuarios de internet son mucho más complejas que las de clientes que acceden a un servidor. Por ejemplo, los enlaces en las páginas web que apuntan a otras páginas web crean una vasta red de documentos conectados entre sí. El desarrollo de técnicas de grafos y de mapeo para representar estas relaciones lógicas es clave para entender el comportamiento de los usua-

rios. Se ha desarrollado una gran variedad de técnicas de rastreo de la web para producir tales grafos. Las técnicas probabilísticas pueden evaluar la importancia relativa de los nodos en estos grafos y, de hecho, juegan un papel central en el funcionamiento de los motores de búsqueda. Nuevas aplicaciones, tales como el intercambio de archivos peer-to-peer y la distribución de contenidos, crean nuevas comunidades con sus propios patrones de interconexión y de grafos. El comportamiento de los usuarios en estas comunidades puede tener un impacto enorme en el volumen, los patrones y la dinámica de los flujos de tráfico en internet. Los métodos probabilísticos están jugando un papel importante en la comprensión de estos sistemas y en la elaboración de métodos para gestionar y controlar los recursos de tal manera que operen de forma fiable y predecible.

Problemas

1.1 Se consideran los tres experimentos aleatorios siguientes:

Experimento 1: Lanzar una moneda al aire.

Experimento 2: Lanzar un dado.

Experimento 3: Seleccionar una bola al azar de una urna que contiene bolas numeradas del 0 al 9.

- Especifica el espacio muestral de cada experimento.
- En cada uno de los experimentos anteriores, encuentra la frecuencia relativa de cada resultado en un gran número de repeticiones del experimento. Explica la respuesta.

1.2 Explica cómo los siguientes experimentos son equivalentes a experimentos aleatorios con urnas:

- Lanzar al aire una moneda equilibrada dos veces.
- Tirar un par de dados equilibrados.
- Robar dos cartas de una baraja de 52 cartas distintas, con reemplazamiento después de la primera extracción; sin reemplazamiento, después de la primera extracción.

1.3 Explica en qué condiciones los siguientes experimentos son equivalentes al lanzamiento aleatorio de una moneda. ¿Cuál es la probabilidad de obtener cara en el experimento?

- Observar un píxel (punto) en un documento escaneado en blanco y negro.
- Recibir una señal binaria en un sistema de comunicación.
- Comprobar si un dispositivo está funcionando.
- Determinar si tu amigo Juan está en línea.
- Determinar si se ha producido un bit error (error de envío de bit) en una transmisión a través de un canal de comunicación ruidoso.

1.4 Una urna contiene tres bolas etiquetadas por vía electrónica, con etiquetas 00, 01 y 10. Se le pide a Lisa, Homer y Bart que caractericen el experimento aleatorio que consiste en seleccionar una bola al azar y leer la etiqueta. El lector de etiquetas de Lisa funciona bien, el lector de Homer tiene la etiqueta del dígito más significativo (el que está más a la izquierda) estancado en 1; la etiqueta del dígito menos significativo del lector de Bart está estancado en 0.

- (a) ¿Cuál es el espacio muestral determinado por Lisa, Homer y Bart?
 (b) ¿Cuáles son las frecuencias relativas observadas por Lisa, Homer y Bart en un gran número de repeticiones del experimento?

1.5 Un experimento aleatorio tiene como espacio muestral $S = \{1, 2, 3, 4\}$ con probabilidades $p_1 = 1/2$, $p_2 = 1/4$, $p_3 = 1/8$ y $p_4 = 1/8$.

- (a) Describe cómo se puede simular este experimento aleatorio usando lanzamientos de una moneda equilibrada.
 (b) Describe cómo se puede simular este experimento aleatorio mediante un experimento con una urna.
 (c) Describe cómo se puede simular este experimento con una baraja de 52 cartas distintas.

1.6 Un experimento consiste en seleccionar al azar dos bolas sucesivamente de una urna que contiene dos bolas negras y una bola blanca.

- (a) Especifica el espacio muestral para este experimento.
 (b) Supongamos que el experimento se ha modificado para que la bola inmediatamente se vuelva a poner en la urna después de la primera selección. ¿Cuál es el espacio muestral ahora?
 (c) ¿Cuál es la frecuencia relativa del resultado (blanco, blanco) para un gran número de repeticiones del experimento del apartado (a)? ¿Y en el caso (b)?
 (d) ¿El resultado de la segunda extracción depende de alguna manera del resultado de la primera extracción en alguno de estos dos experimentos?

1.7 Sea A un suceso asociado a los resultados de un experimento aleatorio, y sea B el suceso definido como “el suceso A no ocurre.” Demuestra que $f_B(n) = 1 - f_A(n)$.

1.8 Sean A , B y C sucesos que no pueden ocurrir simultáneamente por pares o tríos, y sea D el suceso “ A o B o C ocurren.” Demuestra que

$$f_D(n) = f_A(n) + f_B(n) + f_C(n).$$

1.9 La media muestral de una serie de resultados numéricos $X(1), X(2), \dots, X(n)$ de una secuencia de experimentos aleatorios se define como

$$\langle X \rangle_n = \frac{1}{n} \sum_{j=1}^n X(j).$$

Demuestra que la media muestral cumple la fórmula de recurrencia:

$$\langle X \rangle_n = \langle X \rangle_{n-1} + \frac{X(n) - \langle X \rangle_{n-1}}{n}, \quad \langle X \rangle_0 = 0.$$

1.10 Supongamos que se realiza un muestreo de la señal $2 \cos 2\pi t$ en instantes de tiempo aleatorios.

- (a) Encuentra la media muestral a largo plazo.

- (b) Encuentra la frecuencia relativa a largo plazo de los sucesos “la tensión es positiva”; “la tensión es inferior a -2 ”.
- (c) ¿Cambian las respuestas de las partes (a) y (b), si los tiempos de muestreo son periódicos y se toman cada τ segundos?

1.11 Con el fin de generar una secuencia aleatoria de números aleatorios se toma una columna de números telefónicos y se anota un “0” si el último dígito del número de teléfono es par y un “1” si el dígito es impar. Discute cómo se podría determinar si la secuencia resultante de números es efectivamente “aleatoria”. ¿Qué prueba aplicarías a las frecuencias relativas de los resultados individuales? ¿Y a pares de resultados?

2 Conceptos básicos de teoría de la probabilidad

En este capítulo se presentan los conceptos básicos de la teoría de la probabilidad. En general, en el resto de los capítulos, se seguirán desarrollando o elaborando los conceptos básicos presentados aquí. Se estará bien preparado para tratar con el resto del manual, si se tiene una buena comprensión de estos conceptos básicos cuando se complete este capítulo.

Se presentan los siguientes conceptos básicos. En primer lugar, se utiliza la teoría de conjuntos para especificar el espacio muestral y los sucesos de un experimento aleatorio. En segundo lugar, los axiomas de la probabilidad especifican las reglas para el cálculo de las probabilidades de sucesos. En tercer lugar, la noción de probabilidad condicional permite determinar cómo la información parcial sobre los resultados de un experimento afecta a las probabilidades de los sucesos. La probabilidad condicional también permite formular el concepto de “independencia” de sucesos y experimentos. Por último, se consideran experimentos aleatorios “secuenciales” que consisten en realizar una serie de subexperimentos aleatorios simples. Se muestra cómo las probabilidades de sucesos en estos experimentos se pueden derivar de las probabilidades de los subexperimentos simples. A lo largo del manual se demuestra que los experimentos aleatorios complejos pueden ser analizados mediante la descomposición de los mismos en subexperimentos simples.

2.1 Especificación de experimentos aleatorios

Un experimento aleatorio es un experimento en el que el resultado varía de una manera impredecible cuando el experimento se repite en las mismas condiciones. Un experimento aleatorio se especifica al establecer un procedimiento experimental y un conjunto de una o más mediciones y observaciones.

Ejemplo 2.1

Experimento E_1 : Seleccionar una bola de una urna que contiene bolas numeradas del 1 al 50. Anotar el número de la bola.

Experimento E_2 : Seleccionar una bola de una urna que contiene bolas numeradas del 1 al 4. Suponer que las bolas numeradas con 1 y 2 son de color negro y que

las bolas numeradas con 3 y 4 son de color blanco. Anotar el número y el color de la bola que se selecciona.

Experimento E_3 : Lanzar una moneda tres veces y anotar la secuencia de caras y cruces.

Experimento E_4 : Lanzar una moneda tres veces y anotar el número de caras.

Experimento E_5 : Contar el número de paquetes de voz, producido por un grupo de N interlocutores en un período de 10 ms, que contienen sólo silencio.

Experimento E_6 : Un bloque de información se transmite repetidamente a través de un canal ruidoso hasta que un bloque sin errores llega al receptor. Contar el número de transmisiones necesarias.

Experimento E_7 : Elegir al azar un número entre cero y uno.

Experimento E_8 : Medir el tiempo transcurrido entre solicitudes de páginas en un servidor web.

Experimento E_9 : Medir el tiempo de vida de un chip de memoria de un ordenador dado en un entorno determinado.

Experimento E_{10} : Determinar el valor de una señal de audio en un instante de tiempo t_1 .

Experimento E_{11} : Determinar el valor de una señal de audio en los instantes de tiempo t_1 y t_2 .

Experimento E_{12} : Elegir al azar dos números entre cero y uno.

Experimento E_{13} : Escoger al azar un número X entre cero y uno, después escoger al azar un número Y entre cero y X .

Experimento E_{14} : Una componente de un sistema se instala en el instante de tiempo $t = 0$. Sea $X(t) = 1$ para $t \geq 0$ siempre y cuando el componente esté funcionando, y sea $X(t) = 0$ para cualquier instante de tiempo considerado después de que la componente falle.

La especificación de un experimento aleatorio debe incluir una exposición inequívoca de lo que exactamente está medido u observado. Por ejemplo, los experimentos aleatorios pueden constar del mismo procedimiento, pero diferir en las observaciones realizadas, como se ilustra en los experimentos E_3 y E_4 .

Un experimento aleatorio puede involucrar más de una medición u observación, como se ilustra en los experimentos E_2 , E_3 , E_{11} , E_{12} y E_{13} . Un experimento aleatorio incluso puede implicar una continuidad de mediciones, como se muestra en el experimento E_{14} .

Los experimentos E_3 , E_4 , E_5 , E_6 , E_{12} y E_{13} son ejemplos de experimentos secuenciales que pueden ser vistos como una sucesión de subexperimentos simples. ¿Podemos identificar los subexperimentos en cada uno de ellos? Observemos que en el experimento E_{13} el segundo subexperimento depende del resultado del primer subexperimento.

2.1.1 El espacio muestral

Dado que los experimentos aleatorios no siempre producen el mismo resultado, es necesario determinar el conjunto de posibles resultados. Definimos un resultado o elemento muestral de un experimento aleatorio como un resultado que no se puede descomponer en otros resultados. Cuando realizamos un experimento aleatorio, se produce un único resultado. Así, los resultados son mutuamente excluyentes en el sentido de que no pueden ocurrir simultáneamente. El espacio muestral S de un experimento aleatorio se define como el conjunto de todos los posibles resultados.

Denotaremos mediante ζ un resultado o elemento muestral de un experimento, donde ζ es un elemento o punto de S . Cada realización de un experimento aleatorio puede ser visto como la selección al azar de un único punto (resultado) de S .

El espacio muestral S se puede especificar de forma compacta mediante el uso de la notación de conjuntos. Se puede visualizar mediante tablas, diagramas, intervalos de la recta real, o regiones del plano. Hay dos formas básicas para especificar un conjunto:

1. Listar todos los elementos, separados por comas, dentro de un par de llaves:

$$A = \{0, 1, 2, 3\},$$

2. Dar una propiedad que especifique los elementos del conjunto:

$$A = \{x : x \text{ es un entero tal que } 0 \leq x \leq 3\}.$$

Observemos que el orden en el que se enumeran los elementos no cambia el conjunto, por ejemplo, $\{0, 1, 2, 3\}$ y $\{1, 2, 3, 0\}$ son el mismo conjunto.

Los espacios muestrales correspondientes a los experimentos del ejemplo 2.1 se muestran a continuación utilizando notación de conjuntos:

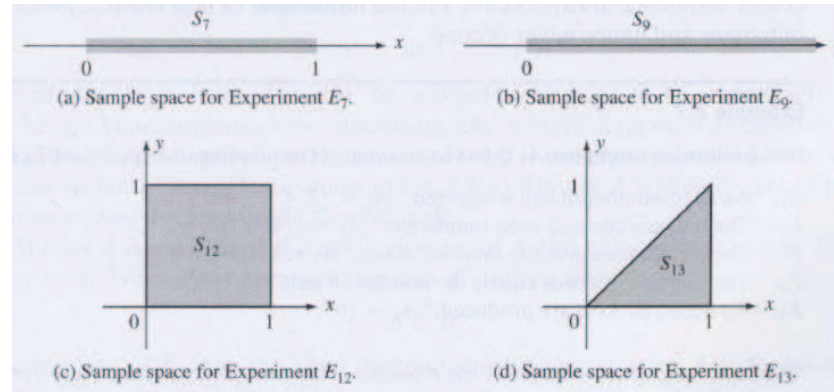


Figura 2.1 Espacios muestrales de los experimentos E_7 , E_9 , E_{12} y E_{13} .

Ejemplo 2.2

$$S_1 = \{1, 2, \dots, 50\}$$

$$S_2 = \{(1, n), (2, n), (3, b), (4, b)\}$$

$$S_3 = \{CCC, CCX, CXC, XCC, XXC, XCX, CXX, XXX\}$$

$$S_4 = \{0, 1, 2, 3\}$$

$$S_5 = \{0, 1, 2, \dots, N\}$$

$$S_6 = \{1, 2, 3, \dots\}$$

$$S_7 = \{x : 0 \leq x \leq 1\} = [0, 1] \text{ ver figura 2.1(a)}$$

$$S_8 = \{t : t \geq 0\} = [0, \infty]$$

$$S_9 = \{t : t \geq 0\} = [0, \infty] \text{ ver figura 2.1(b)}$$

$$S_{10} = \{v : -\infty \leq v \leq \infty\} = (-\infty, \infty)$$

$$S_{11} = \{(v_1, v_2) : -\infty \leq v_1 \leq \infty \text{ y } -\infty \leq v_2 \leq \infty\}$$

$$S_{12} = \{(x, y) : 0 \leq x \leq 1 \text{ y } 0 \leq y \leq 1\} \text{ ver figura 2.1(c)}$$

$$S_{13} = \{(x, y) : 0 \leq y \leq x \leq 1\} \text{ ver figura 2.1(d)}$$

$$S_{14} = \text{conjunto de funciones } X(t) \text{ para las cuales } X(t) = 1 \text{ para } 0 \leq t \leq t_0 \text{ y}$$

$$X(t) = 0 \text{ para } t \geq t_0, \text{ donde } t_0 > 0 \text{ es el instante de tiempo en el que}$$

la componente falla

Los experimentos aleatorios que involucran el mismo procedimiento experimental pueden tener diferentes espacios muestrales, como se muestra en los ex-

perimentos E_3 y E_4 . Por tanto, el propósito de un experimento influye en la elección del espacio muestral.

Hay tres posibilidades para el número de resultados en un espacio muestral. Un espacio muestral puede ser finito, infinito numerable o infinito no numerable. Diremos que S es un espacio muestral discreto si S es numerable, es decir, sus resultados se pueden poner en correspondencia uno a uno con los enteros positivos. Diremos que S es un espacio muestral continuo si S es no numerable. Los experimentos E_1, E_2, E_3, E_4 y E_5 tienen espacios muestrales discretos finitos. El experimento E_6 tiene un espacio muestral discreto infinito numerable. Los experimentos E_7 al E_{13} tienen espacios muestrales continuos.

Dado que el resultado de un experimento puede consistir en una o más observaciones o mediciones, el espacio muestral S puede ser multi-dimensional. Por ejemplo, los resultados en los experimentos E_2, E_{11}, E_{12} y E_{13} son bidimensionales, y los del experimento E_3 son tridimensionales. En algunos casos, el espacio muestral se puede escribir como el producto cartesiano de otros conjuntos. Por ejemplo, $S_{11} = \mathbb{R} \times \mathbb{R}$, donde \mathbb{R} es el conjunto de los números reales, y $S_3 = S \times S \times S$, donde $S = \{C, X\}$.

A veces es conveniente dejar que el espacio muestral incluya resultados que sean imposibles. Por ejemplo, en el experimento E_9 es conveniente definir el espacio muestral como la recta real positiva, a pesar de que un dispositivo no puede tener una duración infinita.

2.1.2 Sucesos

Habitualmente no estamos interesados en la ocurrencia de unos resultados específicos, sino más bien en la ocurrencia de algún evento o suceso (es decir, si el resultado satisface ciertas condiciones). Esto requiere que consideramos subconjuntos de S . Diremos que A es un subconjunto de B si cada elemento de A pertenece también a B . Por ejemplo, en el experimento E_{10} , que involucra la medición de una tensión, podríamos estar interesados en el suceso “la señal de tensión es negativa”. Las condiciones de interés definen un subconjunto del espacio muestral, es decir, el conjunto de puntos ζ de S que satisfacen las condiciones dadas. Por ejemplo, “la tensión es negativa” se corresponde con el conjunto $\{\zeta : -\infty < \zeta < 0\}$. El suceso ocurre si y sólo si el resultado del experimento ζ está en este subconjunto. Por esta razón, los sucesos se corresponden con subconjuntos de S .

Dos sucesos de especial interés son el suceso seguro, S , formado por todos los resultados y, por lo tanto, ocurre siempre, y el suceso imposible o suceso nulo, \emptyset , que no contiene ningún resultado y, por lo tanto, no ocurre nunca.

Ejemplo 2.3

En los ejemplos siguientes, A_k hace referencia a un suceso correspondiente al experimento E_k del ejemplo 2.1.

- E_1 : “Se selecciona una bola con número par,” $A_1 = \{2, 4, \dots, 48, 50\}$.
- E_2 : “La bola es blanca y con número par,” $A_2 = \{(4, b)\}$.
- E_3 : “Los tres lanzamientos dan el mismo resultado,” $A_3 = \{CCC, XXX\}$.
- E_4 : “El número de caras es igual al número de cruces,” $A_4 = \emptyset$.
- E_5 : “No se han producido paquetes activos,” $A_5 = \{0\}$.
- E_6 : “Se requieren menos de 10 transmisiones,” $A_6 = \{1, \dots, 9\}$.
- E_7 : “El número seleccionado no es negativo,” $A_7 = S_7$.
- E_8 : “Han transcurrido menos de t_0 segundos entre las solicitudes de página,”
 $A_8 = \{t : 0 \leq t < t_0\} = [0, t_0)$.
- E_9 : “El chip ha durado más de 1000 horas pero menos de 1500,” $A_9 = \{t : 1000 < t < 1500\} = (1000, 1500)$.
- E_{10} : “El valor absoluto de la tensión es menor que 1 voltio,” $A_{10} = \{v : -1 < v < 1\} = (-1, 1)$.
- E_{11} : “Las dos tensiones tienen polaridades opuestas,” $A_{11} = \{(v_1, v_2) : (v_1 < 0 \text{ y } v_2 > 0) \text{ ó } (v_1 > 0 \text{ y } v_2 < 0)\}$.
- E_{12} : “Los dos números difieren en menos de $1/10$,” $A_{12} = \{(x, y) : (x, y) \in S_{12} \text{ y } |x - y| < 1/10\}$.
- E_{13} : “Los dos números difieren en menos de $1/10$,” $A_{13} = \{(x, y) : (x, y) \in S_{13} \text{ y } |x - y| < 1/10\}$.
- E_{14} : “El sistema está en funcionamiento en el instante de tiempo t_1 ,” A_{14} = subconjunto de S_{14} para el que $X(t_1) = 1$.

Un suceso puede estar formado por un solo resultado, como en A_2 y A_5 . Un suceso procedente de un espacio muestral discreto que consta de un solo resultado se llama suceso elemental. Los sucesos A_2 y A_5 son sucesos elementales. Un suceso también puede estar formado por todo el espacio muestral, como en A_7 . El suceso nulo, \emptyset , surge cuando ninguno de los resultados satisfacen las condiciones que especifican un suceso determinado, como en A_4 .

2.1.3 Revisión de la teoría de conjuntos

En los experimentos aleatorios nos interesa en particular la ocurrencia de sucesos que están representados por conjuntos. Podemos combinar sucesos usando operaciones de conjuntos para obtener a su vez otros sucesos. También se pueden expresar sucesos complejos como combinaciones de sucesos simples. Antes de proceder con un análisis más detallado de los sucesos y los experimentos aleatorios, presentamos algunas conceptos esenciales de la teoría de conjuntos.

Un conjunto es una colección de objetos que denotaremos mediante letras mayúsculas S, A, B, \dots . Se define U como el conjunto universal a aquel que se compone de todos los posibles objetos de interés en un escenario o aplicación dados. En el contexto de los experimentos aleatorios nos referiremos al conjunto universal como el espacio muestral. Por ejemplo, el conjunto universal en el

experimento E_6 es $U = \{1, 2, \dots\}$. Un conjunto A es una colección de objetos de U , y llamaremos a estos objetos elementos o puntos del conjunto A y que denotaremos con letras minúsculas, ζ, a, b, x, y, \dots . Usaremos la notación:

$$x \in A \quad \text{y} \quad x \notin A$$

para indicar que “ x es un elemento de A ” o “ x no es un elemento de A ”, respectivamente.

Usaremos diagramas de Venn cuando tratemos con conjuntos. Un diagrama de Venn permite la ilustración de los conjuntos y sus interrelaciones. El conjunto universal U se representa generalmente como el conjunto de todos los puntos dentro de un rectángulo, como se muestra en la figura 2.2(a). El conjunto A es entonces el conjunto de puntos dentro de una región cerrada dentro del rectángulo.

Diremos que A es un subconjunto de B si cada elemento de A pertenece también a B , es decir, si $x \in A$ implica que $x \in B$. Diremos que “ A está contenido en B ” y denotaremos:

$$A \subset B$$

Si A es un subconjunto de B , entonces el diagrama de Venn muestra la región A dentro de la región B , como se ve en la figura 2.2(e).

Ejemplo 2.4

En el experimento E_6 tres conjuntos de interés podrían ser $A = \{x : x \geq 10\} = \{10, 11, \dots\}$, es decir, se requieren 10 o más transmisiones; $B = \{2, 4, 6, \dots\}$, el número de transmisiones es un número par; y $C = \{x : x \geq 20\} = \{20, 21, \dots\}$. ¿Cuál de estos conjuntos es subconjunto de los demás?

Está claro que C es un subconjunto de A ($C \subset A$). Sin embargo, C no es un subconjunto de B , y B no es un subconjunto de C , ya que ambos conjuntos contienen elementos que el otro conjunto no contiene. Del mismo modo, B no es un subconjunto de A , y A no es un subconjunto de B .

El conjunto vacío \emptyset se define como el conjunto que no tiene elementos. El conjunto vacío \emptyset es subconjunto de cualquier conjunto, es decir, para cualquier conjunto A , se verifica $\emptyset \subset A$.

Diremos que los conjuntos A y B son iguales si contienen los mismos elementos. Puesto que cada elemento en A también está en B , entonces $x \in A$ implica que $x \in B$, por lo que $A \subset B$. Del mismo modo todos los elementos de B también están en A , por lo que $x \in B$ implica que $x \in A$ y así $B \subset A$. Por lo tanto:

$$A = B \quad \text{si y sólo si} \quad A \subset B \quad \text{y} \quad B \subset A$$

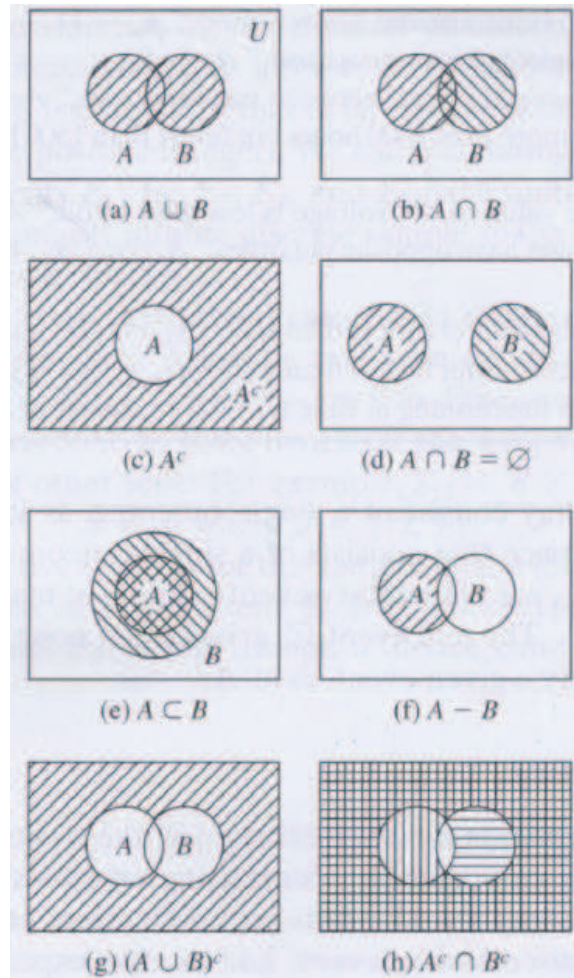


Figura 2.2 Relaciones y operaciones entre conjuntos.

El método estándar para demostrar que dos conjuntos, A y B , son iguales es demostrar que $A \subset B$ y $B \subset A$. Un segundo método consiste en listar todos los elementos de A y todos los elementos de B , y demostrar que todos los elementos son los mismos. Una variante de este segundo método es utilizar un diagrama de Venn para identificar la región que corresponde a A y demostrar entonces que el diagrama de Venn correspondiente a B ocupa la misma región. Veremos ejemplos de ambos métodos más adelante.

Vamos a utilizar tres operaciones básicas entre conjuntos. Las operaciones de unión e intersección se aplican a dos conjuntos y dan como resultado un tercer conjunto. La operación complementario se aplica a un único conjunto y da como resultado otro conjunto.

La unión de dos conjuntos A y B se denota por $A \cup B$ y se define como el

conjunto de resultados que están en A o en B , o en ambos:

$$A \cup B = \{x : x \in A \text{ o } x \in B\}.$$

La operación $A \cup B$ se corresponde con el “o” lógico de las propiedades que definen los conjuntos A y B , es decir, x está en $A \cup B$ si x satisface la propiedad que define a A , o x satisface la propiedad que define a B , o ambas cosas. El diagrama de Venn de $A \cup B$ consiste en la región sombreada de la figura 2.2(a).

La intersección de dos conjuntos A y B se denota por $A \cap B$ y se define como el conjunto de los resultados que están tanto en A como en B :

$$A \cap B = \{x : x \in A \text{ y } x \in B\}.$$

La operación $A \cap B$ se corresponde con el “y” lógico de las propiedades que definen los conjuntos A y B . El diagrama de Venn para $A \cap B$ se compone de la doble región sombreada de la figura 2.2(b). Se dice que dos conjuntos son disjuntos o mutuamente excluyentes si su intersección es el conjunto vacío, $A \cap B = \emptyset$. La figura 2.2(d) muestra dos conjuntos mutuamente excluyentes A y B .

El complementario de un conjunto A se denota por A^c y se define como el conjunto de todos los elementos que no están en A :

$$A^c = \{x : x \notin A\}.$$

La operación A^c se corresponde con el “no” lógico de la propiedad que define el conjunto A . La figura 2.2(c) muestra A^c . Observemos que $S^c = \emptyset$ y $\emptyset^c = S$.

La diferencia de dos conjuntos A y B es el conjunto de elementos de A que no están en B :

$$A - B = \{x : x \in A \text{ y } x \notin B\}.$$

$A - B$ se obtiene mediante la eliminación en A de todos los elementos que también están en B , como se muestra en la figura 2.2(f). Observemos que $A - B = A \cap B^c$. Además, se verifica también que $B^c = S - B$.

Ejemplo 2.5

Sean A, B y C los sucesos del experimento E_6 del ejemplo 2.4. Buscamos los siguientes sucesos: $A \cup B, A \cap B, A^c, B^c, A - B$ y $B - A$.

$$A \cup B = \{2, 4, 6, 8, 10, 11, 12, \dots\};$$

$$A \cap B = \{10, 12, 14, \dots\};$$

$$A^c = \{x : x < 10\} = \{1, 2, \dots, 9\};$$

$$B^c = \{1, 3, 5, \dots\};$$

$$A - B = \{11, 13, 15, \dots\};$$

$$B - A = \{2, 4, 6, 8\}.$$

Las tres operaciones básicas planteadas se pueden combinar para formar otros conjuntos. Las siguientes propiedades de las operaciones de conjuntos son útiles para la obtención de nuevas expresiones para la combinación de conjuntos:

Propiedad conmutativa:

$$A \cup B = B \cup A \quad \text{y} \quad A \cap B = B \cap A. \quad (2.1)$$

Propiedad asociativa:

$$A \cup (B \cup C) = (A \cup B) \cup C \quad \text{y} \quad A \cap (B \cap C) = (A \cap B) \cap C. \quad (2.2)$$

Propiedad distributiva:

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C) \quad \text{y} \quad A \cap (B \cup C) = (A \cap B) \cup (A \cap C). \quad (2.3)$$

Mediante la aplicación de las propiedades anteriores podemos derivar nuevas igualdades. Las leyes de DeMorgan son un ejemplo importante de ello:

Leyes de DeMorgan:

$$(A \cup B)^c = A^c \cap B^c \quad \text{y} \quad (A \cap B)^c = A^c \cup B^c. \quad (2.4)$$

Ejemplo 2.6

Demostración de las leyes de DeMorgan mediante diagramas de Venn y la igualdad de conjuntos. En primer lugar vamos a utilizar un diagrama de Venn para demostrar la primera igualdad. La región sombreada de la figura 2.2(g) muestra el complementario de $A \cup B$, el lado izquierdo de la ecuación. La región rayada de la figura 2.2(h) muestra la intersección de A^c y B^c . Las dos regiones son las mismas y por lo tanto los conjuntos son iguales. El diagrama de Venn para la segunda ecuación se obtiene de forma análoga.

A continuación demostramos las leyes de DeMorgan, verificando la igualdad de conjuntos. La prueba consta de dos partes: en primer lugar, demostramos que $(A \cup B)^c \subset A^c \cap B^c$; a continuación, demostramos que $A^c \cap B^c \subset (A \cup B)^c$. Estos resultados conjuntamente implican que $(A \cup B)^c = A^c \cap B^c$.

En primer lugar, supongamos que $x \in (A \cup B)^c$, entonces $x \notin A \cup B$. En particular, tenemos que $x \notin A$, lo que implica que $x \in A^c$. Del mismo modo, tenemos que $x \in B^c$ y así $x \notin B$. Por lo tanto, x pertenece tanto a A^c como a B^c , es decir, $x \in A^c \cap B^c$. Hemos demostrado entonces que $(A \cup B)^c \subset A^c \cap B^c$.

Para probar la inclusión en el otro sentido, supongamos que $x \in A^c \cap B^c$. Esto implica que $x \in A^c$, por lo que $x \notin A$. Del mismo modo, $x \in B^c$ por lo que $x \notin B$. Por lo tanto, $x \notin (A \cup B)$ y entonces $x \in (A \cup B)^c$. Hemos comprobado con esto que $A^c \cap B^c \subset (A \cup B)^c$. Por lo que queda demostrado que $(A \cup B)^c = A^c \cap B^c$.

Para probar la segunda ley de DeMorgan, aplicamos la primera ley de DeMorgan a A^c y B^c obteniendo:

$$(A^c \cup B^c)^c = (A^c)^c \cap (B^c)^c = A \cap B,$$

donde hemos usado la identidad $A = (A^c)^c$. Tomando ahora el complementario de ambos lados de la anterior ecuación:

$$A^c \cup B^c = (A \cap B)^c.$$

Ejemplo 2.7

Se definen los siguientes conjuntos A, B y C para el experimento E_{10}

$$\begin{aligned} A &= \{v : |v| > 10\}, & \text{“la magnitud de } v \text{ es mayor de 10 voltios”}, \\ B &= \{v : v < -5\}, & \text{“} v \text{ es menor de } -5 \text{ voltios”}, \\ C &= \{v : v > 0\}, & \text{“} v \text{ es positiva”}. \end{aligned}$$

Se puede comprobar que

$$\begin{aligned} A \cup B &= \{v : v < -5 \text{ ó } v > 10\}, \\ A \cap B &= \{v : v < -10\}, \\ C^c &= \{v : v \leq 0\}, \\ (A \cup B) \cap C &= \{v : v > 10\}, \\ A \cap B \cap C &= \emptyset, \\ (A \cup B)^c &= \{v : -5 \leq v \leq 10\}. \end{aligned}$$

Las operaciones de unión e intersección pueden repetirse para un número arbitrario de conjuntos. Así, la unión de n conjuntos

$$\bigcup_{k=1}^n A_k = A_1 \cup A_2 \cup \dots \cup A_n \quad (2.5)$$

es el conjunto que consta de todos los elementos que se encuentran en A_k para al menos un valor de k . La misma definición se aplica a la unión de una sucesión infinita numerable de conjuntos:

$$\bigcup_{k=1}^{\infty} A_k. \quad (2.6)$$

La intersección de n conjuntos

$$\bigcap_{k=1}^n A_k = A_1 \cap A_2 \cap \dots \cap A_n \quad (2.7)$$

es el conjunto que consta de los elementos que están en todos los conjuntos

A_1, \dots, A_n . La misma definición se aplica a la intersección de una sucesión infinita numerable de conjuntos:

$$\bigcap_{k=1}^{\infty} A_k. \quad (2.8)$$

Veremos que las uniones e intersecciones numerables de conjuntos son esenciales en el tratamiento de espacios muestrales que no son finitos.

2.1.4 Clases de sucesos

Hemos presentado el espacio muestral S como el conjunto de todos los posibles resultados del experimento aleatorio. También hemos presentado los sucesos como subconjuntos de S . La teoría de la probabilidad también requiere que establezcamos la clase \mathcal{F} de sucesos de interés. Sólo a los sucesos de esta clase se les asignan probabilidades. Se espera que cualquier operación de conjuntos en \mathcal{F} dé como resultado un conjunto que también sea un suceso de \mathcal{F} .

En particular, recordemos que los conjuntos complementarios, así como las uniones e intersecciones numerables de sucesos de \mathcal{F} , es decir, las ecuaciones (2.1) y (2.5) – (2.8), dan como resultado sucesos de \mathcal{F} . Cuando el espacio muestral S es finito o numerable, simplemente suponemos que \mathcal{F} está formado por todos los subconjuntos de S y se puede proceder sin tener en cuenta mayores consideraciones acerca de \mathcal{F} . Sin embargo, cuando S es la recta real \mathbb{R} (o un intervalo de la recta real), no podemos suponer que \mathcal{F} esté formada por todos los posibles subconjuntos de \mathbb{R} y, al mismo tiempo, satisfaga los axiomas de la probabilidad. Afortunadamente, podemos obtener todos los sucesos de interés práctico definiendo \mathcal{F} como la clase de sucesos obtenidos como complementarios y uniones e intersecciones de intervalos de la recta real, por ejemplo, $(a, b]$ o $(-\infty, b]$. Nos referiremos a este tipo de sucesos como el espacio de Borel. En el resto del curso, haremos referencia a la clase de sucesos \mathcal{F} en contadas ocasiones. Para un curso de probabilidad a nivel introductorio no se necesita saber más de lo establecido en este párrafo.

Cuando hablamos de una clase de sucesos nos referimos a una colección (conjunto) de sucesos (conjuntos), es decir, estamos hablando de un “conjunto de conjuntos”. Nos referimos a la colección de conjuntos como una clase para recordar que los elementos de la clase son conjuntos. Usamos letras góticas mayúsculas para referenciar una clase, por ejemplo, \mathcal{C} , \mathcal{F} , \mathcal{G} . Si la clase \mathcal{C} se compone de la colección de conjuntos de A_1, \dots, A_n entonces escribiremos $\mathcal{C} = \{A_1, \dots, A_n\}$.

Ejemplo 2.8

Sea $S = \{C, X\}$ el resultado del lanzamiento de una moneda. Supongamos que todo subconjunto de S es un suceso. Queremos encontrar todos los posibles sucesos de S . Dado que un suceso es un subconjunto de S , tenemos que encontrar

todos los posibles subconjuntos de S , que son los siguientes:

$$\mathcal{S} = \{\emptyset, \{C\}, \{X\}, \{C, X\}\}$$

Observemos que S incluye tanto el conjunto vacío como S . Sean i_C y i_X números binarios, donde $i = 1$ indica que el correspondiente elemento de S pertenece a un subconjunto dado. Generamos todos los posibles subconjuntos tomando todos los valores posibles del par i_C y i_X . Por tanto, $i_X = 0, i_C = 1$ se corresponde con el conjunto $\{C\}$. Es sencillo ver que hay 2^2 posibles subconjuntos tal y como hemos enumerado anteriormente.

Para un espacio muestral finito, $S = \{1, 2, \dots, k\}$, por lo general supondremos que todos los subconjuntos de S son sucesos. Esta clase de sucesos se denomina conjunto potencia de S y lo denotaremos por \mathcal{S} . Podemos indexar todos los posibles subconjuntos de S con números binarios, i_1, i_2, \dots, i_k , y nos encontramos con que el conjunto potencia de S cuenta con 2^k elementos. Debido a esto, el conjunto potencia también se denota como $\mathcal{S} = 2^S$.

2.2 Los axiomas de la probabilidad

Las probabilidades son números asignados a sucesos que indican cómo de “probable” es que los sucesos ocurran cuando se lleva a cabo un experimento. Una ley de probabilidad para un experimento aleatorio es una regla que asigna probabilidades a los sucesos del experimento que pertenecen a la clase de sucesos \mathcal{F} . Así, una ley de probabilidad es una función que asigna un número a conjuntos (sucesos). En la sección 1.3 encontrábamos una serie de propiedades de la frecuencia relativa que cualquier definición de probabilidad debería satisfacer. Los axiomas de la probabilidad establecen formalmente que una ley de probabilidad debe satisfacer estas propiedades. En esta sección, desarrollamos una serie de resultados que se derivan de este conjunto de axiomas.

Sea E un experimento aleatorio con espacio muestral S y clase de sucesos \mathcal{F} . Una ley de probabilidad para el experimento E es una regla que asigna a cada suceso $A \in \mathcal{F}$ un número $P[A]$, llamado probabilidad de A , que satisface los siguientes axiomas:

- Axioma I* $0 \leq P[A]$
Axioma II $P[S] = 1$
Axioma III Si $A \cap B = \emptyset$, entonces $P[A \cup B] = P[A] + P[B]$.
Axioma III' Si A_1, A_2, \dots es una sucesión de sucesos tales que $A_i \cap A_j = \emptyset$ para todo $i \neq j$, entonces

$$P \left[\bigcup_{k=1}^{\infty} A_k \right] = \sum_{k=1}^{\infty} P[A_k].$$

Los axiomas I, II y III son suficientes para trabajar con experimentos que

tienen asociados espacios muestrales finitos. Con el fin de manejar experimentos con espacios muestrales infinitos, el axioma III necesita ser reemplazado por el axioma III'. Observemos que el axioma III' incluye el axioma III como un caso particular, suponiendo que $A_k = \emptyset$ para todo $k \geq 3$. Así que realmente sólo se necesitan los axiomas I, II y III'. Sin embargo, se obtiene un mayor conocimiento partiendo de los axiomas I, II y III.

Los axiomas nos permiten ver los sucesos como objetos que poseen una propiedad (es decir, su probabilidad) que tiene características similares a la masa física. El axioma I establece que la probabilidad (masa) es no negativa, y el axioma II establece que hay una cantidad fija total de probabilidad (masa), es decir, 1 unidad. El axioma III establece que la probabilidad (masa) total en dos objetos disjuntos es la suma de las probabilidades (masas) individuales.

Los axiomas nos proporciona un conjunto de reglas consistentes que cualquier asignación de probabilidad válida debe satisfacer. A continuación desarrollaremos varias propiedades derivadas de los axiomas que son útiles en el cálculo de probabilidades.

El primer resultado establece que si hacemos una partición del espacio muestral en dos sucesos mutuamente excluyentes A y A^c , entonces las probabilidades de estos dos sucesos suman uno.

COROLARIO 1

$$P[A^c] = 1 - P[A]$$

Demostración Puesto que un suceso A y su complementario A^c son mutuamente excluyentes, $A \cap A^c = \emptyset$, por el axioma III tenemos que

$$P[A \cup A^c] = P[A] + P[A^c].$$

Dado que $S = A \cup A^c$, por el axioma II,

$$1 = P[S] = P[A \cup A^c] = P[A] + P[A^c].$$

El corolario queda demostrado despejando $P[A^c]$. □

El siguiente corolario establece que la probabilidad de un suceso es siempre menor o igual que uno. El corolario 2 en combinación con el axioma I proporcionan una buena herramienta para chequear las soluciones de los problemas: si las probabilidades son negativas o mayores que uno, hay un error en alguna parte.

COROLARIO 2

$$P[A] \leq 1$$

Demostración Por el corolario 1,

$$P[A] = 1 - P[A^c] \leq 1,$$

dado que $P[A^c] \geq 0$ □

El corolario 3 establece que el suceso imposible tiene probabilidad cero.

COROLARIO 3

$$P[\emptyset] = 0$$

Demostración Tomando $A = S$ y $A^c = \emptyset$ en el corolario 1:

$$P[\emptyset] = 1 - P[S] = 0.$$

□

El corolario 4 proporciona el método estándar para calcular la probabilidad de un suceso complejo A . El método consiste en descomponer el suceso A como la unión de sucesos disjuntos A_1, A_2, \dots, A_n . La probabilidad de A es la suma de las probabilidades de los A_k .

COROLARIO 4 Si A_1, A_2, \dots, A_n son mutuamente excluyentes dos a dos, entonces

$$P\left[\bigcup_{k=1}^n A_k\right] = \sum_{k=1}^n P[A_k] \quad \text{para } n \geq 2.$$

Demostración Usaremos inducción matemática. El axioma III implica que el resultado es cierto para $n = 2$. A continuación tenemos que demostrar que si el resultado es verdadero para algún n , entonces también es cierto para $n + 1$. Esto, combinado con el hecho de que el resultado es cierto para $n = 2$, implica que el resultado es cierto para $n \geq 2$.

Spongamos que el resultado es cierto para algún $n \geq 2$, es decir,

$$P\left[\bigcup_{k=1}^n A_k\right] = \sum_{k=1}^n P[A_k], \quad (2.9)$$

y consideremos el caso $n + 1$

$$P\left[\bigcup_{k=1}^{n+1} A_k\right] = P\left[\left\{\bigcup_{k=1}^n A_k\right\} \cup A_{n+1}\right] = P\left[\bigcup_{k=1}^n A_k\right] + P[A_{n+1}], \quad (2.10)$$

donde hemos aplicado el axioma III a la segunda expresión después de observar que la unión de los sucesos A_1 hasta A_n es mutuamente excluyente con A_{n+1} . La propiedad distributiva implica entonces

$$\left\{\bigcup_{k=1}^n A_k\right\} \cap A_{n+1} = \bigcup_{k=1}^n \{A_k \cap A_{n+1}\} = \bigcup_{k=1}^n \emptyset = \emptyset.$$

Sustituyendo la ecuación (2.9) en la ecuación (2.10) obtenemos el caso $n + 1$

$$P\left[\bigcup_{k=1}^{n+1} A_k\right] = \sum_{k=1}^{n+1} P[A_k].$$

□

El corolario 5 proporciona una expresión para la unión de dos sucesos que no son necesariamente mutuamente excluyentes.

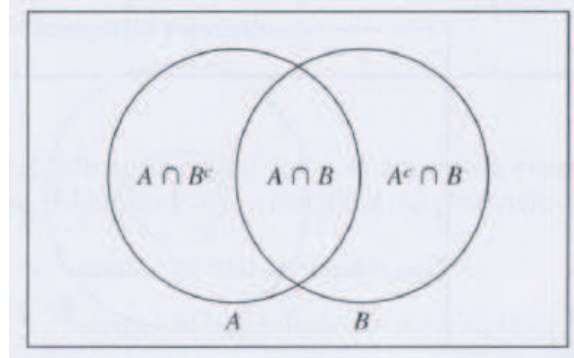


Figura 2.3 Descomposición de $A \cup B$ en tres conjuntos disjuntos.

COROLARIO 5

$$P[A \cup B] = P[A] + P[B] - P[A \cap B].$$

Demostración En primer lugar, se descomponen $A \cup B$, A , y B como unión de sucesos disjuntos. A partir del diagrama de Venn de la figura 2.3,

$$\begin{aligned} P[A \cup B] &= P[A \cap B^c] + P[A^c \cap B] + P[A \cap B], \\ P[A] &= P[A \cap B^c] + P[A \cap B], \\ P[B] &= P[B \cap A^c] + P[A \cap B]. \end{aligned}$$

Sustituyendo la expresión de $P[A \cap B^c]$ y $P[B \cap A^c]$ de las dos últimas ecuaciones en la primera ecuación, se obtiene el corolario. \square

Al observar el diagrama de Venn de la figura 2.3, se comprueba que la suma $P[A] + P[B]$ tiene en cuenta dos veces la probabilidad (masa) del conjunto $A \cap B$. La expresión del corolario 5 realiza la corrección oportuna.

El corolario 5 se generaliza fácilmente al caso de tres sucesos,

$$\begin{aligned} P[A \cup B \cup C] &= P[A] + P[B] + P[C] - P[A \cap B] \\ &\quad - P[A \cap C] - P[B \cap C] + P[A \cap B \cap C], \end{aligned} \quad (2.11)$$

y en general a n sucesos, tal y como se muestra en el corolario 6.

COROLARIO 6

$$P\left[\bigcup_{k=1}^n A_k\right] = \sum_{j=1}^n P[A_j] - \sum_{j < k} P[A_j \cap A_k] + \cdots + (-1)^{n+1} P[A_1 \cap \dots \cap A_n].$$

Demostración Se realiza por inducción. \square

Dado que las probabilidades son no negativas, el corolario 5 implica que la probabilidad de la unión de dos sucesos no es mayor que la suma de las probabilidades de los sucesos individuales

$$P[A \cup B] \leq P[A] + P[B]. \quad (2.12)$$

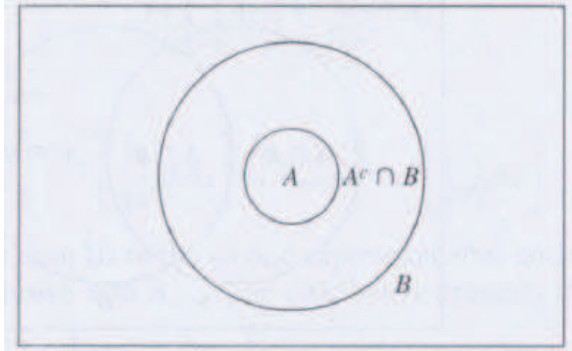


Figura 2.4 Si $A \subset B$, entonces $P[A] \leq P[B]$.

La desigualdad anterior es un caso especial del hecho de que un subconjunto de otro conjunto debe tener menor probabilidad. Este resultado se utiliza con frecuencia para obtener cotas superiores para probabilidades de interés. La situación típica en la que estamos interesados es aquella en la que el suceso A tiene una probabilidad difícil de encontrar; en ese caso buscamos un suceso B para el cual se pueda encontrar su probabilidad y que incluya a A como subconjunto.

COROLARIO 7 Si $A \subset B$, entonces $P[A] \leq P[B]$.

Demostración En la figura 2.4, B es la unión de A y $A^c \cap B$, por tanto

$$P[B] = P[A] + P[A^c \cap B] \geq P[A],$$

dado que $P[A^c \cap B] \geq 0$. □

Los axiomas junto con el corolario nos proporcionan un conjunto de reglas para el cálculo de probabilidades de ciertos sucesos en términos de otros sucesos. Sin embargo, todavía seguimos necesitando una asignación de probabilidad inicial para ciertos conjuntos básicos de sucesos a partir de los que la probabilidad de todos los demás sucesos se puedan calcular. Este problema se aborda en las próximas dos subsecciones.

2.2.1 Espacios muestrales discretos

En esta sección se muestra que la ley de probabilidad para un experimento con un espacio muestral contable se puede especificar proporcionando las probabilidades de los sucesos elementales. En primer lugar, suponemos que el espacio muestral es finito, $S = \{a_1, a_2, \dots, a_n\}$ y suponemos que \mathcal{F} está formado por todos los subconjuntos de S . Todos los diferentes sucesos elementales son mutuamente excluyentes, por lo que por el corolario 4 la probabilidad de cualquier suceso $B = \{a'_1, a'_2, \dots, a'_m\}$ viene dada por

$$P[B] = P[\{a'_1, a'_2, \dots, a'_m\}] = P[\{a'_1\}] + P[\{a'_2\}] + \dots + P[\{a'_m\}]; \quad (2.13)$$

es decir, la probabilidad de un suceso es igual a la suma de las probabilidades de los resultados del suceso. Por tanto, llegamos a la conclusión de que la ley

de probabilidad para un experimento aleatorio con un espacio muestral finito se especifica dando las probabilidades de los sucesos elementales.

Si el espacio muestral tiene n elementos, $S = \{a_1, a_2, \dots, a_n\}$, una asignación de probabilidades de especial interés es el caso en el que los resultados son igualmente probables (equiprobables). La probabilidad de los sucesos elementales es

$$P[\{a_1\}] = P[\{a_2\}] = \dots = P[\{a_n\}] = \frac{1}{n}. \quad (2.14)$$

La probabilidad de cualquier suceso que se compone de k resultados, por ejemplo $B = \{a'_1, a'_2, \dots, a'_k\}$, es

$$P[B] = P[\{a'_1\}] + P[\{a'_2\}] + \dots + P[\{a'_k\}] = \frac{k}{n}. \quad (2.15)$$

De esta forma, si los resultados son equiprobables, entonces la probabilidad de un suceso es igual al número de resultados del suceso dividido por el número total de resultados del espacio muestral.

Consideremos el caso en el que el espacio muestral es infinito numerable, $S = \{a_1, a_2, \dots\}$. Sea \mathcal{S} la clase de sucesos formada por todos los subconjuntos de S . Observemos que \mathcal{F} ahora debe satisfacer la ecuación (2.8) porque los sucesos pueden estar formados por uniones numerables de conjuntos. El axioma III' implica que la probabilidad de un suceso como $D = \{b'_1, b'_2, b'_3, \dots\}$ viene dada por

$$P[D] = P[\{b'_1, b'_2, b'_3, \dots\}] = P[\{b'_1\}] + P[\{b'_2\}] + P[\{b'_3\}] + \dots$$

La probabilidad de un suceso con un espacio muestral infinito numerable se determina a partir de las probabilidades de los sucesos elementales.

Ejemplo 2.9

Una urna contiene 10 bolas idénticas numeradas con las cifras $0, 1, \dots, 9$. Se considera el experimento aleatorio que consiste en seleccionar al azar una bola de la urna y anotar el número de la bola. Queremos encontrar la probabilidad de los siguientes sucesos:

$$\begin{aligned} A &= \text{“el número de la bola seleccionada es impar”}, \\ B &= \text{“el número de la bola seleccionada es múltiplo de 3”}, \\ C &= \text{“el número de la bola seleccionada es menor que 5”}, \end{aligned}$$

y además, $A \cup B$ y $A \cup B \cup C$.

El espacio muestral es $S = \{0, 1, \dots, 9\}$, por lo que los conjuntos de resultados correspondientes a los sucesos anteriores son

$$A = \{1, 3, 5, 7, 9\}, \quad B = \{3, 6, 9\}, \quad \text{y} \quad C = \{0, 1, 2, 3, 4\}.$$

Si suponemos que los resultados son equiprobables, entonces,

$$P[A] = P[\{1\}] + P[\{3\}] + P[\{5\}] + P[\{7\}] + P[\{9\}] = \frac{5}{10},$$

$$P[B] = P[\{3\}] + P[\{6\}] + P[\{9\}] = \frac{3}{10},$$

$$P[C] = P[\{0\}] + P[\{1\}] + P[\{2\}] + P[\{3\}] + P[\{4\}] = \frac{5}{10}.$$

Y por el corolario 5,

$$P[A \cup B] = P[A] + P[B] - P[A \cap B] = \frac{5}{10} + \frac{3}{10} - \frac{2}{10} = \frac{6}{10},$$

donde hemos utilizado el hecho de que $A \cap B = \{3, 9\}$, por lo que $P[A \cap B] = 2/10$.

Finalmente, por el corolario 6,

$$\begin{aligned} P[A \cup B \cup C] &= P[A] + P[B] + P[C] - P[A \cap B] - P[A \cap C] - P[B \cap C] + P[A \cap B \cap C] \\ &= \frac{5}{10} + \frac{3}{10} + \frac{5}{10} - \frac{2}{10} - \frac{2}{10} - \frac{1}{10} + \frac{1}{10} = \frac{9}{10}. \end{aligned}$$

Quedaría para el detalle verificar las respuestas de $P[A \cup B]$ y $P[A \cup B \cup C]$ mediante la enumeración de los resultados de los correspondientes sucesos.

Se pueden concebir muchos modelos de probabilidad para el mismo espacio muestral y los mismos sucesos, variando la asignación de probabilidad; en el caso de los espacios muestrales finitos todo lo que tenemos que hacer es proporcionar, como probabilidades de los sucesos elementales, n números no negativos cuya suma sea uno. Por supuesto, en cualquier situación particular, la asignación de probabilidad debe ser seleccionada de forma que refleje las observaciones experimentales en la medida de lo posible. El siguiente ejemplo muestra que pueden surgir situaciones donde hay más de una asignación de probabilidad “razonable” y donde es necesaria la evidencia experimental para decidir cual es la asignación apropiada.

Ejemplo 2.10

Supongamos que se lanza una moneda tres veces. Si observamos la secuencia de caras y cruces, entonces hay ocho posibles resultados

$$S_3 = \{CCC, CCX, CXC, XCC, XIX, XCX, CXX, XXX\}.$$

Si suponemos que los resultados de S_3 son equiprobables, entonces la probabilidad de cada uno de los ocho sucesos elementales es $1/8$. Esta asignación de probabilidad implica que la probabilidad de obtener dos caras en tres lanzamientos es, por el corolario 3,

$$\begin{aligned} P[\text{“2 caras en 3 lanzamientos”}] &= P[\{CCX, CXC, XCC\}] \\ &= P[\{CCX\}] + P[\{CXC\}] + P[\{XCC\}] = \frac{3}{8} \end{aligned}$$

Supongamos ahora que lanzamos una moneda tres veces pero contamos el número de caras en tres lanzamientos en lugar de observar la sucesión de caras y cruces. El espacio muestral es ahora $S_4 = \{0, 1, 2, 3\}$. Si suponemos que los resultados de S_4 son equiprobables, entonces cada uno de los sucesos elementales de S_4 tiene una probabilidad de $1/4$. Esta segunda asignación de probabilidad predice que la probabilidad de obtener dos caras en tres lanzamientos es

$$P[\text{"2 caras en 3 lanzamientos"}] = P[\{2\}] = \frac{1}{4}.$$

La primera asignación de probabilidad implica que la probabilidad de dos caras en tres lanzamientos es $3/8$, y la segunda asignación de probabilidad prevé que la probabilidad es de $1/4$. Por tanto, las dos asignaciones no son consistentes entre sí. En cuanto a la teoría se refiere, cualquiera de las asignaciones es aceptable. Nos corresponde a nosotros decidir qué asignación es más adecuada. Más adelante en el capítulo veremos que sólo la primera asignación es consistente con el supuesto de que la moneda está equilibrada y que los lanzamientos son "independientes". Esta asignación predice correctamente las frecuencias relativas que se observarían en el experimento real del lanzamiento de una moneda.

Finalmente, consideramos un ejemplo con un espacio muestral infinito numerable.

Ejemplo 2.11

Una moneda equilibrada se lanza repetidamente hasta que aparezca la primera cara; el resultado del experimento es el número de lanzamientos necesarios hasta que aparece la primera cara. Queremos encontrar una ley de probabilidad para este experimento.

Es concebible que sea necesario un número arbitrariamente grande de lanzamientos hasta que aparezca la cara, por lo que el espacio muestral es $S = \{1, 2, 3, \dots\}$. Supongamos que el experimento se repite n veces. Sea N_j el número de ensayos en los que la primera cara aparece en el j -ésimo lanzamiento. Si n es muy grande, se espera que N_1 sea aproximadamente $n/2$ ya que la moneda no está cargada (está equilibrada). Esto implica que es necesario un segundo lanzamiento aproximadamente $n - N_1 \approx n/2$ veces, y de nuevo, se espera que aproximadamente la mitad de estos — es decir, $n/4$ — dará como resultado una cara, y así sucesivamente, como se muestra en la figura 2.5. Por tanto, para n grande, las frecuencias relativas son

$$f_j \approx \frac{N_j}{n} = \left(\frac{1}{2}\right)^j \quad j = 1, 2, \dots$$

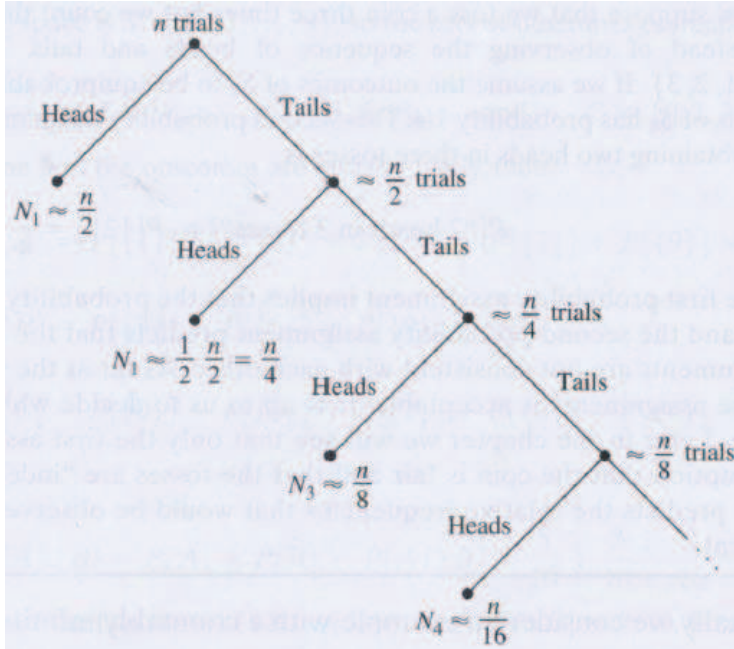


Figura 2.5 En n ensayos aparecen caras en el primer lanzamiento $n/2$ veces aproximadamente, en el segundo lanzamiento $n/4$ veces aproximadamente, y así sucesivamente.

Concluimos por tanto que una ley de probabilidad razonable para este experimento es

$$P[j \text{ lanzamientos hasta la primera cara}] = \left(\frac{1}{2}\right)^j \quad j = 1, 2, \dots \quad (2.16)$$

Se puede comprobar que estas probabilidades suman uno mediante el uso de la serie geométrica para $\alpha = 1/2$

$$\sum_{j=1}^n \alpha^j = \frac{\alpha}{1-\alpha} \Big|_{\alpha=1/2} = 1.$$

2.2.2 Espacios muestrales continuos

Los espacios muestrales continuos aparecen en experimentos en los que los resultados son números que puede asumir un continuo de valores, de forma que se considera que el espacio muestral S es la recta real \mathbb{R} (o un intervalo de la recta real). En este caso, podríamos considerar que la clase de sucesos consistiera en todos los subconjuntos de \mathbb{R} . Pero resulta que esta clase es “demasiado grande”

y es imposible asignar probabilidades a todos los subconjuntos de \mathbb{R} . Afortunadamente, es posible asignar probabilidades a todos los sucesos de una clase más pequeña que incluye todos los sucesos de interés práctico. Esta clase denotada por \mathcal{B} , se llama conjunto de Borel y contiene todos los intervalos abiertos y cerrados de la recta real, así como todos los sucesos que se pueden obtener como uniones, intersecciones y complementarios numerables. El axioma III' vuelve a ser la clave para calcular las probabilidades de sucesos. Sea A_1, A_2, \dots una sucesión de sucesos mutuamente excluyentes que se representan mediante intervalos de la recta real, entonces

$$P\left[\bigcup_{k=1}^{\infty} A_k\right] = \sum_{k=1}^{\infty} P[A_k]$$

donde cada $P[A_k]$ se especifica mediante la ley de probabilidad. Por esta razón, las leyes de probabilidad en experimentos con espacios muestrales continuos especifican una regla para asignar números a intervalos de la recta real.

Ejemplo 2.12

Consideremos el experimento aleatorio “elegir al azar un número x entre cero y uno”. El espacio muestral S para este experimento es el intervalo unitario $[0, 1]$, que es infinito no numerable. Si suponemos que todos los resultados de S tienen la misma probabilidad de ser seleccionados, entonces podríamos conjeturar que la probabilidad de que el resultado esté en el intervalo $[0, 1/2]$ es la misma que la probabilidad de que el resultado esté en el intervalo $[1/2, 1]$. También podríamos conjeturar que la probabilidad de que el resultado sea exactamente igual a $1/2$ es cero, ya que hay un número infinito no numerable de resultados equiprobables.

Consideremos la siguiente ley de probabilidad: “La probabilidad de que el resultado caiga en un subintervalo de S es igual a la longitud del subintervalo”, es decir,

$$P[[a, b]] = (b - a) \quad \text{para } 0 \leq a \leq b \leq 1, \quad (2.17)$$

donde por $P[[a, b]]$ nos referimos a la probabilidad del suceso correspondiente al intervalo $[a, b]$. Claramente, el axioma I se cumple dado que $b \geq a \geq 0$. El axioma II se deriva de que $S = [a, b]$ con $a = 0$ y $b = 1$.

Demostramos ahora que la ley de probabilidad es consistente con las conjeturas anteriores acerca de las probabilidades de los sucesos $[0, 1/2]$, $[1/2, 1]$ y $\{1/2\}$:

$$\begin{aligned} P[[0, 0.5]] &= 0.5 - 0 = 0.5, \\ P[[0.5, 1]] &= 1 - 0.5 = 0.5. \end{aligned}$$

Además, si x_0 es un punto cualquiera de S , entonces $P[[x_0, x_0]] = 0$ ya que los puntos individuales tienen amplitud cero.

Supongamos ahora que estamos interesados en un suceso que es la unión de

varios intervalos; por ejemplo, “el resultado se encuentra, al menos, a 0.3 unidades de distancia del centro del intervalo unidad”, es decir, $A = [0, 0.2] \cup [0.8, 1]$. Puesto que los dos intervalos son disjuntos, tenemos por el axioma III que

$$P[A] = P[[0, 0.2]] + P[[0.8, 1]] = 0.4.$$

El siguiente ejemplo muestra que una asignación de probabilidad inicial que especifique la probabilidad de intervalos semi-infinitos también es suficiente para especificar las probabilidades de todos los sucesos de interés.

Ejemplo 2.13

Supongamos que se mide la vida útil de un chip de memoria de una computadora, y nos encontramos con que “la proporción de chips cuya vida útil es superior a t decrece exponencialmente con una tasa α ”. Buscamos una ley de probabilidad apropiada.

Sea $S = \{(0, \infty)\}$ el espacio muestral de este experimento. Si interpretamos la conclusión anterior como “la probabilidad de que la vida de un chip supera t disminuye exponencialmente a una tasa α ”, se obtiene entonces la siguiente asignación de probabilidades a los sucesos de la forma (t, ∞) :

$$P[(t, \infty)] = e^{-\alpha t} \quad \text{para } t > 0, \quad (2.18)$$

donde $\alpha > 0$. Observemos que la función exponencial es un número entre 0 y 1 para $t > 0$, por lo que se satisface el axioma I. El axioma II también se verifica ya que

$$P[S] = P[(0, \infty)] = 1.$$

La probabilidad de que la vida útil esté en el intervalo $(r, s]$ se obtiene observando en la figura 2.6 que $(r, s] \cup (s, \infty) = (r, \infty)$, por lo que por el axioma III,

$$P[(r, \infty)] = P[(r, s]] + P[(s, \infty)].$$

Reordenando la ecuación anterior obtenemos

$$P[(r, s]] = P[(r, \infty)] - P[(s, \infty)] = e^{-\alpha r} - e^{-\alpha s}.$$

De esta forma se obtiene la probabilidad de intervalos arbitrarios en S .

En ambos ejemplos 2.12 y 2.13, la probabilidad de que el resultado tome un valor específico es igual a cero. Podemos preguntarnos: Si un resultado (o suceso) tiene probabilidad cero, ¿no significa eso que no puede ocurrir? Y, a continuación, nos podemos preguntar: ¿Cómo pueden tener probabilidad cero todos los resultados de un espacio muestral? Se puede explicar esta paradoja

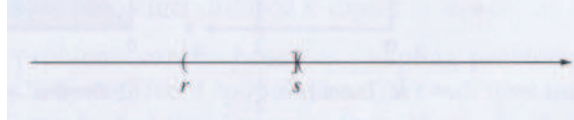


Figura 2.6

$$(r, \infty) = (r, s) \cup [s, \infty).$$

mediante el uso de la interpretación de la probabilidad como una frecuencia relativa. Un suceso que ocurre sólo una vez en un número infinito de pruebas tendrá frecuencia relativa cero. Por lo tanto, el hecho de que un suceso o resultado tenga frecuencia relativa cero, no implica que no pueda ocurrir, sino que se produce con muy poca frecuencia. En el caso de los espacios muestrales continuos, el conjunto de resultados posibles es tan rico que todos los resultados ocurren con muy poca frecuencia, la suficiente como para que sus frecuencias relativas sean cero.

Terminamos esta sección con un ejemplo donde los sucesos son regiones en el plano.

Ejemplo 2.14

Consideremos el experimento E_{12} , donde seleccionábamos dos números x e y al azar entre cero y uno. El espacio muestral es entonces el cuadrado unidad mostrado en la figura 2.7(a). Si suponemos que todos los pares de números en el cuadrado unidad tienen la misma probabilidad de ser seleccionados, entonces es razonable utilizar una asignación de probabilidad en la que la probabilidad de cualquier región R en el interior del cuadrado unidad es igual al área de R . Queremos encontrar la probabilidad de los siguientes sucesos: $A = \{x > 0.5\}$, $B = \{y > 0.5\}$ y $C = \{x > y\}$.

Las figuras 2.7(b) – 2.7(d) muestran las regiones correspondientes a los sucesos A , B y C . Es evidente que cada una de estas regiones tiene un área de $1/2$. Por tanto

$$P[A] = \frac{1}{2}, \quad P[B] = \frac{1}{2}, \quad P[C] = \frac{1}{2}.$$

Reiteramos la forma de proceder desde el planteamiento de un problema hasta su modelo de probabilidad. El planteamiento del problema, implícita o explícitamente define un experimento aleatorio, que especifica un procedimiento experimental y una serie de mediciones y observaciones. Estas mediciones y observaciones determinan el conjunto de todos los resultados posibles y por lo tanto, el espacio muestral S .

A continuación debe determinarse una asignación inicial de probabilidad que especifique la probabilidad de ciertos sucesos. Esta asignación de probabilidad debe satisfacer los axiomas de la probabilidad. Si S es discreto, basta con especificar las probabilidades de sucesos elementales. Si S es continuo, basta con

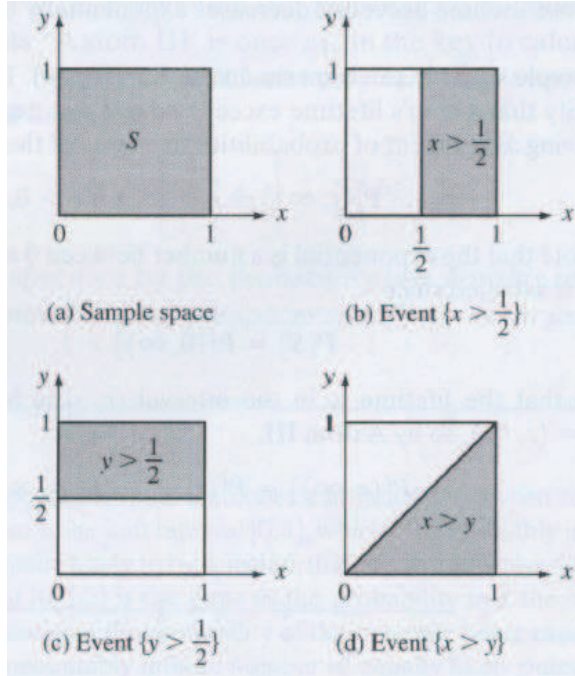


Figura 2.7 Un espacio muestral bidimensional y tres sucesos.

especificar las probabilidades de los intervalos de la recta real o regiones del plano. La probabilidad de otros sucesos de interés, se puede determinar a partir de la asignación inicial de probabilidad y de los axiomas de la probabilidad y sus corolarios. Son posibles muchas asignaciones de probabilidad, por lo que la elección de la asignación de probabilidad debe reflejar las observaciones experimentales y/o la experiencia previa.

2.3 Probabilidad condicionada

Muy a menudo el interés se encuentra en determinar si dos sucesos, A y B , están relacionados en el sentido de que el conocimiento sobre la ocurrencia de uno, digamos B , altera la verosimilitud (posibilidad) de ocurrencia del otro, A . Esto requiere que calculemos la probabilidad condicionada, $P[A|B]$, del suceso A , sabiendo que ha ocurrido el suceso B . La probabilidad condicionada se define como

$$P[A|B] = \frac{P[A \cap B]}{P[B]} \quad \text{para } P[B] > 0. \quad (2.19)$$

El conocimiento de que el suceso B ha ocurrido implica que el resultado del experimento está dentro del conjunto B . Por lo tanto, para calcular $P[A|B]$, podemos suponer que el experimento ahora tiene como espacio muestral reducido el conjunto B , como se muestra en la figura 2.8. El suceso A tiene lugar en el

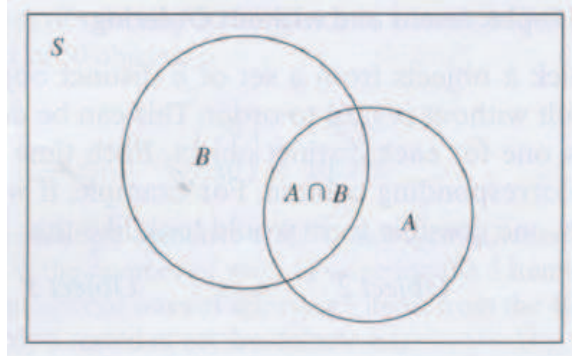


Figura 2.8 Si sabemos que ha ocurrido B , entonces A puede ocurrir solamente si $A \cap B$ ocurre.

espacio muestral reducido si y sólo si el resultado ζ está en $A \cap B$. La ecuación (2.19) simplemente renormaliza la probabilidad de los sucesos que ocurren conjuntamente con B . De esta forma, si suponemos que $A = B$, la ecuación (2.19) queda $P[B|B] = 1$, como cabía esperar. Es fácil demostrar que $P[A|B]$, para B fijado, satisface los axiomas de la probabilidad.

Si se interpreta la probabilidad como una frecuencia relativa, entonces $P[A|B]$ debería ser la frecuencia relativa del suceso $A \cap B$ en aquellos experimentos donde B ha ocurrido. Supongamos que el experimento se realiza n veces, y supongamos que el suceso B ocurre n_B veces, y que el suceso $A \cap B$ ocurre $n_{A \cap B}$ veces. La frecuencia relativa de interés es entonces

$$\frac{n_{A \cap B}}{n_B} = \frac{n_{A \cap B}/n}{n_B/n} \rightarrow \frac{P[A \cap B]}{P[B]},$$

donde hemos supuesto implícitamente que $P[B] > 0$. Esto va en consonancia con la ecuación (2.19).

Ejemplo 2.15

Se selecciona una bola de una urna que contiene dos bolas negras, numeradas con 1 y 2, y dos bolas blancas, numeradas con 3 y 4. Se anota el número y el color de la bola, por lo que el espacio muestral es $\{(1, n), (2, n), (3, b), (4, b)\}$. Suponiendo que los cuatro resultados son equiprobables, queremos calcular $P[A|B]$ y $P[A|C]$, donde A , B , y C son los siguientes sucesos:

$A = \{(1, n), (2, n)\}$, “se ha seleccionado una bola negra”,

$B = \{(2, n), (4, b)\}$, “se ha seleccionado una bola par”,

$C = \{(3, b), (4, b)\}$, “el número de la bola seleccionada es mayor que 2”.

Dado que $P[A \cap B] = P[(2, n)]$ y $P[A \cap C] = P[\emptyset] = 0$, la ecuación (2.24) queda

$$P[A|B] = \frac{P[A \cap B]}{P[B]} = \frac{0.25}{0.5} = 0.5 = P[A],$$

$$P[A|C] = \frac{P[A \cap C]}{P[C]} = \frac{0}{0.5} = 0 \neq P[A].$$

En el primer caso, el conocimiento de B no alteraría la probabilidad de A . En el segundo caso, el conocimiento de C implicaría que A no ocurriera.

Si multiplicamos ambos lados de la definición de $P[A|B]$ por $P[B]$ se obtiene

$$P[A \cap B] = P[A|B]P[B]. \quad (2.20a)$$

Del mismo modo también tenemos que

$$P[A \cap B] = P[B|A]P[A]. \quad (2.20b)$$

En el siguiente ejemplo se muestra cómo esta ecuación es útil para encontrar probabilidades en experimentos secuenciales. El ejemplo también presenta un diagrama de árbol que facilita el cálculo de probabilidades.

Ejemplo 2.16

Una urna contiene dos bolas negras y tres bolas blancas. Se seleccionan dos bolas al azar sin reemplazamiento y se anota la secuencia de colores. Queremos encontrar la probabilidad de que las dos bolas sean de color negro.

Este experimento está compuesto por una secuencia de dos subexperimentos. Podemos imaginar que nos movemos a través del árbol que se muestra en la figura 2.9, desde el nodo superior hasta uno de los nodos de la parte inferior: alcanzaremos el nodo 1 del árbol si el resultado de la primera selección es una bola negra; entonces el siguiente subexperimento consistirá en seleccionar una bola de una urna que contiene una bola negra y tres bolas blancas. Por otro lado, si el resultado de la primera selección es blanco, entonces llegaremos al nodo 2 del árbol y el segundo subexperimento consistirá en seleccionar una bola de una urna que contiene dos bolas negras y dos bolas blancas. Así, si sabemos qué nodo se alcanza después de la primera selección, entonces se pueden establecer las probabilidades de los resultados en el siguiente subexperimento.

Sean B_1 y B_2 los sucesos tales que el resultado es una bola negra en la primera y la segunda selección, respectivamente. De la ecuación (2.20b) tenemos

$$P[B_1 \cap B_2] = P[B_2|B_1]P[B_1].$$

En términos del diagrama de árbol de la figura 2.8, $P[B_1]$ es la probabilidad de alcanzar el nodo 1 y $P[B_2|B_1]$ es la probabilidad de alcanzar el nodo inferior izquierdo desde el nodo 1. Ahora bien, $P[B_1] = 2/5$ dado que la primera selección se realiza en una urna que contiene dos bolas negras y tres bolas blancas; $P[B_2|B_1] = 1/4$, ya que, dado B_1 , la segunda selección se lleva a cabo en una urna que contiene una bola negra y tres bolas blancas. Por tanto,

$$P[B_1 \cap B_2] = \frac{1}{4} \frac{2}{5} = \frac{1}{10}.$$

En general, la probabilidad de cualquier secuencia de colores se obtiene multiplicando las probabilidades correspondientes a las transiciones entre los nodos del

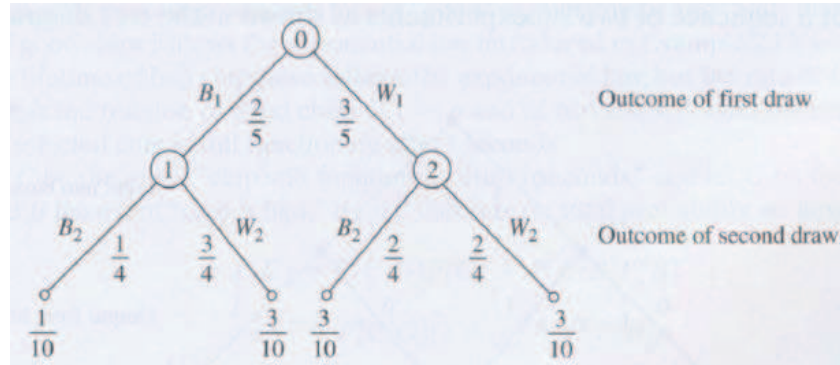


Figura 2.9 Los caminos desde el nodo superior hasta un nodo inferior se corresponden con las posibles respuestas en la extracción de dos bolas de una urna sin reemplazamiento. La probabilidad de un camino es el producto de las probabilidades de las transiciones asociadas.

árbol de la figura 2.9.

Ejemplo 2.17 Sistema binario de comunicación

Muchos de los sistemas de comunicación se puede modelar de la siguiente manera. En primer lugar, el usuario introduce un 0 o un 1 en el sistema, y se transmite una señal correspondiente. En segundo lugar, el receptor toma una decisión acerca de lo que haya sido la entrada al sistema, basándose en la señal que recibe. Supongamos que el usuario envía 0s con probabilidad $1 - p$ y 1s con probabilidad p , y supongamos que el receptor comete errores de decisión aleatorios con una probabilidad ε . Sea A_i el suceso “la entrada era i ”, y sea B_j el suceso “la decisión del receptor fue j ”, con $i = 0, 1$. Queremos encontrar las probabilidades $P[A_i \cap B_j]$ para $i = 0, 1$ y $j = 0, 1$.

El diagrama de árbol para este experimento se muestra en la figura 2.10. Basándonos en él, se pueden obtener fácilmente las probabilidades buscadas.

$$P[A_0 \cap B_0] = (1 - p)(1 - \varepsilon),$$

$$P[A_0 \cap B_1] = (1 - p)\varepsilon,$$

$$P[A_1 \cap B_0] = p\varepsilon,$$

$$P[A_1 \cap B_1] = p(1 - \varepsilon).$$

Sean B_1, B_2, \dots, B_n sucesos mutuamente excluyentes cuya unión es igual al espacio muestral S como se muestra en la figura 2.11. Nos referiremos a estos

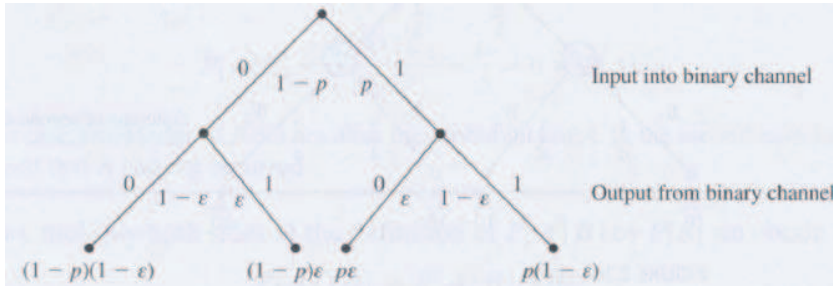


Figura 2.10 Probabilidades de parejas entrada-respuesta en un sistema de transmisión binario.

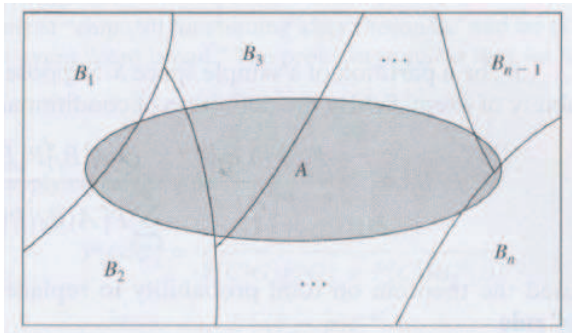


Figura 2.11 Una partición de S en n conjuntos disjuntos.

conjuntos como una partición de S . Cualquier suceso A puede ser representado como la unión de sucesos mutuamente excluyentes de la siguiente manera:

$$\begin{aligned} A &= A \cap S = A \cap (B_1 \cup B_2 \cup \dots \cup B_n) \\ &= (A \cap B_1) \cup (A \cap B_2) \cup \dots \cup (A \cap B_n). \end{aligned}$$

(Ver figura 2.11.) Por el corolario 4, la probabilidad de A es

$$P[A] = P[A \cap B_1] + P[A \cap B_2] + \dots + P[A \cap B_n].$$

Aplicando la ecuación (2.20a) a cada uno de los términos del lado derecho, se obtiene el teorema de la probabilidad total:

$$P[A] = P[A|B_1]P[B_1] + P[A|B_2]P[B_2] + \dots + P[A|B_n]P[B_n]. \quad (2.21)$$

Este resultado es particularmente útil cuando los experimentos se pueden ver como una secuencia de dos subexperimentos como se muestra en el diagrama de árbol de la figura 2.10.

Ejemplo 2.18

En el experimento comentado en el ejemplo 2.16, queremos encontrar la probabilidad del suceso B_2 , que la segunda bola sea de color blanco.

Los sucesos $N_1 = \{(n, n), (n, b)\}$ y $B_1 = \{(b, n), (b, b)\}$ forman una partición del espacio muestral, por lo que aplicando la ecuación (2.21) tenemos

$$P[B_2] = P[B_2|N_1]P[N_1] + P[B_2|B_1]P[B_1] = \frac{3}{4} \frac{2}{5} + \frac{1}{2} \frac{3}{5} = \frac{3}{5}.$$

Es interesante observar que esta probabilidad es la misma que la probabilidad de seleccionar una bola blanca en la primera selección. El resultado tiene sentido porque estamos calculando la probabilidad de una bola blanca en la segunda selección bajo el supuesto de que no tenemos conocimiento del resultado de la primera selección.

Ejemplo 2.19

Un proceso de fabricación produce una mezcla de chips de memoria “buenos” y “malos”. La vida útil de los chips buenos sigue la ley exponencial propuesta en el ejemplo 2.13, con una tasa de fallo α . La vida útil de los chips malos también sigue la ley exponencial, pero la tasa de fallo es 1000α . Supongamos que la fracción de chips buenos es $1 - p$ y de chips malos p . Queremos encontrar la probabilidad de que un chip seleccionado al azar todavía siga funcionando después de t segundos.

Sea C el suceso “el chip sigue funcionando después de t segundos”, y sea B el suceso de “el chip es bueno”, y M el suceso “el chip es malo”. Por el teorema de la probabilidad total tenemos que

$$\begin{aligned} P[C] &= P[C|B]P[B] + P[C|M]P[M] \\ &= P[C|B](1 - p) + P[C|M]p \\ &= (1 - p)e^{-\alpha t} + pe^{-1000\alpha t}. \end{aligned}$$

donde hemos utilizado el hecho de que $P[C|B] = e^{-\alpha t}$ y $P[C|M] = e^{-1000\alpha t}$.

2.3.1 Teorema de Bayes

Sea B_1, B_2, \dots, B_n una partición de un espacio muestral S . Supongamos que ocurre el suceso A ; ¿cuál es la probabilidad del suceso B_j ? Por la definición de probabilidad condicionada tenemos que

$$P[B_j|A] = \frac{P[A \cap B_j]}{P[A]} = \frac{P[A|B_j]P[B_j]}{\sum_{k=1}^n P[A|B_k]P[B_k]}, \quad (2.22)$$

donde hemos utilizado el teorema de la probabilidad total para sustituir $P[A]$. La ecuación (2.22) se conoce como regla o teorema de Bayes.

El teorema de Bayes se aplica a menudo en situaciones como la siguiente. Tenemos un experimento aleatorio en el que los sucesos de interés forman una

partición. Las “probabilidades a priori” de estos sucesos, $P[B_j]$, son las probabilidades de los sucesos antes de que el experimento se lleve a cabo. Supongamos ahora que el experimento se lleva a cabo, y se nos informa de que ha ocurrido un suceso A ; las “probabilidades a posteriori” son las probabilidades de los sucesos en la partición, $P[B_j|A]$, teniendo en cuenta esta nueva información adicional. Los dos ejemplos siguientes ilustran esta situación.

Ejemplo 2.20 Sistema binario de comunicación

En el sistema binario de comunicación del ejemplo 2.17, queremos encontrar qué entrada es más probable sabiendo que el receptor ha emitido como salida un 1. Supongamos que, a priori, la entrada es equiprobable, es decir, tiene la misma probabilidad de ser 0 o 1.

Sea A_k el suceso en el que la entrada era k , con $k = 0, 1$, entonces A_0 y A_1 son una partición del espacio muestral de pares de entrada-salida. Sea B_1 el suceso “la salida del receptor fue un 1”. La probabilidad de B_1 es

$$P[B_1] = P[B_1|A_0]P[A_0] + P[B_1|A_1]P[A_1] = \varepsilon \frac{1}{2} + (1 - \varepsilon) \frac{1}{2} = \frac{1}{2}.$$

Aplicando la regla de Bayes, podemos obtener las probabilidades a posteriori

$$P[A_0|B_1] = \frac{P[B_1|A_0]P[A_0]}{P[B_1]} = \frac{\varepsilon/2}{1/2} = \varepsilon,$$

$$P[A_1|B_1] = \frac{P[B_1|A_1]P[A_1]}{P[B_1]} = \frac{(1 - \varepsilon)/2}{1/2} = (1 - \varepsilon).$$

Por lo tanto, si ε es menor que $1/2$, entonces la entrada 1 es más probable que la entrada 0 cuando se ha observado un 1 en la salida del canal.

Ejemplo 2.21 Control de calidad

Consideremos los chips de memoria estudiados en el ejemplo 2.19. Recordemos que una fracción p de los chips son malos y tienden a fallar mucho más rápido que los chips buenos. Supongamos que con el fin de “cribar” los chips malos, antes de salir de la fábrica, se comprueba cada chip durante t segundos. Los chips que no fallen se descartan y los chips restantes se envían a los clientes. Queremos encontrar el valor de t para el cual el 99% de los chips enviados a los clientes son buenos.

Sea C el suceso “el chip sigue funcionando después de t segundos”, y sean B el suceso “el chip es bueno”, y M el suceso “el chip es malo”. El problema requiere que se encuentre el valor de t para el que

$$P[B|C] = 0.99.$$

Calculamos $P[B|C]$ aplicando la regla de Bayes:

$$\begin{aligned} P[B|C] &= \frac{P[C|B]P[B]}{P[C|B]P[B] + P[C|M]P[M]} \\ &= \frac{(1-p)e^{-\alpha t}}{(1-p)e^{-\alpha t} + pe^{-1000\alpha t}} \\ &= \frac{1}{1 + \frac{pe^{-1000\alpha t}}{(1-p)e^{-\alpha t}}} = 0.99. \end{aligned}$$

La ecuación anterior se puede resolver para t :

$$t = \frac{1}{999\alpha} \ln \left(\frac{99p}{1-p} \right)$$

Por ejemplo, si $1/\alpha = 20000$ horas y $p = 0.10$, entonces $t = 48$ horas.

2.4 Independencia de sucesos

Si el conocimiento de la ocurrencia de un suceso B no altera la probabilidad de otro suceso A , entonces sería natural decir que el suceso A es independiente de B . En términos de probabilidades se produce esta situación cuando

$$P[A] = P[A|B] = \frac{P[A \cap B]}{P[B]}.$$

La ecuación anterior tiene el inconveniente de que el lado derecho no está definido cuando $P[B] = 0$.

Definiremos dos sucesos A y B como independientes si

$$P[A \cap B] = P[A]P[B]. \quad (2.23)$$

La ecuación (2.23) implica entonces tanto

$$P[A|B] = P[A]. \quad (2.24a)$$

como

$$P[B|A] = P[B]. \quad (2.24b)$$

Observemos también que la ecuación (2.24a) implica la ecuación (2.23) cuando $P[B] \neq 0$ y la ecuación (2.24b) implica la ecuación (2.23) cuando $P[A] \neq 0$.

Ejemplo 2.22

Se selecciona una bola de una urna que contiene dos bolas negras, numeradas con 1 y 2, y dos bolas blancas, numeradas con 3 y 4. Se definen lo siguientes

sucesos A , B y C :

$A = \{(1, n), (2, n)\}$, “se ha seleccionado una bola negra”,

$B = \{(2, n), (4, b)\}$, “se ha seleccionado una bola par”,

$C = \{(3, b), (4, b)\}$, “el número de la bola seleccionada es mayor que 2”.

¿Son los sucesos A y B independientes? ¿Son los sucesos A y C independientes?

En primer lugar, consideremos los sucesos A y B . Las probabilidades requeridas por la ecuación (2.23) son

$$P[A] = P[B] = \frac{1}{2},$$

y

$$P[A \cap B] = P[\{(2, n)\}] = \frac{1}{4}.$$

Por tanto

$$P[A \cap B] = \frac{1}{4} = P[A]P[B],$$

y los sucesos A y B son independientes. La ecuación (2.24b) da una visión más clara del significado de independencia:

$$P[A|B] = \frac{P[A \cap B]}{P[B]} = \frac{P[\{(2, n)\}]}{P[\{(2, n), (4, b)\}]} = \frac{1/4}{1/2} = \frac{1}{2},$$

$$P[A] = \frac{P[A]}{P[S]} = \frac{P[\{(1, n), (2, n)\}]}{P[\{(1, n), (2, n), (3, b), (4, b)\}]} = \frac{1/2}{1}.$$

Estas dos ecuaciones implican que $P[A] = P[A|B]$ porque la proporción de resultados en S que conducen a la ocurrencia de A es igual a la proporción de resultados en B que conducen a A . De esta forma, el conocimiento de la ocurrencia de B no altera la probabilidad de la ocurrencia de A .

Los sucesos A y C no son independientes, ya que $P[A \cap C] = P[\emptyset] = 0$, por lo que

$$P[A|C] = 0 \neq P[A] = 0.5.$$

De hecho, A y C son mutuamente excluyentes, ya que $A \cap C = \emptyset$, por lo que la ocurrencia de C implica que definitivamente A no ha ocurrido.

En general, si dos sucesos tienen probabilidad diferente de cero y son mutuamente excluyentes, entonces no pueden ser independientes. Supongamos que fueran independientes y mutuamente excluyentes; entonces

$$0 = P[A \cap B] = P[A]P[B],$$

lo que implicaría que al menos uno de los sucesos tiene probabilidad cero.

Ejemplo 2.23

Se seleccionan dos números x e y al azar entre cero y uno. Se definen los sucesos A, B y C de la siguiente manera:

$$A = \{x < 0.5\}, \quad B = \{y > 0.5\}, \quad C = \{x > y\}.$$

¿Son independientes los sucesos A y B ? ¿Son A y C independientes?

La figura 2.12 muestra las regiones del cuadrado unidad que se corresponden con los sucesos anteriores. Usando la ecuación (2.24a), tenemos

$$P[A|B] = \frac{P[A \cap B]}{P[B]} = \frac{1/4}{1/2} = \frac{1}{2} = P[A],$$

por lo que los sucesos A y B son independientes. Una vez más tenemos que la “proporción” de resultados en S que conducen a A es igual a la “proporción” en B que conducen a A . Usando la ecuación (2.24b), tenemos

$$P[A|C] = \frac{P[A \cap C]}{P[C]} = \frac{3/8}{1/2} = \frac{3}{4} \neq P[A],$$

por lo que los sucesos A y C no son independientes. De hecho en la figura 2.12(b) podemos ver que el conocimiento de el hecho de que x es mayor que y aumenta la probabilidad de que x sea mayor que 0.5.

¿Qué condiciones deben satisfacer tres sucesos A, B y C para ser independientes? En primer lugar, deben ser independientes dos a dos, es decir,

$$P[A \cap B] = P[A]P[B], \quad P[A \cap C] = P[A]P[C], \quad \text{y} \quad P[B \cap C] = P[B]P[C].$$

Además, el conocimiento de la ocurrencia conjunta de cualquier par de ellos, por ejemplo, A y B , no debería afectar la probabilidad del tercero, es decir,

$$P[C|A \cap B] = P[C].$$

Para que esto se verifique, se debe cumplir

$$P[C|A \cap B] = \frac{P[A \cap B \cap C]}{P[A \cap B]} = P[C].$$

Esto a su vez implica que se debe verificar

$$P[A \cap B \cap C] = P[A \cap B]P[C] = P[A]P[B]P[C],$$

donde hemos utilizado el hecho de que A y B son independientes dos a dos. De esta forma, llegamos a la conclusión de que tres sucesos A, B y C son independientes si la probabilidad de la intersección de cualquier par o trío de sucesos es igual al producto de las probabilidades de los sucesos individuales.

El siguiente ejemplo muestra que si tres sucesos son independientes dos a dos, no implica necesariamente que $P[A \cap B \cap C] = P[A \cap B]P[C] = P[A]P[B]P[C]$.

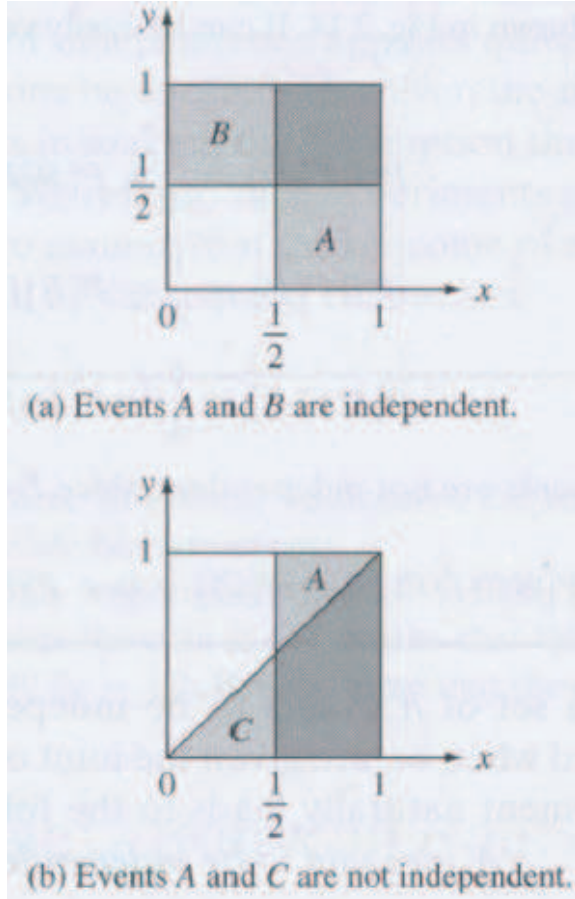


Figura 2.12 Ejemplos de sucesos independientes y dependientes.

Ejemplo 2.24

Consideremos el experimento discutido en el ejemplo 2.23, donde se seleccionan dos números al azar del intervalo unidad. Se definen los siguientes sucesos B , D y F :

$$B = \left\{ y > \frac{1}{2} \right\}, D = \left\{ x < \frac{1}{2} \right\}, F = \left\{ x < \frac{1}{2} \text{ e } y < \frac{1}{2} \right\} \cup \left\{ x > \frac{1}{2} \text{ e } y > \frac{1}{2} \right\}.$$

Los tres sucesos se muestran en la figura 2.13. Se puede verificar fácilmente que cualquier par de estos sucesos es independiente:

$$\begin{aligned} P[B \cap D] &= \frac{1}{4} = P[B]P[D], \\ P[B \cap F] &= \frac{1}{4} = P[B]P[F], \\ P[D \cap F] &= \frac{1}{4} = P[D]P[F]. \end{aligned}$$

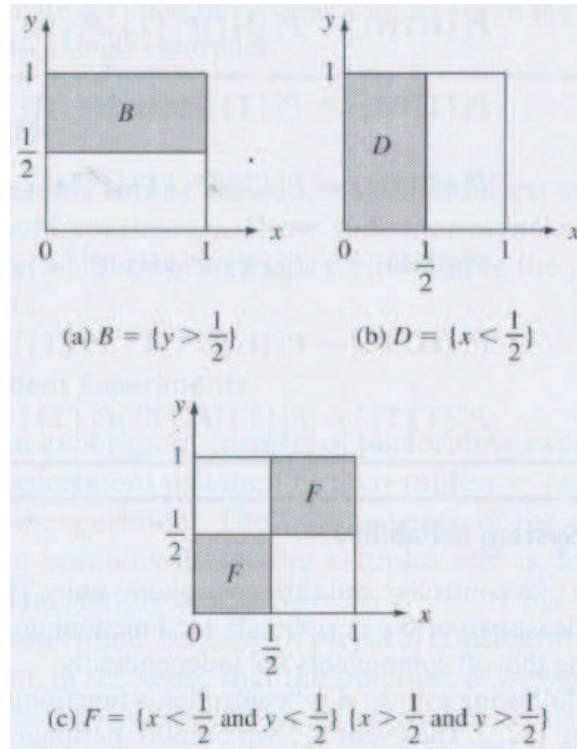


Figura 2.13 Los sucesos B, D y F son independientes dos a dos, pero el triplete B, D, F no está formado por sucesos independientes.

Sin embargo, los tres sucesos no son independientes, ya que $B \cap D \cap F = \emptyset$, por lo que

$$P[B \cap D \cap F] = P[\emptyset] = 0 \neq P[B]P[D]P[F] = \frac{1}{8}.$$

Para que un conjunto de n sucesos sean independientes, la probabilidad de un suceso no debería modificarse cuando se da la ocurrencia conjunta de cualquier subconjunto del resto de sucesos. Este requisito nos lleva de forma natural a la siguiente definición de independencia. Los sucesos A_1, A_2, \dots, A_n se dice que son independientes si para $k = 2, \dots, n$,

$$P[A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}] = P[A_{i_1}]P[A_{i_2}] \dots P[A_{i_k}], \quad (2.25)$$

donde $1 \leq i_1 < i_2 < \dots < i_k \leq n$. Para un conjunto de n sucesos hemos de verificar que las probabilidades de todas las $2^n - n - 1$ intersecciones posibles se factorizan de la forma adecuada.

La definición anterior de independencia parece bastante engorrosa debido a que requiere verificar muchas condiciones. Sin embargo, la aplicación más común del concepto de independencia está en hacer la suposición de que los sucesos de ex-

perimentos separados son independientes. Nos referiremos a tales experimentos como experimentos independientes. Por ejemplo, es común suponer que el resultado del lanzamiento de una moneda es independiente de los resultados de todos los lanzamientos anteriores y posteriores.

Ejemplo 2.25

Supongamos que se lanza una moneda tres veces y se observa la secuencia resultante de caras y cruces. Queremos encontrar la probabilidad de los sucesos elementales.

El espacio muestral de este experimento es

$$S = \{CCC, CCX, CXC, XCC, XXC, XCX, CXX, XXX\}.$$

La suposición de que la moneda está equilibrada significa que los resultados de un único lanzamiento son equiprobables, es decir, $P[C] = P[X] = 1/2$. Si suponemos que los resultados de los lanzamientos de la moneda son independientes, entonces

$$\begin{aligned} P[\{CCC\}] &= P[\{C\}]P[\{C\}]P[\{C\}] = \frac{1}{8}, \\ P[\{CCX\}] &= P[\{C\}]P[\{C\}]P[\{X\}] = \frac{1}{8}, \\ P[\{CXC\}] &= P[\{C\}]P[\{X\}]P[\{C\}] = \frac{1}{8}, \\ P[\{XCC\}] &= P[\{X\}]P[\{C\}]P[\{C\}] = \frac{1}{8}, \\ P[\{XXC\}] &= P[\{X\}]P[\{X\}]P[\{C\}] = \frac{1}{8}, \\ P[\{XCX\}] &= P[\{X\}]P[\{C\}]P[\{X\}] = \frac{1}{8}, \\ P[\{CXX\}] &= P[\{C\}]P[\{X\}]P[\{X\}] = \frac{1}{8}, \\ P[\{XXX\}] &= P[\{X\}]P[\{X\}]P[\{X\}] = \frac{1}{8}. \end{aligned}$$

Ejemplo 2.26 Fiabilidad del sistema

Consideramos un sistema formado por un controlador y tres unidades periféricas. El sistema se dice que está “up” si el controlador y al menos dos de los periféricos están funcionando. Queremos encontrar la probabilidad de que el sistema esté “up”, en el supuesto de que todas las componentes fallen de manera independiente.

Definimos los siguientes sucesos: A es “el controlador está funcionando”, y B_i , es “el periférico i está funcionando”, donde $i = 1, 2, 3$. El suceso F , “dos o más

unidades periféricas están funcionando”, ocurre cuando las tres unidades están funcionando o si exactamente dos unidades están en funcionamiento. Por tanto

$$F = (B_1 \cap B_2 \cap B_3^c) \cup (B_1 \cap B_2^c \cap B_3) \cup (B_1^c \cap B_2 \cap B_3) \cup (B_1 \cap B_2 \cap B_3).$$

Observemos que los sucesos de la anterior unión son mutuamente excluyentes. Así

$$\begin{aligned} P[F] &= P[B_1]P[B_2]P[B_3^c] + P[B_1]P[B_2^c]P[B_3] \\ &\quad + P[B_1^c]P[B_2]P[B_3] + P[B_1]P[B_2]P[B_3] = 3(1-a)^2a + (1-a)^3, \end{aligned}$$

donde hemos supuesto que cada periférico falla con una probabilidad a , de modo que $P[B_i] = 1 - a$ y $P[B_i^c] = a$.

El suceso “el sistema está “up”” es entonces $A \cap F$. Si suponemos que el controlador falla con probabilidad p , entonces

$$\begin{aligned} P[\text{“el sistema está “up””}] &= P[A \cap F] = P[A]P[F] = (1-p)P[F] \\ &= (1-p)\{3(1-a)^2a + (1-a)^3\}. \end{aligned}$$

Sea $a = 10\%$, entonces los tres periféricos están funcionando el $(1-a)^3 = 72.9\%$ del tiempo y dos están funcionando y el otro es “down” el $3(1-a)^2a = 24.3\%$ del tiempo. De esta forma, dos o más periféricos están funcionando el 97.2% del tiempo. Supongamos que el controlador no es muy fiable, por ejemplo que $p = 20\%$, entonces el sistema está “up” sólo el 77.8% del tiempo, debido en gran medida a los fallos del controlador.

Supongamos que se añade al sistema un segundo controlador idéntico con $p = 20\%$, y que el sistema está “up” si al menos uno de los controladores está funcionando y si dos o más de los periféricos están en funcionamiento. Se puede demostrar que al menos uno de los controladores está funcionando el 96% del tiempo, y que el sistema está “up” el $93,3\%$ del tiempo. Este significa un aumento del 16% con respecto al sistema con un solo controlador.

2.5 Experimentos secuenciales

Muchos experimentos aleatorios pueden modelarse como experimentos secuenciales que consisten en una sucesión de subexperimentos más simples. Estos subexperimentos pueden ser independientes o no. En esta sección se estudian métodos para obtener las probabilidades de sucesos en experimentos secuenciales.

2.5.1 Sucesiones de experimentos independientes

Supongamos que un experimento aleatorio consiste en la realización de los experimentos E_1, E_2, \dots, E_n . El resultado de este experimento será entonces una n -tupla $s = (s_1, s_2, \dots, s_n)$, donde s_k es el resultado del subexperimento k -ésimo.

El espacio muestral del experimento secuencial se define como el conjunto que contiene las anteriores n -tuplas y se denota como el producto cartesiano de los espacios muestrales individuales $S_1 \times S_2 \times \cdots \times S_n$.

Por lo general, se puede determinar, debido a consideraciones físicas, cuando los subexperimentos son independientes, en el sentido de que el resultado de cualquier subexperimento dado no puede afectar a los resultados de los otros subexperimentos. Sean A_1, A_2, \dots, A_n sucesos tales que A_k sólo tiene que ver con los resultados del k -ésimo subexperimento. Si los subexperimentos son independientes, entonces es razonable suponer que los sucesos anteriores A_1, A_2, \dots, A_n sean independientes. De esta forma,

$$P[A_1 \cap A_2 \cap \cdots \cap A_n] = P[A_1]P[A_2] \dots P[A_n]. \quad (2.26)$$

Esta expresión nos permite calcular todas las probabilidades de sucesos del experimento secuencial.

Ejemplo 2.27

Supongamos que se seleccionan 10 números al azar en el intervalo $[0, 1]$. Queremos encontrar la probabilidad de que los 5 primeros números sean menores que $1/4$ y los 5 últimos números sean mayores que $1/2$. Sea x_1, x_2, \dots, x_{10} la secuencia de 10 números, entonces los sucesos de interés son

$$A_k = \left\{ x_k < \frac{1}{4} \right\} \quad \text{para } k = 1, \dots, 5,$$

$$A_k = \left\{ x_k > \frac{1}{2} \right\} \quad \text{para } k = 6, \dots, 10.$$

Si suponemos que cada selección de un número es independiente de las otras selecciones, entonces

$$P[A_1 \cap A_2 \cap \cdots \cap A_{10}] = P[A_1]P[A_2] \dots P[A_{10}] = \left(\frac{1}{4}\right)^5 \left(\frac{1}{2}\right)^5.$$

A continuación se deducen varios modelos importantes para los experimentos que consisten en secuencias de subexperimentos independientes.

2.5.2 La ley de probabilidad binomial

Un ensayo o prueba de Bernoulli implica realizar un experimento una vez y anotar si ocurre un determinado suceso A . El resultado de la prueba de Bernoulli se dice que es un “éxito” si ocurre A y un “fracaso” en otro caso. En esta sección estamos interesados en encontrar la probabilidad de k éxitos en n repeticiones independientes de un ensayo de Bernoulli.

Podemos ver el resultado de un único ensayo de Bernoulli como el resultado de un lanzamiento de una moneda para el cual la probabilidad de caras (éxitos) es $p = P[A]$. La probabilidad de k éxitos en n ensayos de Bernoulli es entonces igual a la probabilidad de k caras en n lanzamientos de la moneda.

Ejemplo 2.28

Supongamos que se lanza una moneda tres veces. Si asumimos que los lanzamientos son independientes y que la probabilidad de obtener caras es p , entonces la probabilidad para las secuencias de caras y cruces es

$$\begin{aligned} P[\{CCC\}] &= P[\{C\}]P[\{C\}]P[\{C\}] = p^3, \\ P[\{CCX\}] &= P[\{C\}]P[\{C\}]P[\{X\}] = p^2(1-p), \\ P[\{CXC\}] &= P[\{C\}]P[\{X\}]P[\{C\}] = p^2(1-p), \\ P[\{XCC\}] &= P[\{X\}]P[\{C\}]P[\{C\}] = p^2(1-p), \\ P[\{XXC\}] &= P[\{X\}]P[\{X\}]P[\{C\}] = p(1-p)^2, \\ P[\{XCX\}] &= P[\{X\}]P[\{C\}]P[\{X\}] = p(1-p)^2, \\ P[\{CXX\}] &= P[\{C\}]P[\{X\}]P[\{X\}] = p(1-p)^2, \\ P[\{XXX\}] &= P[\{X\}]P[\{X\}]P[\{X\}] = (1-p)^3. \end{aligned}$$

donde hemos utilizado el hecho de que los lanzamientos son independientes. Sea k el número de caras en tres ensayos, entonces

$$\begin{aligned} P[k=0] &= P[\{XXX\}] = (1-p)^3, \\ P[k=1] &= P[\{XXC, XCX, CXX\}] = 3p(1-p)^2, \\ P[k=2] &= P[\{CCX, CXC, XCC\}] = 3p^2(1-p), \\ P[k=3] &= P[\{CCC\}] = p^3. \end{aligned}$$

El resultado del ejemplo 2.28 es el caso $n = 3$ de la ley de probabilidad binomial.

TEOREMA

Sea k el número de éxitos en n ensayos independientes de Bernoulli, entonces las probabilidades de k vienen dadas por la ley de probabilidad binomial:

$$p_n(k) = \binom{n}{k} p^k (1-p)^{n-k} \quad \text{para } k = 0, \dots, n, \quad (2.27)$$

donde $p_n(k)$ es la probabilidad de k éxitos en n ensayos, y

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \quad (2.28)$$

es el coeficiente binomial.

El término $n!$ en la ecuación (2.28) se llama n factorial y se define por $n! = n(n-1)\dots(2)(1)$. Por definición $0!$ es igual a 1.

A continuación probamos el teorema anterior. Siguiendo el ejemplo 2.26, vemos que cada uno de las secuencias con k éxitos y $n-k$ fracasos tiene la misma probabilidad, es decir, $p^k(1-p)^{n-k}$. Sea $N_n(k)$ el número de secuencias distintas que tienen k éxitos y $n-k$ fracasos, entonces

$$p_n(k) = N_n(k)p^k(1-p)^{n-k}. \quad (2.29)$$

La expresión $N_n(k)$ es el número de formas de elegir para los éxitos k posiciones de entre n . Se puede demostrar que

$$N_n(k) = \binom{n}{k}. \quad (2.30)$$

El teorema se obtiene sustituyendo la ecuación (2.30) en la ecuación (2.29).

Ejemplo 2.29

Queremos verificar que la ecuación (2.27) proporciona las probabilidades encontradas en el ejemplo 2.28.

En el ejemplo 2.28, suponemos que el suceso “el lanzamiento resulta ser cara” se corresponde con un “éxito”, entonces

$$\begin{aligned} p_3(0) &= \frac{3!}{0!3!}p^0(1-p)^3 = (1-p)^3, \\ p_3(1) &= \frac{3!}{1!2!}p^1(1-p)^2 = 3p(1-p)^2, \\ p_3(2) &= \frac{3!}{2!1!}p^2(1-p)^1 = 3p^2(1-p), \\ p_3(3) &= \frac{3!}{3!0!}p^3(1-p)^0 = p^3, \end{aligned}$$

lo que concuerda con nuestros resultados anteriores.

El coeficiente binomial aparece en el cálculo de binomios de la forma

$$(a+b)^n = \sum_{k=0}^n \binom{n}{k} a^k b^{n-k}. \quad (2.31a)$$

Si tomamos $a = b = 1$, entonces

$$2^n = \sum_{k=0}^n \binom{n}{k} = \sum_{k=0}^n N_n(k),$$

lo que concuerda con el hecho de que hay 2^n posibles secuencias distintas de

éxitos y fracasos en n ensayos. Si tomamos $a = p$ y $b = 1 - p$ en la ecuación (2.31a), obtenemos

$$1 = \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=0}^n p_n(k), \quad (2.31b)$$

lo que confirma que las probabilidades de la binomial suman 1.

El término $n!$ crece muy rápidamente a medida que aumenta n , por lo que se encuentran problemas numéricos para valores relativamente pequeños de n si se intenta calcular $p_n(k)$ directamente a través de la ecuación (2.27). La siguiente fórmula recursiva evita la evaluación directa de $n!$ y por lo tanto amplía el rango de n para el que se puede calcular $p_n(k)$ antes de encontrarse con dificultades numéricas:

$$p_n(k+1) = \frac{(n-k)p}{(k+1)(1-p)} p_n(k). \quad (2.32)$$

Más adelante, en el manual, se presentan dos aproximaciones de las probabilidades binomiales para el caso en el que n es grande.

Ejemplo 2.30

Sea k el número de altavoces activos (que no están en silencio) de un grupo de ocho altavoces que no interactúan (es decir, independientes). Supongamos que un altavoz se activa con una probabilidad de $1/3$. Queremos encontrar la probabilidad de que el número de altavoces activos sea superior a seis.

Para $i = 1, \dots, 8$, sea A_i el suceso “el i -ésimo altavoz está activo”. El número de altavoces activos es entonces el número de éxitos en ocho ensayos de Bernoulli con $p = 1/3$. Así, la probabilidad de que más de seis altavoces estén activos es

$$\begin{aligned} P[k = 7] + P[k = 8] &= \binom{8}{7} \left(\frac{1}{3}\right)^7 \left(\frac{2}{3}\right) + \binom{8}{8} \left(\frac{1}{3}\right)^8 \\ &= 0.00244 + 0.00015 = 0.00259 \end{aligned}$$

Ejemplo 2.31 Codificación de corrección de errores

Un sistema de comunicación transmite información binaria a través de un canal que introduce bits erróneos aleatorios con una probabilidad $\varepsilon = 10^{-3}$. El transmisor transmite cada bit de información tres veces, y un decodificador usa un sistema de mayoría de votos sobre los bits recibidos para decidir cuál es el bit transmitido. Queremos encontrar la probabilidad de que el receptor tome una decisión incorrecta.

El receptor puede corregir un error simple, pero tomará una decisión equivocada si el canal presenta dos o más errores. Si consideramos cada transmisión como un ensayo de Bernoulli en el que un “éxito” se corresponde a la introducción de

un error, entonces la probabilidad de dos o más errores en tres ensayos Bernoulli es

$$P[k \geq 2] = \binom{3}{2} 0.001^2 0.999 + \binom{3}{3} 0.001^3 \simeq 3(10^6).$$

2.5.3 La ley de probabilidad multinomial

La ley de probabilidad binomial se puede generalizar al caso en que se observa la ocurrencia de más de un suceso. Sea B_1, B_2, \dots, B_M una partición del espacio muestral S de un experimento aleatorio y sea $P[B_j] = p_j$. Los sucesos son mutuamente excluyentes, por lo que

$$p_1 + p_2 + \dots + p_M = 1.$$

Supongamos que se llevan a cabo n repeticiones independientes del experimento. Sea k_j el número de veces que ocurre el suceso B_j , entonces el vector (k_1, k_2, \dots, k_M) especifica el número de veces que ocurre cada uno de los sucesos B_j . La probabilidad del vector (k_1, \dots, k_M) cumple la ley de probabilidad multinomial:

$$P[(k_1, k_2, \dots, k_M)] = \frac{n!}{k_1! k_2! \dots k_M!} p_1^{k_1} p_2^{k_2} \dots p_M^{k_M}, \quad (2.33)$$

donde $k_1 + k_2 + \dots + k_M = n$. La ley de probabilidad binomial es el caso $M = 2$ de la ley de probabilidad multinomial. La deducción de las probabilidades multinomiales es idéntica a la de las probabilidades binomiales. Tan sólo hay que observar que el número de secuencias diferentes con k_1, k_2, \dots, k_M instancias de sucesos B_1, B_2, \dots, B_M está dada por el coeficiente multinomial de la ecuación (2.26).

Ejemplo 2.32

Se lanza un dardo nueve veces a un blanco que consta de tres áreas. Cada lanzamiento tiene una probabilidad de 0.2, 0.3 y 0.5 de impactar en las áreas 1, 2 y 3, respectivamente. Queremos encontrar la probabilidad de que el dardo caiga exactamente tres veces en cada una de las áreas.

Este experimento consta de nueve repeticiones independientes de un subexperimento que tiene tres resultados posibles. La probabilidad del número de ocurrencias de cada resultado viene dada por las probabilidades multinomiales con parámetros $n = 9$ y $p_1 = 0.2$, $p_2 = 0.3$, y $p_3 = 0.5$:

$$P[(3, 3, 3)] = \frac{9!}{3!3!3!} (0.2)^3 (0.3)^3 (0.5)^3 = 0.04536.$$

Ejemplo 2.33

Supongamos que escogemos al azar 10 números de teléfono de una guía telefónica y anotamos el último dígito de cada uno de los números. ¿Cuál es la probabilidad de que se obtenga cada uno de los números enteros del 0 al 9 sólo una vez?

Las probabilidades para el número de ocurrencias de los números enteros están dadas por las probabilidades multinomiales con parámetros $M = 10$, $n = 10$, y $p_j = 1/10$, siempre que supongamos que los diez enteros en el rango de 0 a 9 son equiprobables. La probabilidad de obtener una vez cada número entero en 10 intentos será entonces

$$\frac{10!}{1!1! \dots 1!} (0.1)^{10} \simeq 3.6(10^{-4}).$$

2.5.4 La ley de probabilidad geométrica

Consideramos un experimento secuencial en el que repetimos ensayos independientes de Bernoulli hasta la aparición del primer éxito. Sea m el resultado de este experimento, el número de ensayos realizados hasta la aparición del primer éxito. El espacio muestral para este experimento es el conjunto de los enteros positivos. La probabilidad, $p(m)$, de que se requieran m ensayos se calcula observando que esto sólo puede ocurrir si los primeros $m - 1$ ensayos han resultado ser fracasos y el m -ésimo ensayo ha resultado ser un éxito. La probabilidad de este suceso es

$$p(m) = P[A_1^c A_2^c \dots A_{m-1}^c A_m] = (1 - p)^{m-1} p \quad m = 1, 2, \dots \quad (2.34a)$$

donde A_i es el suceso “éxito en el i -ésimo ensayo”. La asignación de probabilidad especificada por la ecuación (2.34a) se llama ley de probabilidad geométrica.

Las probabilidades de la ecuación (2.34a) suman 1:

$$\sum_{m=1}^{\infty} p(m) = p \sum_{m=1}^{\infty} q^{m-1} = p \frac{1}{1 - q} = 1, \quad (2.34b)$$

donde $q = 1 - p$, y donde hemos utilizado la fórmula de la suma de una serie geométrica. La probabilidad de que se requieran más de K ensayos antes de que ocurra el éxito se obtiene de forma sencilla:

$$P\{m > K\} = p \sum_{m=K+1}^{\infty} q^{m-1} = pq^K \sum_{j=0}^{\infty} q^j = pq^K \frac{1}{1 - q} = q^K. \quad (2.35)$$

Ejemplo 2.34 Control del error mediante retransmisión

La computadora A envía un mensaje a la computadora B a través de un enlace de radio no fiable. Se codifica el mensaje de tal forma que B puede detectar

cuando se ha introducido errores en el mensaje durante la transmisión. Si B detecta un error, le pide a A que lo retransmita. Si la probabilidad de error en la transmisión de un mensaje es $q = 0.1$, ¿cuál es la probabilidad de que se necesite transmitir un mensaje más de dos veces?

Cada transmisión de un mensaje es una prueba de Bernoulli con probabilidad de éxito $p = 1 - q$. Los ensayos de Bernoulli se repiten hasta el primer éxito (transmisión sin errores). La probabilidad de que se requieran más de dos transmisiones vendrá dada por la ecuación (2.35):

$$P[\{m > 2\}] = q^2 = 10^{-2}.$$

2.5.5 Secuencias de experimentos dependientes

En este apartado se considera una secuencia o “cadena” de subexperimentos en la que el resultado de un subexperimento dado determina qué subexperimento se lleva a cabo a continuación. En primer lugar, vemos un ejemplo sencillo de este tipo de experimento y mostramos cómo se pueden utilizar diagramas para especificar el espacio muestral.

Ejemplo 2.35

Un determinado experimento secuencial consiste en, repetidas veces, sacar una bola de una de dos urnas, anotar el número de la bola y reponer la bola en su urna. La urna 0 contiene una bola con el número 1 y dos bolas con el número 0, y la urna 1 contiene cinco bolas con el número 1 y una bola con el número 0. La urna en la que se hace la primera extracción se selecciona aleatoriamente lanzando una moneda al aire. Se usa la urna 0 si el resultado es cara y la urna 1 si el resultado es cruz. Posteriormente, la urna utilizada en un subexperimento se corresponderá con el número de la bola elegida en el anterior subexperimento.

El espacio muestral de este experimento está formado por secuencias de 0s y 1s. Cada secuencia posible se corresponde con un camino a través del diagrama “enrejado” que se muestra en la figura 2.14(a). Los nodos del diagrama denotan la urna utilizada en el n -ésimo subexperimento, y las etiquetas en las ramas denotan el resultado de un subexperimento. De este modo el camino 0011 se corresponde con la secuencia: El lanzamiento de la moneda fue cara de modo que la primera extracción fue de la urna 0; el resultado de la primera extracción fue 0, por lo que la segunda extracción fue de la urna 0; el resultado de la segunda extracción fue 1, por lo que la tercera extracción fue de la urna 1; y el resultado de la tercera extracción fue 1, por lo que la cuarta extracción es de la urna 1.

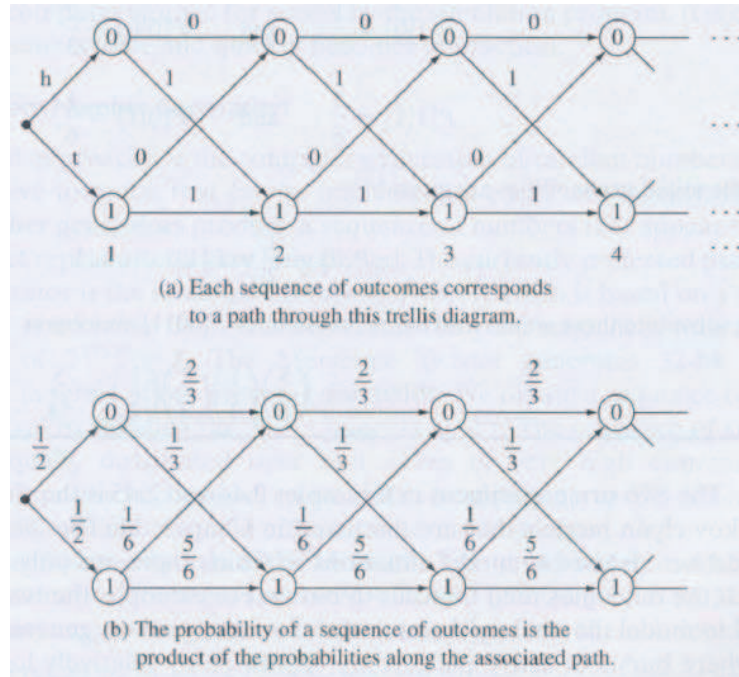


Figura 2.14 Diagrama de enrejado de una cadena de Markov.

Supongamos ahora que queremos calcular la probabilidad de una secuencia particular de resultados, por ejemplo, s_0, s_1, s_2 . Denotamos esta probabilidad como $P[\{s_0\} \cap \{s_1\} \cap \{s_2\}]$. Sea $A = \{s_2\}$ y $B = \{s_0\} \cap \{s_1\}$ entonces, dado que $P[A \cap B] = P[A|B]P[B]$ tenemos que

$$\begin{aligned} P[\{s_0\} \cap \{s_1\} \cap \{s_2\}] &= P[\{s_2\}|\{s_0\} \cap \{s_1\}]P[\{s_0\} \cap \{s_1\}] \\ &= P[\{s_2\}|\{s_0\} \cap \{s_1\}]P[\{s_1\}|\{s_0\}]P[\{s_0\}]. \end{aligned} \quad (2.36)$$

Observemos ahora que en el ejemplo anterior de la urna la probabilidad

$$P[\{s_n\}|\{s_0\} \cap \dots \cap \{s_{n-1}\}]$$

sólo depende de $\{s_{n-1}\}$ ya que el resultado más reciente determina qué subexperimento se lleva a cabo:

$$P[\{s_n\}|\{s_0\} \cap \dots \cap \{s_{n-1}\}] = P[\{s_n\}|\{s_{n-1}\}]. \quad (2.37)$$

Por lo tanto, para la secuencia de interés se tiene que

$$P[\{s_0\} \cap \{s_1\} \cap \{s_2\}] = P[\{s_2\}|\{s_1\}]P[\{s_1\}|\{s_0\}]P[\{s_0\}]. \quad (2.38)$$

Los experimentos secuenciales que satisfacen la ecuación (2.37) se denominan cadenas de Markov. Para estos experimentos, la probabilidad de una secuencia

s_0, s_1, \dots, s_m viene dada por

$$P[s_0, s_1, \dots, s_n] = P[s_n|s_{n-1}]P[s_{n-1}|s_{n-2}] \dots P[s_1|s_0]P[s_0]. \quad (2.39)$$

donde hemos simplificado la notación omitiendo las llaves. Por lo tanto, la probabilidad de la secuencia s_0, \dots, s_n está dada por el producto de la probabilidad del primer resultado s_0 y las probabilidades de todas las transiciones posteriores, de s_0 a s_1 , de s_1 a s_2 , y así sucesivamente. En el capítulo ?? se abordan las cadenas de Markov.

Ejemplo 2.36

Queremos encontrar la probabilidad de la secuencia 0011 para el experimento de la urna introducido en el ejemplo 2.35.

Hay que recordar que la urna 0 contiene dos bolas con la etiqueta 0 y una bola con la etiqueta 1, y que la urna 1 contiene cinco bolas con la etiqueta 1 y una bola con la etiqueta 0. Fácilmente podemos calcular las probabilidades de las secuencias de los resultados mediante el etiquetado de las ramas en el diagrama de enrejado con la probabilidad de la correspondiente transición, como se muestra en la figura 2.14(b). De esta forma, la probabilidad de la secuencia 0011 está dada por

$$P[0011] = P[1|1]P[1|0]P[0|0]P[0],$$

donde las probabilidades de transición vienen dadas por

$$\begin{aligned} P[1|0] &= \frac{1}{3}, & P[0|0] &= \frac{2}{3}, \\ P[1|1] &= \frac{5}{6}, & P[0|1] &= \frac{1}{6}, \end{aligned}$$

y las probabilidades iniciales vienen dadas por

$$P[0] = \frac{1}{2} = P[1].$$

Si sustituimos estos valores en la expresión de $P[0011]$, obtenemos

$$P[0011] = \left(\frac{5}{6}\right) \left(\frac{1}{3}\right) \left(\frac{2}{3}\right) \left(\frac{1}{2}\right) = \frac{5}{54}.$$

El experimento de la doble urna en los ejemplos 2.35 y 2.36 es el ejemplo más simple de los modelos de cadena de Markov que se discuten en el capítulo ?. El experimento de la doble urna discutido aquí se utiliza para modelar situaciones en las que sólo hay dos resultados, y en que los resultados tienden a ocurrir en ráfagas. Por ejemplo, el modelo de dos urnas se usa para modelar la el comportamiento en “ráfaga” de los paquetes de voz generada por un solo altavoz donde las ráfagas de paquetes activos están separados por períodos relativamente largos de

silencio. El modelo también ha sido usado para la secuencia de puntos en blanco y negro que se derivan del escaneado, línea por línea, de imágenes en blanco y negro.

Problemas

2.1 El minutero (aflojado) de un reloj gira con dificultad y se anota la hora a la que la manecilla se detiene.

- (a) ¿Cuál es el espacio muestral?
- (b) Encuentra los conjuntos correspondientes a los sucesos: $A =$ “la manecilla se detiene en las primeras 4 horas”, $B =$ “la manecilla se detiene entre la segunda y octava hora (ambas incluidas)”, y $D =$ “la manecilla se detiene en una hora impar”.
- (c) Encuentra los sucesos: $A \cap B \cap D$, $A^c \cap B$, $A \cup (B \cap D^c)$, $(A \cup B) \cap D^c$.

2.2 Se lanza un dado dos veces y se cuenta y se anota, en el orden de aparición, el número de puntos que aparecen en cada lanzamiento.

- (a) Encuentra el espacio muestral.
- (b) Encuentra el conjunto A correspondiente al suceso “el número de puntos en el primer lanzamiento no es menor que el número de puntos en el segundo lanzamiento”.
- (c) Encuentra el conjunto B correspondiente al suceso “el número de puntos en el primer lanzamiento es 6”.
- (d) ¿ A implica B o B implica A ?
- (e) Encuentra $A \cap B^c$ y describe este suceso con palabras.
- (f) Sea C el suceso “la diferencia de puntos entre los dos dados es 2”. Encuentra $A \cap C$.

2.3 Se lanzan dos dados y se anota la magnitud de la diferencia en el número de puntos obtenidos en los dos dados.

- (a) Encuentra el espacio muestral.
- (b) Encuentra el conjunto A correspondiente al suceso “la magnitud de la diferencia es 3”.
- (c) Expresa cada uno de los sucesos elementales de este experimento como unión de sucesos elementales del problema 2.

2.4 Un sistema binario de comunicación transmite una señal X que o bien es una señal de tensión $+2$ o bien una señal de tensión -2 . Un canal malicioso reduce la magnitud de la señal recibida usando como referencia el número de caras que obtiene en dos lanzamientos de una moneda. Sea Y la señal resultante.

- (a) Encuentra el espacio muestral.
- (b) Encuentra el conjunto de respuestas correspondiente al suceso “la señal transmitida fue $+2$ ”.
- (c) Describe con palabras el suceso correspondiente a la respuesta $Y = 0$.

2.5 Un cajón contiene seis plumas estilográficas, cuatro de las cuales están secas.

- (a) Las plumas son seleccionados al azar, una por una hasta se encuentra una pluma buena. ¿Cuál es el espacio muestral?
- (b) Supongamos que se anota sólo el número, y no la secuencia, de las plumas probadas en el apartado (a). Especifica el espacio muestral.
- (c) Supongamos que las plumas son seleccionadas una por una y se prueban hasta que las dos plumas buenas sean identificadas, y se anota la secuencia de resultados. ¿Cuál es el espacio muestral?
- (d) Especifica el espacio muestral del apartado (c) si sólo se anota el número de plumas probadas.

2.6 Tres amigos (Álvaro, Benito y Carlos) ponen sus nombres en un sombrero y cada uno saca un nombre del sombrero. Suponemos Álvaro elige primero, luego Benito y luego Carlos.

- (a) Encuentra el espacio muestral.
- (b) Encuentra los conjuntos A , B y C que se corresponden con los sucesos “Álvaro elige su propio nombre”, “Benito elige su propio nombre” y “Carlos elige su propio nombre”,
- (c) Encuentra el conjunto correspondiente al suceso, “nadie elige su propio nombre”.
- (d) Encuentra el conjunto correspondiente al suceso, “todos eligen su propio nombre”.
- (e) Encuentra el conjunto correspondiente al suceso, “uno o más eligen su propio nombre”.

2.7 Sea M el número de transmisiones de mensajes en el experimento E_6 .

- (a) ¿Cuál es el conjunto A correspondiente al suceso “ M es par”?
- (b) ¿Cuál es el conjunto B correspondiente al suceso “ M es un múltiplo de 3”?
- (c) ¿Cuál es el conjunto C correspondiente al suceso “se necesitan 6 o menos transmisiones”?
- (d) Encuentra los conjuntos $A \cap B$, $A - B$, $A \cap B \cap C$ y describe los sucesos correspondientes con palabras.

2.8 Se elige un número U al azar en el intervalo unidad. Sean A y B los sucesos: $A = “U$ difiere de $1/2$ en más de $1/4”$ y $B = “1 - U$ es menor que $1/2”$. Encuentra los sucesos $A \cap B$, $A^c \cap B$ y $A \cup B$.

2.9 El espacio muestral de un experimento es la recta real. Sean A y B los sucesos correspondientes a los siguientes subconjuntos de la recta real: $A = (-\infty, r]$ y $B = (-\infty, s]$, donde $r \leq s$. Busca una expresión para el suceso $C = (r, s]$ en términos de A y B . Demuestra que $B = A \cup C$ y $A \cap C = \emptyset$.

2.10 Utiliza los diagramas de Venn para verificar las identidades de conjuntos dadas en las ecuaciones (2.2) y (2.3). Será útil usar diferentes colores o sombreados diferentes para denotar las regiones diferentes con claridad.

2.11 Demuestra que:

- (a) Si el suceso A implica B y B implica C , entonces A implica C .
- (b) Si el suceso A implica B , entonces B^c implica A^c .

2.12 Demuestra que si $A \cup B = A$ y $A \cap B = A$ entonces $A = B$.

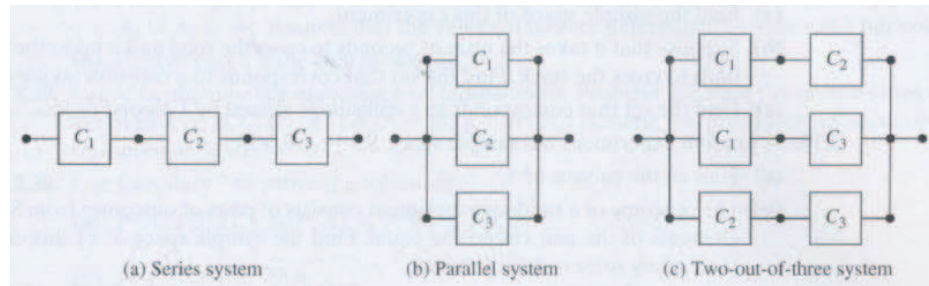


Figura 2.15

2.13 Sean A y B dos sucesos. Encuentra una expresión para el suceso “exactamente uno de los sucesos A y B ocurre.” Dibuja un diagrama de Venn para este suceso.

2.14 Sean A, B , y C sucesos. Encuentra expresiones para los siguientes sucesos:

- Exactamente ocurre uno de los tres sucesos.
- Exactamente ocurren dos de los sucesos.
- Ocurren uno o más sucesos.
- Ocurren dos o más sucesos.
- No ocurre ninguno de los sucesos.

2.15 La figura 2.15 muestra tres sistemas de tres componentes C_1 , C_2 y C_3 . La figura 2.15(a) es un sistema “en serie” en el que el sistema funciona sólo si los tres componentes están funcionando. La figura 2.15(b) es un sistema “en paralelo” en el que el sistema funciona siempre y cuando al menos uno de los tres componentes esté funcionando. La figura 2.15(c) es un sistema “dos de tres” en el que el sistema funciona siempre y cuando al menos dos componentes estén funcionando. Sea A_k el suceso “la componente k está funcionando”. Para cada una de las tres configuraciones de sistema, expresa el suceso “el sistema está funcionando” en términos de A_k .

2.16 Un sistema tiene dos subsistemas clave. El sistema está “up”, si sus dos subsistemas están funcionando. Los sistemas de triple redundancia se configuran para proporcionar una alta fiabilidad. El sistema general está operativo (en funcionamiento), siempre y cuando uno de los tres sistemas esté “up”. Sea A_{jk} el suceso correspondiente al suceso “la unidad k en el sistema j está funcionando”, para $j = 1, 2, 3$ y $k = 1, 2$.

- Escribe una expresión para el suceso “el sistema general está “up””.
- Explica por qué el problema anterior es equivalente al problema de tener una conexión en la red de conmutadores que aparecen en la figura 2.16.

2.17 Entre las 6am y las 6am (periodo de 24h), un estudiante se levanta en el instante de tiempo t_1 y se va a dormir en un instante de tiempo posterior, t_2 .

- El resultado del experimento es el par (t_1, t_2) . Encuentra el espacio muestral y representalo en el plano $x - y$.

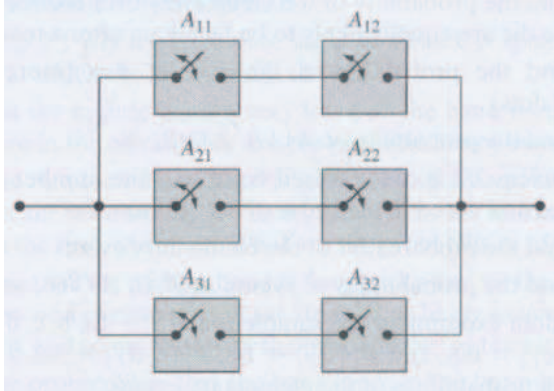


Figura 2.16

- (b) Especifica el suceso $A \equiv$ “el estudiante está dormido a mediodía” y dibuja la región del plano correspondiente.
- (c) Especifica el suceso $B \equiv$ “el estudiante duerme de 7 a 9 am” y dibuja la región del plano correspondiente.
- (c) Dibuja la región correspondiente a $A \cap B$ y describe el suceso correspondiente en palabras.

2.18 Se tira un dado y se anota el número que sale.

- (a) Calcula la probabilidad de los resultados elementales del experimento, suponiendo que el dado está equilibrado.
- (b) Calcula la probabilidad de los sucesos $A = \{\text{más de } 3\}$ y $B = \{\text{número impar}\}$.
- (c) Calcula la probabilidad de $A \cup B$, $A \cap B$ y A^c .

2.19 Demuestra que

$$P[A \cup B \cup C] = P[A] + P[B] + P[C] - P[A \cap B] - P[A \cap C] - P[B \cap C] + P[A \cap B \cap C].$$

2.20 Sea M el número de transmisiones de mensajes del Experimento E_6 . Sean los experimentos:

$A \equiv$ “ M es par”

$B \equiv$ “ M es múltiplo de 3”

$C \equiv$ “se requieren 6 o menos transmisiones”

Calcula la probabilidad de los sucesos A , B , C , C^c , $A \cap B$, $A - B$, $A \cap B \cap C$. Asume que la probabilidad de que una transmisión sea exitosa es $1/2$.

2.21 Se selecciona un número x en el intervalo $[-1, 2]$. Sea $A = \{x < 0\}$, $B = \{|x - 0.5| < 0.5\}$ y $C = \{x > 0.75\}$.

- (a) Calcula la probabilidad de A , B , $A \cap B$ y $A \cap C$.
- (b) Calcula la probabilidad de $A \cup B$, $A \cup C$ y $A \cup B \cup C$ primero, calculando

los sucesos y luego las probabilidades, y segundo, utilizando los axiomas y corolarios apropiados.

2.22 Se lanza un dado dos veces y se anota los números en orden de ocurrencia. Sea A el suceso “el número en la primera tirada no es menor que el número de la segunda tirada” y sea B el suceso “el número en la primera tirada es un seis”. Calcula $P[A|B]$ y $P[B|A]$.

2.23 Utiliza la probabilidad condicionada y un diagrama de árbol para calcular las probabilidades de los resultados elementales de los experimentos aleatorios definidos en los apartados (a)-(d) del problema 2.5.

2.24 (a) Calcula $P[A|B]$ si $A \cap B = \emptyset$; si $A \subset B$; si $B \subset A$.

(b) Demuestra que si $P[A|B] > P[A]$, entonces $P[B|A] > P[B]$.

2.25 En cada lote de 100 productos, se prueban 2 y el lote es rechazado si alguno de los productos probados resulta defectuoso.

(a) Calcula la probabilidad de que un lote con k productos defectuosos sea aceptado.

(b) Supongamos que cuando el proceso de producción no funciona bien, 50 de 100 productos son defectuosos. Con el objetivo de identificar cuando funciona mal el proceso, ¿Cuántos productos tienen que ser probados para que la probabilidad de que uno o más productos resulten defectuosos sea al menos 99 %?

2.26 El transmisor del problema 2.4 transmite $X = +2$ y $X = -2$ con igual probabilidad. El canal malicioso cuenta el número de caras en dos lanzamientos de una moneda equilibrada para decidir cuánto reduce la magnitud del input para producir el resultado Y .

(a) Utiliza un diagrama de árbol del conjunto de posibles pares input-output.

(b) Calcula las probabilidades de los pares input-output.

(c) Calcula las probabilidades de los valores de salida.

(d) Calcula la probabilidad de que el input fuera $X = +2$ dado que $Y = k$.

2.27 Sean $S = \{1, 2, 3, 4\}$ y $A = \{1, 2\}$, $B = \{1, 3\}$, $C = \{1, 4\}$. Asumiendo que los resultados son equiprobables, ¿son A , B y C sucesos independientes?

2.28 Demuestra que si A y B son sucesos independientes, entonces los pares A y B^c ; A^c y B ; y A^c y B^c son también independientes.

2.29 Un experimento consiste en escoger al azar una de dos urnas y luego extraer una bola de la urna y anotar su color (blanco o negro). Sea A el suceso “la urna 1 es escogida” y B el suceso “se observa una bola negra”. ¿Bajo qué condiciones serán independientes A y B ?

2.30 Se repite un experimento muchas veces y se anota la ocurrencia de los sucesos A y B . ¿Cómo se puede comprobar si los sucesos A y B son independientes?

2.31 Se transmite un bloque de 100 bits a través de un canal de comunicación con probabilidad de error en un bit de $p = 10^{-2}$.

(a) Si el bloque tiene 1 o menos errores, el receptor admite el bloque. Calcula la probabilidad de que el bloque sea aceptado.

(b) Si el bloque tiene más de 1 error, entonces hay que volver a transmitir el mensaje. Calcula la probabilidad de que se necesiten M retransmisiones.

2.32 Para que un circuito funcione, siete chips idénticos deben de estar funcionando. Para mejorar la fiabilidad, se incluye un chip adicional, de forma que el diseño permita que reemplace a cualquiera de los siete chips cuando fallen.

(a) Calcula la probabilidad p_b de que el sistema funcione en términos de la probabilidad p de que un chip funcione.

(b) Supongamos que n sistemas como este se colocan en paralelo y que necesitamos un 99.9% de probabilidad de que al menos uno de los sistemas funcione. ¿Cuántos sistemas necesitamos?

2.33 Consideremos una baraja de cartas bien mezclada de 52 cartas diferentes, de las cuales 4 son ases y 4, reyes.

(a) Calcula la probabilidad de obtener un as al sacar la primera carta.

(b) Saca una carta de la baraja y mirala. ¿Cuál es la probabilidad de obtener un as en la siguiente extracción? ¿Cambia tu respuesta si no observas la primera carta?

(c) Supongamos que sacamos 7 cartas de la baraja. ¿Cuál es la probabilidad de que haya tres ases? ¿Cuál es la probabilidad de que haya 2 reyes? ¿Cuál es la probabilidad de que haya 3 ases y/o 2 reyes?

(d) Supongamos que se reparte toda la baraja (el mismo número de cartas) entre cuatro jugadores. ¿Cuál es la probabilidad de que cada jugador tenga un as?

Parte II

Variables aleatorias

3 Variables aleatorias discretas

En la mayoría de experimentos aleatorios estamos interesados en un atributo numérico del resultado del experimento. Una variable aleatoria se define como una función que asigna un valor numérico al resultado del experimento. En este capítulo se introduce el concepto de variable aleatoria y métodos para calcular las probabilidades de sucesos relacionados con una variable aleatoria. Nos centramos en el caso más simple, el de variables aleatorias discretas, y presentamos la función de masa de probabilidad. Definimos el valor esperado de una variable aleatoria y lo relacionamos con nuestra noción intuitiva de la media. También presentamos la función de probabilidad condicionada para el caso en el que se nos da información parcial sobre la variable aleatoria. Estos conceptos y su extensión en el capítulo 4 nos proporcionan las herramientas para calcular las probabilidades y los promedios de interés en el diseño de sistemas con componentes estocásticos. A lo largo del capítulo se introduce algunas variables aleatorias importantes y se discute algunas aplicaciones típicas donde surgen.

3.1 La noción de variable aleatoria

El resultado de un experimento aleatorio no tiene por qué ser numérico. Sin embargo, normalmente no estamos interesados en el resultado del experimento en sí, sino en alguna medida o atributo numérico del resultado. Por ejemplo, en n lanzamientos de una moneda, podemos estar interesados en el número de caras y no en el orden específico en el que salen las caras y las cruces. En una página web elegida al azar, podríamos interesarnos solo por su longitud. En cada uno de estos ejemplos, *una medida asigna un valor numérico al resultado de un experimento aleatorio*. Como los resultados son aleatorios, las medidas de los mismos también serán aleatorias. Por tanto, tiene sentido hablar de las probabilidades de los valores numéricos resultantes. El concepto de variable aleatoria formaliza esta idea.

Una **variable aleatoria** X es una función que asigna un número real, $X(\zeta)$, a cada elemento ζ del espacio muestral de un experimento aleatorio. Recordemos que una función es simplemente una regla para asignar valores numéricos a cada elemento de un conjunto, como se muestra en la Figura 3.1. La especificación de una medida para los sucesos de un experimento aleatorio define una función en el

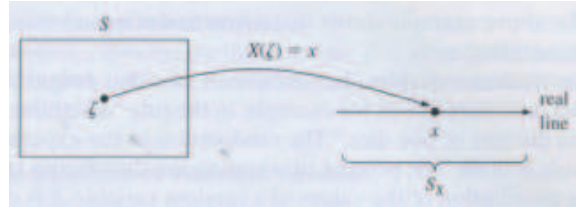


Figura 3.1 Una variable aleatoria asigna un número real $X(\zeta)$ a cada elemento ζ del espacio muestral S de un experimento aleatorio.

espacio muestral y, por consiguiente, una variable aleatoria. El espacio muestral S es el *dominio* de la variable aleatoria y el conjunto S_X de todos los valores que toma X es el *rango* de la variable aleatoria. Entonces S_X es un subconjunto del conjunto de todos los números reales. Utilizaremos la siguiente notación: letras mayúsculas para denotar variables aleatorias, e.g., X o Y , y letras minúsculas para representar posibles valores de las variables aleatorias, e.g., x o y .

Ejemplo 3.1 Lanzamientos de una moneda

Se lanza una moneda tres veces y se anota la secuencia de caras y cruces. El espacio muestral de este experimento es

$$S = \{CCC, CCX, CXC, CXX, XCC, XCX, XXC, XXX\}.$$

Sea X el número de caras en los tres lanzamientos. X asigna un número del conjunto $S_X = \{0, 1, 2, 3\}$ a cada elemento ζ de S . La tabla a continuación contiene los ocho sucesos de S y los correspondientes valores de X .

ζ :	CCC	CCX	CXC	XCC	CXX	XCX	XXC	XXX
$X(\zeta)$:	3	2	2	2	1	1	1	0

Por tanto, X es una variable aleatoria que toma valores en el conjunto $S_X = \{0, 1, 2, 3\}$.

Ejemplo 3.2 Un juego de apuestas

Un jugador apuesta 1.5 euros en el siguiente juego: se lanza una moneda tres veces y se cuenta el número de caras, X . El jugador recibe 1 euro si $X = 2$ y 8 euros si $X = 3$, pero no recibe nada en los demás casos. Sea Y la recompensa para el jugador. Y es una función de la variable aleatoria X y sus valores pueden relacionarse con el espacio muestral del experimento aleatorio subyacente como sigue:

ζ :	CCC	CCX	CXC	XCC	CXX	XCX	XXC	XXX
$X(\zeta)$:	3	2	2	2	1	1	1	0
$Y(\zeta)$:	8	1	1	1	0	0	0	0

Por tanto, Y es una variable aleatoria que toma valores en el conjunto $S_Y = \{0, 1, 8\}$.

El ejemplo anterior muestra que una función de una variable aleatoria produce otra variable aleatoria.

Para una variable aleatoria, la función o regla que asigna valores a cada evento es fija y determinista, como, por ejemplo, en la regla de “suma el resultado del lanzamiento de dos dados.” La aleatoriedad en el experimento es debida a que se realiza los lanzamientos. El proceso de suma de los resultados es determinista. Por lo tanto, la distribución de los valores de una variable aleatoria X viene determinada por la probabilidad de los resultados ζ del experimento aleatorio. En otras palabras, la aleatoriedad en los valores observados de X es *inducida* por el experimento aleatorio subyacente y, por lo tanto, nosotros debemos ser capaces de calcular las probabilidades de los valores de X en términos de las probabilidades de los eventos subyacentes.

Ejemplo 3.3 Lanzamientos de una moneda y apuestas

Sea X el número de caras en tres lanzamientos independientes de una moneda equilibrada. Calcula la probabilidad del suceso $\{X = 2\}$. Calcula la probabilidad de que el jugador del ejemplo 3.2 gane 8 euros.

Solución

Obsérvese que $X(\zeta) = 2$ si y sólo si $\zeta \in \{CCX, CXC, XCC\}$. Por tanto

$$\begin{aligned} P[X = 2] &= P[\{CCX, CXC, XCC\}] \\ &= P[\{CCX\}] + P[\{CXC\}] + P[\{XCC\}] \\ &= 3/8. \end{aligned}$$

El suceso $\{Y = 8\}$ ocurre si y sólo si el resultado ζ es CCC , entonces

$$P[\{Y = 8\}] = P[\{CCC\}] = 1/8.$$

El ejemplo 3.3 ilustra la técnica general para calcular las probabilidades de los sucesos relacionados con la variable aleatoria X . Sean S y \mathcal{F} el espacio muestral y la clase de los sucesos, respectivamente, del experimento aleatorio subyacente. Para calcular la probabilidad de un subconjunto B de \mathbb{R} , e.g., $B = \{x_k\}$, necesitamos encontrar los sucesos de S que tienen como imagen a B , esto es,

$$A = \{\zeta : X(\zeta) \in B\} \tag{3.1}$$

como se muestra en la Figura 3.2. Si ocurre el suceso A entonces $X(\zeta) \in B$, el suceso B ocurre. Por el contrario, si el suceso B ocurre, entonces el valor $X(\zeta)$ implica que $\zeta \in A$, por lo que el suceso A ocurre. Así, la probabilidad de que $X \in B$ viene dada por:

$$P[X \in B] = P[A] = P[\{\zeta : X(\zeta) \in B\}]. \tag{3.2}$$



Figura 3.2
 $P[X \in B] = P[\zeta \in A]$

Diremos que A y B son **sucesos equivalentes**.

En algunos experimentos aleatorios el resultado ζ es el valor numérico en el que estamos interesados. En esos casos, definimos simplemente $X(\zeta) = \zeta$, esto es, la función identidad, para obtener una variable aleatoria.

3.2 Variables aleatorias discretas y la función de probabilidad

Una **variable aleatoria discreta** X se define como una variable aleatoria que toma un conjunto numerable de valores $S_X = \{x_1, x_2, x_3, \dots\}$. Una variable aleatoria discreta es **finita** si su rango es finito, esto es, $S_X = \{x_1, x_2, \dots, x_n\}$. Estamos interesados en calcular la probabilidad de los sucesos relacionados con una variable aleatoria discreta X . Como el espacio muestral S_X es discreto, sólo necesitamos obtener las probabilidades de los sucesos $A_k = \{\zeta : X(\zeta) = x_k\}$ en el experimento aleatorio subyacente. La probabilidad de todos los sucesos relacionados con X se puede obtener conociendo la probabilidad de los A_k .

La **función de probabilidad (pmf) de una variable aleatoria discreta** X se define como:

$$p_X(x) = P[X = x] = P[\{\zeta : X(\zeta) = x\}] \quad x \in \mathbb{R}. \quad (3.3)$$

Obsérvese que $p_X(x)$ es una función de x en la recta real y que puede ser no nula sólo en los valores x_1, x_2, x_3, \dots . Para $x_k \in S_X$, tenemos que $p_X(x_k) = P[A_k]$.

Los sucesos A_1, A_2, \dots forman una partición de S , como se ilustra en la figura 3.3. Para ver esto, primero demostramos que son sucesos disjuntos. Sea $j \neq k$, entonces

$$A_j \cap A_k = \{\zeta : X(\zeta) = x_j, X(\zeta) = x_k\} = \emptyset$$

ya que cada ζ tiene una única imagen en S_X . Ahora demostramos que S es la unión de A_k . Cada $\zeta \in S$ tiene su imagen x_k tal que ζ pertenece a un suceso A_k de la partición. Por tanto:

$$S = A_1 \cup A_2 \cup \dots$$

Todos los sucesos relacionados con la variable aleatoria X se puede expresar como la unión de sucesos A_k . Por ejemplo, supongamos que estamos interesados en el suceso $X \in B = \{x_2, x_5\}$, entonces

$$\begin{aligned} P[X \in B] &= P[\{\zeta : X(\zeta) = x_2\} \cup \{\zeta : X(\zeta) = x_5\}] \\ &= P[A_2 \cup A_5] = P[A_2] + P[A_5] \\ &= p_X(2) + p_X(5). \end{aligned}$$



Figura 3.3 Partición del espacio muestral S asociado a una variable aleatoria discreta.

La pmf $p_X(x)$ satisface tres propiedades que proporcionan toda la información necesaria para calcular la probabilidad de los sucesos relacionados con la variable aleatoria discreta X :

$$(i) \quad p_X(x) \geq 0 \quad \forall x \quad (3.4a)$$

$$(ii) \quad \sum_{x \in S_X} p_X(x) = \sum_k p_X(x_k) = \sum_k P[A_k] = 1 \quad (3.4b)$$

$$(iii) \quad P[X \in B] = \sum_{x \in B} p_X(x) \quad B \in S_X. \quad (3.4c)$$

La propiedad (i) es cierta porque los valores de la pmf se definen como una probabilidad, $p_X(x) = P[X = x]$. La propiedad (ii) se deduce porque los sucesos $A_k = \{X = x_k\}$ forman una partición de S . Obsérvese que las sumas en las ecuaciones de (ii) y (iii) tendrán un número finito o infinito de términos dependiendo de si la variable aleatoria es finita o no. Consideremos ahora la propiedad (iii). Cualquier suceso B relacionado con X es la unión de sucesos elementales, por lo que por el Axioma III', tenemos:

$$P[X \in B] = P \left[\bigcup_{x \in B} \{\zeta : X(\zeta) = x\} \right] = \sum_{x \in B} P[X = x] = \sum_{x \in B} p_X(x).$$

La pmf de X nos da las probabilidades de todos los sucesos elementales de S_X . La probabilidad de cualquier subconjunto de S_X se obtiene sumando las de los sucesos elementales correspondientes. De hecho, tenemos todo lo necesario para especificar una ley de probabilidad para los sucesos en S_X . Si sólo estamos interesados en los sucesos relacionados con X , entonces podemos olvidarnos del experimento aleatorio subyacente y su ley de probabilidad asociada y trabajar sólo con S_X y la pmf de X .

Ejemplo 3.4 Lanzamientos de una moneda y la variable aleatoria Binomial

Sea X el número de caras en tres lanzamientos independientes de una moneda. Calcula la pmf de X .

Solución

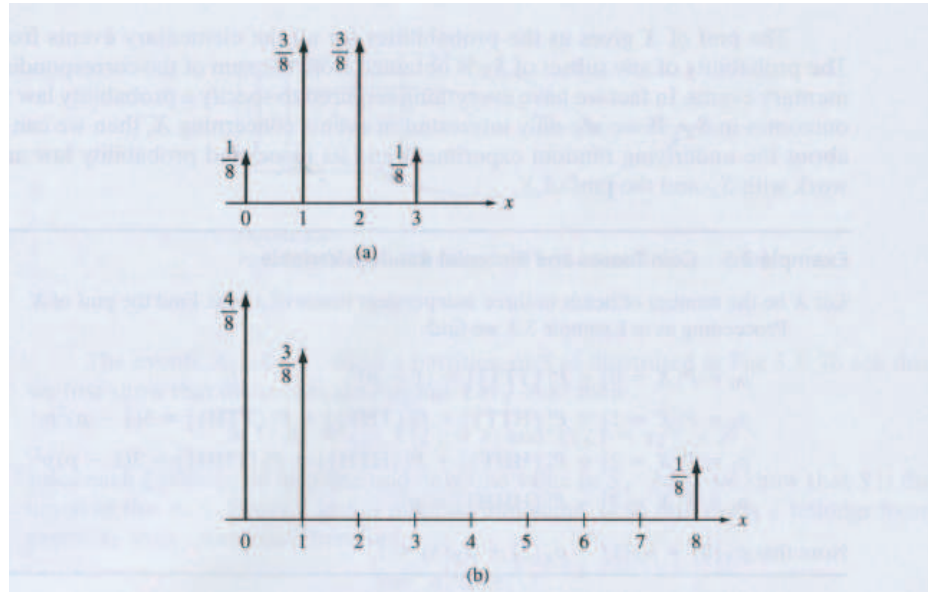


Figura 3.4 (a) Gráfico de la pmf de tres lanzamientos de una moneda; (b) gráfico de la pmf del juego de apuestas.

Procediendo como en el ejemplo 3.3, calculamos:

$$\begin{aligned}
 p_0 &= P[X = 0] = P[\{XXX\}] = (1 - p)^3, \\
 p_1 &= P[X = 1] = P[\{CXX\}] + P[\{XCX\}] + P[\{XXC\}] = 3(1 - p)^2p, \\
 p_2 &= P[X = 2] = P[\{CCX\}] + P[\{CXC\}] + P[\{XCC\}] = 3(1 - p)p^2, \\
 p_3 &= P[X = 3] = P[\{CCC\}] = p^3,
 \end{aligned}$$

Obsérvese que $p_X(0) + p_X(1) + p_X(2) + p_X(3) = 1$.

Ejemplo 3.5 Un juego de apuestas

Un jugador recibe 1 euro si el número de caras en tres lanzamientos de moneda es 2, 8 euros si el número en 3 y nada en el resto de los casos. Calcula la pmf de la recompensa Y .

Solución

$$\begin{aligned}
 p_Y(0) &= P[\zeta \in \{XXX, XXC, XCX, CXX\}] = 4/8 = 1/2 \\
 p_Y(1) &= P[\zeta \in \{XCC, CXC, CCX\}] = 3/8 \\
 p_Y(8) &= P[\zeta \in \{CCC\}] = 1/8.
 \end{aligned}$$

Obsérvese que $p_Y(0) + p_Y(1) + p_Y(8) = 1$.

La figura 3.4 muestra la gráfica de $p_X(x)$ frente a x de las variables aleatorias de los ejemplos 3.4 y 3.5, respectivamente. En general, el gráfico de la pmf de una variable aleatoria discreta está formada por barras verticales de altura $p_X(x_k)$ en los valores x_k de S_X . Podemos interpretar la probabilidad total como una unidad de masa y la $p_X(x)$ como la cantidad de masa de probabilidad situada en cada uno de los puntos discretos x_1, x_2, \dots . Los valores relativos de la pmf en los diferentes puntos dan una idea de las probabilidades relativas de ocurrencia.

Ejemplo 3.6 Generación de números aleatorios

Un generador de números aleatorios produce un número entero X del conjunto $S_X = \{0, 1, 2, \dots, M-1\}$ en el que todos los valores son igualmente probables de ser seleccionados. Calcula la pmf de X .

Solución

Para cada $k \in S_X$ tenemos $p_X(k) = 1/M$. Nótese que

$$p_X(0) + p_X(1) + \dots + p_X(M-1) = 1.$$

X es denominada **variable aleatoria Uniforme** en el conjunto $\{0, 1, 2, \dots, M-1\}$.

Ejemplo 3.7 Variable aleatoria de Bernoulli

Sea A el suceso de interés en un experimento aleatorio, e.g. un aparato no es defectuoso. Decimos que ocurre un “éxito” si A ocurre al realizar el experimento. La variable aleatoria de Bernoulli I_A es igual a 1 si ocurre A y cero en otro caso y viene dada por la *función indicadora* de A :

$$I_A(\zeta) = \begin{cases} 0 & \text{si } \zeta \notin A \\ 1 & \text{si } \zeta \in A. \end{cases} \quad (3.5a)$$

Calcula la pmf de I_A .

Solución

$I_A(\zeta)$ es una variable aleatoria discreta de rango $S_I = \{0, 1\}$ con pmf:

$$\begin{aligned} p_I(0) &= P[\{\zeta : \zeta \in A^c\}] = 1 - p \\ p_I(1) &= P[\{\zeta : \zeta \in A\}] = p. \end{aligned} \quad (3.5b)$$

I_A es conocida como la **variable aleatoria de Bernoulli**. Obsérvese que $p_I(0) + p_I(1) = 1$.

Ejemplo 3.8 Transmisión de mensajes

Sea X el número de veces que necesitamos enviar un mensaje hasta que llegue correctamente a su destino. Calcula la pmf de X . Calcula la probabilidad de que X sea un número par.

Solución

X es una variable aleatoria discreta que toma valores en $S_X = \{1, 2, 3, \dots\}$. El suceso $\{X = k\}$ ocurre si en el experimento subyacente resultan $k - 1$ transmisiones erróneas consecutivas (“fracasos”) seguidos de una sin errores (“éxito”):

$$p_X(k) = P[X = k] = P[00 \dots 01] = (1 - p)^{k-1}p = q^{k-1}p \quad k = 1, 2, \dots \quad (3.6)$$

Diremos que X es una **variable aleatoria Geométrica** y que X se distribuye geoméricamente. En la ecuación (2.34b), vimos que la suma de probabilidades geométricas es 1.

$$P[X \text{ par}] = \sum_{k=1}^{\infty} p_X(2k) = p \sum_{k=1}^{\infty} q^{2k-1} = p \frac{1}{1 - q^2} = \frac{1}{1 - q}.$$

Ejemplo 3.9 Transmisión de errores

Un canal de comunicaciones binario introduce un error en un bit de una transmisión con probabilidad p . Sea X el número de errores en n transmisiones independientes. Calcula la pmf de X . Calcula la probabilidad de a lo sumo un error.

Solución

X toma valores en el conjunto $S_X = \{0, 1, \dots, n\}$. Cada arroja un “0” si no hay errores y “1” si hay un error, $P[\text{“1”}] = p$ y $P[\text{“0”}] = 1 - p$. La probabilidad de k errores en n transmisiones de tamaño un bit es dada por la probabilidad de un patrón de errores que tiene k unos y $n - k$ ceros:

$$p_X(k) = P[X = k] = \binom{n}{k} p^k (1 - p)^{n-k} \quad k = 0, 1, \dots, n. \quad (3.7)$$

X es denominada **variable aleatoria Binomial**, con parámetros n y p . En la ecuación (2.31b), vimos que la suma de probabilidades binomiales es 1.

$$P[X \leq 1] = \binom{n}{0} p^0 (1 - p)^{n-0} + \binom{n}{1} p^1 (1 - p)^{n-1} = (1 - p)^n + np(1 - p)^{n-1}.$$

Por último, vamos a considerar la relación entre las frecuencias relativas y la pmf $p_X(x_k)$. Supongamos que se realizan n repeticiones independientes para obtener n observaciones de la variable aleatoria discreta X . Sea $N_k(n)$ el número de veces que se produce el suceso $X = x_k$ y sea $f_k(n) = N_k(n)/n$ la frecuencia relativa correspondiente. Haciendo n tender a infinito, esperamos que $f_k(n) \rightarrow p_X(x_k)$. Por tanto, el gráfico de las frecuencias relativas debe parecerse al gráfico de la pmf. La figura 3.5(a) muestra el gráfico de las frecuencias relativas de 1000 repeticiones de un experimento que genera una variable aleatoria Uniforme en el conjunto $\{0, 1, \dots, 7\}$ y la pmf correspondiente. La figura 3.5(b) muestra el

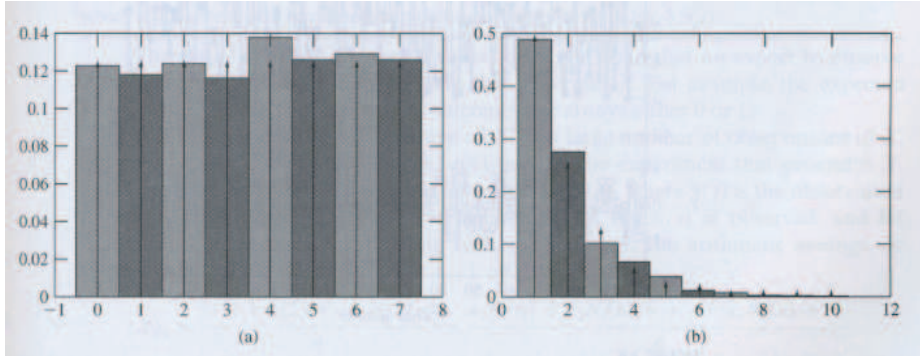


Figura 3.5 (a) Frecuencias relativas y pmf uniforme; (b) Frecuencias relativas y pmf geométrica.

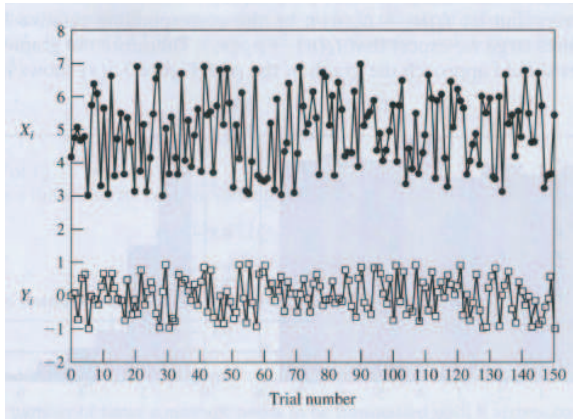


Figura 3.6 Los gráficos muestran 150 repeticiones de los experimentos que generan X e Y . La variable aleatoria X se centra en el valor 5 mientras que Y varía alrededor del valor 0. Es obvio también que X es más dispersa que Y .

gráfico de frecuencias relativas y la pmf para una variable aleatoria Geométrica con $p = 1/2$ y $n = 1000$ repeticiones. En ambos casos vemos que la gráfica de frecuencias relativas se aproxima a la de la pmf.

3.3 Esperanza y momentos de una variable aleatoria discreta

Con el objetivo de describir completamente el comportamiento de una variable aleatoria discreta, debemos definir toda una función, es decir $p_X(x)$. En algunas situaciones estamos interesados en unos pocos parámetros que resumen la información proporcionada por la pmf. Por ejemplo, la figura 3.6 muestra los resultados de muchas repeticiones de un experimento que produce dos variables aleatorias. La variable aleatoria Y varía alrededor del valor 0, mientras que la variable aleatoria X varía alrededor del valor 5. Es obvio también que X es más dispersa que Y . En esta sección se presenta los parámetros que cuantifican estas propiedades.

La **esperanza** o **media** de una **variable aleatoria discreta** X se define

$$m_X = E[X] = \sum_{x \in S_X} xp_X(x) = \sum_k x_k p_X(x_k). \quad (3.8)$$

La esperanza $E[X]$ está definida si la suma anterior es absolutamente convergente, es decir,

$$E[|X|] = \sum_k |x_k| p_X(x_k) < \infty. \quad (3.9)$$

Hay variables aleatorias para las que la ecuación (3.9) no converge. En estos casos, se dice que la esperanza o valor esperado no existe.

Si interpretamos $p_X(x)$ como la distribución de masa en los puntos x_1, x_2, \dots en la recta real, entonces $E[X]$ representa al centro de masa de esta distribución. Por ejemplo, en la figura 3.5(a), podemos ver que la pmf de una variable aleatoria discreta que se distribuye uniformemente en $\{0, \dots, 7\}$ tiene centro de masa 3.5.

Ejemplo 3.10 Media de la variable aleatoria de Bernoulli

Calcula la esperanza de la variable aleatoria de Bernoulli I_A .

Solución

A partir del ejemplo 3.5, tenemos

$$E[I_A] = 0p_I(0) + 1p_I(1) = p.$$

donde p es la probabilidad de éxito en el experimento de Bernoulli.

Ejemplo 3.11 Tres lanzamientos de una moneda y la variable aleatoria Binomial

Sea X el número de caras en tres lanzamientos de una moneda equilibrada. Calcula $E[X]$.

Solución

La ecuación (3.8) y la pmf de X que se calculó en el ejemplo 3.4 dan:

$$E[X] = \sum_{k=0}^3 kp_X(k) = 0\frac{1}{8} + 1\frac{3}{8} + 2\frac{3}{8} + 3\frac{1}{8} = 1.5.$$

Nótese que lo anterior se refiere a una variable aleatoria Binomial con $n = 3$ y $p = 1/2$, que veremos que tiene $E[X] = np$.

Ejemplo 3.12 La media de una variable aleatoria discreta uniforme

Sea X el generador de números aleatorios en el ejemplo 3.6. Calcula $E[X]$.

Solución

A partir del ejemplo 3.4 tenemos que $p_X(j) = 1/M$ para $j = 0, \dots, M - 1$, por

lo que

$$E[X] = \sum_{k=0}^{M-1} k \frac{1}{M} = \frac{1}{M} \{0 + 1 + 2 + \cdots + M - 1\} = \frac{(M-1)M}{2M} = \frac{(M-1)}{2}$$

donde utilizamos el hecho que $1 + 2 + \cdots + L = (L+1)L/2$. Nótese que para $M = 8$, $E[X] = 3.5$, lo cual es consistente con nuestra observación sobre el centro de la masa en la figura 3.5(a).

El uso del término “esperanza” no significa que esperemos observar $E[X]$ cuando llevamos a cabo el experimento que genera X . Por ejemplo, la esperanza de un experimento de Bernoulli es p pero sus resultados son siempre 0 o 1.

La $E[X]$ se corresponde con la “media de X ” en un gran número de observaciones de X . Supongamos que se realiza n repeticiones independientes del experimento que genera X y que registramos los valores observados como $x(1)$, $x(2)$, ..., $x(n)$, donde $x(j)$ es la observación en el j -ésimo experimento. Sea $N_k(n)$ el número de veces que se obtiene x_k , y sea $f_k(n) = N_k(n)/n$ la frecuencia relativa correspondiente. La media aritmética o **media muestral** de las observaciones es la siguiente:

$$\begin{aligned} \langle X \rangle_n &= \frac{x(1) + x(2) + \cdots + x(n)}{n} = \frac{x_1 N_1(n) + x_2 N_2(n) + \cdots + x_k N_k(n) + \cdots}{n} \\ &= x_1 f_1(n) + x_2 f_2(n) + \cdots + x_k f_k(n) + \cdots \\ &= \sum_k x_k f_k(n). \end{aligned} \quad (3.10)$$

El primer numerador suma las observaciones en el orden en que ocurren y el segundo numerador cuenta cuántas veces ocurre cada x_k y luego calcula el total. Al crecer n , se espera que las frecuencias relativas se aproximen a las probabilidades $p_X(x_k)$:

$$\lim_{n \rightarrow \infty} f_k(n) = p_X(x_k) \quad \forall k \quad (3.11)$$

La ecuación (3.10) implica entonces que:

$$\langle X \rangle_n = \sum_k x_k f_k(n) \rightarrow \sum_k x_k p_X(x_k) = E[X]. \quad (3.12)$$

Por lo tanto, esperamos que la media muestral converja a $E[X]$ cuando n se hace grande.

Ejemplo 3.13 Un juego de apuestas

Un jugador en una feria paga 1.50 euros para lanzar una moneda tres veces. El jugador recibe 1 euro si el número de caras es 2, 8 euros si el número es 3, y nada en cualquier otro caso. Calcula la esperanza de la recompensa Y . ¿Cuál es la ganancia esperada?

Solución

La recompensa esperada es:

$$E[Y] = 0p_Y(0) + 1p_Y(1) + 8p_Y(8) = 0\frac{4}{8} + 1\frac{3}{8} + 8\frac{1}{8} = \frac{11}{8}.$$

La ganancia esperada es:

$$E[Y - 1.5] = \frac{11}{8} - \frac{12}{8} = -\frac{1}{8}.$$

Los jugadores pierden 12.5 céntimos por juego en media, por lo que la casa de apuestas obtiene un buen beneficio a largo plazo. En el ejemplo 3.17 veremos que en algunos diseños en ingeniería también se “apuesta” que los usuarios se comportan de cierta manera.

Ejemplo 3.14 La media de una variable aleatoria Geométrica

Sea X el número de bytes en un mensaje, y supongamos que X tiene una distribución geométrica con parámetro p . Calcular la media de X .

Solución

X puede tomar valores arbitrariamente grandes ya que $S_X = \{1, 2, \dots\}$. El valor esperado es:

$$E[X] = \sum_{k=1}^{\infty} kpq^{k-1} = p \sum_{k=1}^{\infty} kq^{k-1}.$$

Esta expresión es fácilmente evaluada mediante

$$\frac{1}{1-x} = \sum_{k=0}^{\infty} x^k \quad (3.13)$$

para obtener

$$\frac{1}{(1-x)^2} = \sum_{k=0}^{\infty} kx^{k-1}. \quad (3.14)$$

Haciendo $x = q$, se obtiene

$$E[X] = p \frac{1}{(1-q)^2} = \frac{1}{p}. \quad (3.15)$$

Vemos que X tiene una esperanza finita siempre que $p > 0$.

Para algunas variables aleatorias los valores altos ocurren con frecuencia suficiente para que la esperanza no exista, como se ilustra en el siguiente ejemplo.

Ejemplo 3.15 La paradoja de San Petersburgo

Se lanza una moneda equilibrada hasta que sale una cruz. Si son necesarios X

lanzamientos, entonces el casino paga al jugador $Y = 2^X$ euros. ¿Cuánto debe estar dispuesto a pagar el jugador para jugar este juego?

Solución

Si el jugador juega a este juego un gran número de veces, entonces el pago debe ser el valor esperado de $Y = 2^X$. Si la moneda es equilibrada, $P[X = k] = (1/2)^k$ y $P[Y = 2^k] = (1/2)^k$, así:

$$E[Y] = \sum_{k=1}^{\infty} 2^k p_Y(2^k) = \sum_{k=1}^{\infty} 2^k \left(\frac{1}{2}\right)^k = 1 + 1 + \dots = \infty.$$

Este juego en efecto, parece ofrecer al jugador un buen trato, ¡por lo que el jugador debe estar dispuesto a pagar cualquier cantidad para jugar! La paradoja está en que una persona cuerda no pagaría mucho por jugar a este juego. ¿Cómo se resuelve esta paradoja?

Las variables aleatorias con una esperanza sin límite no son poco frecuentes y aparecen en los modelos en los que los resultados que tienen valores muy grandes no son tan raros. Por ejemplo, el tamaño de los archivos en las transferencias vía web, la frecuencia de las palabras en los textos grandes y diversos problemas financieros y económicos.

3.3.1 Esperanza de la función de una variable aleatoria

Sea X una variable aleatoria discreta y $Z = g(X)$. Puesto que X es discreta, $Z = g(X)$ tomará un conjunto numerable de valores de la forma $g(x_k)$ donde $x_k \in S_X$. Denotamos el conjunto de los valores que toma $g(X)$ como $\{z_1, z_2, \dots\}$. Una forma de calcular la esperanza de Z es usar la ecuación (3.8), que requiere que calculemos la pmf de Z . Otra forma es utilizar el siguiente resultado:

$$E[Z] = E[g(X)] = \sum_k g(x_k) p_X(x_k). \quad (3.16)$$

Para demostrar la ecuación (3.16) agrupamos los términos x_k asignados a cada valor z_j :

$$\sum_k g(x_k) p_X(x_k) = \sum_j z_j \left\{ \sum_{x_k: g(x_k)=z_j} p_X(x_k) \right\} = \sum_j z_j p_Z(z_j) = E[Z].$$

La suma dentro de las llaves es la probabilidad de todos los términos x_k para los que $g(x_k) = z_j$, que es la probabilidad de que $Z = z_j$, es decir, $p_Z(z_j)$.

Ejemplo 3.16 Dispositivos de ley del cuadrado

Sea X una tensión de ruido que se distribuye uniformemente en

$$S_X = \{-3, -1, +1, +3\}$$

con $p_X(k) = 1/4$ para $k \in S_X$. Calcula $E[Z]$, donde $Z = X^2$.

Solución

Usando el primer método se tiene que la pmf de Z es:

$$p_Z(9) = P[X \in \{-3, +3\}] = p_X(-3) + p_X(3) = 1/2$$

$$p_Z(1) = p_X(-1) + p_X(1) = 1/2$$

y entonces $E[Z] = 1\frac{1}{2} + 9\frac{1}{2} = 5$. El segundo método da:

$$E[Z] = E[X^2] = \sum_k k^2 p_X(k) = \frac{1}{4}\{(-3)^2 + (-1)^2 + 1^2 + 3^2\} = \frac{20}{4} = 5.$$

La ecuación (3.16) implica varios resultados muy útiles. Sea Z la función

$$Z = ag(X) + bh(X) + c$$

donde a , b y c son números reales, entonces:

$$E[Z] = aE[g(X)] + bE[h(X)] + c. \quad (3.17a)$$

Utilizando la ecuación (3.16) se tiene:

$$\begin{aligned} E[Z] &= E[aE[g(X)] + bE[h(X)] + c] = \sum_k (ag(x_k) + bh(x_k) + c)p_X(x_k) \\ &= a \sum_k g(x_k)p_X(x_k) + b \sum_k h(x_k)p_X(x_k) + c \sum_k p_X(x_k) \\ &= aE[g(X)] + bE[h(X)] + c. \end{aligned}$$

La ecuación (3.17a), haciendo a , b y/o c 0 o 1, implica las siguientes expresiones:

$$E[g(X) + h(X)] = E[g(X)] + E[h(X)]. \quad (3.17b)$$

$$E[aX] = aE[X]. \quad (3.17c)$$

$$E[X + c] = E[X] + c. \quad (3.17d)$$

$$E[c] = c \quad (3.17e)$$

Ejemplo 3.17 Dispositivos de ley del cuadrado

La tensión de ruido X del ejemplo anterior se amplifica y desplaza para obtener $Y = 2X + 10$, para producir $Z = Y^2 = (2X + 10)^2$. Calcula $E[Z]$.

Solución

$$\begin{aligned} E[Z] &= E[(2X + 10)^2] = E[4X^2 + 40X + 100] \\ &= 4E[X^2] + 40E[X] + 100 = 4(5) + 40(0) + 100 = 120. \end{aligned}$$

Ejemplo 3.18 Multiplexor de paquetes de voz

Sea X el número de paquetes de voz que contienen lenguaje activo producido por $n = 48$ altavoces independientes en un período de 10 milisegundos como se discutió en la sección 1.4. X es una variable aleatoria Binomial con parámetro n y probabilidad $p = 1/3$. Supongamos que un multiplexor de paquetes transmite hasta $M = 20$ paquetes activos cada 10 ms, y cualquier exceso de paquetes activos es descartado. Sea Z el número de paquetes descartados. Calcula $E[Z]$.

Solución

El número de paquetes descartados cada 10 ms es la siguiente función de X :

$$Z = (X - M)^+ \triangleq \begin{cases} 0 & X \leq M \\ X - M & X > M. \end{cases}$$

$$E[Z] = \sum_{k=20}^{48} (k - 20) \binom{48}{k} \left(\frac{1}{3}\right)^k \left(\frac{2}{3}\right)^{48-k} = 0.182.$$

Cada 10 ms se produce $E[X] = np = 16$ paquetes activos en media, por lo que la fracción de paquetes activos descartados que los clientes van a sufrir es de $0.182/16 = 1,1\%$. Este ejemplo muestra que los sistemas en ingeniería también “apuestan” en juegos en los que las estadísticas favorables son explotadas para utilizar los recursos de manera eficiente. En este ejemplo, el multiplexor transmite 20 paquetes por cada período en lugar de 48 con una reducción de $28/48 = 58\%$.

3.3.2 Varianza de una variable aleatoria

La esperanza $E[X]$, por sí misma, nos proporciona una información limitada acerca de X . Por ejemplo, si sabemos que $E[X] = 0$, entonces podría ser que X fuera igual a cero todo el tiempo. Sin embargo, también es posible que X pueda tomar valores positivos y negativos muy grandes. Estamos interesados, por tanto, no sólo en la media de una variable aleatoria, sino también en la medida de la variación de la variable aleatoria alrededor de su media. Sea $X - E[X]$ la desviación de una variable aleatoria X respecto a su media, que puede tomar valores positivos y negativos. Dado que solamente estamos interesados en la magnitud de las variaciones, es conveniente trabajar con el cuadrado de la desviación, que es siempre positivo, $D(X) = (X - E[X])^2$. La esperanza es una constante, así que lo denotaremos por $m_X = E[X]$. La **varianza de la variable aleatoria** X se define como la esperanza de D :

$$\begin{aligned} \sigma_X^2 &= V[X] = E[(X - m_X)^2] \\ &= \sum_{x \in S_X} (x - m_X)^2 p_X(x) = \sum_{k=1}^{\infty} (x_k - m_X)^2 p_X(x_k). \end{aligned} \quad (3.18)$$

La **desviación típica de la variable aleatoria** X se define como:

$$\sigma_X = STD[X] = V[X]^{1/2}. \quad (3.19)$$

Tomando la raíz cuadrada de la varianza obtenemos una cantidad con las mismas unidades que X .

Una expresión alternativa para la varianza se obtiene como sigue:

$$\begin{aligned} V[X] &= E[(X - m_X)^2] = E[X^2 - 2m_X X + m_X^2] \\ &= E[X^2] - 2m_X E[X] + m_X^2 \\ &= E[X^2] - m_X^2. \end{aligned} \quad (3.20)$$

Se llama **segundo momento de X** a $E[X^2]$. Se define el **n -ésimo momento de X** como $E[X^n]$.

Las ecuaciones (3.17c), (3.17d), y (3.17e) implican las siguientes expresiones para la varianza. Sea $Y = X + c$, entonces

$$\begin{aligned} V[X + c] &= E[(X + c - (E[X] + c))^2] \\ &= E[(X - E[X])^2] = V[X]. \end{aligned} \quad (3.21)$$

Sumar una constante a una variable aleatoria no afecta a la varianza. Sea $Z = cX$, entonces:

$$V[cX] = E[(cX - cE[X])^2] = E[c^2(X - E[X])^2] = c^2 V[X]. \quad (3.22)$$

Multiplicar una variable aleatoria por una constante c multiplica la varianza por c^2 y la desviación típica por $|c|$.

Ahora sea $X = c$, una variable aleatoria constante, entonces

$$V[X] = E[(X - c)^2] = E[0] = 0. \quad (3.23)$$

Una variable aleatoria constante tiene varianza cero.

Ejemplo 3.19 Tres lanzamientos de moneda

Sea X el número de caras en tres lanzamientos de una moneda equilibrada. Calcula $V[X]$.

Solución

$$\begin{aligned} E[X^2] &= 0 \frac{1}{8} + 1^2 \frac{3}{8} + 2^2 \frac{3}{8} + 3^2 \frac{1}{8} = 3 \\ V[X] &= E[X^2] - m_X^2 = 3 - 1.5^2 = 0.75. \end{aligned}$$

Recordemos que ésta es una variable aleatoria Binomial con $n = 3$ y $p = 1/2$. Veremos más tarde que la varianza de la variable aleatoria Binomial es npq .

Ejemplo 3.20 Varianza de variable aleatoria de Bernoulli

Calcula la varianza de la variable aleatoria de Bernoulli I_A .

Solución

$$\begin{aligned} E[I_A^2] &= 0p_I(0) + 1^2p_I(1) = p \\ V[I_A] &= p - p^2 = p(1 - p) = pq. \end{aligned} \quad (3.24)$$

Ejemplo 3.21 Varianza de una variable aleatoria Geométrica

Calcula la varianza de la variable aleatoria Geométrica.

Solución

Desarrollemos el término $(1 - x^2)^{-1}$ en la ecuación (3.14) para obtener

$$\frac{2}{(1 - x)^3} = \sum_{k=0}^{\infty} k(k - 1)x^{k-2}.$$

Sea $x = q$ y multipliquemos ambos lados por pq para obtener:

$$\begin{aligned} \frac{2pq}{(1 - q)^3} &= pq \sum_{k=0}^{\infty} k(k - 1)q^{k-2} \\ &= \sum_{k=0}^{\infty} k(k - 1)pq^{k-1} = E[X^2] - E[X]. \end{aligned}$$

Así que el segundo momento es

$$E[X^2] = \frac{2pq}{(1 - q)^3} + E[X] = \frac{2q}{p^2} + \frac{1}{p} = \frac{1 + q}{p^2}$$

y la varianza

$$V[X] = E[X^2] - E[X]^2 = \frac{1 + q}{p^2} - \frac{1}{p^2} = \frac{q}{p^2}.$$

3.4 Función de probabilidad condicionada

En muchas ocasiones disponemos de información parcial sobre una variable aleatoria X o sobre el resultado del experimento aleatorio subyacente. Estamos interesados en cómo esta información cambia la probabilidad de los sucesos relacionados con la variable aleatoria. La función de probabilidad condicionada aborda esta cuestión para variables aleatorias discretas.

3.4.1 Función de probabilidad condicionada

Sea X una variable aleatoria discreta con pmf $p_X(x)$, y sea C un suceso que tiene probabilidad no nula, $P[C] > 0$. Ver figura 3.7. La **función de probabilidad condicionada** de X se define por la probabilidad condicionada:

$$p_X(x|C) = P[X = x|C], \quad x \in \mathbb{R}. \quad (3.25)$$

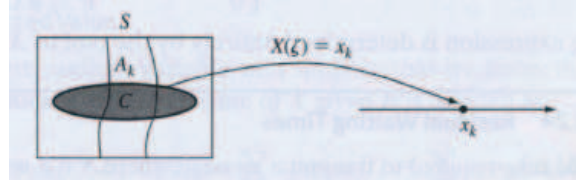


Figura 3.7 pmf condicionada de X dado el suceso C .

Aplicando la definición de probabilidad condicionada se tiene:

$$p_X(x|C) = \frac{P[\{X = x\} \cap C]}{P[C]}. \quad (3.26)$$

La expresión anterior tiene una interpretación intuitiva: la probabilidad condicionada del suceso $\{X = x_k\}$ está dada por las probabilidades de los elementos ζ tal que $X(\zeta) = x_k$ y ζ están en C , normalizada por $P[C]$.

La pmf condicionada satisface las ecuaciones (3.4a) - (3.4c). Consideremos la ecuación (3.4b). El conjunto de sucesos $A_k = \{X = x_k\}$ es una partición de S , por lo que

$$C = \bigcup_k (A_k \cap C),$$

y

$$\begin{aligned} \sum_{x_k \in S_X} p_X(x_k|C) &= \sum_k p_X(x_k|C) = \sum_k \frac{P[\{X = x_k\} \cap C]}{P[C]} \\ &= \frac{1}{P[C]} \sum_k P[A_k \cap C] = \frac{P[C]}{P[C]} = 1. \end{aligned}$$

De forma similar podemos probar que:

$$P[X \in B|C] = \sum_{X \in B} p_X(x|C), \quad B \in S_X.$$

Ejemplo 3.22 Un reloj aleatorio

El minuterero de un reloj se hace girar y el resultado ζ es el minuto donde para el minuterero. Sea X la hora en que el minuterero se detiene. Calcula la pmf de X . Calcula la pmf condicionada de X dado que $B = \{4 \text{ primeras horas}\}$; dado $D = \{1 < \zeta \leq 11\}$.

Solución

Asumamos que el minuterero se detiene en todos los minutos del rango $S = \{1, 2, \dots, 60\}$ con la misma probabilidad, por lo que $P[\zeta = k] = 1/60$ para k en S . X toma valores en $S_X = \{1, 2, \dots, 12\}$ y es fácil demostrar que $p_X(j) = 1/12$

para j en S_X . Puesto que $B = \{1, 2, 3, 4\}$:

$$\begin{aligned} p_X(j|B) &= \frac{P[\{X = j\} \cap B]}{P[B]} = \frac{P[X \in \{j\} \cap \{1, 2, 3, 4\}]}{P[X \in \{1, 2, 3, 4\}]} \\ &= \begin{cases} \frac{P[X=j]}{1/3} = \frac{1}{4} & j \in \{1, 2, 3, 4\} \\ 0 & j \notin \{1, 2, 3, 4\}. \end{cases} \end{aligned}$$

El suceso B anterior está relacionado con X solamente. El suceso D , sin embargo, se expresa en términos de los resultados del experimento subyacente (es decir, minutos, no horas), por lo que la probabilidad de la intersección tiene que ser expresado en consecuencia:

$$\begin{aligned} p_X(j|D) &= \frac{P[\{X = j\} \cap D]}{P[D]} = \frac{P[\zeta : X(\zeta) = j, \zeta \in \{2, \dots, 11\}]}{P[\zeta \in \{2, \dots, 11\}]} \\ &= \begin{cases} \frac{P[\zeta \in \{2, 3, 4, 5\}]}{10/60} = \frac{4}{10} & j = 1 \\ \frac{P[\zeta \in \{6, 7, 8, 9, 10\}]}{10/60} = \frac{5}{10} & j = 2 \\ \frac{P[\zeta \in \{11\}]}{10/60} = \frac{1}{10} & j = 3. \end{cases} \end{aligned}$$

La mayoría de las veces el suceso C se define en términos de X , por ejemplo $C = \{X > 10\}$ o $C = \{a \leq X \leq b\}$. Para x_k en S_X , tenemos el siguiente resultado general:

$$p_X(x_k|C) = \begin{cases} \frac{p_X(x_k)}{P[C]} & x_k \in C \\ 0 & x_k \notin C. \end{cases} \quad (3.27)$$

La expresión anterior está totalmente determinada por la pmf de X .

Ejemplo 3.23 Tiempos de espera residuales

Sea X el tiempo necesario para transmitir un mensaje, donde X es una variable aleatoria Uniforme con $S_X = \{1, 2, \dots, L\}$. Supongamos que un mensaje está siendo transmitido durante m unidades de tiempo, calcula la probabilidad de que el tiempo de transmisión restante sea j unidades de tiempo.

Solución

Tenemos que $C = \{X > m\}$, de modo que $m + 1 \leq m + j \leq L$:

$$\begin{aligned} p_X(m+j|X > m) &= \frac{P[X = m+j]}{P[X > m]} \\ &= \frac{\frac{1}{L}}{\frac{L-m}{L}} = \frac{1}{L-m} \quad m+1 \leq m+j \leq L. \end{aligned} \quad (3.28)$$

X tiene la misma probabilidad de tomar cualquiera de los otros $L - m$ posibles valores. Cuando m crece, $1/(L-m)$ crece implicando que el final de la transmisión

del mensaje es cada vez más probable.

Muchos experimentos aleatorios tiene una partición natural del espacio muestral S mediante la unión de sucesos disjuntos B_1, B_2, \dots, B_n . Sea $p_X(x|B_i)$ la pmf condicionada de X dado B_i . El teorema de la probabilidad total nos permite calcular la pmf de X en términos de las pmf condicionadas:

$$p_X(x) = \sum_{i=1}^n p_X(x|B_i)P[B_i]. \quad (3.29)$$

Ejemplo 3.24 Tiempo de vida de un dispositivo

Una línea de producción produce dos tipos de dispositivos. Los dispositivos de tipo 1 ocurren con probabilidad α y funciona durante un tiempo relativamente corto que es geoméricamente distribuido con parámetro r . Los dispositivos de tipo 2 funcionan durante mucho más tiempo, ocurren con una probabilidad $1 - \alpha$ y tienen una vida útil que es geoméricamente distribuida con parámetro s . Sea X la vida útil de un dispositivo al azar. Calcula la pmf de X .

Solución

El experimento aleatorio que genera X consiste en seleccionar un tipo de dispositivo y observar su vida útil. Podemos dividir el conjunto de los resultados de este experimento en el suceso B_1 , que consiste los resultados en los que el dispositivo es de tipo 1, y B_2 , que consiste en los resultados en que el dispositivo es de tipo 2. La pmf condicionada de X dado el tipo de dispositivo es:

$$p_{X|B_1}(k) = (1 - r)^{k-1}r \quad k = 1, 2, \dots$$

y

$$p_{X|B_2}(k) = (1 - s)^{k-1}s \quad k = 1, 2, \dots$$

Se obtiene la pmf de X usando la ecuación (3.29):

$$\begin{aligned} p_X(k) &= p_X(k|B_1)P[B_1] + p_X(k|B_2)P[B_2] \\ &= (1 - r)^{k-1}r\alpha + (1 - s)^{k-1}s(1 - \alpha) \quad k = 1, 2, \dots \end{aligned}$$

3.4.2 Esperanza condicionada

Sea X una variable aleatoria discreta y supongamos que sabemos que el suceso B ha ocurrido. La **esperanza condicionada de X dado B** se define como:

$$m_{X|B} = E[X|B] = \sum_{x \in S_X} xp_X(x|B) = \sum_k x_k p_X(x_k|B) \quad (3.30)$$

donde se aplica el requisito de convergencia absoluta en la suma. La **varianza condicionada de X dado B** se define como:

$$\begin{aligned} V[X|B] &= E[(X - m_{X|B})^2|B] = \sum_{k=1}^{\infty} (x_k - m_{X|B})^2 p_X(x_k|B) \\ &= E[X^2|B] - m_{X|B}^2. \end{aligned}$$

Obsérvese que la variación se mide con respecto a $m_{X|B}$, no m_X .

Sea B_1, B_2, \dots, B_n la partición de S y sea $p_X(x|B_i)$ la pmf condicionada de X dado B_i . La $E[X]$ puede ser calculada a partir de las esperanzas condicionadas $E[X|B]$:

$$E[X] = \sum_{i=1}^n E[X|B_i]P[B_i]. \quad (3.31a)$$

Por el teorema de probabilidad total tenemos:

$$\begin{aligned} E[X] &= \sum_k k p_X(x_k) = \sum_k k \left\{ \sum_{i=1}^n p_X(x_k|B_i)P[B_i] \right\} \\ &= \sum_{i=1}^n \left\{ \sum_k k p_X(x_k|B_i) \right\} P[B_i] = \sum_{i=1}^n E[X|B_i]P[B_i], \end{aligned}$$

donde expresamos primero $p_X(x_k)$ en términos de la pmf condicionada y luego cambiamos el orden de los sumatorios. Utilizando el mismo enfoque también podemos demostrar

$$E[g(X)] = \sum_{i=1}^n E[g(X)|B_i]P[B_i]. \quad (3.31b)$$

Ejemplo 3.25 Tiempo de vida de un dispositivo

Calcula la media y la varianza de los dispositivos en el ejemplo 3.24.

Solución

La media condicionada y el segundo momento de cada tipo de dispositivo son los de una variable aleatoria Geométrica con el parámetro correspondiente:

$$\begin{aligned} m_{X|B_1} &= 1/r & E[X^2|B_1] &= (1+r)/r^2 \\ m_{X|B_2} &= 1/s & E[X^2|B_2] &= (1+s)/s^2. \end{aligned}$$

La media y el segundo momento de X son entonces:

$$\begin{aligned} m_X &= m_{X|B_1}\alpha + m_{X|B_2}(1-\alpha) = \alpha/r + (1-\alpha)/s \\ E[X^2] &= E[X^2|B_1]\alpha + E[X^2|B_2](1-\alpha) = \alpha(1+r)/r^2 + (1-\alpha)(1+s)/s^2. \end{aligned}$$

Por último, la varianza de X es:

$$V[X] = E[X^2] - m_X^2 = \frac{\alpha(1+r)}{r^2} + \frac{(1-\alpha)(1+s)}{s^2} - \left(\frac{\alpha}{r} + \frac{(1-\alpha)}{s} \right)^2.$$

Obsérvese que *no* usamos las varianzas condicionadas para calcular $V[Y]$ porque la ecuación (3.31b) no se aplica a las varianzas condicionadas. Sin embargo, la ecuación se aplica a los segundos momentos condicionados.

3.5 Variables aleatorias importantes

Algunas variables aleatorias aparecen en aplicaciones muy diversas y no relacionadas. Esto se debe a que dichas variables aleatorias modelan mecanismos fundamentales inherentes al comportamiento aleatorio. En esta sección describimos las variables aleatorias discretas más importantes y discutimos cómo surgen y la relación entre ellas. La tabla 3.1 resume las propiedades básicas de las variables aleatorias discretas aquí presentadas.

Las variables aleatorias discretas surgen mayoritariamente en aplicaciones de conteo. Comenzamos con la variable aleatoria de Bernoulli como modelo para un simple lanzamiento de moneda. Contando los resultados de múltiples lanzamientos de moneda obtenemos las variables aleatorias binomial, geométrica y de Poisson.

3.5.1 Variable aleatoria de Bernoulli

Sea A un suceso relacionado con el resultado de un experimento aleatorio. La variable aleatoria de Bernoulli I_A (definida en el ejemplo 3.5), es igual a 1 si sucede el evento A y 0 en caso contrario. I_A es una variable aleatoria ya que asigna un número a cada suceso de S . Es una variable aleatoria discreta con rango = $\{0, 1\}$ y su pmf es

$$p_I(0) = 1 - p \quad \text{y} \quad p_I(1) = p, \quad (3.32)$$

donde $P[A] = p$.

En el ejemplo 3.10 calculamos la media de I_A :

$$m_I = E[I_A] = p.$$

La media muestral de n experimentos de Bernoulli es simplemente la frecuencia relativa de los éxitos y converge a p cuando n crece:

$$\langle I_A \rangle_n = \frac{0N_0(n) + 1N_1(n)}{n} = f_1(n) \rightarrow p.$$

En el ejemplo 3.20 se calculó la varianza de I_A :

$$\sigma_I^2 = V[I_A] = p(1 - p) = pq.$$

La varianza es cuadrática en p , con valor cero en $p = 0$ y $p = 1$ y máximo en

Cuadro 3.1 Variables aleatorias discretas

Variable aleatoria de Bernoulli

$$S_X = \{0, 1\}$$

$$p_0 = q = 1 - p \quad p_1 = p \quad 0 \leq p \leq 1$$

$$E[X] = p \quad V[X] = p(1 - p) \quad G_X(z) = (q + pz)$$

La variable aleatoria de Bernoulli es el valor de la función indicadora I_A del suceso A ; $X = 1$ si A sucede y 0, en otro caso.

Variable aleatoria binomial

$$S_X = \{0, 1, \dots, n\}$$

$$p_k = \binom{n}{k} p^k (1 - p)^{n-k} \quad k = 0, 1, \dots, n$$

$$E[X] = np \quad V[X] = np(1 - p) \quad G_X(z) = (q + pz)^n$$

X es el número de éxitos en n experimentos de Bernoulli y, por tanto, es la suma de n variables aleatorias de Bernoulli independientes.

Variable aleatoria geométrica

$$S_X = \{1, 2, \dots\}$$

$$p_k = p(1 - p)^{k-1} \quad k = 1, 2, \dots$$

$$E[X] = \frac{1}{p} \quad V[X] = \frac{1-p}{p^2} \quad G_X(z) = \frac{pz}{1-qz}$$

X es el número de intentos de Bernoulli independientes hasta el primer éxito.

La variable aleatoria Geométrica es la única variable aleatoria discreta con la propiedad de no tener memoria.

Variable aleatoria binomial negativa

$S_X = \{r, r + 1, \dots\}$ donde r es un número entero positivo

$$p_k = \binom{k-1}{r-1} p^r (1 - p)^{k-r} \quad k = r, r + 1, \dots$$

$$E[X] = \frac{r}{p} \quad V[X] = \frac{r(1-p)}{p^2} \quad G_X(z) = \left(\frac{pz}{1-qz} \right)^r$$

X es el número de intentos de Bernoulli independientes hasta el r -ésimo éxito.

Variable aleatoria de Poisson

$$S_X = \{0, 1, 2, \dots\}$$

$$p_k = \frac{\alpha^k}{k!} e^{-\alpha} \quad k = 0, 1, \dots \text{ y } \alpha > 0$$

$$E[X] = \alpha \quad V[X] = \alpha \quad G_X(z) = e^{\alpha(z-1)}$$

X es el número de sucesos que ocurren en una unidad de tiempo cuando el tiempo entre sucesos se distribuye exponencialmente con media $1/\alpha$.

Variable aleatoria uniforme

$$S_X = \{1, 2, \dots, L\}$$

$$p_k = \frac{1}{L} \quad k = 1, 2, \dots, L$$

$$E[X] = \frac{L+1}{2} \quad V[X] = \frac{L^2+1}{12} \quad G_X(z) = \frac{z}{L} \frac{1-z^L}{1-z}$$

En la variable aleatoria Uniforme los elementos son equiprobables.

$p = 1/2$. Esto concuerda con la intuición ya que los valores de p cercanos a 0 o a 1 implican una preponderancia de éxitos o fracasos y, por lo tanto, menos variabilidad en los valores observados. La variabilidad máxima se produce cuando $p = 1/2$ que corresponde al caso en el que es más difícil predecir.

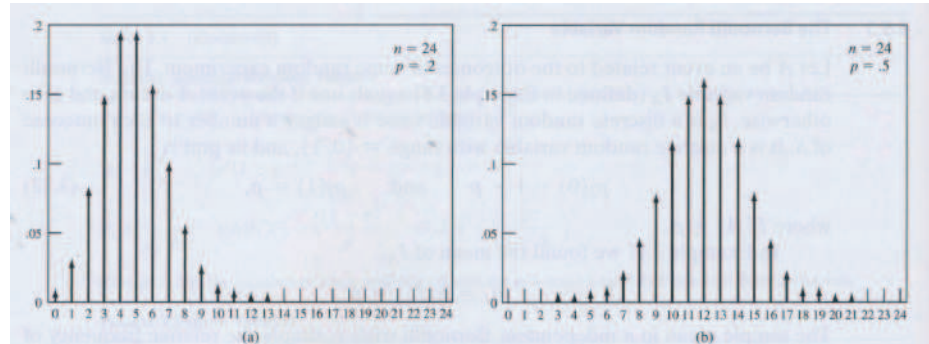


Figura 3.8 Función de probabilidad de la variable aleatoria Binomial (a) $p = 0.2$; (b) $p=0.5$.

Cada experimento de Bernoulli, independientemente de la definición de A , es equivalente al lanzamiento de una moneda no equilibrada con probabilidad de caras p . En este sentido, lanzar una moneda puede ser visto como representativo de un mecanismo fundamental para generar aleatoriedad y la variable aleatoria de Bernoulli es el modelo asociado con él.

3.5.2 Variable aleatoria Binomial

Supongamos que un experimento aleatorio se repite n veces de forma independiente. Sea X el número de veces que un suceso determinado A ocurre en esos n intentos. X es una variable aleatoria de rango $S_X = \{0, 1, \dots, n\}$. Por ejemplo, X puede ser el número de caras en n lanzamientos de una moneda. Sea I_j la función indicadora del suceso A en el lanzamiento j , entonces

$$X = I_1 + I_2 + \dots + I_n,$$

esto es, X es la suma de las variables aleatorias de Bernoulli asociadas con cada uno de los n intentos independientes.

En la Sección 2.5, vimos que la probabilidad de X depende en n y p :

$$P[X = k] = p_X(k) = \binom{n}{k} p^k (1-p)^{n-k} \quad k = 0, 1, \dots, n. \quad (3.33)$$

X es denominada **variable aleatoria Binomial**. La Figura 3.8 muestra la pdf de X para $n = 24$ y $p = 0.2$ y $p = 0.5$. Cabe destacar que $P[X = k]$ alcanza el máximo en $k_{\text{máx}} = \lfloor (n+1)p \rfloor$ donde $\lfloor x \rfloor$ representa al mayor entero menor o igual que x . Cuando $(n+1)p$ es entero, entonces el máximo se alcanza en $k_{\text{máx}}$ y en $k_{\text{máx}} - 1$.

La variable aleatoria Binomial surge en aplicaciones donde hay dos clases de objetos (por ejemplo, caras/cruces, bits correctos/incorrectos, productos correctos/defectuosos) y estamos interesados en el número de objetos de tipo 1 en un

lote de n objetos seleccionados al azar, donde la clase de cada objeto es independiente de la clase de los demás objetos del lote. Vimos ejemplos de variables aleatorias binomiales en la sección 2.5.

Ejemplo 3.26 Media de un variable aleatoria Binomial

La esperanza de X es:

$$\begin{aligned}
 E[X] &= \sum_{k=0}^n kp_X(k) = \sum_{k=0}^n k \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=1}^n k \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \\
 &= np \sum_{k=1}^n \frac{(n-1)!}{(k-1)!(n-k)!} p^{k-1} (1-p)^{n-k} \\
 &= np \sum_{j=0}^{n-1} \frac{(n-1)!}{(n-1-j)!} p^j (1-p)^{n-j-1} = np, \tag{3.34}
 \end{aligned}$$

donde la primera línea use el hecho de que el término para $k = 0$ en el sumatorio es 0, en la segunda línea se elimina los factores k y np fuera del sumatorio y la última línea usa el hecho de que el sumatorio es igual a 1 ya que suma todos los términos de una pmf binomial de parámetros $n - 1$ y p .

La esperanza $E[X] = np$ concuerda con nuestra intuición ya que esperamos una fracción p de los resultados que sean éxitos.

Ejemplo 3.27 Varianza de la variable aleatoria Binomial

Para calcular $E[X^2]$, eliminamos el término $k = 0$ y hacemos $k' = k - 1$:

$$\begin{aligned}
 E[X^2] &= \sum_{k=0}^n k^2 \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} = \sum_{k=1}^n k \frac{n!}{(k-1)!(n-k)!} p^k (1-p)^{n-k} \\
 &= np \sum_{k'=0}^{n-1} (k'+1) \binom{n-1}{k'} p^{k'} (1-p)^{n-1-k} \\
 &= np \left\{ \sum_{k'=0}^{n-1} k' \binom{n-1}{k'} p^{k'} (1-p)^{n-1-k} + \sum_{k'=0}^{n-1} 1 \binom{n-1}{k'} p^{k'} (1-p)^{n-1-k} \right\} \\
 &= np \{ (n-1)p + 1 \} = np(np + q).
 \end{aligned}$$

En la tercera línea vemos que la primera suma es la media de una variable aleatoria Binomial con parámetros $(n - 1)$ y p y entonces es igual a $(n - 1)p$. El segundo sumatorio es la suma de las probabilidades binomiales y, por tanto, iguala a 1.

Obtenemos la varianza como sigue:

$$\sigma_X^2 = E[X^2] - E[X]^2 = np(np + q) - (np)^2 = npq = np(1 - p).$$

Vemos que la varianza de la binomial es n veces la varianza de la variable aleatoria de Bernoulli. Observamos que aquellos valores de p cercanos a 0 o a 1 implican una varianza menor y que la máxima variabilidad se alcanza cuando $p = 1/2$.

Ejemplo 3.28 Sistemas redundantes

Un sistema usa redundancia triple en aras de su fiabilidad: se instala tres microprocesadores y el sistema está diseñado de forma que está operativo siempre que alguno de los procesadores funcione. Supongamos que la probabilidad de que un microprocesador esté activo después de t segundos es $p = e^{-\lambda t}$. Calcula la probabilidad de que el sistema esté operativo después de t segundos.

Solución

Sea X el número de microprocesadores que funcionan en el tiempo t . X es una variable aleatoria Binomial con parámetros $n = 3$ y p . Entonces:

$$P[X \geq 1] = 1 - P[X = 0] = 1 - (1 - e^{-\lambda t})^3.$$

3.5.3 Variable aleatoria Geométrica

La variable aleatoria Geométrica se obtiene contando el número M de experimentos independientes de Bernoulli hasta obtener el primer éxito. M es denominada variable aleatoria Geométrica y toma valores en el conjunto $\{1, 2, \dots\}$. En la Sección 2.5, dedujimos que la pmf de M viene dada por

$$P[M = k] = p_M(k) = (1 - p)^{k-1} p \quad k = 1, 2, \dots, \quad (3.35)$$

donde $p = P[A]$ es la probabilidad de “éxito” en cada intento (experimento de Bernoulli). La Figura 3.5(b) muestra la pmf geométrica para $p = 1/2$. Nótese que $P[M = k]$ decae geoméricamente con k y que el ratio de términos consecutivos es $p_M(k+1)/p_M(k) = (1 - p) = q$. Cuando p crece, la pmf decae más rápido.

La probabilidad de que $M \leq k$ se puede escribir de forma cerrada:

$$P[M \leq k] = \sum_{j=1}^k pq^{j-1} = p \sum_{j'=0}^{k-1} q^{j'} = p \frac{1 - q^k}{1 - q} = 1 - q^k. \quad (3.36)$$

A veces, estaremos interesados en $M' = M - 1$, el número de “fracasos” hasta el primer éxito. M' también es considerada como una variable aleatoria Geométrica. Su pmf es:

$$P[M' = k] = P[M = k + 1] = (1 - p)^k p \quad k = 0, 1, 2, \dots \quad (3.37)$$

En los ejemplos 3.14 y 3.21, calculamos la media y la varianza de la variable aleatoria Geométrica:

$$m_M = E[M] = 1/p \quad V[M] = \frac{1 - p}{p^2}.$$

Vemos que la media y la varianza aumentan cuando p , la probabilidad de éxito, decrece.

La variable aleatoria Geométrica es la única variable aleatoria discreta que satisface la propiedad de no tener memoria:

$$P[M \geq k + j | M > j] = P[M \geq k] \quad \forall j, k > 1.$$

La expresión de arriba establece que si no ha ocurrido ningún éxito en los primeros j intentos, entonces la probabilidad de tener que probar otras k veces es la misma que la probabilidad de necesitar inicialmente al menos k intentos hasta conseguir el primer éxito. Esto implica que cada vez que ocurre un fracaso, el sistema “olvida” y comienza de nuevo, como si el próximo fuera el primer intento.

La variable aleatoria Geométrica surge en aplicaciones en las que uno está interesado en el tiempo (es decir, el número de intentos) que transcurre entre sucesos en una secuencia infinita de experimentos, como en los Ejemplos 2.11 y 2.34. Algunos ejemplos en los que la variable aleatoria Geométrica modificada M' surge son: el número de clientes en espera en un sistemas de colas de espera; el número de puntos blancos entre sucesivos puntos negros en un documento digitalizado en blanco y negro.

3.5.4 Variable aleatoria de Poisson

En muchas aplicaciones estamos interesados en contar el número de veces que ocurre un suceso en un periodo de tiempo concreto o en una región del espacio determinada. La variable aleatoria de Poisson surge en situaciones donde los sucesos ocurren “de forma completamente aleatoria” en el tiempo o en el espacio. Por ejemplo, la variable aleatoria de Poisson aparece en el cómputo de emisiones de sustancias radioactivas, en el conteo de demandas de conexión telefónica y en el recuento de defectos de un chip semiconductor.

La pmf de la **variable aleatoria de Poisson** viene dada por

$$P[N = k] = p_N(k) = \frac{\alpha^k}{k!} e^{-\alpha} \quad k = 0, 1, 2, \dots, \quad (3.38)$$

donde α es la media del número de veces que ocurre un suceso en un determinado intervalo de tiempo o región del espacio. La figura 3.9 muestra la pmf de Poisson para varios valores de α . Para $\alpha < 1$, $P[N = k]$ alcanza el máximo en $k = 0$; para $\alpha > 1$, $P[N = k]$ llega a su máximo con $k = \alpha$ y $k = \alpha - 1$.

La pmf de la variable aleatoria de Poisson suma 1, ya que

$$\sum_{k=0}^{\infty} \frac{\alpha^k}{k!} e^{-\alpha} = e^{-\alpha} \sum_{k=0}^{\infty} \frac{\alpha^k}{k!} = e^{-\alpha} e^{\alpha} = 1,$$

utilizando el hecho de que la suma infinita del segundo término de la igualdad es el desarrollo en serie de e^{α} .

Resulta fácil demostrar que la media y la varianza de una variable aleatoria de Poisson son:

$$E[N] = \alpha \quad \sigma_N^2 = V[N] = \alpha.$$

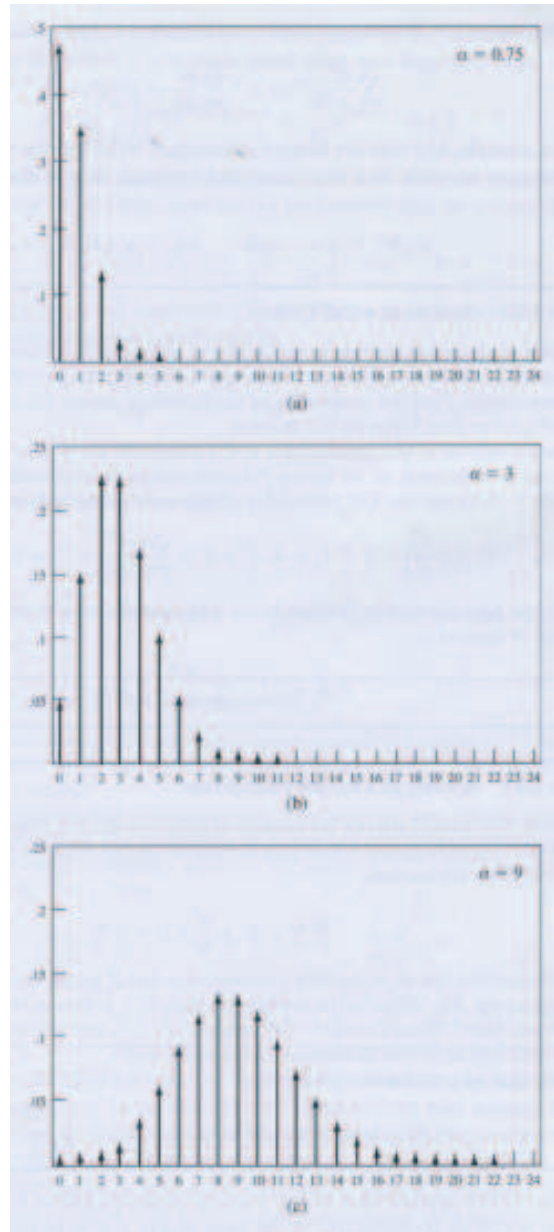


Figura 3.9 Función de probabilidad de la variable aleatoria de Poisson (a) $\alpha = 0.75$; (b) $\alpha = 3$, (c) $\alpha = 9$.

Ejemplo 3.29 Peticiones en un Call Center

El número N de peticiones que llegan a un Call Center en t segundos es una

variable aleatoria de Poisson con $\alpha = \lambda t$ donde λ es la tasa media de llegadas de las consultas por segundo. Supongamos que la tasa de llegada es de cuatro consultas por minuto. Calcula la probabilidad de los siguientes sucesos: (a) más de 4 consultas en 10 segundos, (b) menos de 5 consultas en 2 minutos.

Solución

La tasa de llegadas de las consultas por segundo es $\lambda = 4$ peticiones/60 s = 1/15 consultas/s. En la parte a, el intervalo de tiempo es de 10 segundos, así que tenemos una variable aleatoria de Poisson con $\alpha = (1/15 \text{ consultas/s})(10\text{s}) = 10/15$ consultas. La probabilidad de interés se evalúa numéricamente:

$$P[N > 4] = 1 - P[N \leq 4] = 1 - \sum_{k=0}^4 \frac{(2/3)^k}{k!} e^{-2/3} = 6.33(10^{-4}).$$

En la parte b, el intervalo de tiempo de interés es $t = 120$ segundos, por lo que $\alpha = 1/15 * 120$ segundos = 8. La probabilidad de interés es la siguiente:

$$P[N \leq 5] = \sum_{k=0}^5 \frac{(8)^k}{k!} e^{-8} = 0.10.$$

Ejemplo 3.30 Llegadas a un multiplexor de paquetes

El número N de llegadas de paquetes en t segundos en un multiplexor es una variable aleatoria de Poisson con $\alpha = \lambda t$ donde λ es la tasa media de llegada de paquetes por segundo. Calcula la probabilidad de que no lleguen paquetes en t segundos.

Solución

$$P[N = 0] = \frac{\alpha^0}{0!} e^{-\alpha} = e^{-\alpha}.$$

Esta ecuación tiene una interpretación interesante. Sea Z el tiempo hasta la llegada del primer paquete. Supongamos que nos preguntamos, “¿Cuál es la probabilidad de que $X > t$, es decir, la próxima llegada no se produzca hasta, por lo menos, dentro de t segundos?” Nótese que $\{N = 0\}$ implica que $\{Z > t\}$ y viceversa, por lo que $P[Z > t] = e^{-\lambda t}$. La probabilidad de ninguna llegada disminuye exponencialmente con t .

Obsérvese que también podemos demostrar que

$$P[N(t) \geq n] = 1 - P[N(t) < n] = 1 - \sum_{k=0}^{n-1} \frac{(\lambda t)^k}{k!} e^{-\lambda t}.$$

Una de las aplicaciones de la distribución de Poisson (ecuación (3.38)) es aproximar la distribución binomial cuando p es muy pequeño y n muy grande, esto

es, cuando el suceso A de interés es muy raro pero el número de intentos es muy grande. Demostramos que *si* $\alpha = np$ es fijo, entonces cuando n se hace grande:

$$p_k = \binom{n}{k} p^k (1-p)^{n-k} \simeq \frac{\alpha^k}{k!} e^{-\alpha} \quad k = 0, 1, \dots \quad (3.39)$$

La aproximación de la ecuación (3.39) se obtiene tomando el límite cuando $n \rightarrow \infty$ en la expresión para p_k , manteniendo $\alpha = np$ fijo. Primero, consideremos que la probabilidad de que no ocurra ningún éxito en n intentos:

$$p_0 = (1-p)^n = \left(1 - \frac{\alpha}{n}\right)^n \rightarrow e^{-\alpha} \text{ cuando } n \rightarrow \infty, \quad (3.40)$$

donde el límite en la última expresión es un resultado conocido del cálculo. El resto de probabilidades se obtienen teniendo en cuenta que

$$\begin{aligned} \frac{p_{k+1}}{p_k} &= \frac{\binom{n}{k+1} p^{k+1} q^{n-k-1}}{\binom{n}{k} p^k q^{n-k}} = \frac{k!(n-k)!p}{(k+1)!(n-k-1)!q} \\ &= \frac{(n-k)p}{(k+1)q} = \frac{(1-k/n)\alpha}{(k+1)(1-\alpha/n)} \\ &\rightarrow \frac{\alpha}{k+1} \text{ cuando } n \rightarrow \infty. \end{aligned}$$

Entonces las probabilidades límite satisfacen que

$$p_{k+1} = \frac{\alpha}{k+1} p_k = \left(\frac{\alpha}{k+1}\right) \left(\frac{\alpha}{k}\right) \cdots \left(\frac{\alpha}{1}\right) p_0 = \frac{\alpha^k}{k!} e^{-\alpha} \quad (3.41)$$

Entonces, la pmf de Poisson es el límite de la pmf binomial cuando el número de intentos de Bernoulli n se hace muy grande y la probabilidad de éxito p se mantiene pequeña, tal que $\alpha = np$.

Ejemplo 3.31 Errores en una transmisión óptica

Un sistema de comunicación óptica transmite la información a una velocidad de 10^9 bits /segundo. La probabilidad de error en un bit en el sistema de comunicación óptica es de 10^{-9} . Calcula la probabilidad de cinco o más errores en 1 segundo.

Solución

Cada transmisión de bits corresponde a un ensayo de Bernoulli donde “éxito” corresponde a un error en un bit en la transmisión. La probabilidad de k errores en $n = 10^9$ transmisiones (1 segundo) es entonces dada por la probabilidad binomial con $n = 10^9$ y $p = 10^{-9}$. La aproximación de Poisson utiliza que $\alpha = np = 10^9(10^{-9}) = 1$. Así

$$\begin{aligned} P[N \geq 5] &= 1 - P[N < 5] = 1 - \sum_{k=0}^4 \frac{\alpha^k}{k!} e^{-\alpha} \\ &= 1 - e^{-1} \left\{ 1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \frac{1}{4!} \right\} = 0.00366. \end{aligned}$$

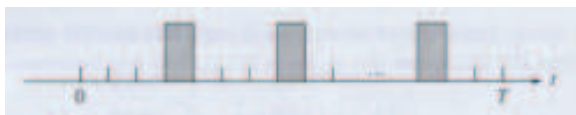


Figura 3.10 Sucesos que ocurren en n subintervalos de $[0, T]$.

La variable aleatoria de Poisson aparece de manera natural en muchas situaciones físicas porque muchos modelos son muy grandes en escala e involucran sucesos muy raros. Por ejemplo, la pmf de Poisson proporciona una predicción precisa de la frecuencia relativa del número de partículas emitidas por una masa radioactiva durante un periodo de tiempo fijo. Esta correspondencia puede explicarse como sigue. Una masa radioactiva se compone de un gran número de átomos, n . En un intervalo de tiempo fijado, cada átomo posee una probabilidad muy pequeña p de desintegración y emisión de una partícula radioactiva. Si los átomos se desintegran independientemente de los demás átomos, entonces el número de emisiones en un intervalo de tiempo puede verse como el número de éxitos en n intentos. Por ejemplo, un microgramo de radio contiene aproximadamente $n = 10^{16}$ átomos y la probabilidad de que uno de esos átomos se desintegre durante un milisegundo de tiempo es $p = 10^{-15}$. Entonces, sería insuficiente afirmar que se cumple las condiciones para la aproximación de la Ecuación (3.39): n es tan grande y p es tan pequeño que podemos argumentar que haciendo el límite $n \rightarrow \infty$ se obtiene que el número de emisiones es exactamente una variable aleatoria de Poisson.

La variable aleatoria de Poisson también surge en situaciones en las que podemos imaginar una secuencia de experimentos de Bernoulli en el tiempo o en el espacio. Supongamos que contamos el número de veces que ocurre un suceso en un intervalo de tiempo de T segundos. Dividamos el intervalo en un número muy grande n de subintervalos, como se muestra en la Figura 3.10. Una pulsación en uno de los subintervalos indica que el suceso ha ocurrido. Cada subintervalo puede verse como un experimento de Bernoulli si se verifican las siguientes condiciones: (1) Puede ocurrir a lo sumo un suceso en un subintervalo, esto es, la probabilidad de que ocurra más de un suceso es despreciable; (2) los resultados en diferentes subintervalos son independientes; y (3) la probabilidad de que ocurra un suceso en un subintervalo es $p = \alpha/n$, donde α es el número medio de sucesos observados en un intervalo de 1 segundo. El número de sucesos N en un segundo es una variable aleatoria Binomial de parámetros n y $p = \alpha/n$. Entonces, cuando $n \rightarrow \infty$, N se convierte en una variable aleatoria de Poisson de parámetro $\lambda = \alpha/T$. En el Capítulo 8 desarrollamos este resultado al tratar el proceso de Poisson.

3.5.5 La variable aleatoria Uniforme

La variable aleatoria Uniforme discreta Y toma valores en un conjunto de números enteros consecutivos $S_Y = \{j + 1, \dots, j + L\}$ con igual probabilidad:

$$p_Y(k) = \frac{1}{L} \text{ para } k \in \{j + 1, \dots, j + L\}. \quad (3.42)$$

Esta variable aleatoria se produce cuando los resultados son igualmente probables, por ejemplo, el lanzamiento de una moneda o un dado, el giro de una flecha en una rueda dividida en segmentos iguales, la selección de números desde una urna. Es fácil demostrar que la media y la varianza son las siguientes:

$$E[Y] = j + \frac{L + 1}{2} \quad V[Y] = \frac{L^2 - 1}{12}.$$

Ejemplo 3.32 Variable aleatoria uniforme discreta en el intervalo unitario

Sea X una variable aleatoria Uniforme en $S_X = \{0, 1, \dots, L - 1\}$. Definimos la variable aleatoria Uniforme discreta en el intervalo unitario como

$$U = \frac{X}{L} \text{ así } S_U = \left\{ 0, \frac{1}{L}, \frac{2}{L}, \frac{3}{L}, \dots, 1 - \frac{1}{L} \right\}.$$

U tiene pmf:

$$p_U\left(\frac{k}{L}\right) = \frac{1}{L} \quad k = 0, 1, \dots, L - 1.$$

La pmf de U pone igual masa de probabilidad $1/L$ en puntos equidistantes $x_k = k/L$ en el intervalo unitario. La probabilidad de un subintervalo del intervalo unitario es igual al número de puntos en el subintervalo multiplicada por $1/L$. Como L es muy grande, esta probabilidad es esencialmente la longitud del subintervalo.

Problemas

3.1 Sea X el número máximo de caras obtenidas cuando Carlos y Miguel lanzan una moneda al aire dos veces.

- Describe el espacio S subyacente a este experimento aleatorio y especifica las probabilidades de sus elementos.
- Muestra la asignación de S a S_X , el rango de X .
- Calcula las probabilidades de los distintos valores de X .

3.2 Una urna contiene 9 billetes de 5 euros y un billete de 50 euros. Sea X la cantidad total resultante de extraer dos billetes de la urna sin reemplazamiento.

- (a) Describe el espacio muestral S subyacente de este experimento aleatorio y especifica las probabilidades de sus sucesos elementales.
- (b) Muestra la asignación de S a S_X , el rango de X .
- (c) Calcula las probabilidades de los distintos valores de X .
- (d) Repite el ejercicio considerando extracciones con reemplazamiento.

3.3 Se necesita una contraseña de m bits para acceder a un sistema. Un hacker trabaja sistemáticamente utilizando todos los posibles patrones de m bits. Sea X el número de patrones de prueba hasta que encuentra la contraseña correcta.

- (a) Describe el espacio muestral S .
- (b) Muestra la asignación de S a S_X , el rango de X .
- (c) Calcula las probabilidades de los distintos valores de X .

3.4 Sea X el máximo de caras de los lanzamientos de la moneda en el problema 3.1.

- (a) Compara la pmf de X con la pmf de Y , el número de caras en dos lanzamientos de una moneda. Explica la diferencia.
- (b) Supongamos que Carlos utiliza una moneda con una probabilidad de cara $p = 3/4$. Calcula la pmf de X .

3.5 Un equipo reserva un path en una red durante 10 minutos. Para ampliar la reserva con éxito, el ordenador debe enviar un mensaje de “actualización” antes del tiempo de caducidad. Sin embargo, los mensajes se pierden con una probabilidad de $1/2$. Supongamos que se necesita 10 segundos para enviar una petición de actualización y recibir un acuse de recibo. ¿Cuándo debe empezar la computadora a enviar mensajes de actualización con el fin de tener un 99% de probabilidad de éxito en la ampliación del tiempo de reserva?

3.6 En el problema 3.4, compara $E[Y]$ con $E[X]$ donde X es el máximo de lanzamientos de la moneda. Compara $V[Y]$ con $V[X]$.

3.7 Considera la paradoja de San Petersburgo en el ejemplo 3.15. Supongamos que el casino cuenta con un total de $M = 2^m$ euros, por lo que sólo puede permitirse un número limitado de lanzamientos de la moneda.

- (a) ¿Cuántos lanzamientos puede permitirse el casino?
- (b) Halla el beneficio esperado para el jugador.
- (c) ¿Cuánto debe estar dispuesto a pagar un jugador para jugar a este juego?

3.8 En el problema 3.4,

- (a) Calcula la pmf condicionada de X , el máximo de lanzamientos de la moneda, dado que $X > 0$.
- (b) Calcula la pmf condicionada de X dado que Miguel saca una cara en dos lanzamientos.

- (c) Calcula la pmf condicionada de X dado que Miguel saca una cara en el primer lanzamiento.
- (d) En el apartado b), calcula la probabilidad de que Carlos consiga el máximo puesto que $X = 2$.

3.9 En el problema 3.2,

- (a) Calcula la pmf condicionada de X dado que la primera extracción fue de k euros.
- (b) Calcula la esperanza condicionada correspondiente a la parte a).
- (c) Calcula $E[X]$, utilizando el resultado de la parte b).
- (d) Calcula $E[X^2]$ y $V[X]$, utilizando el enfoque de las partes b) y c).

3.10 Se selecciona aleatoriamente 8 números del intervalo unidad.

- (a) Calcula la probabilidad de que los primeros cuatro números sean menores que 0.25 y los cuatro últimos mayores que 0.25.
- (b) Calcula la probabilidad de que cuatro números sean menores que 0.25 y otros cuatro sean mayores que 0.25.
- (c) Calcula la probabilidad de que los primeros tres números sean menores que 0.25, los dos siguientes estén entre 0.25 y 0.75 y los tres últimos sean mayores que 0.75.
- (d) Calcula la probabilidad de que tres números sean menores que 0.25, otros dos estén entre 0.25 y 0.75 y otros tres sean mayores que 0.75.
- (e) Calcula la probabilidad de que los primeros cuatro números sean menores que 0.25 y los cuatro últimos sean mayores que 0.75.
- (f) Calcula la probabilidad de que cuatro números sean menores que 0.25 y otros cuatro sean mayores que 0.75.

3.11 Un reproductor de audio utiliza un disco duro de baja calidad. El coste inicial para construir el reproductor es de 50 euros. El disco duro falla después de cada mes de uso con probabilidad $1/2$. El coste de reparar el disco duro es de 20 euros. Si se ofrece una garantía de un año, ¿cuánto debe cobrar el fabricante para que la probabilidad de perder dinero con un reproductor sea del 1% o menos? ¿Cuál es el coste medio por reproductor?

3.12 El número de peticiones esperando a ser procesadas es una variable aleatoria de Poisson con parámetro $\alpha = \lambda/n\mu$, donde λ es el número medio de peticiones que llegan en un día, μ es el número de peticiones que pueden ser procesadas por un empleado en un día y n es el número de empleados. Sea $\lambda = 5$ y $\mu = 1$. Calcula el número de empleados necesarios para que la probabilidad de que haya más de cuatro peticiones en espera sea menor del 10%. ¿Cuál es la probabilidad de que no haya peticiones en espera?

3.13 El número de peticiones de páginas que llegan a un servidor web es una variable aleatoria de Poisson con una media de 6000 peticiones por minuto.

- (a) Calcula la probabilidad de que no haya peticiones en un periodo de 100 ms.
- (b) Calcula la probabilidad de que haya entre 5 y 10 peticiones en un periodo de 100 ms.

3.14 Un sistema de comunicaciones binario tiene una probabilidad de error en un bit de 10^{-6} . Supongamos que las transmisiones ocurren en bloques de 10000 bits. Sea N el número de errores introducidos por el canal en un bloque de transmisión.

- (a) Calcula $P[N = 0]$ y $P[N \leq 3]$.
- (b) ¿Para qué valor de p va a ser la probabilidad de 1 o más errores en un bloque ser del 99%?

3.15 X se distribuye uniformemente en el conjunto $\{-3, \dots, 3.4\}$.

- (a) Calcula la media y la varianza de X .
- (b) Calcula la media y la varianza de $Y = -2X^2 + 3$.
- (c) Calcula la media y la varianza de $W = \cos(\pi X/8)$.
- (d) Calcula la media y la varianza de $Z = \cos^2(\pi X/8)$.

3.16 La proporción de productos defectuosos en una línea de producción es p . Cada producto se somete a una prueba que identifica correctamente los productos defectuosos con probabilidad a .

- (a) Asumiendo que los productos no defectuosos siempre pasan el test, ¿cuál es la probabilidad de que tengamos que realizar la prueba a k productos hasta que encontremos un producto defectuoso?
- (b) Suponiendo que los productos defectuosos se retiren, ¿qué proporción de los productos restantes son defectuosos?
- (c) Ahora supongamos que los productos no defectuosos son identificados como defectuosos con probabilidad b . Repite el apartado b).

3.17 Un sistema de transmisión de datos utiliza mensajes de T segundos de duración. Después de cada transmisión, el transmisor se para y espera la respuesta del receptor durante T segundos. El receptor responde inmediatamente con un mensaje indicando que el mensaje se recibió de forma correcta. El transmisor procede a enviar un mensaje nuevo si recibe una respuesta en menos de T segundos; en otro caso, retransmite el anterior mensaje de nuevo. Supongamos que los mensajes pueden volverse completamente ilegibles mientras que se transmiten y que esto ocurre con una probabilidad p . Calcula la tasa máxima posible de mensajes se pueden transmitir con éxito desde el emisor al receptor.

3.18 Un inspector selecciona cada n -ésimo producto en una línea de producción para una inspección detallada. Supongamos que el tiempo entre llegadas de productos es una variable aleatoria Exponencial con media 1 minuto y supongamos que se tarda 2 minutos en inspeccionar un elemento. Calcula el mínimo valor de

n tal que, con una probabilidad del 90 % o más, la inspección se haya completado antes de la llegada del siguiente producto que requiere una inspección.

3.19 El número X de fotones contados por un receptor en un sistema de comunicación óptico es una variable aleatoria de Poisson con tasa λ_1 en presencia de una señal y es una variable aleatoria de Poisson con tasa $\lambda_0 < \lambda_1$ cuando no hay señales presentes. Supongamos que la presencia de una señal tiene probabilidad p .

- (a) Calcula $P[\text{señal} | X = k]$ y $P[\text{no señal} | X = k]$.
- (b) El receptor utiliza la siguiente regla de decisión: *Si $P[\text{señal} | X = k] > P[\text{no señal} | X = k]$, decide que hay señal; en otro caso, decide ausencia de señal.* Demuestra que esta regla de decisión lleva a la siguiente regla de umbral: *Si $X > T$, decide que hay señal; en otro caso, decide que no la hay.*
- (c) ¿Cuál es la probabilidad de error para la regla de decisión del apartado anterior?

3.20 Una fuente de información binaria (e.g., un escáner de documentos) genera cadenas muy largas de ceros seguidos de unos ocasionales. Supongamos que los símbolos son independientes y que $p = P[\text{símbolo} = 0]$ es muy cercana a uno. Considera el siguiente esquema para codificar la secuencia X de ceros entre dos unos consecutivos:

- (a) Si $X = n$, expresa n como un múltiplo de un número entero $M = 2^m$ y un resto r , esto es, calcula k y r tal que $n = kM + r$, donde $0 \leq r < M - 1$;
- (b) La palabra de código binario para n , entonces, consiste de un prefijo que consiste en k ceros seguido de un 1 y un sufijo que consiste en la representación de m bits del resto r . El decodificador puede deducir el valor de n de esta cadena binaria.
 - (i) Calcula la probabilidad de que el prefijo tenga k ceros, en el supuesto de que $p^M = 1/2$.
 - (ii) Calcula la longitud media de palabra de código cuando $p^M = 1/2$.
 - (iii) Calcula la relación de compresión, que se define como el cociente entre la longitud media entre la longitud media de palabra de código cuando $p^M = 1/2$.

4 Variables aleatorias

En el capítulo 3 introdujimos el concepto de variable aleatoria y se desarrollaron métodos para el cálculo de probabilidades y medias para el caso de que la variable aleatoria es discreta. En este capítulo se examina el caso general, donde la variable aleatoria puede ser discreta, continua o de tipo mixto. Presentamos la función de distribución que se utiliza en la definición formal de una variable aleatoria, y que se puede definir para los tres tipos de variables aleatorias. También presentamos la función de densidad de variables aleatorias continuas. La probabilidad de sucesos relacionados con una variable aleatoria se puede expresar la integral de la función de densidad. La esperanza de las variables aleatorias continuas también se define y se relaciona con nuestra noción intuitiva de la media. Desarrollamos una serie de métodos para el cálculo de probabilidades y medias que son las herramientas básicas en el análisis y diseño de sistemas con componentes estocásticos.

4.1 La función de distribución

La función de probabilidad de una variable aleatoria discreta se define en términos de sucesos de la forma $\{X = b\}$. La función de distribución es un enfoque alternativo que utiliza sucesos de la forma $\{X \leq b\}$. La función de distribución tiene la ventaja de que no se limita a variables aleatorias discretas y se aplica a todos los tipos de variables aleatorias. Comenzamos con una definición formal de variable aleatoria.

DEFINICIÓN 4.1 Consideremos un experimento aleatorio con espacio muestral S y clase de sucesos \mathcal{F} . Una **variable aleatoria** X es una función desde el espacio muestral S a \mathbb{R} con la propiedad de que el conjunto $A_b = \{\zeta : X(\zeta) \leq b\} \in \mathcal{F}$ para cada $b \in \mathbb{R}$.

La definición requiere simplemente que cada conjunto A_b tenga una probabilidad bien definida en el experimento aleatorio subyacente y esto no es un problema en los casos que vamos a considerar. ¿Por qué esta definición utiliza conjuntos de la forma $\{\zeta : X(\zeta) \leq b\}$ y no $\{\zeta : X(\zeta) = x_b\}$? Veremos que todos los sucesos de interés en la recta real se pueden expresar en términos de conjuntos de la forma $\{\zeta : X(\zeta) \leq b\}$.

La **función de distribución** (cdf) de una variable aleatoria X se define como la probabilidad del suceso $\{X \leq x\}$:

$$F_X(x) = P[X \leq x] \quad \text{para } -\infty < x < +\infty, \quad (4.1)$$

esto es, la probabilidad de que la variable aleatoria X tome un valor en el intervalo $(-\infty, x]$. En términos del espacio muestral subyacente, la cdf es la probabilidad del evento $\{\zeta : X(\zeta) \leq x\}$. El suceso $\{X \leq x\}$ y su probabilidad varían con x . En otras palabras, $F_X(x)$ es una función de la variable x .

La cdf es simplemente una forma conveniente de especificar la probabilidad de todos los intervalos semi-infinitos de la recta real de la forma $(-\infty, x]$. Cuando tratamos con números, los eventos de interés son los intervalos de la recta real y sus complementarios, sus uniones e intersecciones. Mostramos más adelante que las probabilidades de todos estos eventos se pueden expresar en términos de la cdf.

La cdf tiene la siguiente interpretación en términos de la frecuencia relativa. Supongamos que el experimento que arroja el resultado ζ , y por tanto $X(\zeta)$, se realiza un número elevado de veces. $F_X(b)$ es la proporción de veces que $X(\zeta) \leq b$.

Antes de enunciar las propiedades generales de la cdf, presentamos ejemplos de los tres tipos de variables aleatorias.

Ejemplo 4.1 Tres lanzamientos de una moneda

La figura 4.1(a) muestra la cdf de X , el número de caras en tres lanzamientos de una moneda equilibrada. A partir del ejemplo 3.1 sabemos que X toma sólo los valores 0, 1, 2 y 3 con probabilidades $1/8, 3/8, 3/8$ y $1/8$, respectivamente, así que $F_X(x)$ es simplemente la suma de las probabilidades de los resultados de $\{0, 1, 2, 3\}$ que son menores o iguales a x . La cdf resultante es una función definida a trozos no decreciente que crece desde 0 a 1. La cdf tiene saltos en los puntos 0, 1, 2, 3 de magnitud $1/8, 3/8, 3/8$ y $1/8$, respectivamente.

Miremos más de cerca una de estas discontinuidades, por ejemplo, en las cercanías de $x = 1$. Para δ positivo y pequeño, tenemos

$$F_X(1 - \delta) = P[X \leq 1 - \delta] = P\{0 \text{ caras}\} = \frac{1}{8}$$

así que el límite de la cdf cuando x se aproxima a 1 por la izquierda es $1/8$. Sin embargo,

$$F_X(1) = P[X \leq 1] = P[0 \text{ o } 1 \text{ caras}] = \frac{1}{8} + \frac{3}{8} = \frac{1}{2},$$

y, además, el límite por la derecha es

$$F_X(1 + \delta) = P[X \leq 1 + \delta] = P[0 \text{ o } 1 \text{ caras}] = \frac{1}{2}.$$

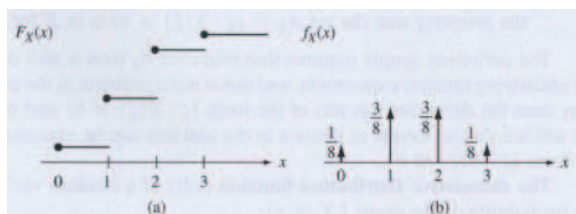


Figura 4.1 cdf (a) y pdf (b) de una variable aleatoria discreta.

Así, la función de distribución es continua por la derecha e igual a $1/2$ en el punto $x = 1$. De hecho, observamos que la magnitud del salto en el punto $x = 1$ es igual a $P[X = 1] = 1/2 - 1/8 = 3/8$. A partir de ahora vamos a utilizar puntos en la gráfica para indicar el valor de la función de distribución en los puntos de discontinuidad.

La cdf se puede escribir de forma compacta en términos de la función escalón unitario:

$$u(x) = \begin{cases} 0 & x < 0 \\ 1 & x \geq 0. \end{cases} \quad (4.2)$$

entonces

$$F_X(x) = \frac{1}{8}u(x) + \frac{3}{8}u(x-1) + \frac{3}{8}u(x-2) + \frac{1}{8}u(x-3).$$

Ejemplo 4.2 Variable aleatoria uniforme en el intervalo unitario

Gira una flecha conectada al centro de un tablero circular. Sea θ el ángulo final de la flecha, donde $0 < \theta \leq 2\pi$. La probabilidad de que θ caiga en un subintervalo de $(0, 2\pi]$ es proporcional a la longitud del subintervalo. La variable aleatoria X se define por $X(\theta) = \theta/2\pi$. Calcula la cdf de X :

Solución

A medida que aumenta θ de 0 a 2π , X aumenta de 0 a 1. Ningún resultado θ conduce a un valor $x \leq 0$, por lo que

$$F_X(x) = P[X \leq x] = P[\emptyset] = 0 \text{ para } x < 0.$$

Para $0 < x \leq 1$, $\{X \leq x\}$ se produce cuando $\{\theta \leq 2\pi x\}$ entonces

$$F_X(x) = P[X \leq x] = P[\{\theta \leq 2\pi x\}] = 2\pi x/2\pi = x \quad 0 < x \leq 1. \quad (4.3)$$

Por último, para $x > 1$, todos los resultados θ llevan a $\{X(\theta) \leq 1 < x\}$, por lo tanto:

$$F_X(x) = P[X \leq x] = P[0 < \theta \leq 2\pi] = 1 \quad x > 1.$$

Decimos que X es una **variable aleatoria Uniforme** en el intervalo unitario. La figura 4.2(a) muestra la función de distribución de la variable aleatoria Uniforme general X . Vemos que $F_X(x)$ es una función continua no decreciente que crece desde 0 hasta 1 cuando x cambia desde su valor mínimo a su valor máximo.

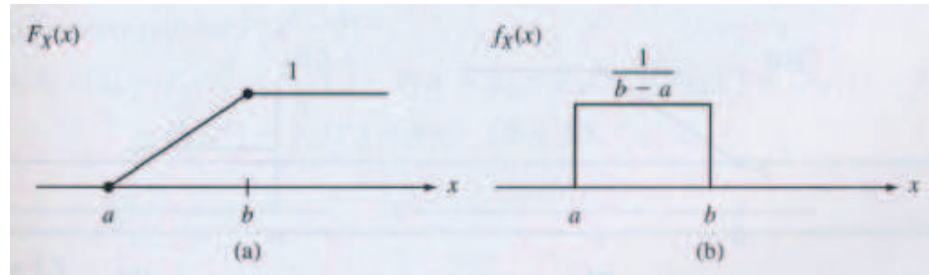


Figura 4.2 cdf (a) y pdf (b) de una variable aleatoria continua.

Ejemplo 4.3 El tiempo de espera X de un cliente en una parada de taxis es cero si el cliente encuentra un taxi estacionado en la parada y es una longitud de tiempo uniformemente distribuida en el intervalo $[0, 1]$ (en horas) si no encuentra un taxi a su llegada. La probabilidad de que un taxi se encuentre en la parada cuando el cliente llega es p . Calcula la cdf de X .

Solución

La cdf se calcula aplicando el teorema de probabilidad total:

$$F_X(x) = P[X \leq x] = P[X \leq x | \text{taxi}]p + P[X \leq x | \text{no taxi}](1 - p).$$

Nótese que $P[X \leq x | \text{taxi}] = 1$ cuando $x \geq 0$ y 0, en caso contrario. Además $P[X \leq x | \text{no taxi}]$ viene dada por la ecuación (4.3), por lo tanto,

$$F_X(x) = \begin{cases} 0 & x < 0 \\ p + (1 - p)x & 0 \leq x \leq 1 \\ 1 & x > 1. \end{cases}$$

La cdf que se muestra en la figura 4.3(a), combina algunas de las propiedades de la cdf del ejemplo 4.1 (discontinuidad en el 0) y la cdf del ejemplo 4.2 (continuidad por intervalos). Obsérvese que $F_X(x)$ se puede expresar como la suma de una función escalón de amplitud p y una función continua de x .

Estamos ya preparados para enunciar las propiedades básicas de la cdf. Los axiomas de probabilidad y sus corolarios implican que la cdf tenga las siguientes propiedades:

- (i) $0 \leq F_X(x) \leq 1$.
- (ii) $\lim_{x \rightarrow \infty} F_X(x) = 1$.
- (iii) $\lim_{x \rightarrow -\infty} F_X(x) = 0$.
- (iv) $F_X(x)$ es una función no decreciente, esto es, si $a < b$ entonces $F_X(a) \leq F_X(b)$.

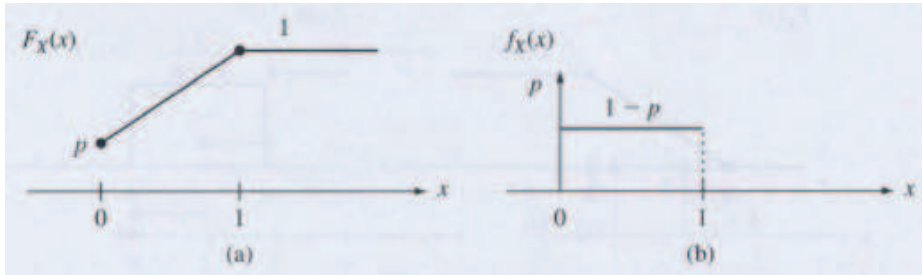


Figura 4.3 cdf (a) y pdf (b) de una variable aleatoria de tipo mixto.

- (v) $F_X(x)$ es continua por la derecha, esto es, para $h > 0$, $F_X(b) = \lim_{h \rightarrow 0} F_X(b+h) = F_X(b^+)$.

Estas cinco propiedades confirman que, en general, la función de distribución es una función no decreciente que crece desde 0 hasta 1 a medida que x aumenta desde $-\infty$ a ∞ . Ya hemos observado estas propiedades en los ejemplos 4.1, 4.2 y 4.3. La propiedad (v) implica que en los puntos de discontinuidad, la cdf es igual al límite por la derecha. Observamos esta propiedad en los ejemplos 4.1 y 4.3. En el ejemplo 4.2, la cdf es continua para todos los valores de x , es decir, la cdf es continua tanto por la derecha y como por la izquierda para todo x .

La cdf tiene las siguientes propiedades que nos permiten calcular la probabilidad de sucesos relacionados con los intervalos y los valores individuales de X :

- (vi) $P[a < X \leq b] = F_X(b) - F_X(a)$.
 (vii) $P[X = b] = F_X(b) - F_X(b^-)$.
 (viii) $P[X > x] = 1 - F_X(x)$.

La propiedad (vii) establece que la probabilidad de que $X = b$ es dada por la magnitud del salto de la cdf en el punto b . En consecuencia, si la cdf es continua en el punto b , entonces $P[X = b] = 0$. Las propiedades (vi) y (vii) se pueden combinar para calcular las probabilidades de otros tipos de intervalos. Por ejemplo, como $\{X \leq a\} \cup \{a < X \leq b\} = \{X \leq b\}$, entonces

$$\begin{aligned} P[a \leq X \leq b] &= P[X = a] + P[a < X \leq b] \\ &= F_X(a) - F_X(a^-) + F_X(b) - F_X(a) = F_X(b) - F_X(a^-) \end{aligned} \quad (4.4)$$

Si la cdf es continua en los extremos de un intervalo, los extremos tienen probabilidad nula y, por tanto, pueden ser incluidos o excluidos del intervalo sin afectar al valor de la probabilidad.

Ejemplo 4.4 Sea X el número de caras en tres lanzamientos de una moneda equilibrada. Usando la cdf, calcula la probabilidad de los sucesos $A = \{1 < X \leq 2\}$, $B = \{0.5 \leq X < 2.5\}$ y $C = \{1 \leq X < 2\}$.

Solución

Por la propiedad (vi) y la figura 4.1 se tiene que

$$P[1 < X \leq 2] = F_X(2) - F_X(1) = 7/8 - 1/2 = 3/8.$$

La cdf es continua en $x = 0.5$ y $x = 2.5$, entonces

$$P[0.5 \leq X < 2.5] = F_X(2.5) - F(0.5) = 7/8 - 1/8 = 6/8.$$

Como $\{1 \leq X < 2\} \cup \{X = 2\} = \{1 \leq X \leq 2\}$, por la ecuación (4.4) tenemos

$$P[1 \leq X < 2] + P[X = 2] = F_X(2) - F_X(1^-),$$

y utilizando la propiedad (vii) para $P[X = 2]$:

$$\begin{aligned} P[1 \leq X < 2] &= F_X(2) - F_X(1^-) - P[X = 2] = F_X(2) - F_X(1^-) - (F_X(2) - F_X(2^-)) \\ &= F_X(2^-) - F_X(1^-) = 4/8 - 1/8 = 3/8. \end{aligned}$$

Ejemplo 4.5 Sea X la variable aleatoria Uniforme del ejemplo 4.2. Utilizando la cdf, calcula la probabilidad de los sucesos $\{-0.5 < X < 0.25\}$, $\{0.3 < X < 0.65\}$ y $\{|X - 0.4| > 0.2\}$.

Solución

La cdf de X es continua en todos los puntos, de modo que se tiene:

$$\begin{aligned} P[-0.55 < X \leq 0.25] &= F_X(0.25) - F_X(-0.5) = 0.25 - 0 = 0.25, \\ P[0.3 < X < 0.65] &= F_X(0.65) - F_X(0.3) = 0.65 - 0.3 = 0.35, \\ P[|X - 0.4| > 0.2] &= P[\{X < 0.2\} \cup \{X > 0.6\}] = P[X < 0.2] + P[X > 0.6] \\ &= F_X(0.2) + (1 - F_X(0.6)) = 0.2 + 0.4 = 0.6. \end{aligned}$$

Ahora consideramos la prueba de las propiedades de la cdf.

- La primera propiedad se deduce por el hecho de que la cdf es una probabilidad y, por consiguiente, tiene que satisfacer el Axioma I y el Corolario 2.
- Para demostrar la propiedad (iv), téngase en cuenta que como $\{X \leq a\}$ es un subconjunto de $\{X \leq b\}$, ha de tener menor o igual probabilidad (Corolario 7).
- Para demostrar la propiedad (vi), téngase en cuenta que $\{X \leq b\}$ se puede escribir como la unión de dos sucesos mutuamente excluyentes: $\{X \leq a\} \cup \{a < X \leq b\}$, y, se tiene, por el Axioma III, que $F_X(a) + P[a < X \leq b] = F_X(b)$.
- La propiedad (viii) se obtiene porque $\{X > x\} = \{X \leq x\}^c$ y el Corolario 1.

4.1.1 Los tres tipos de variables aleatorias

Las variables aleatorias de los ejemplos 4.1, 4.2 y 4.3 representan los tres tipos básicos de variables en los que estamos interesados.

Las **variables aleatorias discretas** tienen una cdf que es una función de x definida a trozos y continua por la derecha, con saltos en un conjunto numerable de puntos x_0, x_1, x_2, \dots . La variable aleatoria del ejemplo 4.1 es un ejemplo típico de variable aleatoria discreta. La cdf $F_X(x)$ de una variable aleatoria discreta es la suma de las probabilidades de los elementos muestrales menores que x y se puede escribir como la suma ponderada de funciones escalón unitario como en el ejemplo 4.1:

$$F_X(x) = \sum_{x_k \leq x} p_X(x_k) = \sum_k p_X(x_k) u(x - x_k), \quad (4.5)$$

donde la pmf $p_X(x_k) = P[X = x_k]$ proporciona la magnitud de los saltos en la cdf. Vemos que la pmf se puede obtener con la cdf y vice versa.

Una **variable aleatoria continua** se define como una variable aleatoria cuya cdf $F_X(x)$ es continua siempre, y que, además, es suficientemente suave como para ser escrita como una integral de una función no negativa $f(x)$:

$$F_X(x) = \int_{-\infty}^x f(t) dt. \quad (4.6)$$

La variable aleatoria del ejemplo 4.2 se puede escribir como la integral de la función mostrada en la figura 4.2(b). Para variables aleatorias continuas, la cdf es continua en todos los puntos, por lo cual, la propiedad (vii) implica que $P[X = x] = 0$ para todo x . ¡Todos los elementos muestrales posibles tienen probabilidad nula! Una consecuencia inmediata es que la pmf no se puede utilizar para caracterizar las probabilidades de X . La comparación entre las ecuaciones (4.5) y (4.6) sugiere cómo podemos proceder para caracterizar las variables aleatorias continuas. Para variables aleatorias discretas (ec. 4.5) calculamos las probabilidades como la suma de masas de probabilidad en puntos. Para variables aleatorias continuas (ec. 4.6) calculamos las probabilidades como integrales de “densidades de probabilidad” en intervalos de la recta real.

Una **variable aleatoria mixta** es una variable aleatoria con un cdf que tiene saltos en un conjunto numerable de puntos x_0, x_1, x_2, \dots pero que también aumenta de forma continua en al menos un intervalo de valores de x . La cdf de estas variables aleatorias tiene la forma

$$F_X(x) = pF_1(x) + (1 - p)F_2(x),$$

donde $0 < p < 1$ y $F_1(x)$ es la cdf de una variable aleatoria discreta y $F_2(x)$ es la cdf de una variable aleatoria continua. La variable aleatoria del ejemplo 4.3 es de tipo mixto.

Las variables aleatorias mixtas pueden ser vistas como producidas por un proceso en dos etapas: se lanza una moneda; si el resultado es cara, se genera

una variable aleatoria discreta de acuerdo con $F_1(x)$; en el otro caso, se genera una variable aleatoria continua con cdf $F_2(x)$.

4.2 La función de densidad

La **función de densidad de X** (pdf), si existe, se define como la derivada de $F_X(x)$:

$$f_X(x) = \frac{dF_X(x)}{dx}. \quad (4.7)$$

En esta sección mostramos que la pdf es una forma alternativa, más útil, de especificar la información contenida en la función de distribución.

La pdf representa la “densidad” de probabilidad en el punto x en el siguiente sentido: La probabilidad de que X esté en un intervalo pequeño en la vecindad de x – es decir, $\{x < X \leq x + h\}$ – es

$$\begin{aligned} P[x < X \leq x + h] &= F_X(x + h) - F_X(x) \\ &= \frac{F_X(x + h) - F_X(x)}{h} h. \end{aligned} \quad (4.8)$$

Si la cdf es derivable en el punto x , entonces cuando h se hace pequeña,

$$P[x < X \leq x + h] \simeq f_X(x)h. \quad (4.9)$$

Entonces $f_X(x)$ representa la “densidad” de probabilidad en el punto x en el sentido de que la probabilidad de que X está en un pequeño intervalo de la vecindad de x es aproximadamente $f_X(x)h$. La derivada de la cdf, cuando existe, es positiva porque la cdf es una función no decreciente de x , entonces

(i)

$$f_X(x) \geq 0. \quad (4.10)$$

Las Ecuaciones (4.9) y (4.10) nos proporcionan un método alternativo para especificar las probabilidades relacionadas con la variable aleatoria X . Podemos empezar por definir una función no negativa $f_X(x)$, llamada función de densidad, que especifique las probabilidades de los sucesos del tipo “ X está en un pequeño intervalo de ancho dx alrededor del punto x ,” como se muestra en la figura 4.4(a). La probabilidad de sucesos relacionados con X se expresan en términos de la pdf sumando las probabilidades de intervalos de ancho dx . Como los anchos de los intervalos se acercan a 0, obtenemos una integral en términos de la pdf. Por ejemplo, la probabilidad de un intervalo $[a, b]$ es

(ii)

$$P[a \leq X \leq b] = \int_a^b f_X(x)dx. \quad (4.11)$$

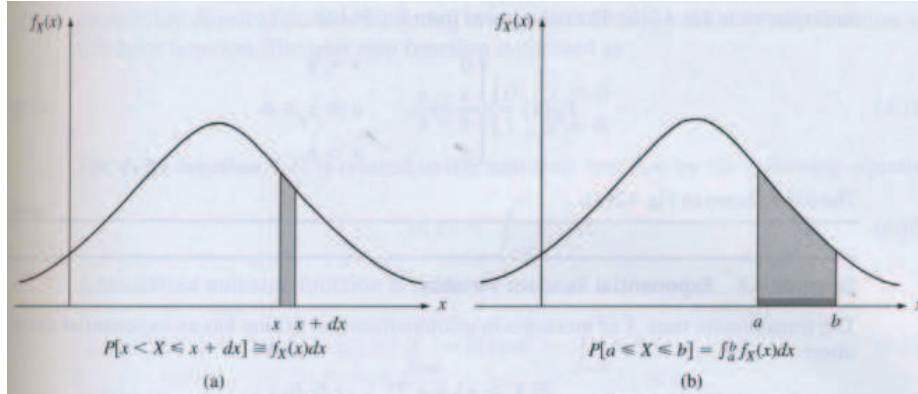


Figura 4.4 (a) La función de densidad especifica la probabilidad de intervalos de anchura infinitesimal. (b) La probabilidad de un intervalo $[a, b]$ es el área por debajo de la pdf en dicho intervalo.

La probabilidad de un intervalo es el área por debajo de $f_X(x)$ en ese intervalo, como se muestra en la figura 4.4(b). La probabilidad de cualquier suceso que consista en la unión de intervalos disjuntos puede ser calculada sumando las integrales de la pdf en cada una de los intervalos.

La cdf de X puede ser obtenida integrando la pdf:

(iii)

$$F_X(x) = \int_{-\infty}^x f_X(t)dt. \quad (4.12)$$

En la sección 4.1, definimos *una variable aleatoria continua* como una variable X cuya cdf es como la descrita en la ecuación (4.12). Como las probabilidades de todos los sucesos relacionados con X pueden ser escritas en función de la cdf, se deduce que éstas también pueden ser escritas en términos de la pdf. Entonces *la pdf especifica completamente el comportamiento de las variables continuas*.

Si hacemos x tender a infinito en la ecuación (4.12), obtenemos una condición de *normalización* para las pdf:

(iv)

$$1 = \int_{-\infty}^{\infty} f_X(t)dt. \quad (4.13)$$

La pdf refuerza la noción intuitiva de probabilidad al tener atributos similares a la “masa física.” La ecuación (4.11) establece que la “masa” de probabilidad en un intervalo es la integral de la “densidad de masa de probabilidad” en el intervalo. La ecuación (4.13) establece que la masa total disponible es una unidad.

Una pdf válida puede estar formada por cualquier función continua a trozos

no negativa $g(x)$ que tenga una integral finita:

$$\int_{-\infty}^{\infty} g(x)dx = c < \infty. \quad (4.14)$$

Definiendo $f_X(x) = g(x)/c$ obtenemos una función que satisface la condición de normalización. Nótese que la pdf debe estar definida para todos los valores reales de x ; si X no toma valores en alguna región de la recta real, definiremos $f_X(x) = 0$ en dicha región.

Ejemplo 4.6 Variable aleatoria uniforme

La pdf de la variable aleatoria Uniforme viene dada por

$$f_X(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & x < a \text{ y } x > b. \end{cases} \quad (4.15a)$$

y se muestra en la figura 4.2(b). La cdf se calcula a partir de la ecuación (4.12):

$$F_X(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \leq x \leq b \\ 1 & x > b. \end{cases} \quad (4.15b)$$

Esta cdf se muestra en la figura 4.2(a).

Ejemplo 4.7 Variable aleatoria exponencial

El tiempo de transmisión de mensajes en un sistema de comunicaciones X tiene una distribución exponencial:

$$P[X > x] = e^{-\lambda x} \quad x > 0.$$

Calcula la cdf y la pdf de X .

Solución

La cdf viene dada por $F_X(x) = 1 - P[X > x]$

$$F_X(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-\lambda x} & x \geq 0. \end{cases} \quad (4.16a)$$

La pdf se obtiene aplicando la ecuación (4.7):

$$f_X(x) = F'_X(x) = \begin{cases} 0 & x < 0 \\ \lambda e^{-\lambda x} & x \geq 0. \end{cases} \quad (4.16b)$$

Ejemplo 4.8 Variable aleatoria de Laplace

La pdf de las muestras de la amplitud de las ondas del habla decae exponencialmente con una tasa α , de modo que se propone la siguiente pdf:

$$f_X(x) = ce^{-\alpha|x|} \quad -\infty < x < \infty. \quad (4.17)$$

Calcula la constante c y luego calcula la probabilidad $P[|X| < v]$.

Solución

Utilizamos la condición de normalización (iv) para hallar c :

$$1 = \int_{-\infty}^{\infty} ce^{-\alpha|x|} dx = 2 \int_0^{\infty} ce^{-\alpha x} dx = \frac{2c}{\alpha}.$$

Entonces $c = \alpha/2$. La probabilidad $P[|X| < v]$ se calcula integrando la pdf:

$$P[|X| < v] = \frac{\alpha}{2} \int_{-v}^v e^{-\alpha|x|} dx = 2 \frac{\alpha}{2} \int_0^v e^{-\alpha x} dx = 1 - e^{-\alpha v}.$$

4.2.1 pdf de variables aleatorias discretas

La derivada de la cdf no existe en aquellos puntos donde la cdf no es continua. En consecuencia, el concepto de pdf, tal y como se define en la Ecuación (4.7) no puede ser aplicado a variables aleatorias discretas en los puntos donde la cdf es discontinua. Podemos generalizar la definición de la función de densidad considerando la relación entre la función escalón unitario y la función delta. La **función escalón unitario** se define como

$$u(x) = \begin{cases} 0 & x < 0 \\ 1 & x \geq 0. \end{cases} \quad (4.18a)$$

La **función delta** $\delta(t)$ se define en función de la función escalón unitario mediante la siguiente ecuación:

$$u(x) = \int_{-\infty}^x \delta(t) dt. \quad (4.18b)$$

Una función escalón unitario trasladada es entonces:

$$u(x - x_0) = \int_{-\infty}^{x-x_0} \delta(t) dt = \int_{-\infty}^x \delta(t' - x_0) dt'. \quad (4.18c)$$

Sustituyendo la ecuación (4.18c) en la cdf de una variable aleatoria discreta:

$$\begin{aligned} F_X(x) &= \sum_k p_X(x_k) u(x - x_k) = \sum_k p_X(x_k) \int_{-\infty}^x \delta(t - x_k) dt \\ &= \int_{-\infty}^x \sum_k p_X(x_k) \delta(t - x_k) dt. \end{aligned} \quad (4.19)$$

Esto sugiere que definamos la **pdf de una variable aleatoria discreta** como

$$f_X(x) = \frac{d}{dx} F_X(x) = \sum_k p_X(x_k) \delta(x - x_k). \quad (4.20)$$

Entonces, la definición generalizada de pdf coloca una función delta de peso $P[X = x_k]$ en los puntos x_k donde la cdf es discontinua.

La pdf de la variable aleatoria discreta del ejemplo 4.1 se muestra en la figura 4.1(b). La pdf de una variable de tipo mixto también contendrá funciones delta en los puntos en los que la cdf no sea continua. La pdf de la variable del ejemplo 4.3 se muestra en la figura 4.3(b).

Ejemplo 4.9 Sea X el número de caras en tres lanzamientos de una moneda como en el ejemplo 4.1. Calcula la pdf de X . Calcula $P[1 < X \leq 2]$ y $P[2 \leq X < 3]$ integrando la pdf.

Solución

En el ejemplo 4.1 vimos que la cdf de X viene dada por

$$F_X(x) = \frac{1}{8}u(x) + \frac{3}{8}u(x-1) + \frac{3}{8}u(x-2) + \frac{1}{8}u(x-3).$$

Entonces, por las ecuaciones (4.18b) y (4.19) se sigue que

$$f_X(x) = \frac{1}{8}\delta(x) + \frac{3}{8}\delta(x-1) + \frac{3}{8}\delta(x-2) + \frac{1}{8}\delta(x-3).$$

Cuando las funciones delta aparecen en los límites de la integral, debemos indicar si se incluyen en la integración. Entonces, en $P[1 < X \leq 2] = P[X \in (1, 2]]$, la función delta localizada en 1 es excluida de la integral y la función delta en el 2 es incluida:

$$P[1 < X \leq 2] = \int_{1+}^{2+} f_X(x)dx = \frac{3}{8}.$$

De forma similar, tenemos que

$$P[2 \leq X < 3] = \int_{2-}^{3-} f_X(x)dx = \frac{3}{8}.$$

4.2.2 cdf y pdf condicionadas

Las cdf condicionadas se pueden definir directamente utilizando el mismo método que utilizamos para las pmf condicionadas. Supongamos que ocurre el suceso C y que $P[C] > 0$. La **cdf de X condicionada a C** se define como

$$F_X(x|C) = \frac{P[\{X \leq x\} \cap C]}{P[C]} \text{ si } P[C] > 0. \quad (4.21)$$

Es fácil de demostrar que $F_X(x|C)$ satisface todas las propiedades de una cdf. La **pdf de X condicionada a C** es a su vez definida como

$$f_X(x|C) = \frac{d}{dx}F_X(x|C). \quad (4.22)$$

Ejemplo 4.10 El tiempo de vida X de una máquina tiene una cdf continua $F_X(x)$. Calcula las cdf y pdf condicionadas, dado el suceso $C = \{X > t\}$ (es decir, “la máquina todavía funciona en el tiempo t ”).

Solución

La cdf condicionada es

$$F_X(x|X > t) = P[X \leq x|X > t] = \frac{P[\{X \leq x\} \cap \{X > t\}]}{P[X > t]}.$$

La intersección de los sucesos del numerador es igual al conjunto vacío cuando $x < t$ e igual a $\{t < X \leq x\}$ cuando $x \geq t$. Entonces

$$F_X(x|X > t) = \begin{cases} 0 & x < t \\ \frac{F_X(x) - F_X(t)}{1 - F_X(t)} & x \geq t. \end{cases}$$

La pdf condicionada se halla derivando con respecto a x :

$$f_X(x|X > t) = \frac{f_X(x)}{1 - F_X(t)} \quad x \geq t.$$

Ahora supongamos que tenemos una partición del espacio muestral S que es la unión de los sucesos disjuntos B_1, B_2, \dots, B_n . Sea $F_X(x|B_i)$ la cdf condicionada de X dado el suceso B_i . El teorema de la probabilidad total nos permite calcular la cdf de X en términos de las cdf condicionadas:

$$F_X(x) = P[X \leq x] = \sum_{i=1}^n P[X \leq x|B_i]P[B_i] = \sum_{i=1}^n F_X(x|B_i)P[B_i]. \quad (4.23)$$

La pdf se obtiene derivando:

$$f_X(x) = \frac{d}{dx}F_X(x) = \sum_{i=1}^n f_X(x|B_i)P[B_i]. \quad (4.24)$$

Ejemplo 4.11 Un sistema de transmisión binario envía un bit “0” transmitiendo una señal de voltaje $-v$ y un bit “1”, transmitiendo un $+v$. La señal recibida se corrompe con un ruido gaussiano y es dada por:

$$Y = X + N$$

donde X es la señal transmitida y N es el voltaje del ruido con pdf $f_N(x)$. Asumamos que $P[\text{“1”}] = p = 1 - P[\text{“0”}]$. Calcula la probabilidad de Y .

Solución

Sea B_0 el suceso se transmite un “0” y B_1 el suceso de que se transmite un “1”,

entonces B_0, B_1 forman una partición y

$$\begin{aligned} F_Y(x) &= F_Y(x|B_0)P[B_0] + F_Y(x|B_1)P[B_1] \\ &= P[Y \leq x|X = -v](1-p) + P[Y \leq x|X = v]p. \end{aligned}$$

Dado que $Y = X + N$, el suceso $\{Y < x|X = v\}$ es equivalente a $\{v + N < x\}$ y $\{N < x - v\}$ y el suceso $\{Y < x|X = -v\}$ es equivalente a $\{N < x + v\}$. Entonces las cdf condicionadas son

$$F_Y(x|B_0) = P[N \leq x + v] = F_N(x + v)$$

y

$$F_Y(x|B_1) = P[N \leq x - v] = F_N(x - v).$$

La cdf es:

$$F_Y(x) = F_N(x + v)(1-p) + F_N(x - v)p.$$

La pdf de Y es, entonces:

$$\begin{aligned} f_Y(x) &= \frac{d}{dx}F_Y(x) \\ &= \frac{d}{dx}F_N(x + v)(1-p) + \frac{d}{dx}F_N(x - v)p \\ &= f_N(x + v)(1-p) + f_N(x - v)p. \end{aligned}$$

La pdf de la variable aleatoria Normal:

$$f_N(x) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-x^2/2\sigma^2} \quad -\infty < x < \infty.$$

Las pdf condicionadas son:

$$f_Y(x|B_0) = f_N(x + v) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-(x+v)^2/2\sigma^2}$$

y

$$f_Y(x|B_1) = f_N(x - v) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-(x-v)^2/2\sigma^2}.$$

La pdf de la señal recibida Y es, entonces:

$$f_Y(x) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-(x+v)^2/2\sigma^2}(1-p) + \frac{1}{\sqrt{2\pi\sigma^2}}e^{-(x-v)^2/2\sigma^2}p.$$

La figura 4.5 muestra ambas pdf condicionadas. Podemos ver que la señal transmitida X cambia el centro de masa de la pdf Normal.

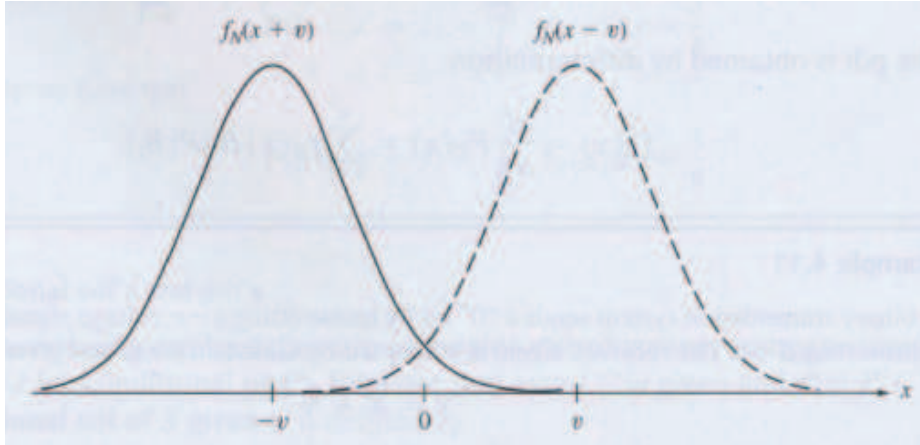


Figura 4.5 La pdf condicionada, dada la señal de entrada.

4.3 La esperanza de X

Vimos la esperanza de una variable aleatoria discreta en la sección 3.3 y mostramos cómo la media muestral de observaciones independientes de una variable aleatoria aproxima $E[X]$. Supongamos que realizamos una serie de experimentos para variables aleatorias continuas. Como las variables aleatorias continuas tienen $P[X = x] = 0$ para cualquier valor específico de x , dividimos la recta real en intervalos pequeños y contamos el número de veces $N_k(n)$ que las observaciones caen en el intervalo $\{x_k < X < x_k + \Delta\}$. Cuando n crece, la frecuencia relativa $f_k(n) = N_k(n)/n$ aproxima $f_X(x_k)\Delta$, la probabilidad del intervalo. Calculamos la media muestral en términos de las frecuencias relativas y hacemos $n \rightarrow \infty$:

$$\langle X \rangle_n = \sum_k x_k f_k(n) \rightarrow \sum_k f_X(x_k) \Delta.$$

La expresión de la derecha aproxima la integral según decrece Δ .

La **esperanza** o **media** de una variable aleatoria X se define como

$$E[X] = \int_{-\infty}^{\infty} t f_X(t) dt. \quad (4.25)$$

La esperanza $E[X]$ está definida si la integral de arriba converge absolutamente, esto es,

$$E[|X|] = \int_{-\infty}^{+\infty} |t| f_X(t) dt < \infty.$$

Si vemos $f_X(x)$ como la distribución de masa en la recta real, entonces $E[X]$ representa el centro de masa de esta distribución.

Ya estudiamos la $E[X]$ en detalle para variables aleatorias discretas pero es conveniente destacar que la definición de la ecuación (4.25) es aplicable en este

caso si expresamos la pdf de la variable aleatoria discreta utilizando funciones delta:

$$\begin{aligned} E[X] &= \int_{-\infty}^{\infty} t \sum_k p_X(x_k) \delta(t - x_k) dt \\ &= \sum_k p_X(x_k) \int_{-\infty}^{+\infty} t \sum_k \delta(t - x_k) dt \\ &= \sum_k p_X(x_k) x_k. \end{aligned}$$

Ejemplo 4.12 Media de la variable aleatoria Uniforme

La media de la variable aleatoria Uniforme viene dada por

$$E[X] = (b - a)^{-1} \int_a^b t dt = \frac{a + b}{2},$$

que es exactamente el punto medio del intervalo $[a, b]$. Los resultados mostrados en la figura 3.6 se obtuvieron repitiendo experimentos cuyos resultados eran variables aleatorias Y y X con cdf en los intervalos $[-1, 1]$ y $[3, 7]$, respectivamente. Las esperanzas correspondientes, 0 y 5, corresponden a los valores alrededor de los cuales X e Y tienden a variar.

El resultado del ejemplo 4.12 se puede calcular inmediatamente sabiendo que $E[X] = m$ cuando la pdf es simétrica alrededor del punto m . Esto es, si

$$f_X(m - x) = f_X(m + x) \text{ para todo } x,$$

entonces, suponiendo que la media existe,

$$0 = \int_{-\infty}^{+\infty} (m - t) f_X(t) dt = m - \int_{-\infty}^{+\infty} t f_X(t) dt.$$

La primera igualdad de arriba se deduce de la simetría de $f_X(t)$ alrededor de $t = m$ y la simetría de $(m - t)$ alrededor del mismo punto. Entonces tenemos que $E[X] = m$.

Ejemplo 4.13 Media de una variable aleatoria Normal

La pdf de una variable aleatoria Normal es simétrica alrededor del punto $x = m$. Entonces $E[X] = m$.

Las expresiones siguientes son útiles cuando X es una variable aleatoria no

negativa:

$$E[X] = \int_0^{\infty} (1 - F_X(t))dt \text{ si } X \text{ es continua y no negativa} \quad (4.26)$$

y

$$E[X] = \sum_{k=0}^{\infty} P[X > k] \text{ si } X \text{ es no negativa y discreta.} \quad (4.27)$$

Ejemplo 4.14 Media de una variable aleatoria Exponencial

El tiempo X entre las llegadas de clientes a una estación de servicio tiene una distribución exponencial. Calcula la media del tiempo entre llegadas.

Solución

Sustituyendo la ecuación (4.17) en la ecuación (4.25) obtenemos

$$E[X] = \int_0^{\infty} t\lambda e^{-\lambda t} dt.$$

Resolvemos la integral usando integración por partes ($\int u dv = uv - \int v du$), con $u = t$ y $dv = \lambda e^{-\lambda t} dt$:

$$\begin{aligned} E[X] &= -te^{-\lambda t} \Big|_0^{\infty} + \int_0^{\infty} e^{-\lambda t} dt \\ &= \lim_{t \rightarrow \infty} te^{-\lambda t} - 0 + \left\{ \frac{-e^{-\lambda t}}{\lambda} \right\}_0^{\infty} \\ &= \lim_{t \rightarrow \infty} \frac{-e^{-\lambda t}}{\lambda} + \frac{1}{\lambda} = \frac{1}{\lambda}, \end{aligned}$$

donde usamos el hecho de que $e^{-\lambda t}$ y $te^{-\lambda t}$ tiende a cero cuando t se acerca a infinito.

Para este ejemplo, la ecuación (4.26) es más fácil de evaluar:

$$E[X] = \int_0^{\infty} e^{-\lambda t} dt = \frac{1}{\lambda}.$$

Recordemos que λ es la tasa de llegadas de clientes medida en clientes por segundo. El resultado es que el tiempo medio entre llegadas $E[X] = 1/\lambda$ segundos por cliente, que tiene sentido intuitivamente.

4.3.1 La esperanza de $Y = g(X)$

Supongamos que nos interesa calcular la esperanza de $Y = g(X)$. Como en el caso de las variables aleatorias discretas (ecuación (3.16)), $E[Y]$ se puede calcular directamente en términos de la pdf de X :

$$E[Y] = \int_{-\infty}^{\infty} g(x)f_X(x)dx. \quad (4.28)$$

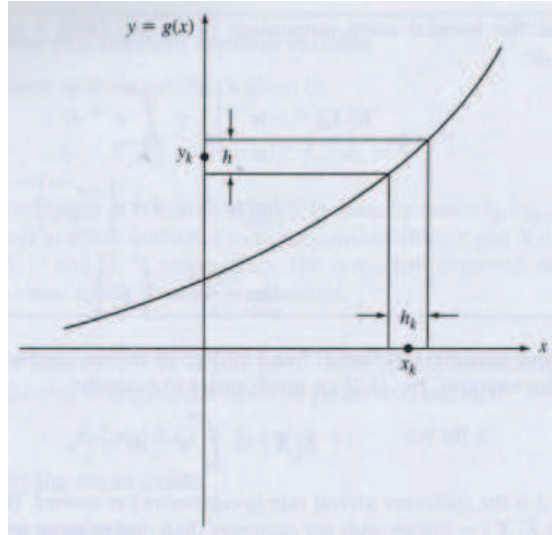


Figura 4.6 Dos sucesos infinitesimales equivalentes.

Para ver cómo se llega a la ecuación (4.28), supongamos que dividimos el eje y en intervalos de longitud h , numeramos los intervalos con el índice k y hacemos que y_k sea el valor central del intervalo k -ésimo. La esperanza de Y es aproximadamente la siguiente suma:

$$E[Y] \simeq \sum_k y_k f_Y(y_k) h.$$

Supongamos que $g(x)$ es estrictamente creciente, entonces al intervalo k -ésimo en el eje y le corresponde un único suceso equivalente de ancho h_k en el eje x como se muestra en la figura 4.6. Sea x_k el valor del k -ésimo intervalo tal que $g(x_k) = y_k$, entonces, como $f_Y(y_k)h = f_X(x_k)h_k$,

$$E[Y] \simeq \sum_k g(x_k) f_X(x_k) h_k.$$

Haciendo h decrecer a cero, obtenemos la ecuación (4.28). Esta ecuación es válida incluso si $g(x)$ no es estrictamente creciente.

Ejemplo 4.15 Esperanza de un senoide con fase aleatoria

Sea $Y = a \cos(\omega t + \Theta)$ donde a , ω y t son constantes y Θ es una variable aleatoria Uniforme en el intervalo $(0, 2\pi)$. La variable aleatoria Y resulta de obtener una muestra de la amplitud de un senoide con fase aleatoria Θ . Calcula la esperanza de Y y de la potencia de Y , Y^2 .

Solución

$$\begin{aligned} E[Y] &= E[a \cos(\omega t + \Theta)] \\ &= \int_0^{2\pi} a \cos(\omega t + \theta) \frac{d\theta}{2\pi} = -a \sin(\omega t + \theta) \Big|_0^{2\pi} \\ &= -a \sin(\omega t + 2\pi) + a \sin(\omega t) = 0. \end{aligned}$$

La potencia media es

$$\begin{aligned} E[Y^2] &= E[a^2 \cos^2(\omega t + \Theta)] = E \left[\frac{a^2}{2} + \frac{a^2}{2} \cos(2\omega t + 2\Theta) \right] \\ &= \frac{a^2}{2} + \frac{a^2}{2} \int_0^{2\pi} \cos(2\omega t + 2\theta) \frac{d\theta}{2\pi} = \frac{a^2}{2}. \end{aligned}$$

donde se ha usado la identidad trigonométrica $\cos^2\alpha = \frac{1+\cos 2\alpha}{2} = \frac{1}{2} + \frac{\cos 2\alpha}{2}$. Nótese que estas respuestas son acordes con el tiempo medio de los sinusoides: el tiempo medio (el valor “dc”) del senoide es cero; el tiempo medio de la potencia es $a^2/2$.

Ejemplo 4.16 Esperanza de la función indicatriz

Sea $g(X) = I_C(X)$ la función indicatriz del suceso $\{X \in C\}$, donde C es un intervalo o la unión de intervalos en la recta real:

$$g(X) = \begin{cases} 0 & X \notin C \\ 1 & X \in C, \end{cases}$$

entonces

$$E[Y] = \int_{-\infty}^{+\infty} g(X) f_X(x) dx = \int_C f_X(x) dx = P[X \in C].$$

Entonces la esperanza de la indicatriz de un suceso es igual a la probabilidad del suceso.

Es fácil demostrar que las ecuaciones (3.17a)-(3.17e) se verifican para las variables aleatorias continuas utilizando la ecuación (4.28). Por ejemplo, sea c una constante, entonces

$$E[c] = \int_{-\infty}^{\infty} c f_X(x) dx = c \int_{-\infty}^{\infty} f_X(x) dx = c \tag{4.29}$$

y

$$E[cX] = \int_{-\infty}^{\infty} cx f_X(x) dx = c \int_{-\infty}^{\infty} x f_X(x) dx = cE[X]. \tag{4.30}$$

La esperanza de la suma de funciones de una variable aleatoria es igual a la suma de la esperanza de las funciones individuales:

$$E[Y] = E \left[\sum_{k=1}^n g_k(X) \right] \quad (4.31)$$

$$= \int_{-\infty}^{\infty} \sum_{k=1}^n g_k(x) f_X(x) dx = \sum_{k=1}^n \int_{-\infty}^{\infty} g_k(x) f_X(x) dx \quad (4.32)$$

$$= \sum_{k=1}^n E[g_k(X)].$$

Ejemplo 4.17 Sea $Y = g(X) = a_0 + a_1X + a_2X^2 + \cdots + a_nX^n$, donde a_k son constantes, entonces

$$E[Y] = E[a_0] + E[a_1X] + \cdots + E[a_nX^n]$$

$$= a_0 + a_1E[X] + a_2E[X^2] + \cdots + a_nE[X^n],$$

donde hemos utilizado la ecuación (4.31) y las ecuaciones (4.29) y (4.30). Un caso especial de este resultado es que

$$E[X + c] = E[X] + c,$$

esto es, podemos desplazar la media de una variable aleatoria añadiéndole una constante.

4.3.2 Varianza de X

La **varianza de la variable aleatoria** X se define como

$$V[X] = E[(X - E[X])^2] = E[X^2] - E[X]^2. \quad (4.33)$$

La **desviación típica de la variable aleatoria** X se define como

$$SD[X] = V[X]^{1/2}. \quad (4.34)$$

Ejemplo 4.18 Varianza de la variable aleatoria Uniforme

Calcula la varianza de la variable aleatoria X que es uniforme en el intervalo $[a, b]$.

Solución

Como la media de X es $(a + b)/2$,

$$V[X] = \frac{1}{b-a} \int_a^b \left(x - \frac{a+b}{2} \right)^2 dx.$$

Sea $y = (x - (a + b)/2)$,

$$V[X] = \frac{1}{b - a} \int_{-(b-a)/2}^{(b-a)/2} y^2 dy = \frac{(b - a)^2}{12}.$$

Las variables aleatorias de la figura 3.6 eran uniformes en los intervalos $[-1, 1]$ y $[3, 7]$, respectivamente. Sus varianzas son $1/3$ y $4/3$. Las desviaciones típicas correspondientes son 0.577 y 1.155 .

Ejemplo 4.19 Varianza de la variable aleatoria Normal

Calcula la varianza de la variable aleatoria Normal.

Solución

Primero multiplicamos la integral de la pdf de X por $\sqrt{2\pi}\sigma$ para obtener

$$\int_{-\infty}^{\infty} e^{-(x-m)^2/2\sigma^2} dx = \sqrt{2\pi}\sigma.$$

Derivamos ambos lados de la ecuación con respecto a σ :

$$\int_{-\infty}^{\infty} \left(\frac{(x - m)^2}{\sigma^3} \right) e^{-(x-m)^2/2\sigma^2} dx = \sqrt{2\pi}.$$

Reorganizando la ecuación obtenemos

$$V[X] = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} (x - m)^2 e^{-(x-m)^2/2\sigma^2} dx = \sigma^2.$$

Este resultado se puede obtener integrando directamente. La figura 4.7 muestra la pdf normal para diferentes valores de σ ; es evidente que el “ancho” de la pdf aumenta con σ .

En la sección 3.3 se derivaron las siguientes propiedades:

$$V[c] = 0 \tag{4.35}$$

$$V[X + c] = V[X] \tag{4.36}$$

$$V[cX] = c^2V[X] \tag{4.37}$$

donde c es constante.

La media y la varianza son los dos parámetros más importantes para resumir la pdf de una variable aleatoria. Otros parámetros se utilizan ocasionalmente. Por ejemplo, la simetría se define como $E[(X - E[X])^3]/SD[X]^3$ y mide el grado de asimetría con respecto a la media. Es fácil demostrar que si una pdf es simétrica alrededor de su media, entonces su simetría es cero. Lo notable de estos parámetros de la pdf es que cada uno utiliza la esperanza de una potencia de X . Estas esperanzas son llamadas momentos de X .

El **momento n -ésimo de la variable aleatoria X** se define como

$$E[X^n] = \int_{-\infty}^{\infty} X^n f_X(x) dx. \tag{4.38}$$

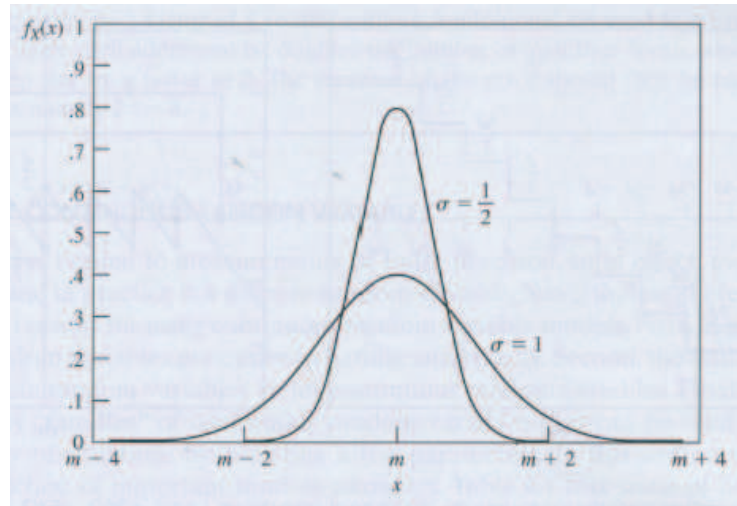


Figura 4.7 Función de densidad de una variable aleatoria Normal.

La media y la varianza se pueden ver como definidas en términos de los dos primeros momentos, $E[X]$ y $E[X^2]$.

4.4 Variables aleatorias continuas importantes

Siempre estamos limitados a mediciones de precisión finita, de modo que, toda variable aleatoria que encontremos en la práctica es una variable aleatoria discreta. No obstante, hay varias razones de peso para usar variables aleatorias continuas. En primer lugar, en general, las variables aleatorias continuas son más fáciles de manejar analíticamente. En segundo lugar, la forma límite de muchas variables aleatorias discretas es una variable aleatoria continua. Finalmente, existe “familias” de variables aleatorias continuas que pueden usarse para modelar una gran variedad de situaciones ajustando unos pocos parámetros. En esta sección continuamos la introducción de variables aleatorias importantes. La tabla 4.1 enumera algunas de las variables aleatorias continuas más importantes.

4.4.1 Variable aleatoria Uniforme

La variable aleatoria Uniforme surge en situaciones en las que todos los valores de un intervalo de la recta real son igualmente probables. La variable aleatoria Uniforme U en el intervalo $[a, b]$ tiene pdf:

$$f_U(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & x < a \text{ y } x > b \end{cases} \quad (4.39)$$

Cuadro 4.1 Variables aleatorias continuas

Variable aleatoria Uniforme

$$S_X = [a, b]$$

$$f_X(x) = \frac{1}{b-a} \quad a \leq x \leq b$$

$$E[X] = \frac{a+b}{2} \quad V[X] = \frac{(b-a)^2}{12} \quad \Phi_X(w) = \frac{e^{jwb} - e^{jwa}}{jw(b-a)}$$

Variable aleatoria Exponencial

$$S_X = [0, \infty)$$

$$f_X(x) = \lambda e^{-\lambda x} \quad x \geq 0 \text{ y } \lambda > 0$$

$$E[X] = \frac{1}{\lambda} \quad V[X] = \frac{1}{\lambda^2} \quad \Phi_X(w) = \frac{\lambda}{\lambda - jw}$$

La variable aleatoria Exponencial es la única variable aleatoria continua sin memoria.

Variable aleatoria Normal o Gaussiana

$$S_X = (-\infty, +\infty)$$

$$f_X(x) = \frac{e^{-(x-m)^2/2\sigma^2}}{\sqrt{2\pi}\sigma} \quad -\infty < x < +\infty \text{ y } \sigma > 0$$

$$E[X] = m \quad V[X] = \sigma^2 \quad \Phi_X(w) = e^{jmw - \sigma^2 w^2/2}$$

Bajo algunas condiciones, X puede ser usada para aproximación de la suma de un gran número de variables aleatorias independientes.

Variable aleatoria Gamma

$$S_X = (0, +\infty)$$

$$f_X(x) = \frac{\lambda(\lambda x)^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)} \quad x > 0 \text{ y } \alpha > 0, \lambda > 0$$

donde $\Gamma(z)$ es la función Gamma (ecuación 4.56).

$$E[X] = \alpha/\lambda \quad V[X] = \alpha/\lambda^2 \quad \Phi_X(w) = \frac{1}{(1-jw/\lambda)^\alpha}$$

Casos especiales de la variable aleatoria Gamma: Erlang con $m = 1$ y Chi cuadrado.

Variable aleatoria Beta

$$S_X = (0, 1)$$

$$f_X(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1} \quad 0 < x < 1 \text{ y } \alpha > 0, \beta > 0$$

$$E[X] = \frac{\alpha}{\alpha+\beta} \quad V[X] = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

Variable aleatoria de Cauchy

$$S_X = (-\infty, +\infty)$$

$$f_X(x) = \frac{\alpha/\pi}{x^2 + \alpha^2} \quad -\infty < x < +\infty \text{ y } \alpha > 0$$

La media y la varianza no existen. $\Phi_X(w) = e^{-\alpha|w|}$

Variable aleatoria de Pareto

$$S_X = [x_m, \infty) \quad x_m > 0$$

$$f_X(x) = \alpha \frac{x_m^\alpha}{x^{\alpha+1}} \quad x \geq x_m$$

$$E[X] = \frac{\alpha x_m}{\alpha-1} \text{ para } \alpha > 1 \quad V[X] = \frac{\alpha x_m^2}{(\alpha-2)(\alpha-1)^2} \text{ para } \alpha > 2$$

y cdf

$$F_U(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \leq x \leq b \\ 1 & x > b. \end{cases} \quad (4.40)$$

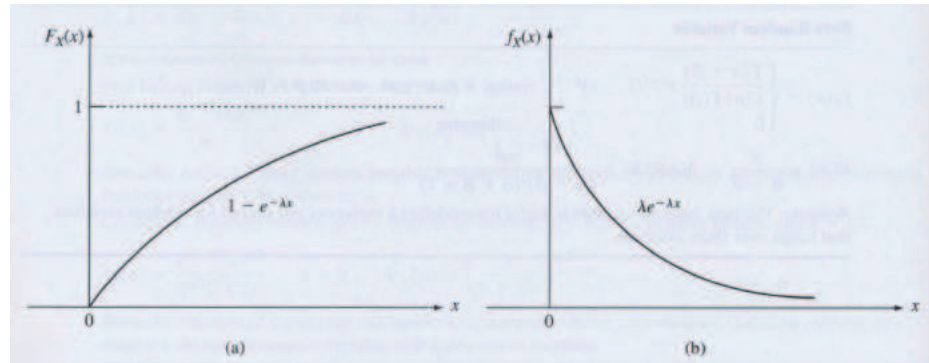


Figura 4.8 Un ejemplo de variable aleatoria continua – la variable aleatoria Exponencial. La parte (a) es la cdf y la parte (b) es la pdf.

Ver la figura 4.2. La media y la varianza de U vienen dadas por:

$$E[U] = \frac{a+b}{2} \text{ y } V[U] = \frac{(b-a)^2}{12}. \quad (4.41)$$

La variable aleatoria Uniforme aparece en muchas situaciones que incluyen variables aleatorias continuas igualmente probables. Obviamente U solo puede ser definida sobre intervalos finitos.

4.4.2 Variable aleatoria Exponencial

La variable aleatoria Exponencial surge al modelar el tiempo entre dos eventos (por ejemplo, el tiempo entre dos peticiones de conexión de dos clientes) y al modelar el tiempo de vida de sistemas y dispositivos. La **variable aleatoria Exponencial** X de parámetro λ tiene pdf

$$f_X(x) = \begin{cases} 0 & x < 0 \\ \lambda e^{-\lambda x} & x \geq 0 \end{cases} \quad (4.42)$$

y cdf

$$F_X(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-\lambda x} & x \geq 0. \end{cases} \quad (4.43)$$

La cdf y la pdf de X se muestran en la figura 4.8.

El parámetro λ es la tasa de incidencia del evento y, por ello, en la ecuación (4.43) la probabilidad de que ocurra un suceso en el tiempo x aumenta si lo hace la tasa λ . Recordemos que, por el ejemplo 3.30, los tiempos entre ocurrencias de sucesos en un proceso de Poisson (figura 3.10) es una variable aleatoria Exponencial.

La media y la varianza de X vienen dadas por:

$$E[U] = \frac{1}{\lambda} \text{ y } V[X] = \frac{1}{\lambda^2}. \quad (4.44)$$

En situaciones de llegadas entre sucesos, λ se mide en sucesos/segundo y $1/\lambda$ se mide en unidades de segundos entre ocurrencias de dos sucesos.

La distribución exponencial **no tiene memoria**:

$$P[X > t + h | X > t] = P[X > h]. \quad (4.45)$$

La expresión de la izquierda es la probabilidad de tener que esperar por lo menos h segundos más dado que ya hemos esperado t segundos. La expresión de la derecha es la probabilidad de esperar al menos h segundos desde el comienzo. Entonces, la probabilidad de esperar al menos h segundos más es la misma, ¡sin importar el tiempo que llevemos esperando! Veremos más tarde que la falta de memoria de la variable aleatoria Exponencial es la propiedad fundamental utilizada en las cadenas de Markov, usadas exhaustivamente para evaluar el comportamiento de sistemas de computador y redes de comunicación.

Ahora demostramos dicha propiedad:

$$\begin{aligned} P[X > t + h | X > t] &= \frac{P[\{X > t + h\} \cap \{X > t\}]}{P[X > t]} \text{ para } h > 0 \\ &= \frac{P[X > t + h]}{P[X > t]} = \frac{e^{-\lambda(t+h)}}{e^{-\lambda t}} \\ &= e^{-\lambda h} = P[X > h]. \end{aligned}$$

Se puede demostrar que la variable aleatoria Exponencial es la única variable aleatoria continua sin memoria.

Los ejemplos 2.13, 2.19 y 2.21 están dedicados a la variable aleatoria Exponencial.

4.4.3 Variable aleatoria Normal

Existen muchas situaciones donde tenemos que lidiar con una variable aleatoria X que consiste en la suma de un gran número de variables aleatorias “pequeñas”. La descripción exacta de la pdf de X en términos de las variables aleatorias que la componen puede ser compleja. Sin embargo, bajo condiciones muy generales, cuando el número de componentes crece, la cdf de X se aproxima a aquella de la **variable aleatoria Normal***. Esta variable aleatoria aparece tan frecuentemente en problemas que involucran aleatoriedad que se la conoce como variable aleatoria Normal.

La pdf de la variable aleatoria Normal X viene dada por

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-m)^2/2\sigma^2} \quad -\infty < x < \infty, \quad (4.46)$$

donde m y $\sigma > 0$ son números reales, que ya demostramos en los ejemplos 4.14 y 4.19 que correspondían a la media y la desviación típica de X . La figura 4.7 muestra que la pdf normal es una curva con forma de campana centrada y simétrica alrededor de m y cuyo “ancho” aumenta con σ .

* Este resultado, llamado teorema central del límite, se estudia en el capítulo 7.

La cdf de la variable aleatoria Normal viene dada por

$$P[X \leq x] = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^x e^{-(x'-m)^2/2\sigma^2} dx'. \quad (4.47)$$

El cambio de variable $t = (x' - m)/\sigma$ resulta en

$$F_X(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(x-m)/\sigma} e^{-t^2/2} dt \quad (4.48)$$

$$= \Phi\left(\frac{x-m}{\sigma}\right) \quad (4.49)$$

donde $\Phi(x)$ es la cdf de la variable aleatoria Normal de media $m = 0$ y $\sigma = 1$:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt. \quad (4.50)$$

Entonces, cualquier probabilidad relativa a una variable aleatoria Normal arbitraria se puede expresar en términos de $\Phi(x)$.

En ingeniería eléctrica se acostumbra a trabajar con la función Q , que se define por

$$Q(x) = 1 - \Phi(x) \quad (4.51)$$

$$= \frac{1}{\sqrt{2\pi}} \int_x^{\infty} e^{-t^2/2} dt. \quad (4.52)$$

$Q(x)$ es simplemente la probabilidad de la “cola” de la pdf. La simetría de la pdf implica que

$$Q(0) = 1/2 \text{ y } Q(-x) = 1 - Q(x). \quad (4.53)$$

La integral de la ecuación (4.50) no tiene una expresión en forma cerrada. Tradicionalmente, las integrales han sido evaluadas mirando las tablas de los valores de $Q(x)$ o mediante el uso de aproximaciones que requieren una evaluación numérica. La siguiente expresión proporciona con buena precisión una aproximación a $Q(x)$ en todo el rango $0 < x < \infty$:

$$Q(x) \simeq \left[\frac{1}{(1-a)x + a\sqrt{x^2 + b}} \right] \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad (4.54)$$

donde $a = 1/\pi$ y $b = 2\pi$. En algunos problemas, estamos interesados en encontrar el valor de x para que $Q(x) = 10^{-k}$. La tabla 4.2 da dichos valores para $k = 1, \dots, 10$.

La variable aleatoria Normal juega un papel muy importante en los sistemas de comunicación, donde las señales de transmisión están dañadas por las tensiones de ruido resultantes del movimiento térmico de los electrones. Se puede demostrar a partir de principios físicos que estas tensiones tienen una pdf normal.

Ejemplo 4.20 Un sistema de comunicación acepta un voltaje positivo V como entrada y devuelve un voltaje $Y = \alpha V + N$, donde $\alpha = 10^{-2}$ y N es una variable

k	$x = Q^{-1}(10^{-k})$
1	1.2815
2	2.3263
3	3.0902
4	3.7190
5	4.2649
6	4.7535
7	5.1993
8	5.6120
9	5.9978
10	6.3613

Cuadro 4.2 $Q(x) = 10^{-k}$

aleatoria Normal con parámetros $m = 0$ y $\sigma = 2$. Calcula el valor de V que da $P[Y < 0] = 10^{-6}$.

Solución

La probabilidad $P[Y < 0]$ se escribe en términos de N como sigue:

$$\begin{aligned} P[Y < 0] &= P[\alpha V + N < 0] \\ &= P[N < -\alpha V] = \Phi\left(\frac{-\alpha V}{\sigma}\right) = Q\left(\frac{\alpha V}{\sigma}\right) = 10^{-6}. \end{aligned}$$

Por la tabla 4.2 vemos que el argumento de la función Q debe ser $\alpha V/\sigma = 4.753$. Entonces $V = (4.753)\sigma/\alpha = 950.6$.

4.4.4 Variable aleatoria Gamma

La variable aleatoria Gamma es una variable aleatoria versátil que aparece en muchas aplicaciones. Por ejemplo, se utiliza para modelar el tiempo necesario para servir a los clientes en la cola de espera de un sistema, la vida útil de los dispositivos y sistemas en los estudios de fiabilidad, y el comportamiento de agrupación por defecto en los chips VLSI.

La pdf de la **variable aleatoria Gamma** tiene dos parámetros, $\alpha > 0$ y $\lambda > 0$, y viene dada por

$$f_X(x) = \frac{\lambda(\lambda x)^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)} \quad 0 < x < \infty, \quad (4.55)$$

donde $\Gamma(z)$ es la función gamma, que se define por la integral

$$\Gamma(z) = \int_0^{\infty} x^{z-1} e^{-x} dx \quad z > 0. \quad (4.56)$$

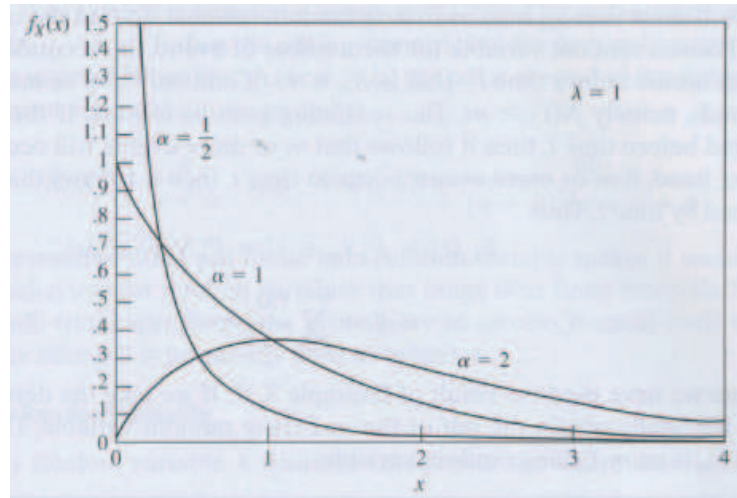


Figura 4.9 Función de densidad de la variable aleatoria Gamma.

La función gamma tiene las siguientes propiedades:

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi},$$

$$\Gamma(z + 1) = z\Gamma(z) \text{ para } z > 0, \text{ y}$$

$$\Gamma(m + 1) = m! \text{ para un entero no negativo } m.$$

La versatilidad de la variable aleatoria Gamma se debe a la riqueza de la función gamma $\Gamma(z)$. La pdf de la variable aleatoria Gamma puede asumir gran variedad de formas, como se muestra en la figura 4.9. Mediante la variación de los parámetros α y λ , es posible para adaptar la pdf de la gamma a muchos tipos de datos experimentales. Además, muchas variables aleatorias son casos especiales de la variable aleatoria Gamma. La variable aleatoria Exponencial se obtiene con $\alpha = 1$. Si $\lambda = 1/2$ y $\alpha = k/2$, donde k es un entero positivo, se obtiene la **variable aleatoria chi-cuadrado**, que aparece en ciertos problemas estadísticos. La **variable aleatoria Erlang** de parámetro m se obtiene cuando $\alpha = m$, un entero positivo. La variable aleatoria Erlang se utiliza en los modelos de fiabilidad de sistemas y en los modelos de sistemas de cola de espera. Ambas variables son discutidos en ejemplos posteriores.

Ejemplo 4.21 Demuestra que la integral de la pdf de una variable aleatoria Gamma es uno.

Solución

La integral de la pdf es

$$\begin{aligned}\int_0^{\infty} f_X(x)dx &= \int_0^{\infty} \frac{\lambda(\lambda x)^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)} dx \\ &= \frac{\lambda^\alpha}{\Gamma(\alpha)\lambda^\alpha} \int_0^{\infty} y^{\alpha-1} e^{-y} dy = 1,\end{aligned}$$

donde utilizamos el hecho de que la integral es igual a $\Gamma(\alpha)$.

En general, la función de distribución de la variable aleatoria Gamma no tiene una expresión en forma cerrada. Vamos a demostrar que el caso especial de la variable aleatoria Erlang de parámetro m sí tiene una expresión en forma cerrada de la función de distribución mediante el uso de su estrecha relación con las variables aleatorias exponencial y de Poisson. La cdf también se puede obtener mediante la integración de la pdf.

Consideremos una vez más el procedimiento de tomar límites utilizado para obtener la distribución de la variable aleatoria de Poisson. Supongamos que observamos el tiempo que transcurre hasta la ocurrencia del suceso m -ésimo, S_m . Los tiempos entre sucesos X_1, X_2, \dots, X_m son variables aleatorias exponenciales, por lo que se tiene que

$$S_m = X_1 + X_2 + \dots + X_m.$$

Vamos a demostrar que S_m es una variable aleatoria Erlang. Para hallar la función de distribución de S_m , sea $N(t)$ la variable aleatoria de Poisson para el número de sucesos en t segundos. Observemos que el suceso m -ésimo se produce antes del tiempo t , es decir, $S_m \leq t$ —si y sólo si m o más sucesos han sucedido en t segundos, es decir, $N(t) \geq m$. El razonamiento es el siguiente. Si el m -ésimo suceso se ha producido antes del tiempo t , entonces se sigue que m o más sucesos han ocurrido en el tiempo t . Por otro lado, si m o más sucesos han ocurrido en el tiempo t , entonces se deduce que el m -ésimo evento ocurrió en el momento t . Así

$$F_{S_m}(t) = P[S_m \leq t] = P[N(t) \geq m] \quad (4.57)$$

$$= 1 - \sum_{k=0}^{m-1} \frac{(\lambda t)^k}{k!} e^{-\lambda t}, \quad (4.58)$$

donde hemos usado el resultado del ejemplo 3.30. Si tomamos la derivada de la anterior cdf, se obtiene finalmente la pdf de la variable aleatoria de Erlang con parámetro m . Así, hemos demostrado que S_m es una variable aleatoria de Erlang.

Ejemplo 4.22 Una fábrica cuenta con dos piezas de recambio de un componente crítico del sistema que tiene una vida media de $1/\lambda = 1$ mes. Calcula la

probabilidad de que los tres componentes (el operativo y las dos piezas de recambio) duren más de 6 meses. Supongamos que la vida útil de los componentes son variables aleatorias exponenciales.

Solución

La vida útil restante de los componentes del servicio es una variable aleatoria Exponencial con una tasa λ por la propiedad de no tener memoria. Así, la vida total de los tres componentes X es la suma de tres variables aleatorias exponenciales con parámetro $\lambda = 1$. Por lo tanto X tiene una distribución de Erlang con $m = 3$ y $\lambda = 1$. De la ecuación (4.58) la probabilidad de que X sea mayor que 6 es

$$\begin{aligned} P[X > 6] &= 1 - P[X \leq 6] \\ &= \sum_{k=0}^2 \frac{6^k}{k!} e^{-6} = 0.06197. \end{aligned}$$

4.4.5 Variable aleatoria Beta

La variable aleatoria Beta X toma valores en un intervalo cerrado y tiene pdf:

$$f_X(x) = cx^{a-1}(1-x)^{b-1} \text{ para } 0 < x < 1 \quad (4.59)$$

donde la constante de normalización es la inversa de la función beta

$$\frac{1}{c} = B(a, b) = \int_0^1 x^{a-1}(1-x)^{b-1} dx$$

y donde la función beta está relacionada con la función gamma por la siguiente expresión:

$$B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}.$$

Cuando $a = b = 1$, tenemos la variable aleatoria Uniforme. Otras elecciones de a y b llevan a pdf en intervalos finitos que pueden diferir notablemente de la uniforme. Si $a = b > 1$, entonces la pdf es simétrica respecto de $x = 1/2$ y concentrada en $x = 1/2$ también. Cuando $a = b < 1$, entonces la pdf es simétrica, pero la densidad se concentra en los bordes del intervalo. Cuando $a < b$ (o $a > b$) la pdf es asimétrica hacia la derecha (o la izquierda).

La media y la varianza vienen dadas por:

$$E[X] = \frac{a}{a+b} \text{ y } V[X] = \frac{ab}{(a+b)^2(a+b+1)}. \quad (4.60)$$

La versatilidad de la pdf de la variable aleatoria Beta hace que sea útil para modelar gran variedad de comportamientos de variables aleatorias que toman valores en intervalos finitos. Por ejemplo, en un experimento de Bernoulli, la probabilidad de éxito p podría ser una variable aleatoria. La pdf beta se utiliza con frecuencia para modelar p .

4.4.6 Variable aleatoria de Cauchy

La variable aleatoria de Cauchy X toma valores en toda la recta real y tiene pdf:

$$f_X(x) = \frac{1/\pi}{1+x^2}. \quad (4.61)$$

Es fácil verificar que la integral de esta pdf es 1. Sin embargo, X no tiene ningún momento ya que las integrales asociadas no convergen. La variable aleatoria de Cauchy surge como la tangente de una variable aleatoria Uniforme en el intervalo unidad.

4.4.7 Variable aleatoria de Pareto

La variable aleatoria de Pareto se plantea en el estudio de la distribución de la riqueza en el que se ha observado una tendencia a que una pequeña porción de la población posea una gran parte de la riqueza. Recientemente, se ha descubierto que la distribución de Pareto captura el comportamiento de muchas cantidades de interés en el estudio del comportamiento de Internet, por ejemplo, el tamaño de los archivos, los retrasos de paquetes, las preferencias del título de audio y video, etc. La variable aleatoria de Pareto puede ser vista como la versión continua de la variable aleatoria discreta de Zipf.

La variable aleatoria de Pareto X toma valores en el rango $x > x_m$, donde x_m es un número real positivo. X tiene una cdf complementaria con parámetro de forma $\alpha > 0$ dada por:

$$P[X > x] = \begin{cases} 1 & x < x_m \\ \frac{x_m^\alpha}{x^\alpha} & x \geq x_m. \end{cases} \quad (4.62)$$

La cola de X decae algebraicamente con x que es bastante lentamente en comparación con las variables aleatorias exponencial y normal. La variable aleatoria de Pareto es el ejemplo más destacado de variables aleatorias con “largas colas”.

La cdf y la pdf de la X son:

$$F_X(x) = \begin{cases} 0 & x < x_m \\ 1 - \frac{x_m^\alpha}{x^\alpha} & x \geq x_m. \end{cases} \quad (4.63)$$

Debido a su larga cola, la cdf de X se aproxima a 1 poco a poco a medida que x aumenta.

$$f_X(x) = \begin{cases} 0 & x < x_m \\ \alpha \frac{x_m^\alpha}{x^{\alpha+1}} & x \geq x_m \end{cases} \quad (4.64)$$

Ejemplo 4.23 Media y varianza de la variable aleatoria de Pareto

Calcular la media y la varianza de la variable aleatoria de Pareto.

Solución

$$E[X] = \int_{x_m}^{\infty} t \alpha \frac{x_m^\alpha}{t^{\alpha+1}} dt = \int_{x_m}^{\infty} \alpha \frac{x_m^\alpha}{t^\alpha} dt = \frac{\alpha}{\alpha-1} \frac{x_m^\alpha}{x_m^{\alpha-1}} = \frac{\alpha x_m}{\alpha-1} \text{ para } \alpha > 1 \quad (4.65)$$

donde la integral está definida para $\alpha > 1$ y

$$E[X^2] = \int_{x_m}^{\infty} t^2 \alpha \frac{x_m^\alpha}{t^{\alpha+1}} dt = \int_{x_m}^{\infty} \alpha \frac{x_m^\alpha}{t^{\alpha-1}} dt = \frac{\alpha}{\alpha-2} \frac{x_m^\alpha}{x_m^{\alpha-2}} = \frac{\alpha x_m^2}{\alpha-2} \text{ para } \alpha > 2$$

donde el segundo momento está definido para $\alpha > 2$.

La varianza de X es entonces:

$$V[X] = \frac{\alpha x_m^2}{\alpha-2} - \left(\frac{\alpha x_m}{\alpha-1} \right)^2 = \frac{\alpha x_m^2}{(\alpha-2)(\alpha-1)^2} \text{ para } \alpha > 2. \quad (4.66)$$

4.5 Funciones de una variable aleatoria

Sea X una variable aleatoria y sea $g(x)$ una función real definida sobre la recta real. Definamos $Y = g(X)$, es decir, Y se determina mediante la evaluación de la función $g(x)$ en el valor tomado por la variable aleatoria X . Entonces Y es también una variable aleatoria. Las probabilidades con las que Y toma los diferentes valores dependen de la función $g(x)$, así como de la función de distribución de X . En esta sección consideramos el problema de encontrar la cdf y la pdf de Y .

Ejemplo 4.24 Sea la función $h(x) = (x)^+$, definida como sigue:

$$(x)^+ = \begin{cases} 0 & x < 0 \\ x & x \geq 0. \end{cases}$$

Por ejemplo, sea X el número de altavoces activos en un grupo de N altavoces y sea Y el número de altavoces activos por encima de M , entonces $Y = (X - M)^+$. En otro ejemplo, sea X una entrada de tensión de un rectificador halfwave, entonces $Y = (X)^+$ es la salida.

Ejemplo 4.25 Consideremos la función lineal $c(x) = ax + b$, donde a y b son constantes. Esta función se presenta en muchas situaciones. Por ejemplo, $c(x)$ podría ser el costo asociado con la cantidad x , con la constante a representando el costo por unidad de x y b siendo un componente de costos fijos. En un contexto de procesamiento de señales, $c(x) = ax$ podría ser la versión ampliada (si $a > 1$) o la versión atenuada (si $a < 1$) de la tensión de x .

La probabilidad de un suceso C que involucra a Y es igual a la probabilidad del suceso equivalente B de los valores de X tal que $g(X)$ se encuentra en C :

$$P[Y \in C] = P[g(X) \in C] = P[X \in B].$$

Tres tipos de sucesos equivalentes son útiles para determinar la cdf y la pdf de $Y = g(X)$: (1) El suceso $\{g(X) = y_k\}$ se utiliza para determinar la magnitud del salto en un punto y_k donde la cdf de Y se sabe discontinua; (2) el suceso $\{g(X) \leq y\}$ se utiliza para calcular la función de distribución de Y directamente; y (3) el suceso $\{y < g(X) \leq y + h\}$ es útil para determinar la pdf de Y . Vamos a demostrar el uso de estos tres métodos en una serie de ejemplos.

Los dos ejemplos siguientes demuestran cómo la pmf se calcula en los casos en los que $Y = g(X)$ es discreta. En el primer ejemplo, X es discreta. En el segundo ejemplo, X es continua.

Ejemplo 4.26 Sea X el número de altavoces activos en un grupo de N altavoces independientes. Sea p la probabilidad de que un altavoz esté activo. En el ejemplo 2.30 se demostró que X tiene una distribución binomial con parámetros N y p . Supongamos que un sistema de transmisión de voz puede transmitir hasta M señales de voz a la vez y que cuando X supera M , $X - M$ señales seleccionadas al azar se descartan. Sea Y el número de señales descartadas, entonces

$$Y = (X - M)^+.$$

Y toma valores en el conjunto $S_Y = \{0, 1, \dots, N - M\}$. Y será igual a cero cuando X sea menor o igual a M , e Y igual a $k > 0$ cuando X es igual a $M + k$. Por lo tanto

$$P[Y = 0] = P[X \in \{0, 1, \dots, M\}] = \sum_{j=0}^M p_j$$

y

$$P[Y = k] = P[X = M + k] = p_{M+k} \quad 0 < k \leq N - M.$$

donde p_j es la pmf de X .

Ejemplo 4.27 Sea X una muestra de la tensión de una onda del habla y supongamos que X tiene una distribución uniforme en el intervalo $[-4d, 4d]$. Sea $Y = q(X)$, donde el cuantificador de entrada-salida característico es el que se muestra en la figura 4.9. Calcula la pmf de Y .

Solución

El suceso $\{Y = q\}$ para $q \in S_Y$ es equivalente al suceso $\{X \in I_q\}$ donde I_q es un intervalo de puntos mapeado a la representación del punto q . La pmf de Y es por lo tanto calculada mediante la evaluación de

$$P[Y = q] = \int_{I_q} f_X(t) dt.$$

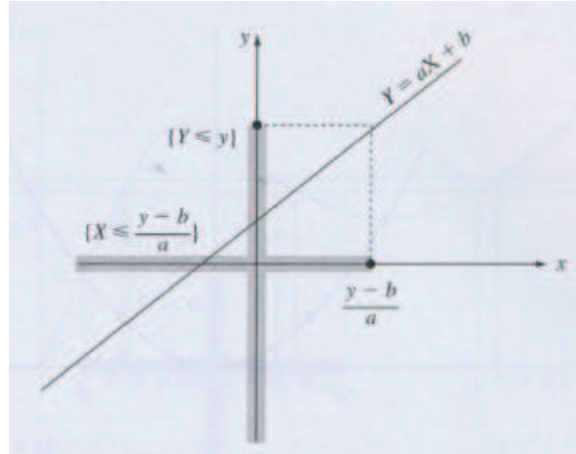


Figura 4.10 El suceso equivalente para $\{Y \leq y\}$ es el suceso $\{X \leq (y - b)/a\}$, si $a > 0$.

Es fácil ver que el punto de representación tiene un intervalo de longitud d mapeados a él. Así, los ocho resultados posibles son igualmente probables, es decir, $P[Y = q] = 1/8$ para $q \in S_Y$.

En el ejemplo 4.27, cada trozo constante de la función $g(X)$ produce una función delta en la pdf de Y . En general, si la función $g(X)$ es constante durante ciertos intervalos y si la pdf de X es distinto de cero en estos intervalos, entonces la pdf de Y contendrá funciones delta. Y entonces será discreta o de tipo mixto.

La cdf de Y se define como la probabilidad del suceso $\{Y \leq y\}$. En principio, siempre se puede obtener mediante el cálculo de la probabilidad del suceso equivalente $\{g(X) \leq y\}$ como se muestra en los siguientes ejemplos.

Ejemplo 4.28 Una función lineal

Sea Y la variable aleatoria definida por

$$Y = aX + b,$$

donde a es una constante diferente de cero. Suponiendo que X tiene función de distribución $F_X(x)$, calcula $F_Y(y)$.

Solución

El suceso $\{Y \leq y\}$ se produce cuando $A = \{aX + b \leq y\}$ ocurre. Si $a > 0$, entonces $A = \{X \leq (y - b)/a\}$ (ver fig. 4.10), y por lo tanto

$$F_Y(y) = P\left[X \leq \frac{y - b}{a}\right] = F_X\left(\frac{y - b}{a}\right) \quad a > 0.$$

Por otro lado, si $a < 0$, entonces $A = \{X \geq (y - b)/a\}$ y

$$F_Y(y) = P\left[X \geq \frac{y - b}{a}\right] = 1 - F_X\left(\frac{y - b}{a}\right) \quad a < 0.$$

Se puede obtener la pdf de Y derivando con respecto a y . Para ello tenemos que utilizar la regla de la cadena:

$$\frac{dF}{dy} = \frac{dF}{du} \frac{du}{dy},$$

donde u es el argumento de F . En este caso, $u = (y - b)/a$, y se obtiene entonces

$$f_Y(y) = \frac{1}{a} f_X\left(\frac{y - b}{a}\right) \quad a > 0$$

y

$$f_Y(y) = \frac{1}{-a} f_X\left(\frac{y - b}{a}\right) \quad a < 0.$$

Estos dos resultados se puede escribir de forma compacta como

$$f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y - b}{a}\right). \quad (4.67)$$

Ejemplo 4.29 Una función lineal de una variable aleatoria Normal

Sea X una variable aleatoria con una pdf normal con media m y desviación típica σ :

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-m)^2/2\sigma^2} \quad -\infty < x < \infty. \quad (4.68)$$

Sea $Y = aX + b$ calcula la pdf de Y .

Solución

Sustituyendo la ecuación (4.68) en la ecuación (4.67) se obtiene

$$f_Y(y) = \frac{1}{\sqrt{2\pi}|a|\sigma} e^{-(y-b-am)^2/2(a\sigma)^2}.$$

Nótese que Y también tiene una distribución normal con media $b + am$ y desviación típica $|a|\sigma$. Por lo tanto, *una función lineal de una variable aleatoria Normal es también una variable aleatoria Normal*.

Ejemplo 4.30 Sea Y la variable aleatoria definida por

$$Y = X^2,$$

donde X es una variable aleatoria continua. Calcula la cdf y la pdf de Y .

Solución

El suceso $\{Y \leq y\}$ ocurre cuando $\{X^2 \leq y\}$ o, equivalentemente, cuando

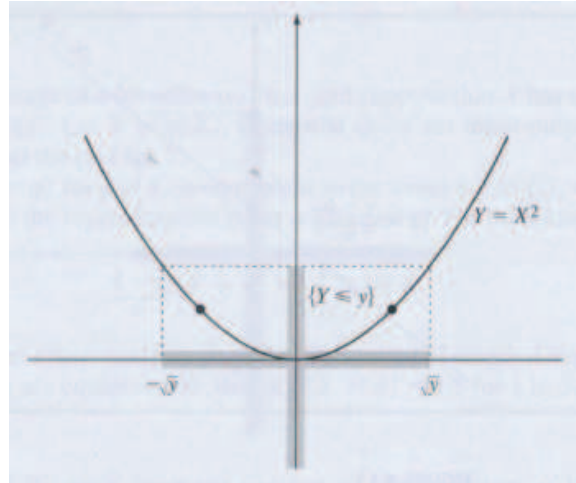


Figura 4.11 El suceso equivalente para $\{Y \leq y\}$ es el suceso $\{-\sqrt{y} \leq X \leq \sqrt{y}\}$, si $y \geq 0$.

$\{-\sqrt{y} \leq X \leq \sqrt{y}\}$ para y no negativo, ver fig. 4.11. El suceso es el conjunto vacío cuando y es negativo. Así

$$F_Y(y) = \begin{cases} 0 & y < 0 \\ F_X(\sqrt{y}) - F_X(-\sqrt{y}) & y > 0 \end{cases}$$

y derivando con respecto a y ,

$$\begin{aligned} f_Y(y) &= \frac{f_X(\sqrt{y})}{2\sqrt{y}} - \frac{f_X(-\sqrt{y})}{-2\sqrt{y}} \quad y > 0 \\ &= \frac{f_X(\sqrt{y})}{2\sqrt{y}} + \frac{f_X(-\sqrt{y})}{2\sqrt{y}}. \end{aligned} \quad (4.69)$$

Ejemplo 4.31 Variable aleatoria chi-cuadrado

Sea X una variable aleatoria Normal con media $m = 0$ y la desviación típica $\sigma = 1$. Se dice que X es una variable aleatoria Normal estándar. Sea $Y = X^2$. Calcula la pdf de la Y .

Solución

Sustituyendo la ecuación (4.68) en la ecuación (4.69) se obtiene

$$f_Y(y) = \frac{e^{-y/2}}{\sqrt{2y\pi}} \quad y \geq 0. \quad (4.70)$$

Vemos que $f_Y(y)$ es la pdf de una *variable aleatoria chi-cuadrado con un grado de libertad*.

El resultado del ejemplo 4.30 sugiere que si la ecuación $y_0 = g(x)$ tiene n soluciones, x_0, x_1, \dots, x_n , entonces $f_Y(y_0)$ será igual a n términos del tipo de los que aparecen en el lado derecho de la ecuación (4.69). Vamos a demostrar que

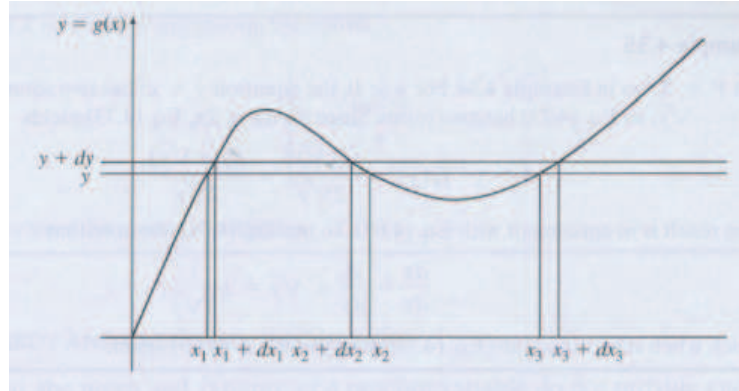


Figura 4.12 El suceso equivalente de $\{y < Y < y + dy\}$ es $\{x_1 < X < x_1 + dx_1\} \cup \{x_2 + dx_2 < X < x_2\} \cup \{x_3 < X < x_3 + dx_3\}$.

esto es cierto en general mediante el uso de un método para obtener directamente la pdf de Y en términos de la pdf de X .

Consideremos una función no lineal $Y = g(X)$ como la que se muestra en la figura 4.12. Consideraremos el suceso $C_y = \{y < Y < y + dy\}$ y sea B_y su suceso equivalente. Para la y representada en la figura, la ecuación $g(x) = y$ tiene tres soluciones x_1 , x_2 y x_3 y el suceso equivalente B_y tiene un segmento que corresponde a cada solución:

$$B_y = \{x_1 < X < x_1 + dx_1\} \cup \{x_2 + dx_2 < X < x_2\} \cup \{x_3 < X < x_3 + dx_3\}.$$

La probabilidad de que el suceso C_y es aproximadamente

$$P[C_y] = f_Y(y)|dy|, \quad (4.71)$$

donde $|dy|$ es la longitud del intervalo $y < Y \leq y + dy$. Del mismo modo, la probabilidad del suceso B_y es aproximadamente

$$P[B_y] = f_X(x_1)|dx_1| + f_X(x_2)|dx_2| + f_X(x_3)|dx_3|. \quad (4.72)$$

Como C_y y B_y son sucesos equivalentes, sus probabilidades deben ser iguales. Al equiparar las ecuaciones (4.71) y (4.72) obtenemos

$$f_Y(y) = \sum_k \frac{f_X(x)}{|dy/dx|} \Big|_{x=x_k} \quad (4.73)$$

$$= \sum_k f_X(x) \left| \frac{dx}{dy} \right| \Big|_{x=x_k}. \quad (4.74)$$

Es claro que si la ecuación $g(x) = y$ tiene n soluciones, la expresión para la pdf de Y en esos puntos está dado por las ecuaciones (4.73) y (4.74) y contiene n términos.

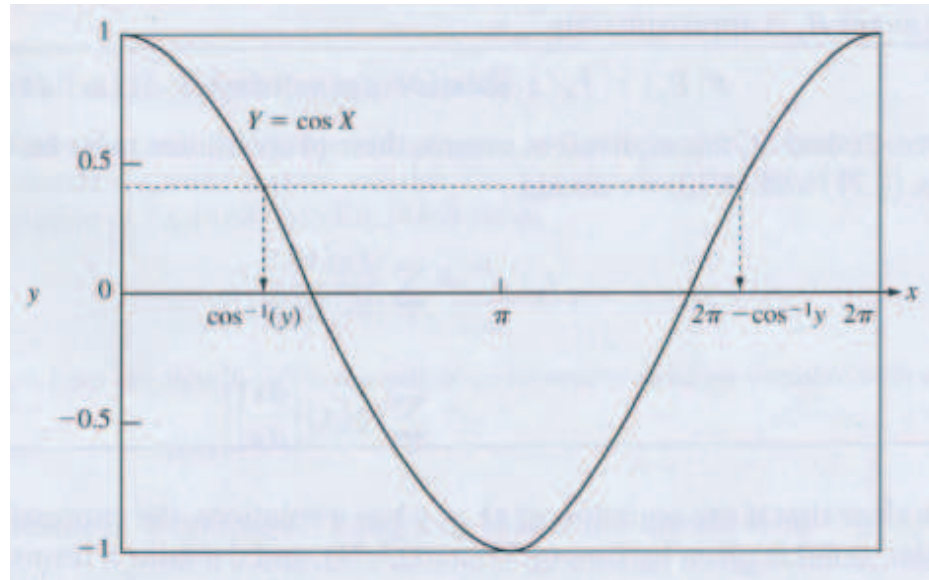


Figura 4.13 $y = \cos x$ tiene dos raíces en el intervalo $(0, 2\pi)$.

Ejemplo 4.32 Sea $Y = X^2$ como en el ejemplo 4.31. Para $y \geq 0$, la ecuación $y = x^2$ tiene dos soluciones, $x_0 = \sqrt{y}$ y $x_1 = -\sqrt{y}$, por tanto la ecuación (4.73) tiene dos términos. Puesto que $dy/dx = 2x$, la ecuación (4.73) lleva a

$$f_Y(y) = \frac{f_X(\sqrt{y})}{2\sqrt{y}} + \frac{f_X(-\sqrt{y})}{2\sqrt{y}}.$$

Este resultado es acorde con la ecuación (4.69). Para utilizar la ecuación (4.74), observamos que

$$\frac{dx}{dy} = \frac{d}{dy} \pm \sqrt{y} = \pm \frac{1}{2\sqrt{y}},$$

que cuando se sustituye en la ecuación (4.74) entonces se obtiene la ecuación (4.69) de nuevo.

Ejemplo 4.33 Muestras de la amplitud de una onda sinusoidal

Sea $Y = \cos(X)$, donde X se distribuye uniformemente en el intervalo $(0, 2\pi]$. Y puede ser visto como la muestra de una forma de onda sinusoidal en un instante de tiempo aleatorio que se distribuye uniformemente a lo largo el período de la senoide. Calcula la pdf de Y .

Solución

Se puede observar en la figura 4.13 que para $-1 < y < 1$ la ecuación $y = \cos(x)$

tiene dos soluciones en el intervalo de interés, $x_0 = \cos^{-1}(y)$ y $x_1 = 2\pi - x_0$. Como

$$\frac{dy}{dx}|_{x_0} = -\sin(x_0) = -\sin(\cos^{-1}(y)) = -\sqrt{1-y^2},$$

y puesto que $f_X(x) = 1/2\pi$ en el intervalo de interés, la ecuación (4.73) lleva a

$$\begin{aligned} f_Y(y) &= \frac{1}{2\pi\sqrt{1-y^2}} + \frac{1}{2\pi\sqrt{1-y^2}} \\ &= \frac{1}{\pi\sqrt{1-y^2}} \text{ para } -1 < y < 1. \end{aligned}$$

La cdf de Y se calcula integrando lo anterior:

$$F_Y(y) = \begin{cases} 0 & y < -1 \\ \frac{1}{2} + \frac{\sin^{-1}y}{\pi} & -1 \leq y \leq 1 \\ 1 & y > 1. \end{cases}$$

Se dice que Y tiene la **distribución arco seno**.

4.6 Las desigualdades de Chebyshev y Markov

En general, la media y la varianza de una variable aleatoria no proporcionan información suficiente para determinar la cdf o la pdf. Sin embargo, la media y la varianza de una variable aleatoria X nos permiten obtener los límites de las probabilidades de la forma $P[|X| \geq t]$. Supongamos primero que X es una variable aleatoria no negativa con media $E[X]$. La **desigualdad de Markov** establece que

$$P[X \geq a] \leq \frac{E[X]}{a} \text{ para } X \text{ no negativa.} \quad (4.75)$$

Obtenemos la ecuación (4.75) de la siguiente manera:

$$\begin{aligned} E[X] &= \int_0^a t f_X(t) dt + \int_a^\infty t f_X(t) dt \geq \int_a^\infty t f_X(t) dt \\ &\geq \int_a^\infty a f_X(t) dt = aP[X \geq a]. \end{aligned}$$

La primera desigualdad resulta de descartar la integral de cero a a ; la segunda desigualdad resulta de la sustitución de t con el número más pequeño a .

Ejemplo 4.34 La altura media de los niños en una clase de guardería es de 3 pies y 6 pulgadas. Calcula la cota de la probabilidad de que un niño de la clase mida más de 9 pies. La desigualdad de Markov da $P[H \geq 9] \leq 42/108 = 0.389$.

La cota en el ejemplo anterior parece ser ridícula. Sin embargo, una cota, por naturaleza, debe tomar el peor de los casos en consideración. Podemos construir una variable aleatoria para la cual la cota dada por la desigualdad de Markov es exacta. La razón por la que sabemos que la cota del ejemplo anterior es ridícula es porque conocemos la variabilidad de la altura de los niños respecto a su media.

Supongamos ahora que la media $E[X] = m$ y que la varianza $V[X] = \sigma^2$ de la variable aleatoria son conocidas y que estamos interesados en acotar $P[|X - m| \geq a]$. La **desigualdad de Chebyshev** establece que

$$P[|X - m| \geq a] \leq \frac{\sigma^2}{a^2}. \quad (4.76)$$

La desigualdad de Chebyshev es una consecuencia de la desigualdad de Markov. Sea $D^2 = (X - m)^2$ la desviación con respecto a la media al cuadrado. Entonces, la desigualdad de Markov aplicada a D^2 da

$$P[D^2 \geq a^2] \leq \frac{E[(X - m)^2]}{a^2} = \frac{\sigma^2}{a^2}.$$

La ecuación (4.76) se obtiene cuando se observa que $\{D^2 \geq a^2\}$ y $\{|X - m| \geq a\}$ son sucesos equivalentes.

Supongamos que una variable aleatoria X tiene varianza cero, entonces la desigualdad de Chebyshev implica que

$$P[X = m] = 1, \quad (4.77)$$

es decir, la variable aleatoria es igual a su media con probabilidad uno. En otras palabras, X es igual a la constante m en casi todos los experimentos.

Ejemplo 4.35 El tiempo medio de respuesta y la desviación típica en un sistema informático multiusuario son 15 segundos y 3 segundos, respectivamente. Estima la probabilidad de que el tiempo de respuesta supere en más de 5 segundos a la media.

Solución

La desigualdad de Chebyshev con $m = 15$ segundos, $\sigma = 3$ segundos y $a = 5$ segundos da

$$P[|X - 15| \geq 5] \leq \frac{9}{25} = 0.36.$$

Ejemplo 4.36 Si X tiene media m y varianza σ^2 , la desigualdad de Chebyshev para $a = k\sigma$ da

$$P[|X - m| \geq k\sigma] \leq \frac{1}{k^2}.$$

Supongamos ahora que sabemos que X es una variable aleatoria Normal, entonces para $k = 2$, $P[|X - m| \geq 2\sigma] = 0.0456$, mientras que la desigualdad de Chebyshev da el límite superior de 0.25.

Ejemplo 4.37 La cota de Chebyshev es ajustada

Sea X la variable aleatoria tal que $P[X = -v] = P[X = v] = 0.5$. La media es cero y la varianza es $V[X] = E[X^2] = (-v)^2 0.5 + v^2 0.5 = v^2$.

Nótese que $P[|X| \geq v] = 1$. La desigualdad de Chebyshev establece que:

$$P[|X| \geq v] \leq 1 - \frac{V[X]}{v^2} = 1.$$

Vemos que la cota y el valor exacto coinciden, por lo que la cota es ajustada.

Vemos en el ejemplo 4.35 que para ciertas variables aleatorias, la desigualdad de Chebyshev puede dar lugar a cotas grandes. Sin embargo, la desigualdad es útil en situaciones en la que no tenemos conocimiento sobre la distribución de una variable aleatoria dada y solo conocemos su media y su varianza. En la sección 7.2, vamos a utilizar la desigualdad de Chebyshev para demostrar que la media aritmética de las mediciones independientes de la misma variable aleatoria es muy probable que se aproxime al valor esperado de la variable aleatoria cuando el número de mediciones es grande.

Si hay más información disponible más que la media y la varianza, entonces es posible obtener cotas más ajustadas que las desigualdades de Markov y Chebyshev. Consideremos la desigualdad de Markov de nuevo. La región de interés es $A = \{t \geq a\}$, así que sea $I_A(t)$ la función indicatriz, es decir, $I_A(t) = 1$ si $t \in A$ y $I_A(t) = 0$ en caso contrario. El paso clave en la derivación es tener en cuenta que $t/a \geq 1$ en la región de interés. En efecto, se acota $I_A(t)$ por t/a como se muestra en la figura 4.14. Tenemos entonces:

$$P[X \geq a] = \int_0^{\infty} I_A(t) f_X(t) dt \leq \int_0^{\infty} \frac{t}{a} f_X(t) dt = \frac{E[X]}{a}.$$

Al cambiar el límite superior de $I_A(t)$, podemos obtener distintos límites de $P[X \geq a]$. Consideremos el límite $I_A(t) \leq e^{s(t-a)}$, también en la figura 4.14, donde $s > 0$. La cota resultante es:

$$\begin{aligned} P[X \geq a] &= \int_0^{\infty} I_A(t) f_X(t) dt \leq \int_0^{\infty} e^{s(t-a)} f_X(t) dt \\ &= e^{-sa} \int_0^{\infty} e^{st} f_X(t) dt = e^{-sa} E[e^{sX}]. \end{aligned} \quad (4.78)$$

Esta cota se llama **cota de Chernoff**, que puede verse como dependiente de la esperanza de una función exponencial de X . Esta función se llama el función generadora de momentos.

Problemas

4.1 Una fuente de información produce pares binarios que designamos como $S_X = \{1, 2, 3, 4\}$ con las siguientes pmf:

- (i) $p_k = p_1/k$ para todo $k \in S_X$.
- (ii) $p_{k+1} = p_k/2$ para $k = 2, 3, 4$.

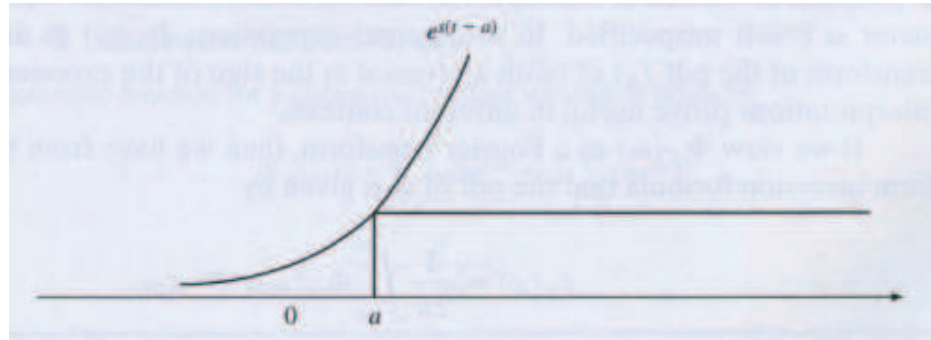


Figura 4.14 Límites en la función indicatriz para $A = \{t \geq a\}$.

(iii) $p_{k+1} = p_k/2^k$ para $k = 2, 3, 4$.

(a) Dibuja la cdf de estas tres variables aleatorias.

(b) Utiliza la cdf para calcular la probabilidad de los sucesos: $\{X \leq 1\}$, $\{X > 2.5\}$, $\{0.5 < X \leq 2\}$ y $\{1 < X < 4\}$.

4.2 Una variable aleatoria X tiene cdf:

$$F_X(x) = \begin{cases} 0 & x < 0 \\ 1 - \frac{1}{4}e^{-2x} & x \geq 0. \end{cases}$$

(a) Dibuja la cdf e identifica el tipo de variable aleatoria.

(b) Calcula $P[X \leq 2]$, $P[X = 0]$, $P[X < 0]$, $P[2 < X < 6]$ y $P[X > 10]$.

4.3 Una variable aleatoria X tiene pdf:

$$f_X(x) = \begin{cases} c(1-x^2) & -1 \leq x \leq 1 \\ 0 & \text{en otro caso.} \end{cases}$$

(a) Calcula c y dibuja la pdf.

(b) Dibuja la cdf de X .

(c) Calcula $P[X = 0]$, $P[0 < X < 0.5]$, y $P[|X - 0.5| < 0.25]$.

4.4 Calcula y dibuja la pdf del problema 4.2. Utiliza la pdf para calcular $P[X = 0]$ y $P[X > 8]$.

4.5 Sea X la variable aleatoria Exponencial.

(a) Calcula y dibuja $F_X(x|X > t)$. ¿En qué se diferencia $F_X(x|X > t)$ de $F_X(x)$?

(b) Calcula y dibuja $f_X(x|X > t)$.

(c) Demuestra que $P[X > t + x|X > t] = P[X > x]$. Explica por qué esto es llamado la propiedad de no tener memoria.

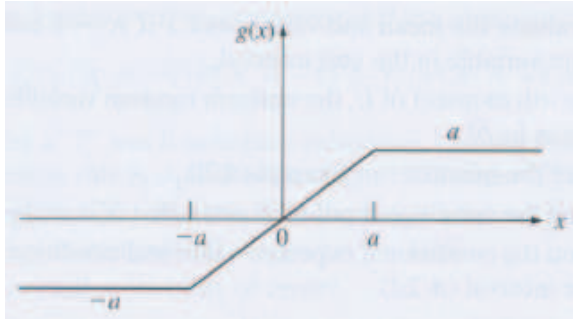


Figura 4.15

4.6 Un sistema de transmisión binaria envía un bit “0” bit usando una señal de tensión -1 y un bit “1” mediante la transmisión de un $+1$. La señal recibida está dañada por el ruido N que tiene una distribución de Laplace con parámetro α . Asume que “0” y “1” son equiprobables.

- Calcula la pdf de la señal recibida $Y = X + N$, donde X es la señal transmitida, dado que se transmite un “0” y, luego, dado que un “1” se transmite.
- Supongamos que el receptor decide que se transmitió un “0” si $Y < 0$ y que se envió un “1” si $Y \geq 0$. ¿Cuál es la probabilidad de que el receptor cometa un error dado que se transmitió un $+1$? ¿Y dado que se transmitió un -1 ?
- ¿Cuál es la probabilidad global de error?

4.7 Calcula la media y la varianza de X en el problema 4.3.

4.8 En la figura 4.15 se muestra un limitador.

- Calcula la expresión de la media y la varianza de $Y = g(X)$ para una variable aleatoria continua X arbitraria.
- Evalúa la media y la varianza si X es una variable aleatoria de Laplace con $\lambda = a = 1$.
- Repite la parte b) si X es la variable del problema 4.3 con $a = 1/2$.
- Evalúa la media y la varianza si $X = U^3$ donde U es una variable aleatoria Uniforme en el intervalo unidad, $[-1, 1]$ y $a = 1/2$.

4.9 Sea X una variable aleatoria Normal con media $m = 5$ y $\sigma^2 = 16$.

- Calcula $P[X > 4]$, $P[X \geq 7]$, $P[6.72 < X < 10.16]$, $P[2 < X < 7]$ y $P[6 \leq X \leq 8]$.
- Calcula a tal que $P[X < a] = 0.8869$.
- Calcula b tal que $P[X > b] = 0.11131$.
- Calcula c tal que $P[13 < X \leq c] = 0.0123$.

4.10 Se considera dos chips para uso en cierto sistema. El tiempo de vida del chip 1 se modela con una variable aleatoria Normal con media 20000 horas y desviación típica de 5000 horas. (La probabilidad de un tiempo de vida negativo es despreciable.) El tiempo de vida del chip 2 es también una variable aleatoria

Normal pero con media 22000 horas y una desviación típica 1000 horas. ¿Qué chip es preferible si el tiempo de vida objetivo del sistema es 20000 horas? ¿Y si es 24000 horas?

4.11 Los pasajeros llegan a una parada de taxi a una velocidad de un pasajero por minuto. El conductor del taxi no sale hasta que llegan siete pasajeros para llenar el taxi. Supongamos que el tiempo entre llegadas son variables aleatorias exponenciales y sea X el tiempo que se tarda en llenar el taxi. Calcula la probabilidad de que el taxi tarde más de 10 minutos en llenarse.

4.12 Dibuja la pdf de la variable aleatoria Beta con: $a = b = 1/4, 1, 4, 8$; $a = 5, b = 1$; $a = 1, b = 3$ y $a = 2, b = 5$.

4.13 El beneficio neto en una transacción es dado por $Y = 2 - 4X$ donde X es la variable aleatoria del problema 4.2. Calcula la cdf y la pdf de Y .

4.14 Sea X es el número de éxitos en n intentos de Bernoulli donde la probabilidad de éxito es p . Sea $Y = X/n$ el número medio de éxitos por intento. Aplica la desigualdad de Chebyshev al suceso $\{|Y - p| > a\}$. ¿Qué pasa cuando $n \rightarrow \infty$?

4.15 Tres tipos de clientes llegan a una estación de servicio. El tiempo de servicio para los clientes tipo 1 es una variable aleatoria Exponencial con media 2. Los de tipo 2 tienen una distribución de Pareto con $\alpha = 3$ y $x_m = 1$. Los clientes de tipo 3 requieren un tiempo de servicio constante de 2 segundos. Supongamos que la proporción de clientes de tipo 1, 2 y 3 es $1/2, 1/8$ y $3/8$, respectivamente. Calcula la probabilidad de que un cliente arbitrario necesite más de 15 segundos de tiempo de servicio. Compara la probabilidad anterior con la cota proporcionada por la desigualdad de Markov.

4.16 El tiempo de vida X de una bombilla es una variable aleatoria con

$$P[X > t] = 2/(2 + t) \text{ para } t > 0.$$

Supongamos que se instala tres nuevas bombillas en el tiempo $t = 0$. En $t = 1$ todas las bombillas aun funcionan. Calcula la probabilidad de que al menos una bombilla esté funcionando en el tiempo $t = 9$.

4.17 La variable aleatoria X es Uniforme en el intervalo $[0, a]$. Supongamos que a es desconocido, entonces estimamos a como el valor máximo observado en n repeticiones independientes del experimento, esto es, estimamos a por $Y = \max\{X_1, X_2, \dots, X_n\}$.

(a) Calcula $P[Y \leq y]$.

(b) Calcula la media y la varianza de Y y explica por qué Y es un buen estimador para a cuando N es grande.

4.18 La muestra X de una señal es una variable aleatoria Normal con $m = 0$ y $\sigma^2 = 1$. Supongamos que X se cuantifica con un cuantificador que consiste en cuatro intervalos: $(-\infty, -a]$, $(-a, 0]$, $(0, a]$ y (a, ∞) .

- (a) Calcula el valor de a para que la X sea equiprobable en los cuatro intervalos.
 (b) Halla el punto de representación $x_i = q(X)$ para X en $(0, a]$ que minimiza el error cuadrático medio, esto es,

$$\int_0^a (x - x_1)^2 f_X(x) dx$$

se minimiza. Pista: Deriva la expresión anterior con respecto a x_i . Calcula los puntos de representación para los otros intervalos.

- (c) Evalúa el error cuadrático medio del cuantificador $E[(X - q(X))^2]$.

4.19 La salida Y de un sistema binario de comunicación es una variable aleatoria Normal de varianza 1 con media 0 cuando la entrada es “0” y con media 1 cuando la entrada es “1”. Asumimos que la entrada es “1” con probabilidad p .

- (a) Calcula $P[\text{entrada es } 1 | y < Y < y + h]$ y $P[\text{entrada es } 0 | y < Y < y + h]$.
 (b) El receptor utiliza la siguiente regla de decisión: “Si $P[\text{entrada es } 1 | y < Y < y + h] > P[\text{entrada es } 0 | y < Y < y + h]$, decide que la entrada era un 1; en otro caso, decide que la entrada fue 0”. Demuestra que esta regla de decisión lleva al siguiente umbral de decisión: “Si $Y > T$, decide que la entrada era 1; en otro caso, decide que la entrada era 0”.
 (c) ¿Cuál es la probabilidad de error para la regla de decisión anterior?

5 Variables aleatorias bidimensionales

Muchos experimentos aleatorios implican varias variables aleatorias. En algunos experimentos se miden diferentes cantidades. Por ejemplo, puede interesarnos el voltaje de la señal en varios puntos de un circuito en un determinado tiempo. Otros experimentos implican medidas repetidas de cierta cantidad como por ejemplo el muestreo de la amplitud de una señal de video o audio que varía en el tiempo. En el capítulo 4 desarrollamos técnicas para calcular la probabilidad de sucesos relacionados con una variable aleatoria aislada. En este capítulo, extenderemos estos conceptos a dos variables aleatorias:

- Utilizaremos las pmf, cdf y pdf conjuntas para calcular las probabilidades de sucesos que implican un comportamiento conjunto de dos variables aleatorias.
- Utilizaremos la esperanza para definir momentos conjuntos que resuman el comportamiento de dos variables aleatorias.
- Determinaremos cuándo dos variables aleatorias son independientes y cuantificaremos su grado de “correlación” cuando no sean independientes.
- Obtendremos probabilidades condicionadas que involucren dos variables aleatorias.

De algún modo, ya hemos cubierto todos los conceptos fundamentales de la probabilidad y las variables aleatorias y desarrollaremos simplemente el caso de dos o más variables aleatorias. Sin embargo, existen técnicas analíticas importantes que necesitamos aprender, por ejemplo, sumatorios dobles de pmf e integración doble de pdf, así que primero discutiremos el caso de dos variables aleatorias en detalle porque es el caso más intuitivo ya que podemos dibujarlo fácilmente. El capítulo 6 considera el caso general de variables aleatorias multidimensionales o vectores de variables aleatorias.

5.1 Dos variables aleatorias

La noción de variable aleatoria como una función (un mapeo) es fácilmente generalizable al caso en el que existen dos cantidades de interés. Consideremos un experimento aleatorio con espacio muestral S y clases de sucesos \mathcal{F} . Estamos interesados en una función que asigne un par de números reales

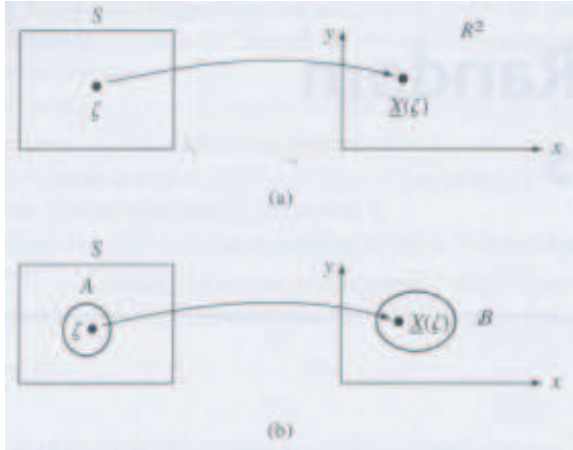


Figura 5.1 (a) Una función que asigna un par de números reales a cada elemento de S . (b) Sucesos equivalentes para dos variables aleatorias.

$\mathbf{X}(\zeta) = (X(\zeta), Y(\zeta))$ a cada elemento muestral ζ de S . Esencialmente, estamos definiendo una función vectorial que mapea S a \mathbb{R}^2 , el plano real, como se muestra en la figura 5.1. En última instancia, estamos interesados en sucesos relacionados con el par (X, Y) .

Ejemplo 5.1 Sea un experimento aleatorio que consiste en seleccionar aleatoriamente el nombre de un estudiante. Sea ζ el resultado del experimento y definimos las siguientes funciones:

$$H(\zeta) = \text{altura en centímetros del estudiante } \zeta$$

$$W(\zeta) = \text{peso en centímetros del estudiante } \zeta$$

$(H(\zeta), W(\zeta))$ asigna un par de números a cada ζ en S .

Estamos interesados en sucesos relacionados con el par (H, W) . Por ejemplo, el suceso $B = \{H \leq 183, W \leq 82\}$ representa a los estudiantes que miden menos que 183 cm y pesan menos de 82 kg.

Ejemplo 5.2 Una página web proporciona al usuario la posibilidad de elegir entre ver un anuncio corto o ir directamente a la página. Sean ζ los patrones de llegada de usuarios en T segundos, e.g. número de llegadas, lista de tiempo de llegadas y tipos de llegada. Sea $N_1(\zeta)$ el número de veces que la página web es requerida directamente y sea $N_2(\zeta)$ el número de veces que se elige el anuncio. $(N_1(\zeta), N_2(\zeta))$ asigna un par de números enteros no negativos a cada ζ de S . Supongamos que una petición de tipo 1 aporta 0.001 céntimos en ingresos y una solicitud de tipo 2 aporta 1 céntimo. Da una expresión para el suceso “los ingresos de T segundos son inferiores a 100 euros”.

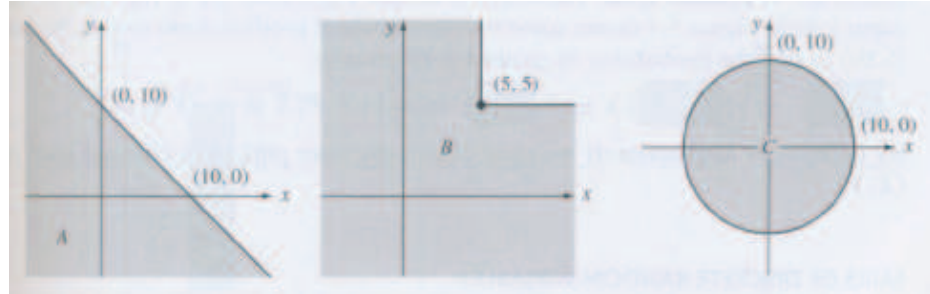


Figura 5.2 Ejemplos de sucesos bidimensionales.

Solución

El total de ingresos en T segundos es $0.001N_1 + 1N_2$ y, por tanto, el suceso de interés es $B = \{0.001N_1 + 1N_2 < 10000\}$.

Ejemplo 5.3 Sea ζ la longitud de un mensaje seleccionado aleatoriamente. Supongamos que los mensajes están divididos en paquetes de longitud máxima M bytes. Sea Q el número de paquetes enteros en un mensaje y R el número de bytes restantes. $(Q(\zeta), R(\zeta))$ asigna un par de números a cada ζ de S . Q toma valores en el rango $0, 1, 2, \dots$ y R toma valores en el rango $0, 1, \dots, M - 1$. Un suceso de interés podría ser $B = \{R < M/2\}$, “el último paquete está lleno hasta menos de la mitad”.

Ejemplo 5.4 Sea $\zeta = (\zeta_1, \zeta_2)$ el resultado de un experimento aleatorio que consiste en hacer girar dos veces de forma independiente una rueda de la fortuna. Cada “tirada” de la rueda da como resultado un número en el intervalo $(0, 2\pi]$. Definimos un par de números (X, Y) en el plano como sigue:

$$X(\zeta) = \left(2 \ln \frac{2\pi}{\zeta_1}\right)^{1/2} \cos \zeta_2 \quad Y(\zeta) = \left(2 \ln \frac{2\pi}{\zeta_1}\right)^{1/2} \sin \zeta_2.$$

La función vectorial $(X(\zeta), Y(\zeta))$ asigna un par de números en el plano a cada ζ de S . El término de la raíz cuadrada corresponde al radio y el de ζ_2 al ángulo.

Veremos que (X, Y) modela las tensiones de ruido de los sistemas de comunicación digital. Un suceso de interés aquí podría ser $B = \{X^2 + Y^2 < r^2\}$, “la potencia total del ruido es inferior a r^2 ”.

Los sucesos relacionados con un par de variables aleatorias (X, Y) se especifican por las condiciones en las que estamos interesados y son representados por

regiones en el plano. La figura 5.2 muestra tres ejemplos de sucesos:

$$\begin{aligned} A &= \{X + Y \leq 10\}, \\ B &= \{\min(X, Y) \leq 5\}, \\ C &= \{X^2 + Y^2 \leq 100\}. \end{aligned}$$

El suceso A divide el plano en dos regiones mediante una línea recta. Obsérvese que el suceso del ejemplo 5.2 es de este tipo. El suceso C define un disco centrado en el origen y se corresponde con el suceso del ejemplo 5.4. El suceso B se calcula observando que $\{\min(X, Y) \leq 5\} = \{X \leq 5\} \cup \{Y \leq 5\}$, esto es, el mínimo de X e Y es menor o igual que 5 si, o bien X y/o Y es/son menores o iguales a 5.

Para determinar la probabilidad de que el par $\mathbf{X} = (X, Y)$ esté en una región B en el plano, procedemos como en el capítulo 3 cuando hallábamos el suceso equivalente de B en el espacio muestral S :

$$A = \mathbf{X}^{-1}(B) = \{\zeta : (X(\zeta), Y(\zeta)) \in B\}. \quad (5.1a)$$

La relación entre $A = \mathbf{X}^{-1}(B)$ y B se muestra en la figura 5.1(b). Si A está en \mathcal{F} , entonces tiene una probabilidad asignada a él y obtenemos:

$$P[X \in B] = P[A] = P[\{\zeta : (X(\zeta), Y(\zeta)) \in B\}]. \quad (5.1b)$$

El método es idéntico al que utilizamos en el caso de variables aleatorias unidimensionales. La única diferencia es que consideramos *el comportamiento conjunto de X e Y* que es inducido por el experimento aleatorio subyacente.

Para deducir el comportamiento conjunto de dos variables aleatorias podemos usar el diagrama de dispersión. Un diagrama de dispersión coloca un punto en cada par observado (x, y) , resultante del experimento generador de (X, Y) . La figura 5.3 muestra los diagramas de dispersión de 200 observaciones de cuatro diferentes pares de variables aleatorias. El par de la figura 5.3(a) parece distribuirse uniformemente en el cuadrado unidad. El par de la figura 5.3(b) está claramente confinado a un disco de radio 1 y parece estar más concentrado alrededor del origen. El par de la figura 5.3(c) se concentra alrededor del origen y parece tener simetría circular, pero no está acotado a una región cerrada. El par en la figura 5.3(d) también se concentra en el origen y parece poseer una clara relación lineal de algún tipo, esto es, los valores altos de x tienden a ser directamente proporcionales a los valores de y . Presentaremos más adelante algunas funciones y momentos para caracterizar el comportamiento de variables aleatorias bivariantes ilustradas en estos ejemplos.

La función de masa de probabilidad conjunta, la función de distribución conjunta y la función de densidad conjunta proporcionan métodos para especificar la ley de probabilidad que gobierna el comportamiento del par (X, Y) . Nuestro método general es como sigue. Primero nos centramos en sucesos que corresponden a rectángulos en el plano:

$$B = \{X \in A_1\} \cap \{Y \in A_2\} \quad (5.2)$$

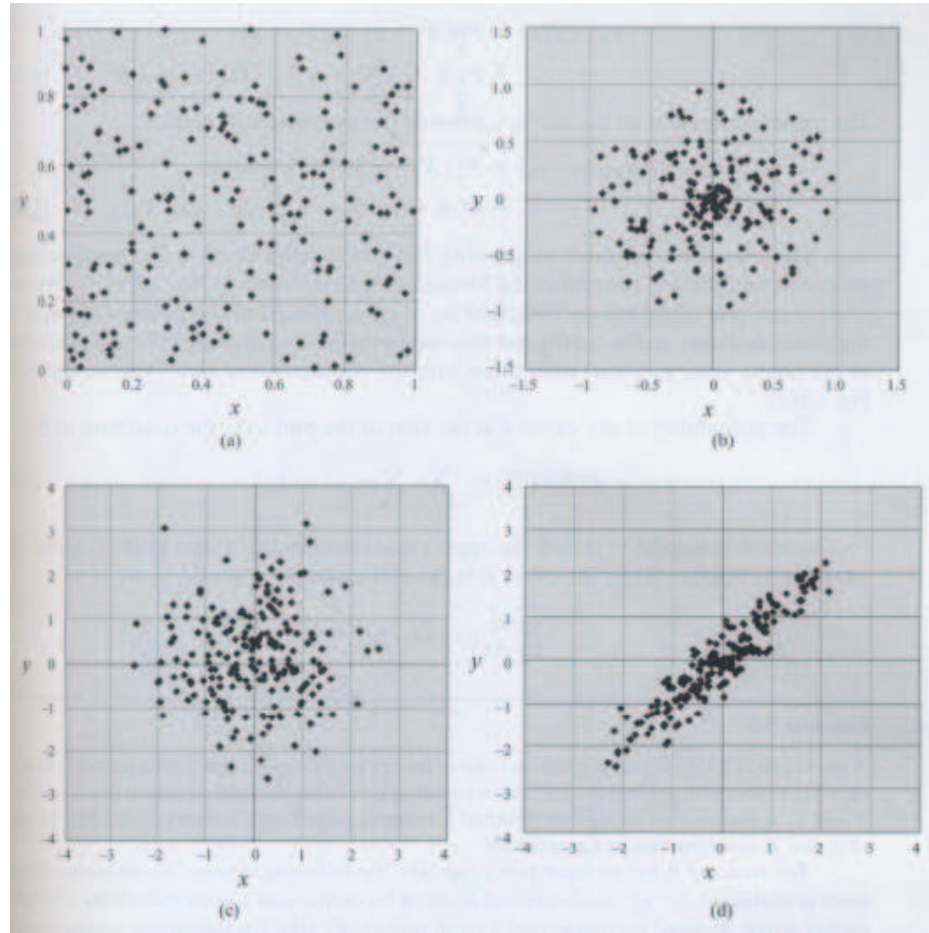


Figura 5.3 Diagramas de dispersión de 200 observaciones de cuatro diferentes pares de variables aleatorias.

donde A_k es un suceso unidimensional (i.e. un subconjunto de la recta real). Decimos que estos sucesos están en **forma de producto**. El suceso B se verifica cuando ambos, $\{X \in A_1\}$ y $\{Y \in A_2\}$, ocurren a la vez. La figura 5.4 muestra algunos sucesos bidimensionales en forma de producto:

$$P[B] = P[\{X \in A_1\} \cap \{Y \in A_2\}] \triangleq P[X \in A_1, Y \in A_2]. \quad (5.3)$$

Definiendo A de manera apropiada obtenemos la pmf conjunta, la cdf conjunta y la pdf conjunta de (X, Y) .

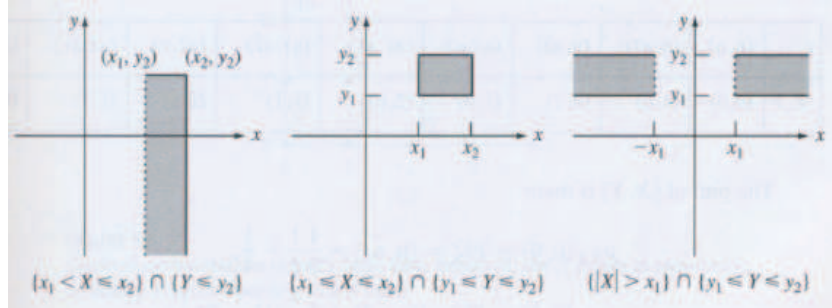


Figura 5.4 Algunos sucesos bidimensionales en forma de producto.

5.2 Pares de variables aleatorias discretas

Sea $\mathbf{X} = (X, Y)$ un vector de variables aleatorias que toma valores en un conjunto numerable $S_{X,Y} = \{(x_j, y_k), j = 1, 2, \dots, k = 1, 2, \dots\}$. La **función de masa de probabilidad conjunta** de \mathbf{X} proporciona el valor de las probabilidades de los sucesos $\{X = x\} \cup \{Y = y\}$:

$$\begin{aligned} p_{X,Y}(x, y) &= P[\{X = x\} \cup \{Y = y\}] \\ &\triangleq P[X = x, Y = y] \quad \text{para } (x, y) \in \mathbb{R}^2. \end{aligned} \quad (5.4a)$$

Los valores de la pmf en el conjunto $S_{X,Y}$ proporcionan la información necesaria:

$$\begin{aligned} p_{X,Y}(x_j, y_k) &= P[\{X = x_j\} \cup \{Y = y_k\}] \\ &\triangleq P[X = x_j, Y = y_k] \quad \text{para } (x_j, y_k) \in S_{X,Y}. \end{aligned} \quad (5.4b)$$

Existen varias formas de mostrar gráficamente la pmf: (1) Para espacios muestrales pequeños podemos representar la pmf en forma de tabla como se muestra en la figura 5.5(a). (2) Podemos representar la pmf utilizando flechas de altura $p_{X,Y}(x_j, y_k)$ colocadas en los puntos $\{(x_j, y_k)\}$ del plano, como se muestra en la figura 5.5(b), pero puede ser difícil de dibujar. (3) Otra opción es colocar puntos en $\{(x_j, y_k)\}$ y etiquetarlos con el valor de la pmf correspondiente como se muestra en la figura 5.5(c).

La probabilidad de cualquier suceso B es la suma de la pmf en los elementos muestrales en B :

$$P[\mathbf{X} \in B] = \sum_{(x_j, y_k) \in B} p_{X,Y}(x_j, y_k). \quad (5.5)$$

Con frecuencia es útil dibujar la región que contiene los puntos en B , como se muestra, por ejemplo, en la figura 5.6. Cuando el suceso B es todo el espacio muestral $S_{X,Y}$, tenemos:

$$\sum_{j=1}^{\infty} \sum_{k=1}^{\infty} p_{X,Y}(x_j, y_k) = 1. \quad (5.6)$$

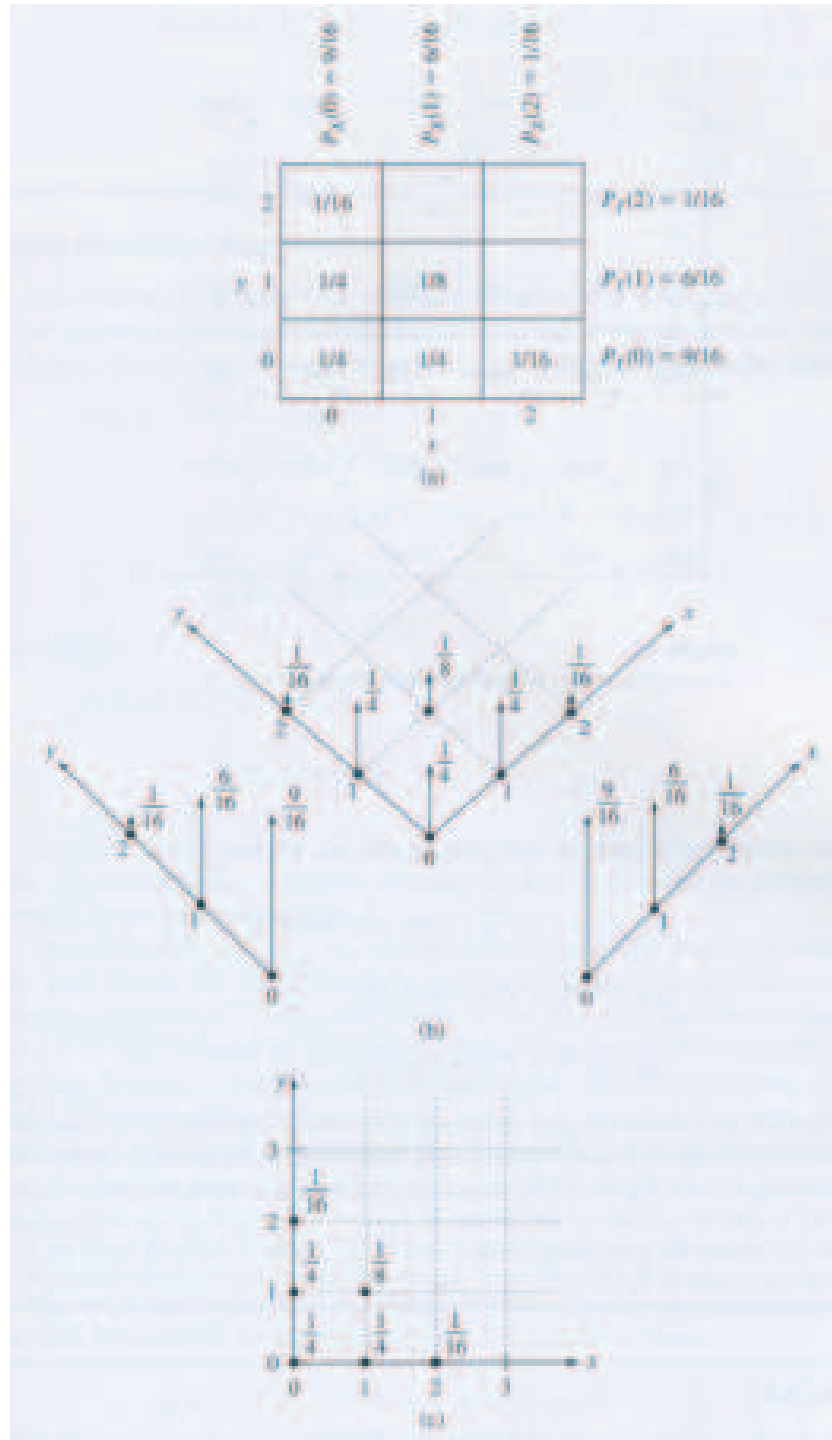


Figura 5.5 Representaciones gráficas de la pmf: (a) en forma de tabla; (b) uso de flechas para mostrar altura; (c) puntos etiquetados con el valor de la pmf.

Ejemplo 5.5 Un conmutador de paquetes tiene dos puertos de entrada y dos puertos de salida. En un intervalo de tiempo dado un paquete llega a cada puerto de entrada con una probabilidad de $1/2$, y es igualmente probable que sea destinado al puerto de salida de 1 o 2. Sean X e Y el número de paquetes destinados a las salidas 1 y 2, respectivamente. Buscar la pmf de X e Y , y mostrarla gráficamente.

Solución

El resultado I_j para el puerto de entrada j puede tomar los siguientes valores: “n”, ninguna llegada de paquetes (con una probabilidad de $1/2$), “a1”, llegada de un paquete destinado al puerto de salida 1 (con una probabilidad de $1/4$), “a2”, llegada de un paquete destinado al puerto de salida 2 (con una probabilidad de $1/4$). El espacio muestral S subyacente consiste en el par de los resultados de entrada $\zeta = (I_1, I_2)$. La asignación de (X, Y) se muestra en la siguiente tabla:

ζ	(n,n)	(n,a1)	(n,a2)	(a1,n)	(a1,a1)	(a1,a2)	(a2,n)	(a2,a1)	(a2,a2)
X, Y	(0,0)	(1,0)	(0,1)	(1,0)	(2,0)	(1,1)	(0,1)	(1,1)	(0,2)

La pmf de (X, Y) es entonces:

$$\begin{aligned}
 p_{X,Y}(0,0) &= P[\zeta = (n,n)] = \frac{1}{2} \frac{1}{2} = \frac{1}{4}, \\
 p_{X,Y}(0,1) &= P[\zeta \in \{(n,a2), (a2,n)\}] = 2 \frac{1}{8} = \frac{1}{4}, \\
 p_{X,Y}(1,0) &= P[\zeta \in \{(n,a1), (a1,n)\}] = \frac{1}{4}, \\
 p_{X,Y}(1,1) &= P[\zeta \in \{(a1,a2), (a2,a1)\}] = \frac{1}{8}, \\
 p_{X,Y}(0,2) &= P[\zeta = (a2,a2)] = \frac{1}{16}, \\
 p_{X,Y}(2,0) &= P[\zeta = (a1,a1)] = \frac{1}{16}.
 \end{aligned}$$

La figura 5.5(a) muestra la pmf en forma de tabla, donde el número de filas y columnas contienen el rango de X e Y , respectivamente. Cada entrada en la tabla da el valor de la pmf para el valor de x e y correspondiente. La figura 5.5(b) muestra la pmf utilizando flechas en el plano. Se coloca una flecha de altura $p_{X,Y}(j,k)$ en cada uno de los puntos de $S_{X,Y} = \{(0,0), (0,1), (1,0), (1,1), (0,2), (2,0)\}$. La figura 5.5(c) muestra la pmf con puntos etiquetados en el plano. Se coloca un punto con la etiqueta $p_{X,Y}(j,k)$ en cada uno de los puntos de $S_{X,Y}$.

Ejemplo 5.6 Un experimento aleatorio consiste en lanzar dos dados “trucados”. La pmf conjunta $p_{X,Y}(j,k)$ para $j = 1, \dots, 6$ y $k = 1, \dots, 6$ viene dada por la tabla de doble entrada que se muestra en la figura 5.6. La posición (j,k) de la tabla contiene el valor de $p_{X,Y}(j,k)$. Calcula la $P[\text{mín}(X, Y) = 3]$.

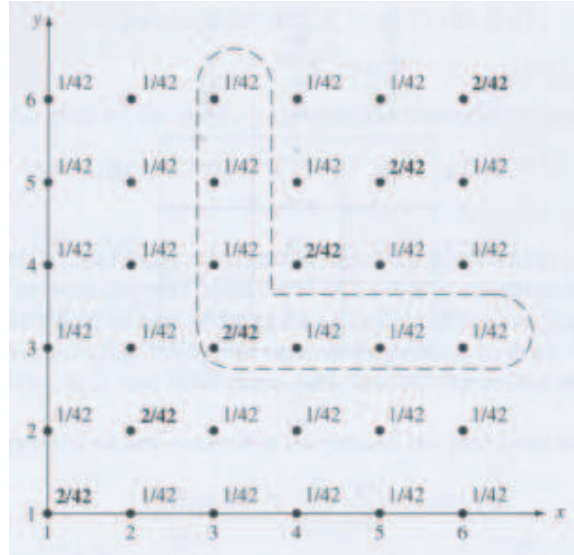


Figura 5.6 Mostrando la pmf a través de un esquema que contiene los puntos en B .

Solución

La figura 5.6 muestra la región que corresponde al conjunto $\{\min(x, y) = 3\}$. La probabilidad de este suceso está dada por:

$$\begin{aligned} P[\min(X, Y) = 3] &= p_{X,Y}(6, 3) + p_{X,Y}(5, 3) + p_{X,Y}(4, 3) \\ &\quad + p_{X,Y}(3, 3) + p_{X,Y}(3, 4) + p_{X,Y}(3, 5) + p_{X,Y}(3, 6) \\ &= 6 \left(\frac{1}{42} \right) + \frac{2}{42} = \frac{8}{42}. \end{aligned}$$

5.2.1 Función de masa de probabilidad marginal

La pmf conjunta de \mathbf{X} proporciona la información sobre el comportamiento conjunto de X e Y . También estamos interesados en la probabilidad de sucesos relacionados con cada una de las variables aleatorias de manera aislada. Estos se pueden calcular en términos de las **funciones de la masa de probabilidad marginales**:

$$\begin{aligned} p_X(x_j) &= P[X = x_j] \\ &= P[X = x_j, Y = \text{cualquiera}] \\ &= P[\{X = x_j \text{ y } Y = y_1\} \cup \{X = x_j \text{ y } Y = y_2\} \cup \dots] \\ &= \sum_{k=1}^{\infty} p_{X,Y}(x_j, y_k), \end{aligned} \tag{5.7a}$$

y de forma similar

$$\begin{aligned} p_Y(y_k) &= P[Y = y_k] \\ &= \sum_{j=1}^{\infty} p_{X,Y}(x_j, y_k). \end{aligned} \quad (5.7b)$$

Las pmf marginales satisfacen todas las propiedades de las pmf unidimensionales y proveen de la información necesaria para calcular la probabilidad de sucesos relacionados con la variable aleatoria correspondiente.

La probabilidad $p_{X,Y}(x_j, y_k)$ se puede interpretar como el límite de la frecuencia relativa del suceso conjunto $\{X = X_j\} \cap \{Y = Y_k\}$ en una serie de repeticiones del experimento aleatorio. La ecuación (5.7a) corresponde al hecho de que la frecuencia relativa del suceso $\{X = X_j\}$ se calcula sumando las frecuencias relativas de todos los pares de resultados en los que aparece X_j . En general, es imposible deducir las frecuencias relativas de pares de valores X e Y a partir de las frecuencias relativas de X e Y aisladas. Lo mismo se verifica para las pmf: en general, conocer las pmf marginales es insuficiente para especificar la pmf conjunta.

Ejemplo 5.7 Calcula la pmf marginal de los puertos de salida (X, Y) del ejemplo 5.5.

Solución

La figura 5.5(a) muestra que la pmf marginal se calcula sumando las filas o las columnas de la tabla. Por ejemplo, sumando la columna correspondiente a $x = 1$ se tiene:

$$p_X(1) = P[X = 1] = p_{X,Y}(1, 0) + p_{X,Y}(1, 1) = \frac{1}{4} + \frac{1}{8} = \frac{3}{8}.$$

De forma similar, sumando la fila de $y = 0$:

$$p_Y(0) = P[Y = 0] = p_{X,Y}(0, 0) + p_{X,Y}(1, 0) + p_{X,Y}(2, 0) = \frac{1}{4} + \frac{1}{4} + \frac{1}{16} = \frac{9}{16}.$$

La figura 5.5(b) muestra la pmf marginal utilizando flechas en la recta real.

Ejemplo 5.8 Calcula la pmf marginal en el experimento con los dados no equilibrados del ejemplo 5.6.

Solución

La probabilidad de que $X = 1$ se calcula sumando la primera fila:

$$P[X = 1] = \frac{2}{42} + \frac{1}{42} + \cdots + \frac{1}{42} = \frac{1}{6}.$$

De forma similar, concluimos que $P[X = j] = 1/6$ para $j = 2, \dots, 6$. La probabilidad de que $Y = k$ se calcula sumando la columna k -ésima. Vemos entonces que $P[Y = k] = 1/6$ para $k = 1, 2, \dots, 6$. Entonces, cada dado de forma aislada

parece equilibrado en el sentido que cada cara es igualmente probable. Si sólo conociéramos las pmf marginales no tendríamos idea alguna de que los dados están “trucados”.

Ejemplo 5.9 En el ejemplo 5.3, sea N el número de bytes en un mensaje con distribución geométrica de parámetro $1 - p$ y rango $S_N = \{0, 1, 2, \dots\}$. Calcula la pmf conjunta y la pmf marginales de Q y R .

Solución

Si un mensaje tiene N bytes, entonces el número de paquetes enteros es el cociente, Q , de la división de N entre M y el número de bytes restantes, R , es el resto. La probabilidad del par $\{(q, r)\}$ viene dada por

$$P[Q = q, R = r] = P[N = qM + r] = (1 - p)p^{qM+r},$$

La pmf marginal de Q es

$$\begin{aligned} P[Q = q] &= P[N \in \{qM, qM + 1, \dots, qM + (M - 1)\}] \\ &= \sum_{k=0}^{M-1} (1 - p)p^{qM+k} \\ &= (1 - p)p^{qM} \frac{1 - p^M}{1 - p} = (1 - p^M)(p^M)^q \quad q = 0, 1, 2, \dots \end{aligned}$$

La pmf marginal de Q es geométrica con parámetro p^M . La pmf marginal de R es:

$$\begin{aligned} P[R = r] &= P[N \in \{r, M + r, 2M + r, \dots\}] \\ &= \sum_{q=0}^{\infty} (1 - p)p^{qM+r} = \frac{(1 - p)}{1 - p^M} p^r \quad r = 0, 1, \dots, M - 1. \end{aligned}$$

R tiene una pmf geométrica truncada. Como ejercicio, queda verificar que las pmf marginales de arriba suman 1.

5.3 La cdf conjunta de X e Y

En el capítulo 3 vimos que los intervalos semi-infinitos de la forma $(-\infty, x]$ son los ladrillos básicos con los cuales se construyen otros sucesos unidimensionales. Definiendo la cdf $F_X(x)$ como la probabilidad de $(-\infty, x]$, éramos capaces de expresar las probabilidades de otros sucesos en términos de la cdf. En esta sección repetimos aquella sección para el caso de variables aleatorias bidimensionales.

La pieza básica para sucesos relacionados con variables aleatorias bidimensionales es el rectángulo semi-infinito definido por $\{(x, y) : x \leq x_1 \text{ e } y \leq y_1\}$, como se muestra en la figura 5.7. Usaremos la siguiente notación $\{x \leq x_1, y \leq y_1\}$ para

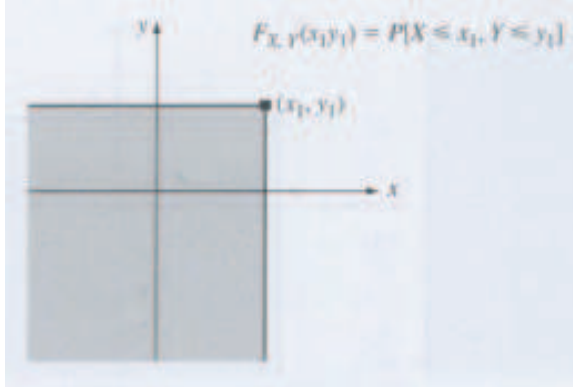


Figura 5.7 La función de distribución conjunta se define como la probabilidad del rectángulo semi-infinito definido por el punto (x_1, y_1) .

referirnos a esta región. La **función de distribución conjunta de X e Y** se define como la probabilidad del suceso $\{X \leq x_1\} \cap \{Y \leq y_1\}$:

$$F_{X,Y}(x_1, y_1) = P[X \leq x_1, Y \leq y_1]. \quad (5.8)$$

En términos de la frecuencia relativa, $F_{X,Y}(x_1, y_1)$ representa el límite de la proporción de tiempo en el que el resultado del experimento aleatorio arroja un punto X que cae en la región rectangular de la figura 5.7. En términos de probabilidad, $F_{X,Y}(x_1, y_1)$ representa la cantidad de masa contenida en la región rectangular.

La cdf conjunta satisface las siguientes propiedades.

(i) La cdf conjunta es una función no decreciente de x e y :

$$F_{X,Y}(x_1, y_1) \leq F_{X,Y}(x_2, y_2) \text{ si } x_1 \leq x_2 \text{ y } y_1 \leq y_2, \quad (5.9a)$$

(ii)

$$F_{X,Y}(x_1, -\infty) = 0, \quad F_{X,Y}(-\infty, y_1) = 0, \quad F_{X,Y}(\infty, \infty) = 1. \quad (5.9b)$$

(iii) Obtenemos las **funciones de distribución marginales** eliminando la restricción en una de las variables. Las cdf marginales son las probabilidades de las regiones representadas en la figura 5.8:

$$F_X(x_1) = F_{X,Y}(x_1, \infty) \text{ y } F_Y(y_1) = F_{X,Y}(\infty, y_1). \quad (5.9c)$$

(iv) La cdf conjunta es continua por el “norte” y por el “este”, esto es,

$$\lim_{x \rightarrow a^+} F_{X,Y}(x, y) = F_{X,Y}(a, y) \text{ y } \lim_{x \rightarrow b^+} F_{X,Y}(x, y) = F_{X,Y}(x, b). \quad (5.9d)$$

(v) La probabilidad del rectángulo $\{x_1 < X \leq x_2, y_1 < Y \leq y_2\}$ viene dada por:

$$\begin{aligned} P[x_1 < X \leq x_2, y_1 < Y \leq y_2] = \\ F_{X,Y}(x_2, y_2) - F_{X,Y}(x_2, y_1) - F_{X,Y}(x_1, y_2) + F_{X,Y}(x_1, y_1). \end{aligned} \quad (5.9e)$$

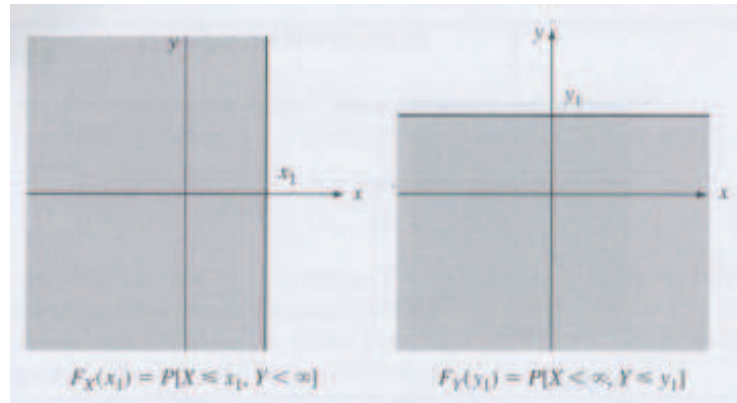


Figura 5.8 Las cdf marginales son las probabilidades de estos semiplanos.

La propiedad (i) se deduce porque el rectángulo semi-infinito definido por (x_1, y_1) está contenido en el definido por (x_2, y_2) y aplicando el corolario 7. Las propiedades (ii) a (iv) se obtienen tomando límites. Por ejemplo, la secuencia $\{x \leq x_1 \text{ y } y \leq -n\}$ es decreciente y se aproxima al conjunto vacío \emptyset , entonces

$$F_{X,Y}(x_1, -\infty) = \lim_{n \rightarrow \infty} F_{X,Y}(x_1, -n) = P[\emptyset] = 0.$$

Para la propiedad (iii) tomamos la secuencia $\{x \leq x_1 \text{ e } y \leq n\}$ que aumenta hasta $\{x \leq x_1\}$, entonces

$$\lim_{n \rightarrow \infty} F_{X,Y}(x_1, n) = P[X \leq x_1] = F_X(x_1).$$

Para la propiedad (v), nótese en la figura 5.9(a) que $B = \{x_1 < x \leq x_2, y \leq y_1\} = \{X \leq x_2, Y \leq y_1\} - \{X \leq x_1, Y \leq y_1\}$, de modo que $P[B] = P[x_1 < x \leq x_2, y \leq y_1] = F_{X,Y}(x_2, y_1) - F_{X,Y}(x_1, y_1)$. En la figura 5.9(b), nótese que $F_{X,Y}(x_2, y_2) = P[A] + P[B] + F_{X,Y}(x_1, y_2)$. La propiedad (v) se obtiene resolviendo $P[A]$ y sustituyendo en la expresión para $P[B]$.

Ejemplo 5.10 Dibuja la cdf de X e Y del ejemplo 5.6. Calcula la cdf marginal de X .

Solución

Para calcular la cdf de \mathbf{X} , identificamos las regiones en el plano de acuerdo a qué puntos de $S_{X,Y}$ están incluidos en la región rectangular definida por (x, y) . Por ejemplo,

- Las regiones fuera del primer cuadrante no incluyen ninguno de los puntos, por lo que $F_{X,Y}(x, y) = 0$.
- La región $\{0 \leq x < 1, 0 \leq y < 1\}$ contiene el punto $(0, 0)$, por lo que $F_{X,Y}(x, y) = 1/4$.

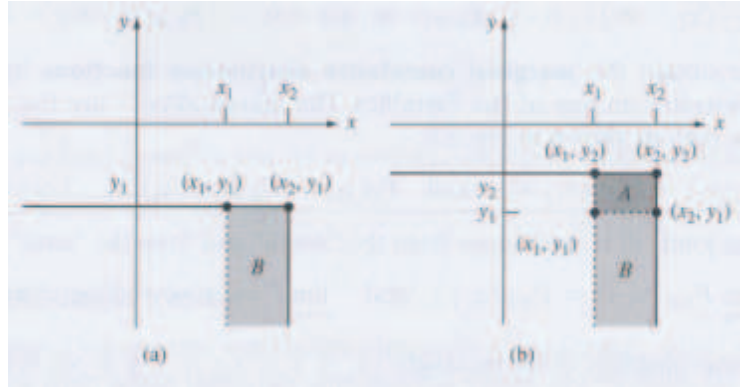


Figura 5.9 La cdf conjunta se puede usar para determinar la probabilidad de varios sucesos.

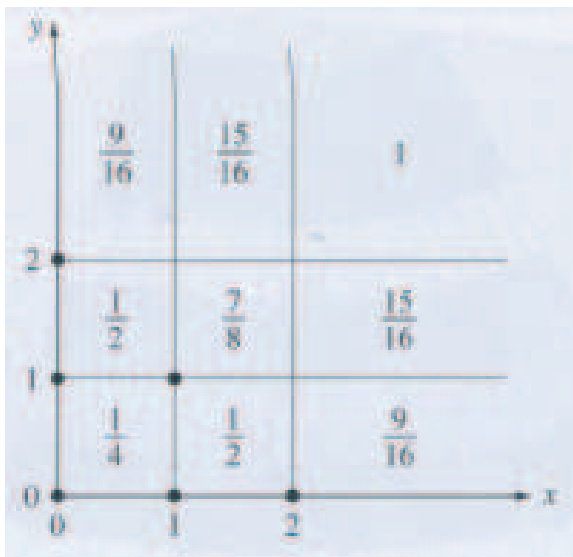


Figura 5.10 cdf conjunta del ejemplo del conmutador de paquetes.

La figura 5.10 muestra la cdf después de examinar todas las regiones posibles.

Tenemos que considerar varios casos para calcular $F_X(x)$. Para $x < 0$, se tiene $F_X(x) = 0$. Para $0 \leq x < 1$, tenemos $F_X(x) = F_{X,Y}(x, \infty) = 9/16$. Para $1 \leq x < 2$, tenemos $F_X(x) = F_{X,Y}(x, \infty) = 15/16$. Por último, para $x \geq 2$, se tiene $F_X(x) = F_{X,Y}(x, \infty) = 1$. Por lo tanto, $F_X(x)$ es una función definida a trozos y X es una variable aleatoria discreta con $p_X(0) = 9/16$, $p_X(1) = 6/16$ y $p_X(2) = 1/16$.

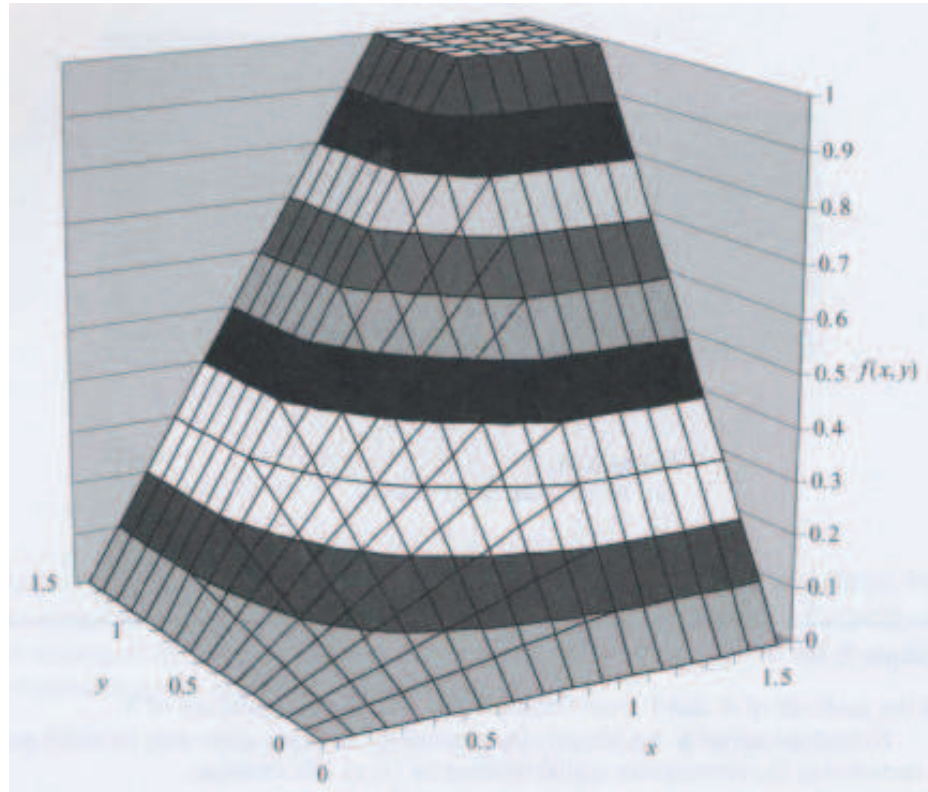


Figura 5.11 cdf conjunta de dos variables aleatorias uniformes.

Ejemplo 5.11 La cdf conjunta para el par de variables aleatorias $\mathbf{X} = (X, Y)$ viene dada por

$$F_{X,Y}(x, y) = \begin{cases} 0 & x < 0 \text{ o } y < 0 \\ xy & 0 \leq x \leq 1, 0 \leq y \leq 1 \\ x & 0 \leq x \leq 1, y > 1 \\ y & 0 \leq y \leq 1, x > 1 \\ 1 & x \geq 1, y \geq 1. \end{cases} \quad (5.10)$$

Dibuja la cdf conjunta y calcula la cdf marginal de X .

Solución

La figura 5.11 muestra una gráfica de la cdf conjunta de X e Y . $F_{X,Y}(x, y)$ es continua para todos los puntos del plano. $F_{X,Y}(x, y) = 1$ para todo $x \geq 1$ e $y \geq 1$, lo que implica que tanto X como Y asumen valores menores o iguales a uno.

La cdf marginal de X es:

$$F_X(x) = F_{X,Y}(x, \infty) = \begin{cases} 0 & x < 0 \\ x & 0 \leq x \leq 1 \\ 1 & x \geq 1. \end{cases}$$

X se distribuye uniformemente en el intervalo unitario.

Ejemplo 5.12 La cdf conjunta para el vector aleatorio $\mathbf{X} = (X, Y)$ viene dada por

$$F_{X,Y}(x, y) = \begin{cases} (1 - e^{-\alpha x})(1 - e^{-\beta y}) & x \geq 0, y \geq 0 \\ 0 & \text{en otro caso.} \end{cases}$$

Calcula las cdf marginales.

Solución

La cdf marginal se obtiene haciendo que una de las variables tienda a infinito:

$$\begin{aligned} F_X(x) &= \lim_{y \rightarrow \infty} F_{X,Y}(x, y) = 1 - e^{-\alpha x} \quad x \geq 0 \\ F_Y(y) &= \lim_{x \rightarrow \infty} F_{X,Y}(x, y) = 1 - e^{-\beta y} \quad y \geq 0. \end{aligned}$$

X e Y de forma individual tienen una distribución exponencial con parámetros α y β , respectivamente.

Ejemplo 5.13 Calcula la probabilidad de los sucesos $A = \{X \leq 1, Y \leq 1\}$, $B = \{X > x, Y > y\}$, donde $x > 0$ y $y > 0$, y $D = \{1 < X \leq 2, 2 < Y \leq 5\}$ en el ejemplo 5.12.

Solución

La probabilidad de A viene dada directamente por la cdf:

$$P[A] = P[X \leq 1, Y \leq 1] = F_{X,Y}(1, 1) = (1 - e^{-\alpha})(1 - e^{-\beta}).$$

La probabilidad de B requiere más trabajo. Por la ley de De Morgan:

$$B^c = (\{X > x\} \cap \{Y > y\})^c = \{X \leq x\} \cup \{Y \leq y\}.$$

El corolario 5 en la sección 2.2 proporciona la probabilidad de la unión de dos sucesos:

$$\begin{aligned} P[B^c] &= P[X \leq x] + P[Y \leq y] - P[X \leq x, Y \leq y] \\ &= (1 - e^{-\alpha x}) + (1 - e^{-\beta y}) - (1 - e^{-\alpha x})(1 - e^{-\beta y}) \\ &= 1 - e^{-\alpha x} e^{-\beta y}. \end{aligned}$$

Por último, calculamos la probabilidad de B :

$$P[B] = 1 - P[B^c] = e^{-\alpha x} e^{-\beta y}.$$

Un ejercicio interesante sería dibujar la región B en el plano e identificar los sucesos relacionados con el cálculo de la probabilidad de B^c .

La probabilidad del suceso D se calcula aplicando la propiedad (vi) de la cdf conjunta:

$$\begin{aligned} P[1 < X \leq 2, 2 < Y \leq 5] \\ &= F_{X,Y}(2, 5) - F_{X,Y}(2, 2) - F_{X,Y}(1, 5) + F_{X,Y}(1, 2) \\ &= (1 - e^{-2\alpha})(1 - e^{-5\beta}) - (1 - e^{-2\alpha})(1 - e^{-2\beta}) \\ &\quad - (1 - e^{-\alpha})(1 - e^{-5\beta}) + (1 - e^{-\alpha})(1 - e^{-2\beta}). \end{aligned}$$

5.3.1 Variables aleatorias de diferente tipo

En algunos problemas es necesario trabajar con pares de variables aleatorias que difieren en tipo, es decir, una es discreta y la otra es continua. Por lo general, es bastante penoso trabajar con la cdf conjunta y, por tanto, es preferible trabajar o bien con $P[X = k, Y \leq y]$ o con $P[X = k, y_1 < Y \leq y_2]$. Estas probabilidades son suficientes para calcular la cdf conjunta en caso de que tengamos que hacerlo.

Ejemplo 5.14 Canal de comunicación con entrada discreta y salida continua

La entrada X a un canal de comunicación toma los valores $+1$ voltio o -1 voltio con igual probabilidad. La salida Y del canal es la entrada más un ruido de voltaje N que se distribuye uniformemente en el intervalo $(-2 \text{ volt.}, +2 \text{ volt.})$. Calcula $P[X = +1, Y \leq 0]$.

Solución

Este problema se presta para el uso de la probabilidad condicionada:

$$P[X = +1, Y \leq y] = P[Y \leq y | X = +1]P[X = +1],$$

donde $P[X = +1] = 1/2$. Cuando la entrada $X = 1$, la salida Y se distribuye uniformemente en el intervalo $[-1, 3]$, por lo que

$$P[Y \leq y | X = +1] = \frac{y+1}{4} \text{ for } -1 \leq y \leq 3.$$

Entonces $P[X = +1, Y \leq 0] = P[Y \leq 0 | X = +1]P[X = +1] = (1/2)(1/4) = 1/8$.

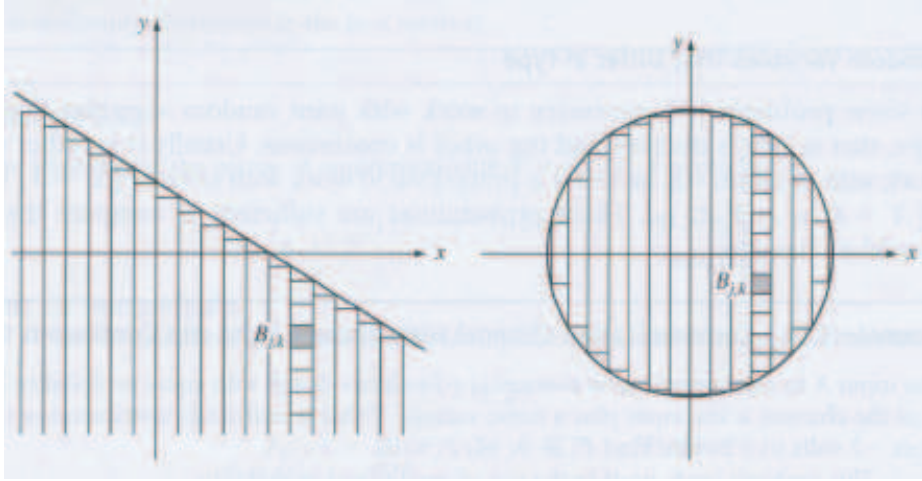


Figura 5.12 Sucesos bi-dimensionales que no atienden a la forma de producto.

5.4 La pdf conjunta de dos variables aleatorias continuas

La cdf conjunta nos permite calcular la probabilidad de sucesos que corresponden a formas “rectangulares” en el plano. Para calcular la probabilidad de sucesos asociados a regiones no rectangulares, aclaramos que cualquier forma razonable (i.e., discos, polígonos o semi-planos) puede ser aproximada por la unión de rectángulos infinitesimales disjuntos, $B_{j,k}$. Por ejemplo, la figura 5.12 muestra cómo los sucesos $A = \{X + Y \leq 1\}$ y $B = \{X^2 + Y^2 \leq 1\}$ son aproximados con rectángulos de ancho infinitesimal. La probabilidad de tales sucesos, por tanto, puede ser aproximada por la suma de las probabilidades de rectángulos infinitesimales y, si la cdf es suficientemente continua, la probabilidad de cada rectángulo se puede expresar en términos de una función de densidad:

$$P[B] \approx \sum_j \sum_k P[B_{j,k}] = \sum_{(x_j, y_k) \in B} f_{X,Y}(x_j, y_k) \Delta x \Delta y.$$

Cuando Δx y Δy tienden a cero, la ecuación anterior se convierte en la integral de una función de densidad en la región B .

Decimos que las variables aleatorias X e Y **son continuas conjuntamente** si las probabilidades de los sucesos relacionados con (X, Y) se pueden expresar como la integral de una función de densidad. En otras palabras, existe una función no negativa $f_{X,Y}(x, y)$, llamada la **función de densidad conjunta**, definida en el plano real, tal que para cada suceso B , subconjunto del plano,

$$P[\mathbf{X} \in B] = \int_B \int f_{X,Y}(x', y') dx' dy', \quad (5.11)$$

como se muestra en la figura 5.13. Nótese la similitud con la ecuación (5.5) para variables aleatorias discretas. Cuando B es el plano completo, la integral debe

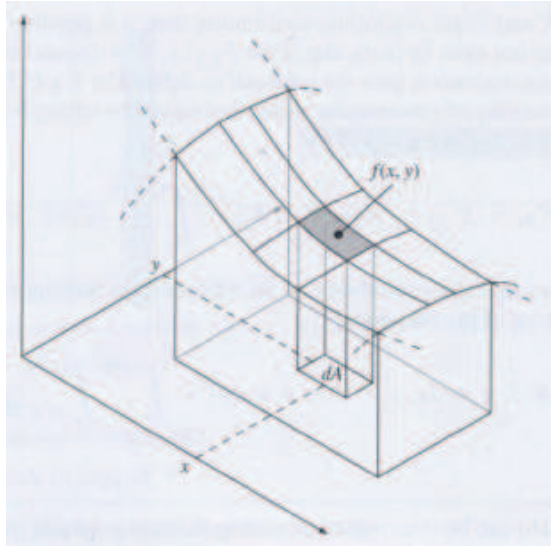


Figura 5.13 La probabilidad de A es la integral de $f_{X,Y}(x, y)$ en la región definida por A .

ser igual a uno:

$$1 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x', y') dx' dy'. \quad (5.12)$$

Las ecuaciones (5.11) y (5.12) de nuevo sugieren que la “masa” de probabilidad de un suceso se calcula integrando la densidad de masa de probabilidad sobre la región correspondiente al suceso.

La cdf conjunta se obtiene en términos de la pdf conjunta de variables aleatorias conjuntamente continuas integrando en el rectángulo semi-infinito definido por (x, y) :

$$F_{X,Y}(x, y) = \int_{-\infty}^x \int_{-\infty}^y f_{X,Y}(x', y') dx' dy'. \quad (5.13)$$

Se sigue entonces que si X e Y son variables aleatorias conjuntamente continuas, entonces la pdf se obtiene derivando la cdf:

$$f_{X,Y}(x, y) = \frac{\delta^2 F_{X,Y}(x, y)}{\delta x \delta y}. \quad (5.14)$$

Nótese que si X e Y no son conjuntamente continuas es posible que la derivada parcial anterior no exista. En particular, si $F_{X,Y}(x, y)$ es discontinua o sus derivadas parciales lo son, entonces la pdf conjunta definida por la ecuación (5.14) no existe.

La probabilidad de una región rectangular se obtiene haciendo $B = \{(x, y) :$

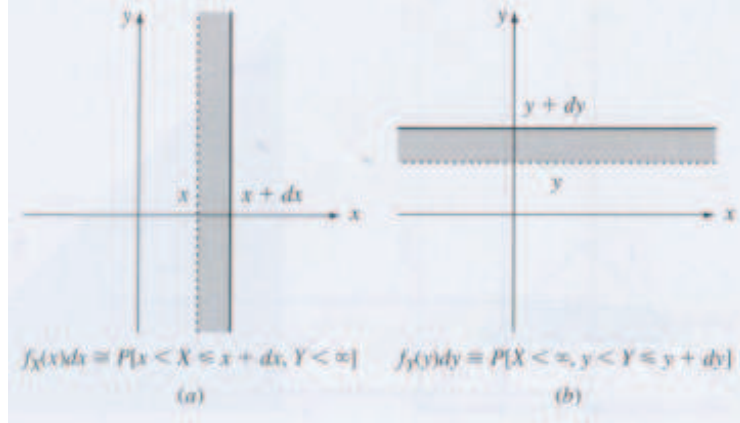


Figura 5.14 Interpretación de las pdf marginales.

$a_1 < x \leq b_1$ y $a_2 < y \leq b_2$ en la ecuación (5.11):

$$P[a_1 < X \leq b_1, a_2 < Y \leq b_2] = \int_{a_1}^{b_1} \int_{a_2}^{b_2} f_{X,Y}(x', y') dx' dy'. \quad (5.15)$$

Se sigue entonces que la probabilidad de un rectángulo infinitesimal es el producto de la pdf y el área del rectángulo:

$$\begin{aligned} P[x < X \leq x + dx, y < Y \leq y + dy] &= \int_x^{x+dx} \int_y^{y+dy} f_{X,Y}(x', y') dx' dy' \\ &\simeq f_{X,Y}(x, y) dx dy. \end{aligned} \quad (5.16)$$

La ecuación (5.16) se interpreta afirmando que la pdf conjunta especifica la probabilidad de sucesos en forma de producto

$$\{x < X \leq x + dx\} \cap \{y < Y \leq y + dy\}.$$

Las **pdf marginales** $f_X(x)$ y $f_Y(y)$ se obtienen tomando derivadas en las cdf marginales correspondientes, $F_X(x) = F_{X,Y}(x, \infty)$ y $F_Y(y) = F_{X,Y}(\infty, y)$. Entonces

$$\begin{aligned} f_X(x) &= \frac{d}{dx} \int_{-\infty}^x \left\{ \int_{-\infty}^{\infty} f_{X,Y}(x', y') dy' \right\} dx' \\ &= \int_{-\infty}^{\infty} f_{X,Y}(x, y') dy'. \end{aligned} \quad (5.17a)$$

De forma similar,

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x', y) dx'. \quad (5.17b)$$

Entonces, las pdf marginales se obtienen integrando las variables que no son de interés.

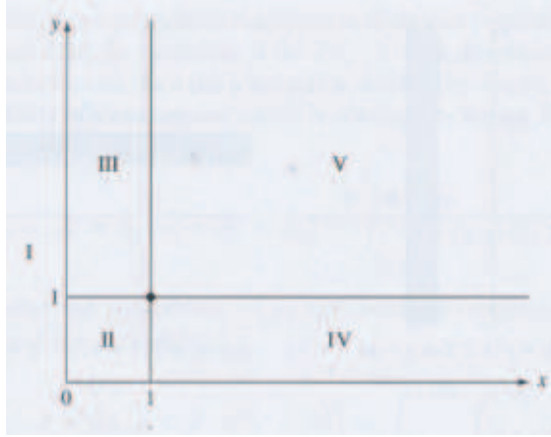


Figura 5.15 Regiones que deben ser consideradas de forma separada en el cálculo de la cdf del ejemplo 5.15.

Aclaremos que $f_X(x)dx \simeq P[x < X \leq x + dx, Y < \infty]$ es la probabilidad de la franja infinitesimal mostrada en la figura 5.14(a). Esto nos recuerda la interpretación de las pmf marginales como las probabilidades de columnas y filas en el caso de variables aleatorias discretas. No es sorprendente entonces que las ecuaciones (5.17a) y (5.17b) para las pdf marginales y las ecuaciones (5.7a) y (5.7b) para las pmf marginales sean idénticas excepto por el hecho de que una contenga una integral y la otra, un sumatorio. Como en el caso de las pmf, aclaramos que, en general, la pdf conjunta no se puede obtener de las pdf marginales.

Ejemplo 5.15 Variables aleatorias conjuntamente uniformes

Un punto (X, Y) seleccionado al azar en el cuadrado unidad tiene una pdf conjunta uniforme dada por

$$f_{X,Y}(x, y) = \begin{cases} 1 & 0 \leq x \leq 1 \text{ y } 0 \leq y \leq 1 \\ 0 & \text{en otro caso.} \end{cases}$$

El diagrama de dispersión en la figura 5.3(a) corresponde a este par de variables aleatorias. Calcula la cdf conjunta de X e Y .

Solución

La cdf se calcula evaluando la ecuación (5.13). Debemos ser cuidadosos con los límites de la integral: los límites deben definir la intersección del rectángulo semi-infinito definido por (x, y) y la región donde la pdf es no negativa. Hay cinco casos en este problema que se corresponden con las cinco regiones mostradas en la figura 5.15.

1. Si $x < 0$ o $y < 0$, la pdf es cero y la ecuación (5.14) implica

$$F_{X,Y}(x, y) = 0.$$

2. Si (x, y) está dentro del intervalo unidad,

$$F_{X,Y}(x, y) = \int_0^x \int_0^y 1 dx' dy' = xy.$$

3. Si $0 \leq x \leq 1$ y $y > 1$,

$$F_{X,Y}(x, y) = \int_0^x \int_0^1 1 dx' dy' = x.$$

4. De manera similar, si $x > 1$ y $0 \leq y \leq 1$,

$$F_{X,Y}(x, y) = y.$$

5. Por último, si $x > 1$ y $y > 1$,

$$F_{X,Y}(x, y) = \int_0^1 \int_0^1 1 dx' dy' = 1.$$

Vemos que ésta es la cdf conjunta del ejemplo 5.11.

Ejemplo 5.16 Calcula la constante de normalización c y las pdf marginales para la siguiente pdf conjunta:

$$f_{X,Y}(x, y) = \begin{cases} ce^{-x}e^{-y} & 0 \leq y \leq x < \infty \\ 0 & \text{en otro caso.} \end{cases}$$

Solución

La pdf es no negativa en la región sombreada de la figura 5.16(a). La constante c se calcula con la condición de normalización especificada en la ecuación (5.12):

$$1 = \int_0^{\infty} \int_0^x ce^{-x}e^{-y} dy dx = \int_0^{\infty} ce^{-x}(1 - e^{-x}) dx = \frac{c}{2}.$$

Entonces $c = 2$. Las pdf marginales se calculan evaluando las ecuaciones (5.17a) y (5.17b):

$$f_X(x) = \int_0^{\infty} f_{X,Y}(x, y) dy = \int_0^x 2e^{-x}e^{-y} dy = 2e^{-x}(1 - e^{-x}) \quad 0 \leq x < \infty$$

y

$$f_Y(y) = \int_0^{\infty} f_{X,Y}(x, y) dx = \int_y^{\infty} 2e^{-x}e^{-y} dx = 2e^{-2y} \quad 0 \leq y < \infty.$$

Queda como ejercicio completar los pasos en la evaluación de las integrales así como verificar que las pdf marginales integran a 1.

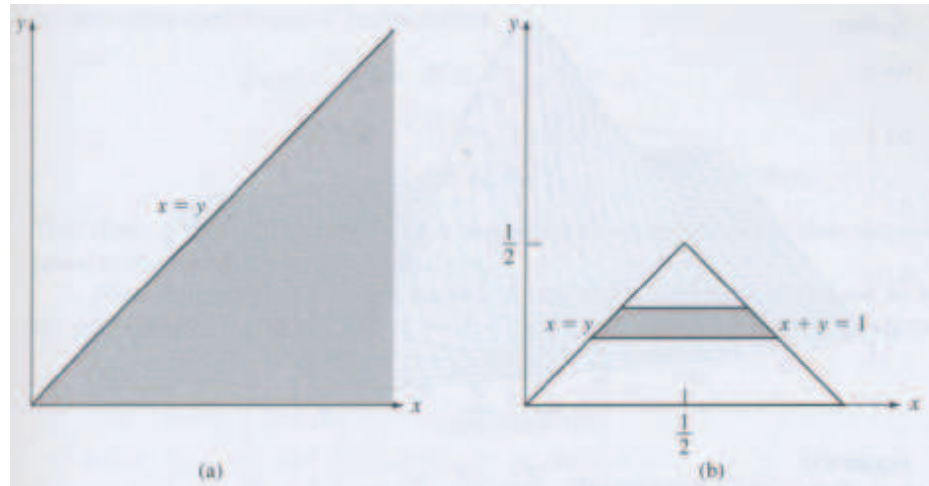


Figura 5.16 Las variables aleatorias X e Y de los ejemplos 5.16 y 5.17 tienen una pdf que es no negativa solo en la región sombreada mostrada en la parte (a).

Ejemplo 5.17 Calcula $P[X + Y \leq 1]$ en el ejemplo 5.16.

Solución

La figura 5.16(b) muestra la intersección del suceso $\{X + Y \leq 1\}$ y la región donde la pdf es distinta de cero. Se obtiene la probabilidad del suceso “sumando” (en realidad integrando) rectángulos infinitesimales de anchura dy , como se indica en la figura:

$$\begin{aligned} P[X + Y \leq 1] &= \int_0^{0.5} \int_y^{1-y} 2e^{-x}e^{-y} dx dy = \int_0^{0.5} 2e^{-y}[e^{-y} - e^{-(1-y)}] dy \\ &= 1 - 2e^{-1}. \end{aligned}$$

Ejemplo 5.18 Variables aleatorias conjuntamente normales La pdf conjunta de X e Y , que se muestra en la figura 5.17, es

$$f_{X,Y}(x, y) = \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-(x^2-2\rho xy+y^2)/2(1-\rho^2)} \quad -\infty < x, y < \infty. \quad (5.18)$$

Decimos que X e Y son conjuntamente normales*. Calcula las pdf marginales.

Solución

La pdf marginal de X se calcula integrando $f_{X,Y}(x, y)$ en y :

$$f_X(x) = \frac{e^{x^2/2(1-\rho^2)}}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{\infty} e^{-(y^2-2\rho xy)/2(1-\rho^2)} dy.$$

Completamos el cuadrado del argumento de la exponencial sumando y restando

* Se trata de un importante caso especial de variables aleatorias conjuntamente normales. El caso general se analiza en la Sección 5.9.

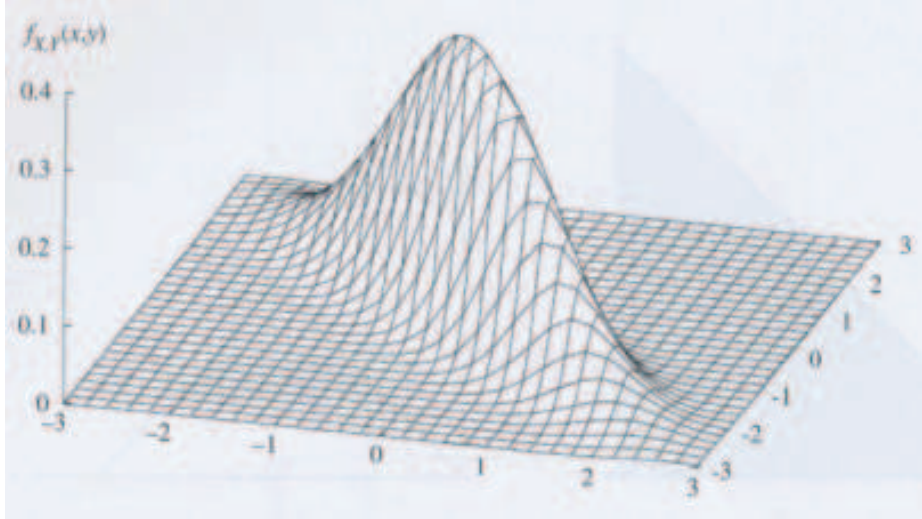


Figura 5.17 pdf conjunta de dos variables aleatorias normales.

$\rho^2 x^2$, esto es, $y^2 - 2\rho xy + \rho^2 x^2 - \rho^2 x^2 = (y - \rho x)^2 - \rho^2 x^2$. Entonces

$$\begin{aligned} f_X(x) &= \frac{e^{-x^2/2(1-\rho^2)}}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{\infty} e^{-[(y-\rho x)^2 - \rho^2 x^2]/2(1-\rho^2)} dy \\ &= \frac{e^{-x^2/2}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{e^{-(y-\rho x)^2/2(1-\rho^2)}}{\sqrt{2\pi(1-\rho^2)}} dy \\ &= \frac{e^{-x^2/2}}{\sqrt{2\pi}}, \end{aligned}$$

donde hemos utilizado el hecho de que la última integral es igual a uno, ya que su integrando es una pdf normal de media ρx y varianza $1 - \rho^2$. La pdf marginal de X es por tanto una pdf normal unidimensional con media 0 y varianza 1. Como $f_{X,Y}(x, y)$ es simétrica en x e y , podemos concluir que la pdf marginal de Y también es una pdf normal de una dimensión con media cero y varianza uno.

5.5 Independencia de dos variables aleatorias

X e Y son variables aleatorias independientes si *cualquier* suceso A_1 definido en términos de X es independiente de *cualquier* suceso A_2 definido en términos de Y , esto es,

$$P[X \in A_1, Y \in A_2] = P[X \in A_1]P[Y \in A_2]. \quad (5.19)$$

En esta sección presentamos un conjunto de condiciones para determinar cuándo X e Y son independientes.

Supongamos que X e Y son un par de variables aleatorias discretas y supongamos que queremos calcular la probabilidad del suceso $A = A_1 \cap A_2$, donde A_1 está relacionado con X y A_2 se refiere solo a Y . En particular, si X e Y son independientes, entonces A_1 y A_2 son sucesos independientes. Si definimos $A_1 = \{X = x_j\}$ y $A_2 = \{Y = y_k\}$, entonces la independencia de X e Y implica que

$$\begin{aligned} p_{X,Y}(x_j, y_k) &= P[X = x_j, Y = y_k] \\ &= P[X = x_j]P[Y = y_k] \\ &= p_X(x_j)p_Y(y_k) \quad \forall x_j, y_k. \end{aligned} \quad (5.20)$$

Por tanto, si X e Y son variables aleatorias discretas independientes, entonces la pmf conjunta es igual al producto de las pmf marginales.

Ahora supongamos que no sabemos si X e Y son independientes, pero que sí sabemos que la pmf satisface la ecuación (5.20). Sea $A = A_1 \cap A_2$ un suceso en forma de producto como el anterior, entonces

$$\begin{aligned} P[A] &= \sum_{x_j \in A_1} \sum_{y_k \in A_2} p_{X,Y}(x_j, y_k) \\ &= \sum_{x_j \in A_1} \sum_{y_k \in A_2} p_X(x_j)p_Y(y_k) \\ &= \sum_{x_j \in A_1} p_X(x_j) \sum_{y_k \in A_2} p_Y(y_k) \\ &= P[A_1]P[A_2], \end{aligned} \quad (5.21)$$

que implica que A_1 y A_2 son sucesos independientes. Por tanto, si la pmf conjunta de X e Y es igual al producto de las pmf marginales, entonces X e Y son independientes. Hemos demostrado que la afirmación “ X e Y son independientes” es equivalente a la afirmación “la pmf conjunta es igual al producto de las pmf marginales”. En lenguaje matemático decimos que “las variables aleatorias discretas X e Y son independientes si y solo si la pmf conjunta es el producto de las pmf marginales para todo x_j, y_k ”.

Ejemplo 5.19 ¿Es la pmf del ejemplo 5.6 consistente con un experimento que consiste en tiradas independientes de dos dados equilibrados?

Solución

La probabilidad de cada cara en el resultado de una tirada es $1/6$. Si tiramos dos dados equilibrados y las tiradas son independientes, la probabilidad de cualquier par de caras, j y k , es:

$$P[X = j, Y = k] = P[X = j]P[Y = k] = \frac{1}{36}.$$

Entonces, todos los pares de resultados posibles deberían ser equiprobables. Este

no es el caso de la pmf conjunta dada en el ejemplo 5.6. Entonces, las tiradas del ejemplo 5.6 no son independientes.

Ejemplo 5.20 ¿Son independientes las variables Q y R del ejemplo 5.9?

Solución

En el ejemplo 5.9 tenemos

$$\begin{aligned} P[Q = q]P[R = r] &= (1 - p^M)(p^M)^q \frac{(1 - p)}{1 - p^M} p^r \\ &= (1 - p)p^{Mq+r} \\ &= P[Q = q, R = r] \quad \forall q = 0, 1, \dots; r = 0, \dots, M - 1. \end{aligned}$$

Entonces, Q y R son independientes.

En general, se puede demostrar que las variables aleatorias X e Y son independientes si y solo si su cdf conjunta es igual al producto de sus cdf marginales:

$$F_{X,Y}(x, y) = F_X(x)F_Y(y) \quad \forall x, y. \quad (5.22)$$

De forma similar, si X e Y son conjuntamente continuas, entonces, X e Y son independientes si y solo si su pdf conjunta es igual al producto de sus pdf marginales:

$$f_{X,Y}(x, y) = f_X(x)f_Y(y) \quad \forall x, y. \quad (5.23)$$

La ecuación (5.23) se obtiene de la ecuación (5.22) aplicando la derivada. Por otro lado, la ecuación (5.22) se obtiene de la ecuación (5.23) calculando la integral.

Ejemplo 5.21 ¿Son independientes las variables aleatorias X e Y del ejemplo 5.16?

Solución

Nótese que $f_X(x)$ y $f_Y(y)$ son diferentes de cero para todo $x > 0$ y todo $y > 0$. Por lo tanto, $f_X(x)f_Y(y)$ es distinto de cero en todo el cuadrante positivo. Sin embargo, $f_{X,Y}(x, y)$ es distinto de cero sólo en la región $y < x$ dentro del cuadrante positivo. Por lo tanto la ecuación (5.23) no se verifica para todo x, y y las variables aleatorias no son independientes. Debe destacarse que en este ejemplo la pdf conjunta parece factorizarse, pero sin embargo, no es el producto de las pdf marginales.

Ejemplo 5.22 ¿Son independientes las variables aleatorias X e Y en el ejemplo 5.18?

Solución

El producto de las pdf marginales de X e Y en el ejemplo 5.18 es

$$f_X(x)f_Y(y) = \frac{1}{2\pi}e^{-(x^2+y^2)/2} \quad -\infty < x, y < \infty.$$

Comparandolo con la ecuación (5.18) vemos que el producto de las marginales es igual a la pdf conjunta, si y sólo si $\rho = 0$. Por tanto, las variables aleatorias normales conjuntas X e Y son independientes si y sólo si $\rho = 0$. Veremos en una sección posterior que ρ es el *coeficiente de correlación* entre X e Y .

Ejemplo 5.23 ¿Son independientes las variables aleatorias X e Y del ejemplo 5.12?

Solución

Si multiplicamos las cdf marginales calculadas en el ejemplo 5.12, tenemos

$$F_X(x)F_Y(y) = (1 - e^{-\alpha x})(1 - e^{-\beta y}) = F_{X,Y}(x, y) \quad \forall x, y.$$

Por lo tanto, la ecuación (5.22) se cumple y se sigue que X e Y son independientes.

Si X e Y son variables aleatorias independientes, entonces las variables aleatorias definidas por cualquier par de funciones $g(X)$ y $h(Y)$ también son independientes. Para demostrar esto, consideremos los sucesos unidimensionales A y B . Sea A' el conjunto de todos los valores de x tal que si x pertenece a A' , entonces $g(x)$ pertenece a A , y sea B' el conjunto de todos los valores de y tal que si y pertenece a B' , entonces $h(y)$ pertenece a B . (En el capítulo 3 denominamos a A' y B' los sucesos equivalentes de A y B .) A continuación,

$$\begin{aligned} P[g(X) \in A, h(Y) \in B] &= P[X \in A', Y \in B'] \\ &= P[X \in A']P[Y \in B'] \\ &= P[g(X) \in A]P[h(Y) \in B]. \end{aligned} \quad (5.24)$$

La primera y la tercera igualdad se derivan del hecho de que A y A' y B y B' son sucesos equivalentes. La segunda igualdad se sigue de la independencia de X e Y . Por lo tanto $g(X)$ y $h(Y)$ son variables aleatorias independientes.

5.6 Momentos conjuntos y esperanza de una función de dos variables aleatorias

La esperanza de X representa el centro de masa de la distribución de X . La varianza, que se define como la esperanza de $(X - m)^2$, proporciona una medida de la dispersión de la distribución. En el caso de dos variables aleatorias nos interesa conocer cómo X e Y varían conjuntamente. Concretamente, estamos interesados en saber si X e Y están correladas. Por ejemplo, si X crece ¿tiende

Y a crecer o a decrecer? Los momentos conjuntos de X e Y , que se definen como las esperanzas de funciones de X e Y , proporcionan esta información.

5.6.1 Esperanza de una función de dos variables aleatorias

El problema de calcular la esperanza de una función de dos variables aleatorias es parecido al de calcular la esperanza de una función de una variable aleatoria. Se puede demostrar que la esperanza de $Z = g(X, Y)$ se calcula utilizando las siguientes expresiones:

$$E[Z] = \begin{cases} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f_{X,Y}(x, y) dx dy & X, Y \text{ conjuntamente continuas} \\ \sum_i \sum_n g(x_i, y_n) p_{X,Y}(x_i, y_n) & X, Y \text{ discretas.} \end{cases} \quad (5.25)$$

Ejemplo 5.24 Suma de variables aleatorias

Sea $Z = X + Y$. Calcula $E[Z]$.

Solución

$$\begin{aligned} E[Z] &= E[X + Y] \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x' + y') f_{X,Y}(x', y') dx' dy' \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x' f_{X,Y}(x', y') dy' dx' + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y' f_{X,Y}(x', y') dx' dy' \\ &= \int_{-\infty}^{\infty} x' f_X(x') dx' + \int_{-\infty}^{\infty} y' f_Y(y') dy' = E[X] + E[Y]. \end{aligned} \quad (5.26)$$

Entonces, la esperanza de la suma de dos variables aleatorias es igual a la suma de las esperanzas individuales. Nótese que X e Y no necesitan ser independientes.

El resultado del ejemplo 5.24 junto con un argumento de inducción demuestra que *la esperanza de la suma de n variables aleatorias es la suma de las esperanzas*:

$$E[X_1 + X_2 + \cdots + X_n] = E[X_1] + \cdots + E[X_n]. \quad (5.27)$$

Nótese que las variables aleatorias no tienen que ser independientes.

Ejemplo 5.25 Producto de funciones de variables aleatorias independientes

Supongamos que X e Y son variables aleatorias independientes y sea $g(X, Y) = g_1(X)g_2(Y)$. Calcula $E[g(X, Y)] = E[g_1(X)g_2(Y)]$.

Solución

$$\begin{aligned} E[g_1(X)g_2(Y)] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g_1(x')g_2(y')f_X(x')f_Y(y')dx'dy' \\ &= \left\{ \int_{-\infty}^{\infty} g_1(x')f_X(x')dx' \right\} \left\{ \int_{-\infty}^{\infty} g_2(y')f_Y(y')dy' \right\} \\ &= E[g_1(X)]E[g_2(Y)]. \end{aligned}$$

5.6.2 Momentos conjuntos, correlación y covarianza

Los momentos conjuntos de dos variables aleatorias X e Y resumen la información sobre su comportamiento conjunto. El **momento conjunto** jk -ésimo de \mathbf{X} e \mathbf{Y} se define como

$$E[X^j Y^k] = \begin{cases} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^j y^k f_{X,Y}(x,y) dx dy & X, Y \text{ conjuntamente continuas} \\ \sum_i \sum_n x_i^j y_n^k p_{X,Y}(x_i, y_n) & X, Y \text{ discretas.} \end{cases} \quad (5.28)$$

Si $j = 0$, obtenemos los momentos de Y y si $k = 0$, obtendremos los momentos de X . En ingeniería, se acostumbra a llamar al momento con $j = 1$ y $k = 1$, $E[XY]$, la **correlación de \mathbf{X} e \mathbf{Y}** . Si $E[XY] = 0$, decimos que \mathbf{X} e \mathbf{Y} son **ortogonales**.

El **momento central** jk -ésimo de \mathbf{X} e \mathbf{Y} se define como el momento conjunto de las variables aleatorias centradas, $X - E[X]$ e $Y - E[Y]$:

$$E[(X - E[X])^j (Y - E[Y])^k].$$

Nótese que $j = 2$ y $k = 0$ dan $V[X]$ y que $j = 0$ y $k = 2$, $V[Y]$.

La **covarianza de \mathbf{X} e \mathbf{Y}** se define como el momento central con $j = k = 1$:

$$Cov(X, Y) = E[(X - E[X])(Y - E[Y])]. \quad (5.29)$$

A menudo resulta más conveniente trabajar con la siguiente forma para la $Cov(X, Y)$:

$$\begin{aligned} Cov(X, Y) &= E[XY - XE[Y] - YE[X] + E[X]E[Y]] \\ &= E[XY] - 2E[X]E[Y] + E[X]E[Y] \\ &= E[XY] - E[X]E[Y]. \end{aligned} \quad (5.30)$$

Nótese que $Cov(X, Y) = E[XY]$ si alguna de las variables aleatorias tiene media cero.

Ejemplo 5.26 Covarianza de variables aleatorias independientes

Sean X e Y variables aleatorias independientes. Calcula su covarianza.

Solución

$$\begin{aligned} \text{Cov}(X, Y) &= E[(X - E[X])(Y - E[Y])] \\ &= E[X - E[X]]E[Y - E[Y]] \\ &= 0, \end{aligned}$$

donde la segunda igualdad se sigue del hecho que X e Y son independientes y la tercera de que $E[X - E[X]] = E[X] - E[X] = 0$. Entonces *un par de variables aleatorias independientes tienen covarianza cero*.

Vamos a ver cómo la covarianza mide la correlación entre X e Y . La covarianza mide la desviación con respecto a $m_X = E[X]$ y $m_Y = E[Y]$. Si un valor positivo de $(X - m_X)$ tiende a ir acompañado de un valor positivo de $(Y - m_Y)$ y $(X - m_X)$ negativos tienden a ser acompañados por $(Y - m_Y)$, entonces $(X - m_X)(Y - m_Y)$ tenderá a ser un valor positivo y su valor esperado, $\text{Cov}(X, Y)$, será positivo. Este es el caso del diagrama de dispersión de la figura 5.3(d) donde los puntos observados tienden a agruparse a lo largo de una recta con pendiente positiva. Por otro lado, si $(X - m_X)$ y $(Y - m_Y)$ tienden a tener signos opuestos, entonces, la $\text{Cov}(X, Y)$ será negativa. Un diagrama de dispersión para este caso tendría la nube de puntos a lo largo de una línea de pendiente negativa. Por último, si $(X - m_X)$ y $(Y - m_Y)$ tienen el mismo signo en algunas ocasiones y en otras tienen signos opuestos, entonces la $\text{Cov}(X, Y)$ se acercará a cero. Los tres diagramas de dispersión en las figuras 5.3(a), (b) y (c) entran en esta categoría.

Multiplicando X o Y por un número alto se incrementará la covarianza, por lo que encontramos la necesidad de normalizar la covarianza para medir la correlación en una escala absoluta. El **coeficiente de correlación de X e Y** se define por

$$\rho_{X,Y} = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{E[XY] - E[X]E[Y]}{\sigma_X \sigma_Y}, \quad (5.31)$$

donde $\sigma_X = \sqrt{V(X)}$ y $\sigma_Y = \sqrt{V(Y)}$ son las desviaciones típicas de X e Y , respectivamente.

El coeficiente de correlación es un número cuya magnitud es a lo sumo uno:

$$-1 \leq \rho_{X,Y} \leq 1. \quad (5.32)$$

Para demostrar la ecuación (5.32), partimos de una desigualdad que resulta del

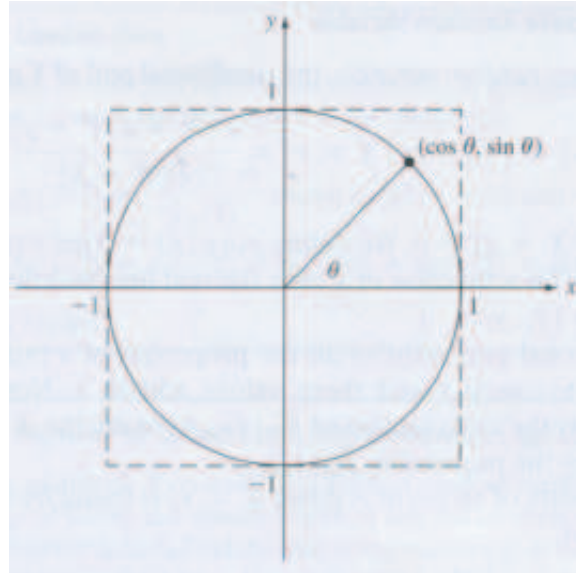


Figura 5.18 (X, Y) es un punto seleccionado de forma aleatoria en la circunferencia de radio 1. X e Y están incorreladas pero no son independientes.

hecho de que la esperanza del cuadrado de una variable aleatoria es no negativa:

$$\begin{aligned} 0 &\leq E \left\{ \left(\frac{X - E[X]}{\sigma_X} \pm \frac{Y - E[Y]}{\sigma_Y} \right)^2 \right\} \\ &= 1 \pm 2\rho_{X,Y} + 1 \\ &= 2(1 \pm \rho_{X,Y}). \end{aligned}$$

La última ecuación implica la ecuación (5.32).

Los valores extremos de $\rho_{X,Y}$ se obtienen cuando X e Y están linealmente relacionadas, $Y = aX + b$; $\rho_{X,Y} = 1$ si $a > 0$ y $\rho_{X,Y} = -1$ si $a < 0$.

Se dice que X e Y están **incorreladas** si $\rho_{X,Y} = 0$. Si X e Y son independientes, la $Cov(X, Y) = 0$ y, por tanto, $\rho_{X,Y} = 0$. Entonces, *si X e Y son variables aleatorias independientes, X e Y están incorreladas*. En el ejemplo 5.22, vimos que si X e Y son normales conjuntas y $\rho_{X,Y} = 0$, entonces X e Y son variables aleatorias independientes. El ejemplo 5.27 muestra que esto no siempre es cierto para variables aleatorias no normales: es posible que X e Y estén incorreladas pero que no sean independientes.

Ejemplo 5.27 Variables aleatorias dependientes no correladas

Sea Θ uniformemente distribuida en el intervalo $(0, 2\pi)$. Sean

$$X = \cos(\Theta) \quad \text{y} \quad Y = \sin(\Theta).$$

El punto (X, Y) se corresponde con un punto en la circunferencia de radio 1 especificado por el ángulo Θ , como se muestra en la figura 5.18. En el ejemplo 4.33 vimos que las pdf marginales de X e Y son pdf arco-seno, que es no nula

en el intervalo $(-1, 1)$. El producto de las marginales es no nulo en el cuadrado definido por $-1 \leq x \leq 1$ y $-1 \leq y \leq 1$, por tanto, si X e Y fueran independientes el punto (X, Y) asumiría todos los valores en este cuadrado. Este no es el caso, por tanto, X e Y son dependientes.

Ahora demostramos que X e Y están incorreladas:

$$\begin{aligned} E[XY] &= E[\sin \Theta \cos \Theta] = \frac{1}{2\pi} \int_0^{2\pi} \sin \phi \cos \phi d\phi \\ &= \frac{1}{4\pi} \int_0^{2\pi} \sin 2\phi d\phi = 0. \end{aligned}$$

Como $E[X] = E[Y] = 0$, la ecuación (5.30) implica entonces que X e Y están incorreladas.

Ejemplo 5.28 Sean X e Y las variables aleatorias del ejemplo 5.16. Calcula $E[XY]$, $Cov(X, Y)$ y $\rho_{X, Y}$.

Solución

Las ecuaciones (5.30) y (5.31) requieren que calculemos la media, la varianza y la correlación de X e Y . Utilizando las pdf marginales de X e Y calculadas en el ejemplo 5.16, llegamos a $E[X] = 3/2$ y $V[X] = 5/4$, y $E[Y] = 1/2$ y $V[Y] = 1/4$. La correlación de X e Y es

$$\begin{aligned} E[XY] &= \int_0^{\infty} \int_0^x xy 2e^{-x} e^{-y} dy dx \\ &= \int_0^{\infty} 2x 2e^{-x} (1 - e^{-x} - xe^{-x}) dx = 1. \end{aligned}$$

Entonces el coeficiente de correlación viene dado por

$$\rho_{X, Y} = \frac{1 - \frac{3}{2} \frac{1}{2}}{\sqrt{\frac{5}{4}} \sqrt{\frac{1}{4}}} = \frac{1}{\sqrt{5}}.$$

5.7 Probabilidad condicionada y esperanza condicionada

Muchas variables aleatorias de interés práctico no son independientes: la salida Y de un canal de comunicación debe depender de la entrada X con el fin de transmitir la información; es probable que las muestras consecutivas de una señal que varía lentamente estén cerca de un valor y, por lo tanto, no son independientes. En esta sección nos interesa calcular la probabilidad de sucesos relacionados con la variable aleatoria Y , dado que sabemos que $X = x$. También nos interesa la esperanza de Y dado que $X = x$. Se muestra que las nociones de probabilidad condicionada y de esperanza condicionada son herramientas muy útiles para resolver problemas, incluso en situaciones en las que sólo nos ocupa una de las variables aleatorias.

5.7.1 Probabilidad condicionada

La definición de probabilidad condicionada de la sección 2.3 permite calcular la probabilidad de que Y pertenezca a A dado que sabemos que $X = x$:

$$P[Y \in A | X = x] = \frac{P[Y \in A, X = x]}{P[X = x]} \text{ para } P[X = x] > 0. \quad (5.33)$$

Caso 1: X es una variable aleatoria discreta

Para las variables aleatorias discretas X e Y , la **pmf condicionada de Y dado que $X = x$** se define como:

$$p_Y(y|x) = P[Y = y | X = x] = \frac{P[X = x, Y = y]}{P[X = x]} = \frac{p_{X,Y}(x, y)}{p_X(x)} \quad (5.34)$$

para x tal que $P[X = x] > 0$. Definimos $p_Y(y|x) = 0$ para x tal que $P[X = x] = 0$. Nótese que $p_Y(y|x)$ es una función de y en la recta real y que $p_Y(y|x) > 0$ sólo para y en un conjunto discreto $\{y_1, y_2, \dots\}$.

La pmf condicional satisface todas las propiedades de una pmf, esto es, asigna valores no negativos a cada y y estos valores suman 1. Véase que de la ecuación (5.34) se deduce que $p_Y(y|x_k)$ es simplemente la sección transversal de $p_{X,Y}(x_k, y)$ a lo largo de la columna $X = x_k$ representada en la figura 5.6, normalizada por la probabilidad $p_X(x_k)$.

La probabilidad de un suceso A dado que $X = x_k$ se calcula sumando los valores de la pmf en los elementos de A :

$$P[Y \in A | X = x_k] = \sum_{y_j \in A} p_Y(y_j | x_k). \quad (5.35)$$

Si X e Y son independientes, entonces utilizando la ecuación (5.20)

$$p_Y(y_j | x_k) = \frac{P[X = x_k, Y = y_j]}{P[X = x_k]} = P[Y = y_j] = p_Y(y_j). \quad (5.36)$$

En otras palabras, saber que $X = x_k$ no afecta a la probabilidad de sucesos A relacionados con Y .

La ecuación (5.34) implica que la pmf conjunta $p_{X,Y}(x, y)$ se puede expresar como el producto de una pmf condicional y una pmf marginal:

$$p_{X,Y}(x_k, y_j) = p_Y(y_j | x_k) p_X(x_k) \text{ y } p_{X,Y}(x_k, y_j) = p_X(x_k | y_j) p_Y(y_j). \quad (5.37)$$

Esta expresión es muy útil ya que podemos ver al par (X, Y) como generado de forma secuencial, por ejemplo, primero X y luego Y dado que $X = x$. Calculamos

la probabilidad de $Y \in A$ como sigue:

$$\begin{aligned}
 P[Y \in A] &= \sum_{x_k, y_j} \sum_{\in A} p_{X,Y}(x_k, y_j) \\
 &= \sum_{x_k, y_j} \sum_{\in A} p_Y(y_j | x_k) p_X(x_k) \\
 &= \sum_{x_k} p_X(x_k) \sum_{y_j \in A} p_Y(y_j | x_k) \\
 &= \sum_{x_k} P[Y \in A | X = x_k] p_X(x_k). \tag{5.38}
 \end{aligned}$$

La ecuación (5.38) es simplemente una reformulación del teorema de la probabilidad total discutido en el capítulo 2. En otras palabras, para calcular $P[Y \in A]$ podemos calcular primero $P[Y \in A | X = x_k]$ y luego la “media” con respecto a X_k .

Ejemplo 5.29 Dado no equilibrado

Calcula $p_Y(y|5)$ en el dado “trucado” considerado en los ejemplos 5.6 y 5.8.

Solución

En el ejemplo 5.8 calculamos que $p_X(5) = 1/6$. Entonces:

$$\begin{aligned}
 p_Y(y|5) &= \frac{p_{X,Y}(5, y)}{p_X(5)} \text{ y entonces } p_Y(5|5) = 2/7 \text{ y} \\
 p_Y(1|5) &= p_Y(2|5) = p_Y(3|5) = p_Y(4|5) = p_Y(6|5) = 1/7.
 \end{aligned}$$

Claramente, este dado no es equilibrado.

Ejemplo 5.30 Número de defectos en una región; generación aleatoria de un conteo Poisson

El número total de defectos X en un chip es una variable aleatoria de Poisson con media α . Cada defecto tiene una probabilidad p de caer en una región específica R y el emplazamiento de cada defecto es independiente del de los otros defectos. Calcula la pmf del número de defectos Y que caen en la región R .

Solución

Imaginemos que llevamos a cabo un experimento de Bernoulli cada vez que un defecto ocurre, definiendo que ocurre un “éxito” cuando el defecto cae en la región R . Si el número total de defectos es $X = k$, entonces Y es una variable aleatoria Binomial con parámetros k y p :

$$p_Y(j|k) = \begin{cases} 0 & j > k \\ \binom{k}{j} p^j (1-p)^{k-j} & 0 \leq j \leq k. \end{cases}$$

De la ecuación (5.38) y teniendo en cuenta que $k \geq j$, se tiene que

$$\begin{aligned} p_Y(j) &= \sum_{k=0}^{\infty} p_Y(j|k)p_X(k) = \sum_{k=j}^{\infty} \frac{k!}{j!(k-j)!} p^j (1-p)^{k-j} \frac{\alpha^k}{k!} e^{-\alpha} \\ &= \frac{(\alpha p)^j e^{-\alpha}}{j!} \sum_{k=j}^{\infty} \frac{\{(1-p)\alpha\}^{k-j}}{(k-j)!} \\ &= \frac{(\alpha p)^j e^{-\alpha}}{j!} e^{(1-p)\alpha} = \frac{(\alpha p)^j}{j!} e^{-\alpha p}. \end{aligned}$$

Entonces Y es una variable aleatoria de Poisson con media αp .

Supongamos que Y es una variable aleatoria continua. La ecuación (5.33) se puede utilizar para definir la **cdf condicionada de Y dado que $X = x_k$** :

$$F_Y(y|x_k) = \frac{P[Y \leq y, X = x_k]}{P[X = x_k]}, \quad \text{para } P[X = x_k] > 0. \quad (5.39)$$

Resulta fácil demostrar que $F_Y(y|x_k)$ satisface todas las propiedades de una cdf. La **pdf condicionada de Y dado que $X = x_k$** , si la derivada existe, viene dada por

$$f_Y(y|x_k) = \frac{d}{dy} F_Y(y|x_k). \quad (5.40)$$

Si X e Y son independientes, $P[Y \leq y, X = x_k] = P[Y \leq y]P[X = x_k]$ y entonces $F_Y(y|x) = F_Y(y)$ y $f_Y(y|x) = f_Y(y)$. La probabilidad de un suceso A dado $X = x_k$ se obtiene integrando la pdf condicionada:

$$P[Y \in A|X = x_k] = \int_{y \in A} f_Y(y|x_k) dy. \quad (5.41)$$

Obtenemos $P[Y \in A]$ utilizando la ecuación (5.38).

Ejemplo 5.31 Sistema de comunicaciones binario

La entrada X a un canal de comunicación asume los valores $+1$ o -1 con probabilidad $1/3$ y $2/3$. La salida Y del canal viene dada por $Y = X + N$, donde N es una variable aleatoria Normal con media 0 y varianza 1 . Calcula la pdf condicionada de Y dado $X = +1$ y dado $X = -1$. Calcula $P[X = +1|Y > 0]$.

Solución

La cdf condicionada de Y dada $X = +1$ es:

$$\begin{aligned} F_Y(y|+1) &= P[Y \leq y|X = +1] = P[N + 1 \leq y] \\ &= P[N \leq y - 1] = \int_{-\infty}^{y-1} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \end{aligned}$$

donde destacamos que si $X = +1$, entonces $Y = N + 1$ y Y depende sólo en N .

Por tanto, si $X = +1$, entonces Y es una variable aleatoria Normal con media 1 y varianza 1. De forma similar, si $X = -1$, entonces Y es normal con media -1 y varianza 1.

Las probabilidades de que $Y > 0$ dado que $X = +1$ y $X = -1$ son:

$$P[Y > 0|X = +1] = \int_0^{\infty} \frac{1}{\sqrt{2\pi}} e^{-(x-1)^2/2} dx = \int_{-1}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt = 1 - Q(1) = 0.841.$$

$$P[Y > 0|X = -1] = \int_0^{\infty} \frac{1}{\sqrt{2\pi}} e^{-(x+1)^2/2} dx = \int_1^{\infty} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt = Q(1) = 0.159.$$

Aplicando la ecuación (5.38), obtenemos:

$$P[Y > 0] = P[Y > 0|X = +1]\frac{1}{3} + P[Y > 0|X = -1]\frac{2}{3} = 0.386.$$

Con el teorema de Bayes calculamos:

$$P[X = +1|Y > 0] = \frac{P[Y > 0|X = +1]P[X = +1]}{P[Y > 0]} = \frac{(1 - Q(1))/3}{(1 + Q(1))/3} = 0.726.$$

Concluimos que si $Y > 0$ entonces $X = +1$ es más probable que $X = -1$. Por tanto el receptor debe decidir que la entrada es $X = +1$ cuando observa $Y > 0$.

En el ejemplo anterior, hemos dado un paso interesante que merece ser estudiado en más profundidad porque aparece frecuentemente: $P[Y \leq y|X = +1] = P[N + 1 \leq y]$, donde $Y = X + N$. Vamos a examinarlo más de cerca:

$$\begin{aligned} P[Y \leq z|X = x] &= \frac{P[\{X + N \leq z\} \cap \{X = x\}]}{P[X = x]} = \frac{P[\{x + N \leq z\} \cap \{X = x\}]}{P[X = x]} \\ &= P[x + N \leq z|X = x] = P[N \leq z - x|X = x]. \end{aligned}$$

En la primera línea, los sucesos $\{X + N \leq z\}$ y $\{x + N \leq z\}$ son diferentes. El primero involucra a las dos variables aleatorias X y N , mientras que el segundo sólo involucra a N y, consecuentemente, es mucho más simple. Podemos aplicar entonces una expresión como la ecuación (5.38) para obtener $P[Y \leq z]$. El paso que hacemos en el ejemplo, sin embargo, es más interesante aún. Como X y N son variables aleatorias independientes, podemos llevar la expresión un paso más lejos:

$$P[Y \leq z|X = 0] = P[N \leq z - x|X = x] = P[N \leq z - x].$$

La independencia de X y N permite que nos olvidemos de condicionar todo a x .

Caso 2: X es una variable aleatoria continua

Si X es una variable aleatoria continua, entonces $P[X = x] = 0$ y entonces la ecuación (5.33) no está definida para ningún x . Si X e Y tienen una pdf conjunta continua y no nula en alguna región del plano, definimos **cdf de Y condicionada a $X = x$** con el siguiente límite:

$$F_Y(y|x) = \lim_{h \rightarrow 0} F_Y(y|x < X \leq x + h). \quad (5.42)$$

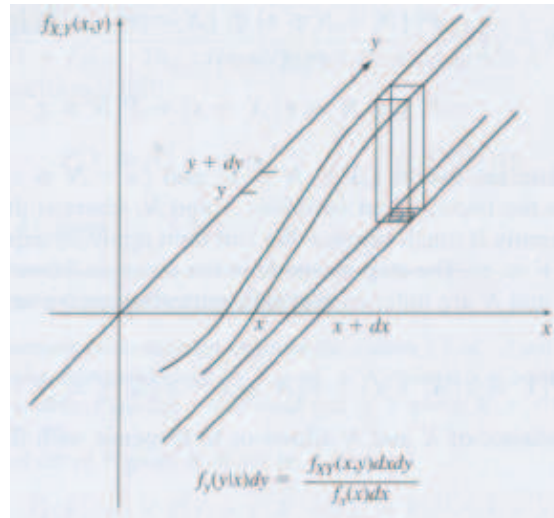


Figura 5.19 Interpretación de la pdf condicionada.

La cdf condicionada de la parte derecha de la ecuación (5.42) es:

$$\begin{aligned} F_Y(y|x < X \leq x+h) &= \frac{P[Y \leq y, x < X \leq x+h]}{P[x < X \leq x+h]} \\ &= \frac{\int_{-\infty}^y \int_x^{x+h} f_{X,Y}(x',y') dx' dy'}{\int_x^{x+h} f_X(x') dx'} = \frac{\int_{-\infty}^y f_{X,Y}(x,y') dy' h}{f_X(x) h}. \end{aligned} \quad (5.43)$$

Cuando h tiende a cero, las ecuaciones (5.42) y (5.43) implican que

$$F_Y(y|x) = \frac{\int_{-\infty}^y f_{X,Y}(x,y') dy'}{f_X(x)}. \quad (5.44)$$

La pdf condicionada de Y dado $X = x$ es entonces:

$$f_Y(y|x) = \frac{d}{dy} F_Y(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}. \quad (5.45)$$

Es fácil demostrar que $f_Y(y|x)$ satisface las propiedades de una pdf. Podemos interpretar $f_Y(y|x)dy$ como la probabilidad de que Y esté en la franja infinitesimal definida por $(y, y + dy)$ dado que X está en la franja infinitesimal definida por $(x, x + dx)$, como se muestra en la figura 5.19.

La probabilidad del suceso A dado $X = x$ se obtiene como sigue:

$$P[Y \in A | X = x] = \int_{y \in A} f_Y(y|x) dy. \quad (5.46)$$

Existe un gran parecido entre la ecuación (5.34) para el caso discreto y la ecuación (5.45) para el caso continuo. De hecho, verifican las mismas propiedades. Por

ejemplo, se obtiene la regla de la multiplicación de la ecuación (5.45):

$$f_{X,Y}(x, y) = f_Y(y|x)f_X(x) \text{ y } f_{X,Y}(x, y) = f_X(x|y)f_Y(y). \quad (5.47)$$

Si X e Y son independientes, $f_{X,Y}(x, y) = f_X(x)f_Y(y)$ y $f_Y(y|x) = f_Y(y)$, $f_X(x|y) = f_X(x)$, $F_Y(y|x) = F_Y(y)$ y $F_X(x|y) = F_X(x)$.

Combinando las ecuaciones (5.46) y (5.47), podemos demostrar que:

$$P[Y \in A] = \int_{-\infty}^{\infty} P[Y \in A|X = x]f_X(x)dx. \quad (5.48)$$

Se puede pensar en la ecuación (5.48) como la versión “continua” del teorema de la probabilidad total. Los siguientes ejemplos muestran la utilidad de los resultados anteriores en el cálculo de probabilidades de sucesos complejos.

Ejemplo 5.32 Sea X e Y las variables aleatorias del ejemplo 5.8. Calcula $f_X(x|y)$ y $f_Y(y|x)$.

Solución

Utilizando las pdf marginales obtenidas en el ejemplo 5.8, tenemos

$$f_X(x|y) = \frac{2e^{-x}e^{-y}}{2e^{-2y}} = e^{-(x-y)} \text{ para } x \geq y.$$

$$f_Y(y|x) = \frac{2e^{-x}e^{-y}}{2e^{-x}(1-e^{-x})} = \frac{e^{-y}}{1-e^{-x}} \text{ para } 0 < y < x.$$

La pdf condicionada de X es una pdf exponencial trasladada a la derecha por y . La pdf condicionada de Y es una pdf exponencial truncada en el intervalo $[0, x]$.

Ejemplo 5.33 Número de llegadas durante el tiempo de servicio del cliente

El número N de clientes que llegan a una estación de servicio durante un tiempo t es una variable aleatoria de Poisson con parámetro βt . El tiempo T necesario para atender a cada cliente es una variable aleatoria Exponencial con parámetro α . Calcula la pmf para el número N que llega durante el tiempo de servicio T de un cliente específico. Supongamos que las llegadas de los clientes son independientes del tiempo de servicio al cliente.

Solución

La ecuación (5.48) se verifica incluso si Y es una variable aleatoria discreta, por lo tanto

$$P[N = k] = \int_0^{\infty} P[N = k|T = t]f_T(t)dt$$

$$= \int_0^{\infty} \frac{(\beta t)^k}{k!} e^{-\beta t} \alpha e^{-\alpha t} dt$$

$$= \frac{\alpha \beta^k}{k!} \int_0^{\infty} t^k e^{-(\alpha+\beta)t} dt.$$

Sea $r = (\alpha + \beta)t$, entonces

$$\begin{aligned} P[N = k] &= \frac{\alpha\beta^k}{k!(\alpha + \beta)^{k+1}} \int_0^\infty r^k e^{-r} dr \\ &= \frac{\alpha\beta^k}{(\alpha + \beta)^{k+1}} = \left(\frac{\alpha}{\alpha + \beta}\right) \left(\frac{\beta}{\alpha + \beta}\right)^k, \end{aligned}$$

donde hemos utilizado el hecho de que la última integral es una función gamma y es igual a $k!$. Por lo tanto N es una variable aleatoria Geométrica con probabilidad de “éxito” $\alpha/(\alpha + \beta)$. Cada vez que un cliente llega, podemos imaginar que un nuevo ensayo de Bernoulli comienza, donde el “éxito” se produce si el tiempo de servicio al cliente se completa antes de la próxima llegada.

Ejemplo 5.34 Sea X un número del intervalo unitario que se escoge aleatoriamente; entonces Y es seleccionado aleatoriamente del intervalo $(0, X)$. Calcula la cdf de Y .

Solución

Cuando $X = x$, Y es uniforme en el intervalo $(0, x)$, de forma que la cdf condicionada a que $X = x$ es

$$P[Y \leq y | X = x] = \begin{cases} y/x & 0 \leq y \leq x \\ 1 & x < y. \end{cases}$$

La ecuación (5.48) y la cdf anterior implican:

$$\begin{aligned} F_Y(y) &= P[Y \leq y] = \int_0^1 P[Y \leq y | X = x] f_X(x) dx = \\ &= \int_0^y 1 dx' + \int_y^1 \frac{y}{x'} dx' = y - y \ln y. \end{aligned}$$

La pdf correspondiente se obtiene tomando la derivada de la cdf:

$$f_Y(y) = -\ln y \quad 0 \leq y \leq 1.$$

Ejemplo 5.35 Receptor máximo a posteriori

Para el sistema de comunicaciones en el ejemplo 5.31, calcula la probabilidad de que la entrada fuera $X = +1$ dado que la salida del canal es $Y = y$.

Solución

Ésta es una versión difícil de la regla de Bayes. Condicionando al suceso $\{y <$

$Y \leq y + \Delta$ en lugar de $\{Y = y\}$:

$$\begin{aligned} P[X = +1|y < Y < y + \Delta] &= \frac{P[y < Y < y + \Delta|X = +1]P[X = +1]}{P[y < Y < y + \Delta]} \\ &= \frac{f_Y(y|+1)\Delta(1/3)}{f_Y(y|+1)\Delta(1/3) + f_Y(y|-1)\Delta(2/3)} \\ &= \frac{\frac{1}{\sqrt{2\pi}}e^{-(y-1)^2/2}(1/3)}{\frac{1}{\sqrt{2\pi}}e^{-(y-1)^2/2}(1/3) + \frac{1}{\sqrt{2\pi}}e^{-(y+1)^2/2}(2/3)} \\ &= \frac{e^{-(y-1)^2/2}}{e^{-(y-1)^2/2} + 2e^{-(y+1)^2/2}} = \frac{1}{1 + 2e^{-2y}}. \end{aligned}$$

La expresión anterior es igual a $1/2$ cuando $y_T = 0.3466$. Para $y > y_T$, $X = +1$ es más probable, y para $y < y_T$, $X = -1$ es más probable. Un receptor que selecciona la entrada X que es más probable, dado que $Y = y$, se llama *receptor máximo a posteriori*.

5.7.2 Esperanza condicionada

La **esperanza condicionada de Y dado que $X = x$** se define como

$$E[Y|x] = \int_{-\infty}^{\infty} y f_Y(y|x) dy. \quad (5.49a)$$

En el caso especial en el que X e Y son variables aleatorias discretas tenemos:

$$E[Y|x_k] = \sum_{y_j} y_j p_Y(y_j|x_k). \quad (5.49b)$$

Claramente, $E[Y|x]$ es simplemente el centro de masa asociado a la pmf o la pdf condicionada.

La esperanza condicionada $E[Y|x]$ puede ser vista como la definición de una función de x : $g(x) = E[Y|x]$. Por lo tanto, tiene sentido hablar de la variable aleatoria $g(X) = E[Y|X]$. Podemos imaginar que se realiza un experimento aleatorio obteniendo un valor para X , por ejemplo $X = x_0$, y luego se calcula el valor $g(x_0) = E[Y|x_0]$. Queremos conocer $E[g(X)] = E[E[Y|X]]$. En particular, se demuestra que

$$E[Y] = E[E[Y|X]], \quad (5.50)$$

donde el lado derecho de la ecuación es

$$E[E[Y|X]] = \int_{-\infty}^{\infty} E[Y|x] f_X(x) dx \quad X \text{ continua} \quad (5.51a)$$

$$E[E[Y|X]] = \sum_{x_k} E[Y|x_k] p_X(x_k) \quad X \text{ discreta} \quad (5.51b)$$

Probamos la ecuación (5.50) para el caso en que X e Y son variables aleatorias

conjuntamente continuas, entonces

$$\begin{aligned}
 E[E[Y|X]] &= \int_{-\infty}^{\infty} E[Y|x]f_X(x)dx \\
 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} yf_Y(y|x)dyf_X(x)dx \\
 &= \int_{-\infty}^{\infty} y \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx dy \\
 &= \int_{-\infty}^{\infty} yf_Y(y)dy = E[Y].
 \end{aligned}$$

El resultado anterior también es válido para el valor esperado de una función de Y :

$$E[h(Y)] = E[E[h(Y)|X]].$$

En particular, el momento k -ésimo de Y viene dado por

$$E[Y^k] = E[E[Y^k|X]].$$

Ejemplo 5.36 Número medio de defectos en una región

Calcula la media de Y en el ejemplo 5.30 utilizando la esperanza condicionada.

Solución

$$E[Y] = \sum_{k=0}^{\infty} E[Y|X = k]P[X = k] = \sum_{k=0}^{\infty} kpP[X = k] = pE[X] = p\alpha.$$

La segunda igualdad utiliza el hecho de que $E[Y|X = k] = kp$ ya que Y es binomial con parámetros k y p . Nótese que la penúltima igualdad se verifica para *cualquier* pmf de X . El hecho de que X sea Poisson de media α no se utiliza hasta la última igualdad.

Ejemplo 5.37 Canal de comunicación binario

Calcula la media de la salida Y del canal de comunicaciones en el ejemplo 5.31.

Solución

Como Y es una variable aleatoria Normal con media $+1$ cuando $X = +1$, y -1 cuando $X = -1$, los valores esperados de Y dado X son los siguientes:

$$E[Y|+1] = 1 \text{ y } E[Y|-1] = -1.$$

La ecuación (5.51b) implica

$$E[Y] = \sum_{k=0}^{\infty} E[Y|X = k]P[X = k] = +1(1/3) - 1(2/3) = -1/3.$$

La media es negativa debido a que la entrada $X = -1$ se produce dos veces más que $X = +1$.

Ejemplo 5.38 Número medio de llegadas durante el tiempo de servicio

Calcula la media y la varianza del número de llegadas N durante el tiempo de servicio de un cliente del ejemplo 5.33.

Solución

N es una variable aleatoria de Poisson con parámetro βt cuando $T = t$ viene dado, entonces, los dos primeros momentos condicionados son:

$$E[N|T = t] = \beta t \quad E[N^2|T = t] = (\beta t) + (\beta t)^2.$$

Los dos primeros momentos de N se obtienen de la ecuación (5.50):

$$\begin{aligned} E[N] &= \int_0^\infty E[N|T = t]f_T(t)dt = \int_0^\infty \beta t f_T(t)dt = \beta E[T] \\ E[N^2] &= \int_0^\infty E[N^2|T = t]f_T(t)dt = \int_0^\infty \{\beta t + \beta^2 t^2\}f_T(t)dt \\ &= \beta E[T] + \beta^2 E[T^2]. \end{aligned}$$

La varianza de N es entonces

$$\begin{aligned} V[N] &= E[N^2] - (E[N])^2 \\ &= \beta^2 E[T^2] + \beta E[T] - \beta^2 (E[T])^2 \\ &= \beta^2 V[T] + \beta E[T]. \end{aligned}$$

Nótese que si T no es aleatoria (i.e., $E[T] = cte.$ y $V[T] = 0$) entonces la media y la varianza de N son aquellas de la variable aleatoria de Poisson con parámetro $\beta E[T]$. Cuando T es aleatoria, la media de N sigue siendo la misma pero la varianza de N crece con el término $\beta^2 V[T]$, esto es, la variabilidad de T causa mayor variabilidad en N . Hasta este punto, hemos evitado intencionadamente utilizando el hecho de que T tiene una distribución exponencial para enfatizar que los resultados anteriores se verifican para *cualquier* distribución $f_T(t)$ para el tiempo de servicio. Si T es exponencial con parámetro α , entonces $E[T] = 1/\alpha$ y $V[T] = 1/\alpha^2$, entonces

$$E[N] = \frac{\beta}{\alpha} \quad \text{y} \quad V[N] = \frac{\beta^2}{\alpha^2} + \frac{\beta}{\alpha}.$$

5.8 Funciones de dos variables aleatorias

En muchas ocasiones trabajamos con una o más funciones de variables aleatorias asociadas a algún experimento. Por ejemplo, si realizamos medidas repetidas de la misma cantidad aleatoria, podríamos estar interesados en el valor máximo

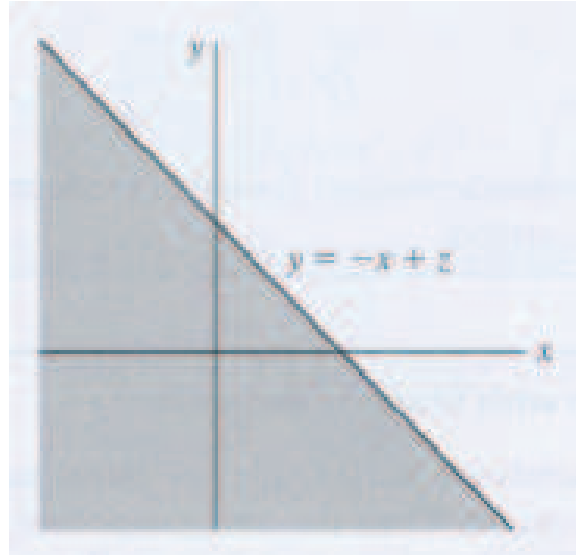


Figura 5.20
 $P[Z \leq z] = P[X + Y \leq z]$.

o el mínimo del conjunto, así como en la media muestral y la varianza muestral. En esta sección presentamos métodos para determinar las probabilidades de sucesos que involucran funciones de dos variables aleatorias.

5.8.1 Una función de dos variables aleatorias

Sea Z una variable aleatoria definida como una función de dos variables aleatorias:

$$Z = g(X, Y). \quad (5.52)$$

La cdf de Z se calcula hallando primero el suceso equivalente de $\{Z \leq z\}$, esto es, el conjunto $R_z = \{\mathbf{x} = (x, y) \text{ tal que } g(\mathbf{x}) \leq z\}$, entonces

$$F_z(z) = P[\mathbf{X} \in R_z] = \int \int_{(x,y) \in R_z} f_{X,Y}(x', y') dx' dy'. \quad (5.53)$$

La pdf de Z se calcula entonces tomando la derivada de $F_z(z)$.

Ejemplo 5.39 Suma de dos variables aleatorias

Sea $Z = X + Y$. Calcula $F_Z(z)$ y $f_Z(z)$ en términos de la pdf conjunta de X y Y .

Solución

La cdf de Z se calcula integrando la pdf conjunta de X e Y en la región del plano correspondiente al suceso $\{Z \leq z\}$, como se observa en la figura 5.20.

$$F_Z(z) = \int_{-\infty}^{\infty} \int_{-\infty}^{z-x'} f_{X,Y}(x', y') dy' dx'.$$

La pdf de Z es

$$f_Z(z) = \frac{d}{dz} F_Z(z) = \int_{-\infty}^{\infty} f_{X,Y}(x', z - x') dx'. \quad (5.54)$$

Entonces la pdf de la suma de dos variables aleatorias viene dada por una integral de la superposición.

Si X e Y son variables aleatorias independientes, entonces por la ecuación (5.23) la pdf viene dada por la *integral de la convolución* de las pdf marginales de X e Y :

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x') f_Y(z - x') dx'. \quad (5.55)$$

En el capítulo 7 mostramos cómo los métodos de transformación se usan para evaluar integrales de convolución como la de la ecuación (5.55).

Ejemplo 5.40 Suma de variables aleatorias normales no independientes

Calcula la pdf de la suma $Z = X + Y$ de dos variables aleatorias normales con media 0, varianza 1 y coeficiente de correlación $\rho = -1/2$.

Solución

La pdf conjunta para este par de variables se dio en el ejemplo 5.18. La pdf de Z se calcula substituyendo la pdf de las variables aleatorias normales bivariantes en la integral de superposición calculada en el ejemplo 5.39:

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{\infty} f_{X,Y}(x', z - x') dx' \\ &= \frac{1}{2\pi(1 - \rho^2)^{1/2}} \int_{-\infty}^{\infty} e^{-[x'^2 - 2\rho x'(z - x') + (z - x')^2]/2(1 - \rho^2)} dx' \\ &= \frac{1}{2\pi(3/4)^{1/2}} \int_{-\infty}^{\infty} e^{-(x'^2 - x'z + z^2)/2(3/4)} dx'. \end{aligned}$$

Después de completar el cuadrado del argumento de la exponencial obtenemos

$$f_Z(z) = \frac{e^{-z^2/2}}{\sqrt{2\pi}}.$$

Entonces, la suma de estas dos variables aleatorias normales no independientes es también una variable aleatoria Normal con media 0 y varianza 1.

Ejemplo 5.41 Un sistema con redundancia en espera

Un sistema con redundancia en espera tiene un componente clave funcionando y un duplicado de ese componente en modo de espera. Cuando el primer componente falla, el segundo componente se pone operativo. Calcula la pdf del tiempo de vida del sistema en espera si los componentes tienen tiempos de vida independientes distribuidos exponencialmente con la misma media.

Solución

Sean T_1 y T_2 los tiempos de vida de dos componentes, entonces el tiempo de vida del sistema es $T = T_1 + T_2$ y la pdf de T viene dada por la ecuación (5.55). Los términos del integrando son

$$f_{T_1}(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

$$f_{T_2}(z-x) = \begin{cases} \lambda e^{-\lambda(z-x)} & z-x \geq 0 \\ 0 & x > z. \end{cases}$$

Nótese que la primera ecuación establece el límite inferior de integración a 0 y la segunda ecuación establece el límite superior para z . La ecuación (5.55) se queda como

$$f_T(z) = \int_0^z \lambda e^{-\lambda x} \lambda e^{-\lambda(z-x)} dx$$

$$= \lambda^2 e^{-\lambda z} \int_0^z dx = \lambda^2 z e^{-\lambda z}.$$

Entonces T es una variable aleatoria de Erlang con parámetro $m = 2$.

La pdf condicionada se puede utilizar para calcular la pdf de una función de varias variables aleatorias. Sea $Z = g(X, Y)$ y supongamos que sabemos que $Y = y$, entonces $Z = g(X, y)$ es una función de una variable aleatoria. Por tanto, podemos utilizar los métodos desarrollados en la sección 4.5 para variables aleatorias simples para calcular la pdf de Z dado $Y = y$: $f_Z(z|Y = y)$. La pdf de Z se calcula entonces con

$$f_Z(z) = \int_{-\infty}^{\infty} f_Z(z|y') f_Y(y') dy'.$$

Ejemplo 5.42 Sea $Z = X/Y$. Calcula la pdf de Z si X e Y son independientes y distribuidas exponencialmente con media uno.

Solución

Asumamos que $Y = y$, entonces $Z = X/y$ es simplemente una versión de X . Por tanto, por el ejemplo 4.28

$$f_Z(z|y) = |y| f_X(yz|y).$$

La pdf de Z es por tanto

$$f_Z(z) = \int_{-\infty}^{\infty} |y'| f_X(y'z|y') f_Y(y') dy' = \int_{-\infty}^{\infty} |y'| f_{X,Y}(y'z, y') dy'.$$

Ahora utilizamos el hecho de que X e Y son independientes y con distribución exponencial con media 1:

$$\begin{aligned} f_Z(z) &= \int_0^\infty y' f_X(y'z) f_Y(y') dy' \quad z > 0 \\ &= \int_0^\infty y' e^{-y'z} e^{-y'} dy' \\ &= \frac{1}{(1+z)^2} \quad z > 0. \end{aligned}$$

5.8.2 Transformaciones de dos variables aleatorias

Sea X e Y dos variables aleatorias asociadas a algún experimento y sea Z_1 y Z_2 dos variables aleatorias definidas por dos funciones de $\mathbf{X} = (X, Y)$:

$$Z_1 = g_1(\mathbf{X}) \quad \text{y} \quad Z_2 = g_2(\mathbf{X}).$$

Ahora consideremos el problema de calcular la cdf y la pdf conjuntas de Z_1 y Z_2 .

La cdf conjunta de Z_1 y Z_2 en el punto $\mathbf{z} = (z_1, z_2)$ es igual a la probabilidad de la región de \mathbf{x} donde $g_k(\mathbf{x}) \leq z_k$ para $k = 1, 2$:

$$F_{Z_1, Z_2}(z_1, z_2) = P[g_1(\mathbf{X}) \leq z_1, g_2(\mathbf{X}) \leq z_2]. \quad (5.56a)$$

Si X, Y tiene pdf conjunta, entonces

$$F_{Z_1, Z_2}(z_1, z_2) = \int \int_{\mathbf{x}': g_k(\mathbf{x}') \leq z_k} f_{X, Y}(x', y') dx' dy'. \quad (5.56b)$$

Ejemplo 5.43 Sea W y Z variables aleatorias definidas por

$$W = \min(X, Y) \quad \text{y} \quad Z = \max(X, Y).$$

Calcula la cdf conjunta de W y Z en términos de la cdf conjunta de X e Y .

Solución

La ecuación (5.56a) implica que

$$F_{W, Z}(w, z) = P[\{\min(X, Y) \leq w\} \cap \{\max(X, Y) \leq z\}].$$

La región correspondiente a este suceso se muestra en la figura 5.21. Observando la figura, queda claro que si $z > w$, la probabilidad anterior es la probabilidad del rectángulo semi-infinito definido por el punto (z, z) menos la región cuadrada denotada como A . Entonces si $z > w$,

$$\begin{aligned} F_{W, Z}(w, z) &= F_{X, Y}(z, z) - P[A] \\ &= F_{X, Y}(z, z) - \{F_{X, Y}(z, z) - F_{X, Y}(w, z) - F_{X, Y}(z, w) + F_{X, Y}(w, w)\} \\ &= F_{X, Y}(w, z) + F_{X, Y}(z, w) - F_{X, Y}(w, w). \end{aligned}$$

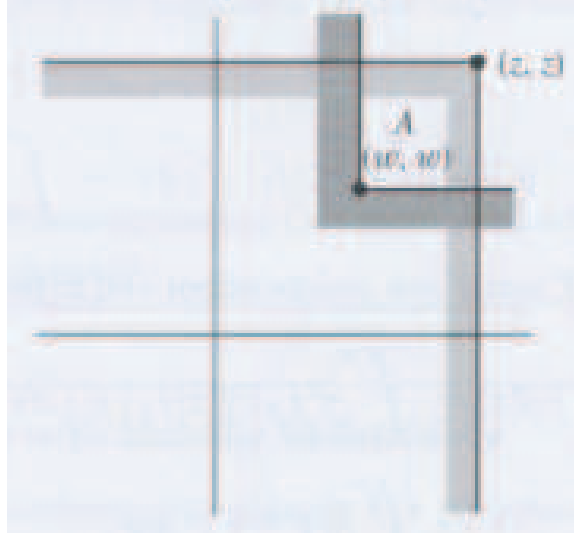


Figura 5.21 $\{\min(X, Y) \leq w = \{X \leq w\} \cup \{Y \leq w\}\}$ y $\{\max(X, Y) \leq z = \{X \leq z\} \cap \{Y \leq z\}\}$.

Si $z < w$ entonces

$$F_{W,Z}(w, z) = F_{X,Y}(z, z).$$

Ejemplo 5.44 Radio y ángulo de variables aleatorias normales independientes

Sea X e Y variables aleatorias normales independientes de media 0 y varianza 1. Calcula la cdf y la pdf conjuntas de R y Θ , el radio y el ángulo del punto (X, Y) :

$$R = (X^2 + Y^2)^{1/2} \quad \Theta = \tan^{-1}(Y/X).$$

Solución

La cdf conjunta de R y Θ es:

$$F_{R,\Theta}(r_0, \theta_0) = P[R \leq r_0, \Theta \leq \theta_0] = \int \int_{(x,y) \in R_{(r_0, \theta_0)}} \frac{e^{-(x^2+y^2)/2}}{2\pi} dx dy$$

donde

$$R_{(r_0, \theta_0)} = \{(x, y) : \sqrt{x^2 + y^2} \leq r_0, 0 < \tan^{-1}(Y/X) \leq \theta_0\}.$$

La región $R_{(r_0, \theta_0)}$ es la región en forma de trozo de tarta en la figura 5.22. Cambiamos las variables de coordenadas cartesianas a polares para obtener:

$$\begin{aligned} F_{R,\Theta}(r_0, \theta_0) &= P[R \leq r_0, \Theta \leq \theta_0] = \int_0^{r_0} \int_0^{\theta_0} \frac{e^{-r^2/2}}{2\pi} r dr d\theta \\ &= \frac{\theta_0}{2\pi} (1 - e^{-r_0^2/2}), 0 < \theta_0 < 2\pi, 0 < r_0 < \infty. \end{aligned} \quad (5.57)$$

R y Θ son variables aleatorias independientes, donde R tiene una distribución de

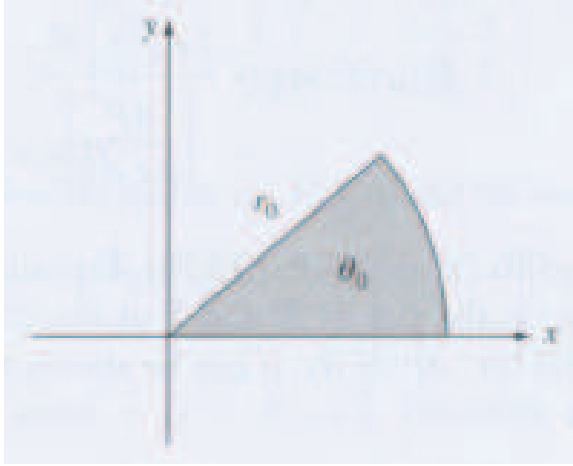


Figura 5.22 Región de integración $R_{(r_0, \theta_0)}$ en el ejemplo 5.44.

Rayleigh y Θ se distribuye uniformemente en el intervalo $(0, 2\pi)$. La pdf conjunta se obtiene tomando derivadas parciales con respecto a r y θ :

$$\begin{aligned} f_{R, \Theta}(r, \theta) &= \frac{\delta^2}{\delta r \delta \theta} \frac{\theta}{2\pi} (1 - e^{-r^2/2}) \\ &= \frac{1}{2\pi} (r e^{-r^2/2}), 0 < \theta < 2\pi, 0 < r < \infty. \end{aligned}$$

Esta transformación conecta todos los puntos del plano de coordenadas cartesianas a polares. Podemos también deshacer la transformación partiendo de las coordenadas polares a cartesianas. Primero generamos variables aleatorias independientes, Rayleigh para R y uniforme para Θ . Entonces transformamos R y Θ a coordenadas cartesianas para obtener un par de variables aleatorias independientes normales de media 0 y varianza 1.

5.8.3 pdf de transformaciones lineales

La pdf conjunta de \mathbf{Z} se puede calcular directamente en términos de la pdf conjunta de \mathbf{X} calculando los sucesos equivalentes de los rectángulos infinitesimales. Consideramos la **transformación lineal** de dos variables aleatorias:

$$V = aX + bY$$

$$W = cX + eY$$

o

$$\begin{bmatrix} V \\ W \end{bmatrix} = \begin{bmatrix} a & b \\ c & e \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix}.$$

Denotemos A a la matriz de arriba. Vamos a asumir que A tiene inversa, esto es, que su determinante $|ae - bc| \neq 0$, de modo que a cada punto (v, w) le

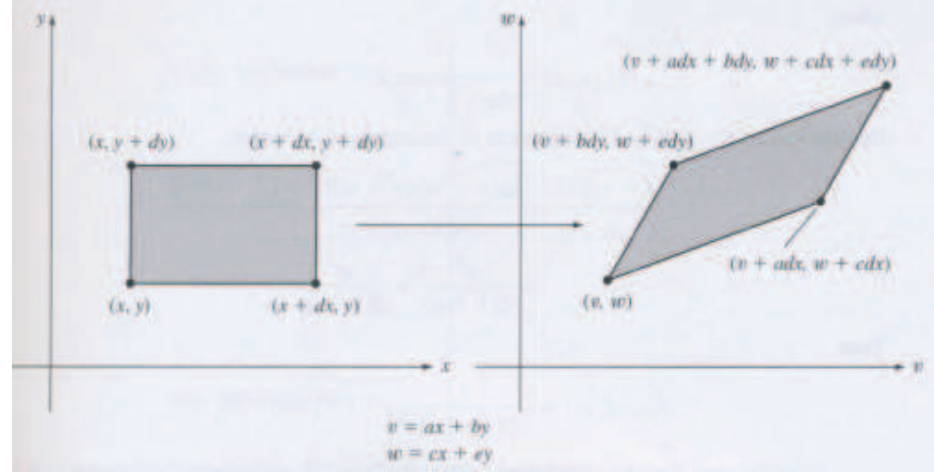


Figura 5.23 Imagen de un rectángulo infinitesimal bajo una transformación lineal.

corresponde un único punto (x, y) obtenido de

$$\begin{bmatrix} x \\ y \end{bmatrix} = A^{-1} \begin{bmatrix} v \\ w \end{bmatrix}. \quad (5.58)$$

Consideremos el rectángulo infinitesimal mostrado en la figura 5.23. Los puntos de este rectángulo se llevan al paralelogramo que se muestra en la figura. El rectángulo infinitesimal y el paralelogramo son sucesos equivalentes, por tanto, sus probabilidades deben ser iguales. Entonces

$$f_{X,Y}(x, y)dxdy \simeq f_{V,W}(v, w)dP$$

donde dP es el área del paralelogramo. La pdf conjunta de V y W viene dada en consecuencia por

$$f_{V,W}(v, w) = \frac{f_{X,Y}(x, y)}{\left| \frac{dP}{dxdy} \right|}, \quad (5.59)$$

donde x e y se relacionan con (v, w) mediante la ecuación (5.58). La ecuación (5.59) establece que la pdf conjunta de V y W en (v, w) es la pdf de X e Y en el punto correspondiente (x, y) , pero reescalada por el factor $dP/dxdy$. Se puede demostrar que $dP = (|ae - bc|)dxdy$, y por tanto el factor es

$$\left| \frac{dP}{dxdy} \right| = \frac{|ae - bc|(dxdy)}{(dxdy)} = |ae - bc| = |A|,$$

donde $|A|$ es el determinante de A .

El resultado anterior se puede escribir de forma compacta utilizando la notación matricial. Sea el vector \mathbf{Z}

$$\mathbf{Z} = \mathbf{A}\mathbf{X},$$

donde A es una matriz invertible de dimensión $n \times n$. La pdf conjunta de \mathbf{Z} es entonces

$$f_{\mathbf{Z}}(\mathbf{z}) = \frac{f_{\mathbf{X}}(A^{-1}\mathbf{z})}{|A|}. \quad (5.60)$$

Ejemplo 5.45 Transformación lineal de variables aleatorias conjuntamente normales

Sea X e Y las variables aleatorias conjuntamente normales del ejemplo 5.18. Obtenemos V y W a partir de (X, Y) por

$$\begin{bmatrix} V \\ W \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix} = A \begin{bmatrix} X \\ Y \end{bmatrix}.$$

Calcula la pdf conjunta de V y W .

Solución

El determinante de la matriz es $|A| = 1$ y la aplicación inversa está dada por

$$\begin{bmatrix} X \\ Y \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} V \\ W \end{bmatrix},$$

entonces $X = (V - W)/\sqrt{2}$ e $Y = (V + W)/\sqrt{2}$. Por lo tanto la pdf de V y W es

$$f_{V,W}(v, w) = f_{X,Y}\left(\frac{v-w}{\sqrt{2}}, \frac{v+w}{\sqrt{2}}\right),$$

donde

$$f_{X,Y}(x, y) = \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-(x^2 - 2\rho xy + y^2)/2(1-\rho^2)}.$$

Sustituyendo por x e y , el argumento del exponente queda como sigue

$$\frac{(v-w)^2/2 - 2\rho(v-w)(v+w)/2 + (v+w)^2/2}{2(1-\rho^2)} = \frac{v^2}{2(1+\rho)} + \frac{w^2}{2(1-\rho)}.$$

Entonces

$$f_{V,W}(v, w) = \frac{1}{2\pi(1-\rho^2)^{1/2}} e^{-\{[v^2/2(1+\rho)] + [w^2/2(1-\rho)]\}}.$$

Se puede observar que las variables transformadas V y W son normales independientes de media cero y varianza $1+\rho$ y $1-\rho$, respectivamente. La figura 5.24 muestra un mapa de contorno de la pdf conjunta de (X, Y) . Se puede observar que la pdf tiene simetría elíptica alrededor el origen de los ejes principales a 45° con respecto a los ejes del plano. En la sección 5.9 se muestra que la transformación lineal anterior corresponde a una rotación del sistema de coordenadas para que los ejes del plano estén alineados con los ejes de la elipse.

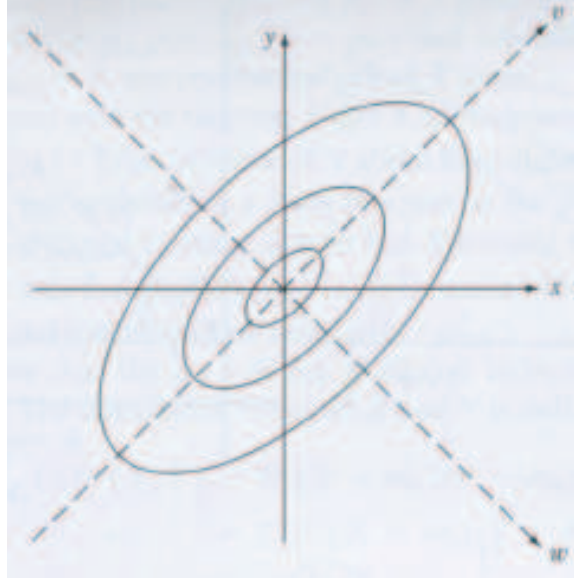


Figura 5.24 Mapa de isobaras de la pdf normal conjunta del ejemplo 5.45.

5.9 Variable aleatoria normal bivalente

Las variable aleatoria Normal bivalente aparecen en numerosas aplicaciones en ingeniería. Se utiliza con frecuencia para modelar las señales en las aplicaciones de procesamiento de señales y son el modelo más importante utilizado en sistemas de comunicación que involucran el tratamiento de señales en presencia de ruido. También juega un papel central en muchos métodos estadísticos.

Se dice que las variables aleatorias X e Y tienen una distribución normal bivalente si su pdf conjunta tiene la forma

$$f_{X,Y}(x,y) = \frac{\exp\left\{\frac{-1}{2(1-\rho_{X,Y}^2)}\left[\left(\frac{x-m_1}{\sigma_1}\right)^2 - 2\rho_{X,Y}\left(\frac{x-m_1}{\sigma_1}\right)\left(\frac{x-m_2}{\sigma_2}\right) + \left(\frac{x-m_2}{\sigma_2}\right)^2\right]\right\}}{2\pi\sigma_1\sigma_2\sqrt{1-\rho_{X,Y}^2}} \quad (5.61a)$$

para $-\infty < x < \infty$ y $-\infty < y < \infty$.

La pdf se centra en el punto (m_1, m_2) y tiene forma de campana que depende de los valores de σ_1 , σ_2 y $\rho_{X,Y}$ como se muestra en la figura 5.25. Como se muestra en la figura, la pdf es constante para los valores x e y para los que el argumento de la exponencial es constante:

$$\left[\left(\frac{x-m_1}{\sigma_1}\right)^2 - 2\rho_{X,Y}\left(\frac{x-m_1}{\sigma_1}\right)\left(\frac{x-m_2}{\sigma_2}\right) + \left(\frac{x-m_2}{\sigma_2}\right)^2\right] = cte. \quad (5.61b)$$

La figura 5.26 muestra la orientación de estos contornos elípticos para varios valores de σ_1 , σ_2 y $\rho_{X,Y}$. Cuando $\rho_{X,Y} = 0$, es decir, cuando X e Y son independientes, las líneas de contorno de la pdf son elipses con los ejes principales

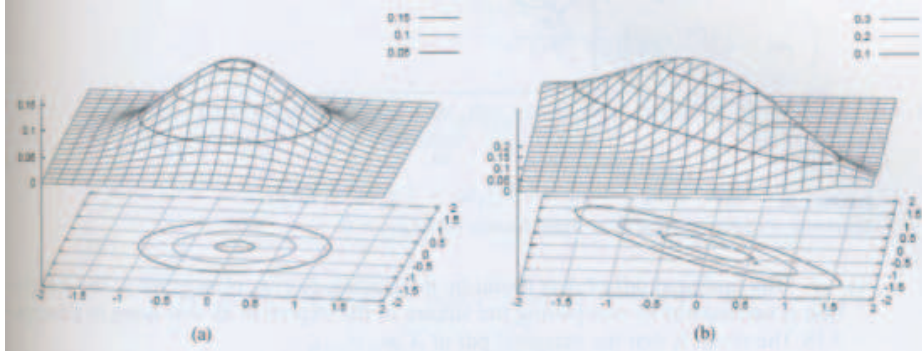


Figura 5.25 pdf de la normal bivalente (a) $\rho = 0$ (b) $\rho = -0.9$.

alineados con los ejes x e y . Cuando $\rho_{X,Y} \neq 0$, el eje mayor de la elipse se orienta a lo largo del ángulo

$$\theta = \frac{1}{2} \arctan^{-1} \tan \left(\frac{2\rho_{X,Y}\sigma_1\sigma_2}{\sigma_1^2 - \sigma_2^2} \right). \quad (5.62)$$

Observese que el ángulo es de 45° cuando las varianzas son iguales.

La pdf marginal de X se calcula integrando $f_{X,Y}(x,y)$ en todos los y . La integral se lleva a cabo completando el cuadrado en la exponencial como se hizo en el ejemplo 5.18. El resultado es que la pdf marginal de X es

$$f_X(x) = \frac{e^{-(x-m_1)^2/2\sigma_1^2}}{\sqrt{2\pi}\sigma_1}, \quad (5.63)$$

es decir, X es una variable aleatoria Normal con media m_1 y varianza σ_1^2 . Del mismo modo, se concluye que la pdf marginal de Y es normal con media m_2 y varianza σ_2^2 .

Las pdf condicionadas $f_X(x|y)$ y $f_Y(y|x)$ nos dan información acerca de la interrelación entre X e Y . La pdf condicionada de X dado que $Y = y$ es

$$\begin{aligned} f_X(x|y) &= \frac{f_{X,Y}(x,y)}{f_Y(y)} \\ &= \frac{\exp \left\{ \frac{-1}{2(1-\rho_{X,Y}^2)\sigma_1^2} \left[x - \rho_{X,Y} \frac{\sigma_1}{\sigma_2} (y - m_2) - m_1 \right]^2 \right\}}{\sqrt{2\pi\sigma_1^2(1-\rho_{X,Y}^2)}}. \end{aligned} \quad (5.64)$$

La ecuación (5.64) muestra que la pdf condicionada de X dado $Y = y$ es también normal pero con media condicionada $m_1 + \rho_{X,Y}(\sigma_1/\sigma_2)(y - m_2)$ y varianza condicionada $\sigma_1^2(1 - \rho_{X,Y}^2)$. Nótese que cuando $\rho_{X,Y} = 0$, la pdf condicionada de X dado que $Y = y$ es igual a la pdf marginal de X . Esto es consistente con el hecho de que X e Y sean independientes cuando $\rho_{X,Y} = 0$. Por otra parte, a medida que $|\rho_{X,Y}| \rightarrow 1$ la varianza de X alrededor de la media condicionada se aproxima

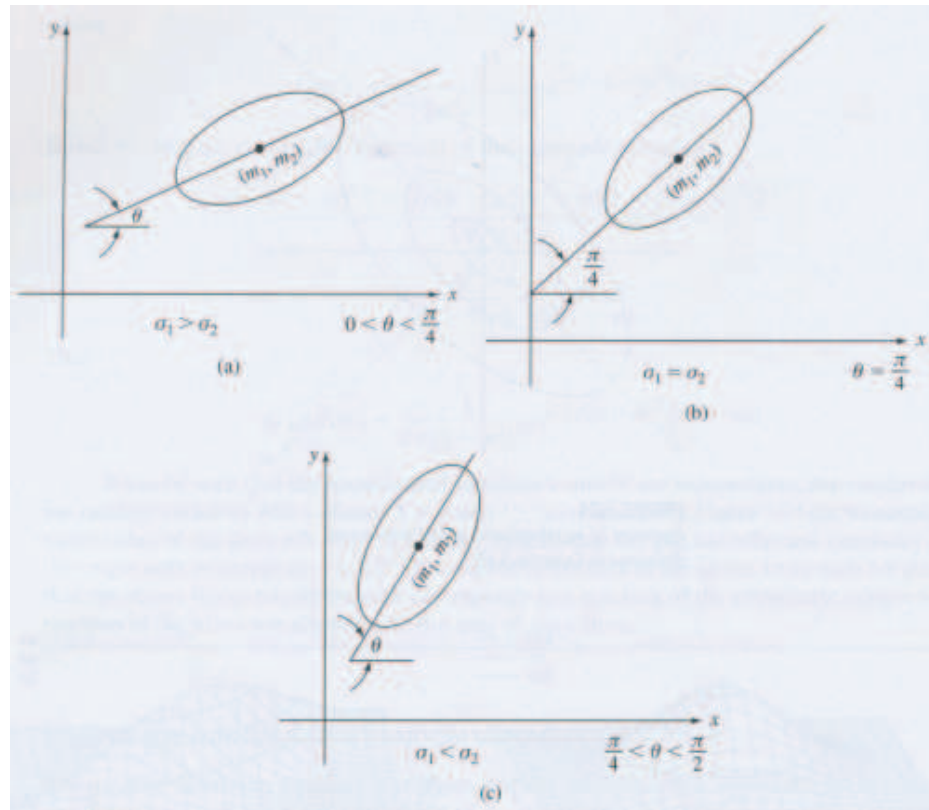


Figura 5.26 Orientación de las líneas de contorno de la pdf de la normal bivalente para $\rho_{X,Y} > 0$.

a cero, por lo que la pdf condicionada tiende a una función delta en la media condicionada. Así, cuando $|\rho_{X,Y}| = 1$, la varianza condicionada es igual a cero y X es igual a la media condicionada con probabilidad uno. Observamos que de manera similar $f_Y(y|x)$ es normal con media condicionada $m_2 + \rho_{X,Y}(\sigma_2/\sigma_1)(x - m_1)$ y varianza condicionada $\sigma_2^2(1 - \rho_{X,Y}^2)$.

Vamos a demostrar que la $\rho_{X,Y}$ de la ecuación (5.61a) es de hecho el coeficiente de correlación entre X e Y . La covarianza entre X e Y se define por

$$\begin{aligned} \text{Cov}(X, Y) &= E[(X - m_1)(Y - m_2)] \\ &= E[E[(X - m_1)(Y - m_2)|Y]]. \end{aligned}$$

Ahora la esperanza condicionada de $(X - m_1)(Y - m_2)$ dado $Y = y$ es

$$\begin{aligned} E[(X - m_1)(Y - m_2)|Y = y] &= (y - m_2)E[X - m_1|Y = y] \\ &= (y - m_2)(E[X|Y = y] - m_1) \\ &= (y - m_2) \left(\rho_{X,Y} \frac{\sigma_1}{\sigma_2} (y - m_2) \right), \end{aligned}$$

donde hemos utilizado el hecho de que la media condicionada de X dado $Y = y$ es $m_1 + \rho_{X,Y}(\sigma_1/\sigma_2)(y - m_2)$. Por lo tanto

$$E[(X - m_1)(Y - m_2)|Y] = \rho_{X,Y} \frac{\sigma_1}{\sigma_2} (Y - m_2)^2$$

y

$$\begin{aligned} \text{Cov}(X, Y) &= E[E[(X - m_1)(Y - m_2)|Y]] = \rho_{X,Y} \frac{\sigma_1}{\sigma_2} E[(Y - m_2)^2] \\ &= \rho_{X,Y} \sigma_1 \sigma_2. \end{aligned}$$

La ecuación anterior es consistente con la definición del coeficiente de correlación, $\rho_{X,Y} = \text{Cov}(X, Y)/\sigma_1 \sigma_2$. Así, el $\rho_{X,Y}$ en la ecuación (5.61a) es de hecho el coeficiente de correlación entre X e Y .

Ejemplo 5.46 La cantidad de lluvia anual en las ciudades 1 y 2 es modelada por una variable aleatoria Normal bivalente, X e Y , con pdf dada por la ecuación (5.61a). Calcula el valor más probable de X dado que sabemos que $Y = y$.

Solución

El valor más probable de X dado $Y = y$ es el valor x para el que $f_X(x|y)$ alcanza su máximo. La pdf condicionada de X dado $Y = y$ viene dada por la ecuación (5.64), que es máxima en la media condicionada

$$E[X|y] = m_1 + \rho_{X,Y} \frac{\sigma_1}{\sigma_2} (y - m_2).$$

Nótese que éste estimador de “máxima verosimilitud” es una función lineal de las observaciones y .

Ejemplo 5.47 La estimación de la señal en presencia de ruido

Sea $Y = X + N$ donde X (la “señal”) y N (el “ruido”) son variables aleatorias normales independientes de media cero con varianzas diferentes. Calcula el coeficiente de correlación entre la señal observada Y y la señal deseada X . Calcula el valor x que maximiza $f_X(x|y)$.

Solución

La media y la varianza de Y y la covarianza de X e Y son:

$$\begin{aligned} E[Y] &= E[X] + E[N] = 0 \\ \sigma_Y^2 &= E[Y^2] = E[(X + N)^2] = E[X^2 + 2XN + N^2] = E[X^2] + E[N^2] = \sigma_X^2 + \sigma_N^2. \\ \text{Cov}(X, Y) &= E[(X - E[X])(Y - E[Y])] = E[XY] = E[X(X + N)] = \sigma_X^2. \end{aligned}$$

Por lo tanto, el coeficiente de correlación es:

$$\rho_{X,Y} = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{\sigma_X}{\sigma_Y} = \frac{\sigma_X}{(\sigma_X^2 + \sigma_N^2)^{1/2}} = \frac{1}{\left(1 + \frac{\sigma_N^2}{\sigma_X^2}\right)^{1/2}}.$$

Nótese que $\rho_{X,Y}^2 = \sigma_X^2/\sigma_Y^2 = 1 - \sigma_N^2/\sigma_Y^2$.

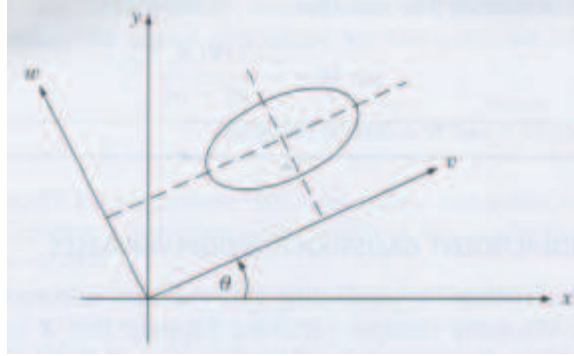


Figura 5.27 Una rotación del sistema de coordenadas transforma un par de variables aleatorias normales dependientes en un par de variables aleatorias normales independientes.

Para calcular la pdf conjunta de X e Y consideremos la transformación lineal siguiente:

$$\begin{aligned} X &= X & \text{cuya inversa es} & & X &= X \\ Y &= X + N & & & N &= -X + Y. \end{aligned}$$

De la ecuación (5.52) tenemos:

$$\begin{aligned} f_{X,Y}(x,y) &= \frac{f_{X,N}(x,y)}{\det A} \Big|_{x=x, n=y-x} = \frac{e^{-x^2/2\sigma_X^2} e^{-n^2/2\sigma_N^2}}{\sqrt{2\pi}\sigma_X \sqrt{2\pi}\sigma_N} \Big|_{x=x, n=y-x} \\ &= \frac{e^{-x^2/2\sigma_X^2} e^{-(y-x)^2/2\sigma_N^2}}{\sqrt{2\pi}\sigma_X \sqrt{2\pi}\sigma_N} \end{aligned}$$

La pdf condicionada de la señal X dado la observación Y es, entonces:

$$\begin{aligned} f_X(x|y) &= \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{e^{-x^2/2\sigma_X^2} e^{-(y-x)^2/2\sigma_N^2} \sqrt{2\pi}\sigma_Y}{\sqrt{2\pi}\sigma_X \sqrt{2\pi}\sigma_N e^{-y^2/2\sigma_Y^2}} \\ &= \frac{\exp\left\{-\frac{1}{2}\left(\left(\frac{x}{\sigma_X}\right)^2 + \left(\frac{y-x}{\sigma_N}\right)^2 - \left(\frac{y}{\sigma_Y}\right)^2\right)\right\}}{\sqrt{2\pi}\sigma_N\sigma_X/\sigma_Y} = \frac{\exp\left\{-\frac{1}{2}\frac{\sigma_Y^2}{\sigma_X^2\sigma_N^2}\left(x - \frac{\sigma_X^2}{\sigma_X^2 + \sigma_N^2}y\right)^2\right\}}{\sqrt{2\pi}\sigma_N\sigma_X/\sigma_Y} \\ &= \frac{\exp\left\{-\frac{1}{2(1-\rho_{X,Y}^2)\sigma_X^2}\left(x - \left(\frac{\sigma_X^2}{\sigma_X^2 + \sigma_N^2}\right)y\right)^2\right\}}{\sqrt{1-\rho_{X,Y}^2}\sigma_X}. \end{aligned}$$

Esta pdf alcanza su máximo cuando el argumento del exponente es cero, es decir,

$$x = \left(\frac{\sigma_X^2}{\sigma_X^2 + \sigma_N^2}\right)y = \left(\frac{1}{1 + \frac{\sigma_N^2}{\sigma_X^2}}\right)y.$$

La relación señal-ruido (SNR) se define como el cociente entre la varianza de X y la varianza de N . Para SNR alta este estimador da $x \approx y$, y a muy baja relación señal-ruido, da $x \approx 0$.

Ejemplo 5.48 Rotación de una variable aleatoria Normal bivalente

La elipse correspondiente a una normal bivalente forma un ángulo

$$\theta = \frac{1}{2} \arctan \left(\frac{2\rho\sigma_1\sigma_2}{\sigma_1^2 - \sigma_2^2} \right)$$

en relación con el eje x . Supongamos que se define un nuevo sistema de coordenadas cuyos ejes están alineados con los de la elipse, como se muestra en la figura 5.27. Esto se logra mediante el uso de la matriz de rotación siguiente:

$$\begin{bmatrix} V \\ W \end{bmatrix} = \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix}.$$

Para demostrar que las nuevas variables aleatorias son independientes basta con demostrar que tienen covarianza cero:

$$\begin{aligned} Cov(X, Y) &= E[(V - E[V])(W - E[W])] \\ &= E[\{(X - m_1) \cos \theta + (Y - m_2) \sin \theta\} \\ &\quad \times \{-(X - m_1) \sin \theta + (Y - m_2) \cos \theta\}] \\ &= -\sigma_1^2 \sin \theta \cos \theta + Cov(X, Y) \cos^2 \theta \\ &\quad - Cov(X, Y) \sin^2 \theta + \sigma_2^2 \sin \theta \cos \theta \\ &= \frac{(\sigma_2^2 - \sigma_1^2) \sin 2\theta + 2Cov(X, Y) \cos 2\theta}{2} \\ &= \frac{\cos 2\theta [(\sigma_2^2 - \sigma_1^2) \tan 2\theta + 2Cov(X, Y)]}{2}. \end{aligned}$$

Si hacemos que el ángulo de rotación θ sea tal que

$$\tan 2\theta = \frac{2Cov(X, Y)}{\sigma_1^2 - \sigma_2^2},$$

entonces la covarianza de V y W es igual a cero como se requería.

Problemas

5.1 Sea X el máximo e Y el mínimo del número total de caras obtenidas por Carlos y Miguel al lanzar una moneda equilibrada dos veces cada uno.

- Describe el espacio S subyacente de este experimento aleatorio y muestra el rango del par (X, Y) .
- Calcula las probabilidades de todos los valores de (X, Y) .
- Calcula $P[X = Y]$.
- Repite las partes b y c considerando que la moneda de Carlos está trucada, con una $P[C] = 3/4$.

5.2 Sea H y W la estatura y el peso en el ejemplo 5.1. El índice de masa corporal es una medida de la grasa corporal y se define como $bmi = W/H^2$ donde W está en kilogramos y H en metros. Determina y dibuja en el plano los

siguientes sucesos: $A = \{\text{"obeso"} , bmi \geq 30\}$; $B = \{\text{"sobrepeso"} , 25 \leq bmi < 30\}$; $C = \{\text{"normal"} , 18.5 \leq bmi < 25\}$ y $A = \{\text{"por debajo"} , bmi < 18.5\}$.

5.3 (a) Calcula y dibuja $p_{X,Y}(x,y)$ del problema 5.1 cuando la moneda es equilibrada.

(b) Calcula $p_X(x)$ y $p_Y(y)$.

(c) Repite la parte a) y la b) si Carlos usa la moneda trucada.

5.4 (a) Dibuja la cdf conjunta del par (X,Y) del problema 5.1 y verifica que las propiedades de la cdf conjunta se satisfacen. Encontrarás útil dividir primero el plano en regiones donde la cdf es constante.

(b) Calcula la cdf marginal de X e Y .

5.(a) El par (X,Y) tiene cdf conjunta dada por:

$$F_{X,Y}(x,y) = \begin{cases} (1 - 1/x^2)(1 - 1/y^2) & x > 1, y > 1 \\ 0 & \text{en otro caso.} \end{cases}$$

(a) Dibuja la cdf conjunta.

(b) Calcula la cdf marginal de X y de Y .

(c) Calcula la probabilidad de los siguientes sucesos: $\{X < 3, Y \leq 5\}$, $\{X > 4, Y > 3\}$.

5.5 ¿Es válida la siguiente cdf? ¿Por qué?

$$F_{X,Y}(x,y) = \begin{cases} (1 - 1/x^2y^2) & x > 1, y > 1 \\ 0 & \text{en otro caso.} \end{cases}$$

5.6 El número N de usuarios que han iniciado sesión en un sistema y el tiempo T hasta que el siguiente usuario termina su sesión tienen una probabilidad conjunta dada por:

$$P[N = n, T \leq t] = (1 - p)p^{n-1}(1 - e^{-n\lambda t}) \quad n = 1, 2, \dots \quad t > 0.$$

(a) Dibuja la probabilidad conjunta.

(b) Calcula la pmf marginal de N .

(c) Calcula la cdf marginal de T .

(d) Calcula $P[N \leq 3, T > 3/\lambda]$.

5.7 Sea la pdf conjunta de X e Y :

$$f_{X,Y}(x,y) = k(x+y) \quad 0 \leq x \leq 1, 0 \leq y \leq 1.$$

(a) Calcula k .

- (b) Calcula la cdf conjunta de (X, Y) .
- (c) Calcula la pdf marginal de X y de Y .
- (d) Calcula $P[X < Y]$, $P[Y < X^2]$, $P[X + Y > 0.5]$.

5.8 Sea X el número de pares enteros de puntos observados en la tirada de un dado y sea Y el resto de puntos. ¿Son X e Y variables aleatorias independientes?

5.9 Marta toma el autobús de las 7:30 todas las mañanas. El tiempo de llegada del autobús a la parada es uniforme en el intervalo $[7:27, 7:37]$. La llegada de Marta a la parada de autobús es también uniforme en el intervalo $[7:25, 7:40]$. Asumiendo que las llegadas de Marta y del autobús son independientes.

- (a) ¿Cuál es la probabilidad de que Marta llegue más de cinco minutos antes que el autobús?
- (b) ¿Cuál es la probabilidad de que Marta pierda el autobús?

5.10 ¿Son independientes X e Y del problema 5.4?

5.11 ¿Son independientes X e Y del problema 5.7?

- 5.12** (a) Calcula $E[(X + Y)^2]$.
- (b) Calcula la varianza de $X + Y$.
 - (c) ¿Bajo qué condición es la varianza de la suma igual a la suma de las varianzas individuales?

5.13 Sea N y T el número de usuarios que iniciaron sesión y el tiempo hasta el siguiente cierre de sesión del problema 5.6. Calcula la correlación y la covarianza de N y T e indica si las variables aleatorias son independientes, ortogonales y/o incorreladas.

5.14 Sea X e Y dos variables aleatorias Normales independientes con media 0 y varianza 1. Demuestra que $Z = X/Y$ es una variable aleatoria de Cauchy.

5.15 Las variables aleatorias X e Y tienen pdf conjunta:

$$f_{X,Y}(x, y) = c \sin(x + y) \quad 0 \leq x \leq \pi/2, 0 \leq y \leq \pi/2.$$

- (a) Calcula el valor de la constante c .
- (b) Calcula la cdf conjunta de X e Y .
- (c) Calcula las pdf marginales de X y de Y .
- (d) Calcula la media, varianza y covarianza de X e Y .

5.16 Un inspector selecciona un producto para su inspección de acuerdo con el resultado del lanzamiento al aire de una moneda: el producto es inspeccionado si el resultado es cara. Supongamos que el tiempo entre llegadas de los productos

es una variable aleatoria exponencial con una media uno. Supongamos que el tiempo para inspeccionar un producto es un valor constante t .

- (a) Calcula la pmf para el número de llegadas de productos entre inspecciones consecutivas.
- (b) Calcula la pdf del tiempo entre inspecciones, X . Sugerencia: Utiliza la esperanza condicionada.
- (c) Determine el valor de p , de modo que con una probabilidad del 90 %, una inspección se completa antes de que el siguiente punto sea seleccionado para su inspección.

5.17 El tiempo de vida X de un dispositivo es una variable aleatoria exponencial con media $= 1/R$. Supongamos que, debido a irregularidades en el proceso de producción, el parámetro R es aleatorio y tiene una distribución gamma.

- (a) Calcula la pdf conjunta de X y R .
- (b) Calcula la pdf de X .
- (c) Calcula la media y la varianza de X .

5.18 Sea X e Y muestras de una señal aleatoria en dos instantes de tiempo. Supongamos que X e Y son variables aleatorias Normales independientes de media cero con la misma varianza. Cuando la señal “0” está presente la varianza es σ_0^2 y cuando la señal “1” se presenta la varianza es $\sigma_1^2 > \sigma_0^2$. Supongamos que las señales 0 y 1 ocurren con probabilidades p y $1 - p$, respectivamente. Sea $R^2 = X^2 + Y^2$ la energía total de las dos observaciones.

- (a) Calcula la pdf de R^2 cuando la señal 0 está presente y cuando signal 1 está presente. Calcula la pdf de R^2 .
- (b) Supongamos que usamos la siguiente regla de “detección de señales”: si $R^2 > T$, decidimos que la señal 1 está presente, de lo contrario, decidimos que es la señal 0. Calcula la expresión para la probabilidad de error en términos de T .
- (c) Determine el valor de T que minimiza la probabilidad de error.

5.19 Sea U_0, U_1, \dots una secuencia de variables aleatorias Normales independientes de media cero y varianza uno. Un “filtro de paso bajo” toma la secuencia U_i y produce como salida la secuencia $X_n = (U_n + U_{n-1})/2$ y un “filtro de paso alto” produce la secuencia $Y_n = (U_n - U_{n-1})/2$ como salida.

- (a) Calcula la pdf conjunta de X_n y X_{n-1} ; de X_n y X_{n+m} , $m > 1$.
- (b) Repite la parte a) para Y_n .
- (c) Calcula la pdf conjunta de X_n e Y_m .

6 Variables aleatorias multidimensionales

En el capítulo anterior hemos presentado métodos para trabajar con dos variables aleatorias. En este capítulo se extienden estos métodos para el caso de n variables aleatorias de las siguientes maneras:

- Mediante la representación de n variables aleatorias como un vector se obtiene una notación compacta para las pmf, cdf y pdf conjuntas, así como para las distribuciones marginales y condicionadas.
- Presentamos un método general para calcular la pdf de transformaciones de variables aleatorias multivariantes.
- La descripción de la distribución de una variable aleatoria multivariante es dada por un vector de valor esperado y una matriz de covarianza.
- Utilizamos transformaciones lineales y funciones características para calcular representaciones alternativas de vectores aleatorios y sus probabilidades.
- Desarrollamos estimadores óptimos para estimar el valor de una variable aleatoria basado en las observaciones de otras variables aleatorias.
- Se muestra cómo la variable aleatoria Normal multivariante tiene una función característica y una pdf compactas y fáciles de utilizar.

6.1 Variables aleatorias multidimensionales

La noción de variable aleatoria se generaliza fácilmente al caso en que varias cantidades sean de interés. Una **variable aleatoria multivariante** \mathbf{X} es una función que asigna un vector de números reales a cada elemento $\zeta \in S$ del espacio muestral de un experimento aleatorio. Utilizaremos letras mayúsculas en negrita para las variables aleatorias vectores. Por convención, \mathbf{X} es un vector columna (n filas por una columna), por lo que la variable aleatoria multidimensional con componentes X_1, X_2, \dots, X_n se corresponde con

$$\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} = [X_1, X_2, \dots, X_n]^T,$$

donde “ T ” denota la traspuesta de una matriz o un vector. De vez en cuando escribiremos $\mathbf{X} = (X_1, X_2, \dots, X_n)$ para ahorrar espacio y omitiremos la traspuesta excepto cuando trabajemos con matrices. Los posibles valores que toma la variable aleatoria vectorial se denotan como $\mathbf{x} = (x_1, x_2, \dots, x_n)$ donde x_i se corresponde con el valor de X_i .

Ejemplo 6.1 Llegadas a un conmutador de paquetes

Los paquetes llegan a cada uno de los tres puertos de entrada de un conmutador de paquetes de acuerdo a ensayos de Bernoulli independientes con $p = 1/2$. Cada paquete que llega tiene la misma probabilidad de ser destinado a cualquiera de los tres puertos de salida. Sea $\mathbf{X} = (X_1, X_2, X_3)$ donde X_i es el número total de paquetes que llegan al puerto de salida i . \mathbf{X} es una variable aleatoria multidimensional cuyos valores son determinados por el patrón de llegadas a los puertos de entrada.

Ejemplo 6.2 Conteos de Poisson multivariantes

Un experimento aleatorio consiste en contar el número de defectos en un chip semiconductor y en identificar su ubicación. El resultado de este experimento consiste en el vector $\zeta = (n, \mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n)$, donde el primer componente especifica el número total de defectos y el resto de los componentes especifican las coordenadas de su ubicación. Supongamos que el chip se compone de M regiones. Sea $N_1(\zeta), N_2(\zeta), \dots, N_M(\zeta)$ el número de defectos en cada una de estas regiones, es decir, $N_k(\zeta)$ es el número de \mathbf{y} que están en la región k . El vector $\mathbf{N}(\zeta) = (N_1, N_2, \dots, N_M)$ es entonces un vector aleatorio.

Ejemplo 6.3 Muestras de una señal de audio

Sea la señal de audio $X(t)$ el resultado ζ de un experimento aleatorio. Sea la variable aleatoria $X_k = X(kT)$ la muestra de la señal tomada en el instante kT . Un codec mp3 procesa el audio en bloques de n muestras $\mathbf{X} = (X_1, X_2, \dots, X_n)$. \mathbf{X} es una variable aleatoria multivariante.

6.1.1 Sucesos y probabilidades

A cada suceso A relacionado con $\mathbf{X} = (X_1, X_2, \dots, X_n)$ le corresponde una región en el espacio real de n dimensiones \mathbb{R}^n . Como antes, utilizaremos conjuntos “rectangulares” en forma producto como nuestros elementos básicos de construcción. Para la variable aleatoria multidimensional $\mathbf{X} = (X_1, X_2, \dots, X_n)$, estamos interesados en sucesos en su **forma producto**

$$A = \{X_1 \in A_1\} \cap \{X_2 \in A_2\} \cap \dots \cap \{X_n \in A_n\}, \quad (6.1)$$

donde cada A_k es un suceso de dimensión uno (i.e., un subconjunto de la recta real) relacionado sólo con X_k . El suceso A se verifica cuando todos los sucesos $\{X_k \in A_k\}$ ocurren.

Estamos interesados en calcular las probabilidades de estos sucesos en forma producto:

$$\begin{aligned} P[A] &= P[\mathbf{X} \in A] = P[\{X_1 \in A_1\} \cap \{X_2 \in A_2\} \cap \dots \cap \{X_n \in A_n\}] \\ &\triangleq P[X_1 \in A_1, X_2 \in A_2, \dots, X_n \in A_n]. \end{aligned} \quad (6.2)$$

En principio, la probabilidad en la ecuación (6.2) se obtiene calculando la probabilidad de los sucesos equivalentes en el espacio muestral subyacente, esto es,

$$\begin{aligned} P[A] &= P[\{\zeta \in S : \mathbf{X}(\zeta) \in A\}] \\ &= P[\{\zeta \in S : X_1(\zeta) \in A_1, X_2(\zeta) \in A_2, \dots, X_n(\zeta) \in A_n\}]. \end{aligned} \quad (6.3)$$

La ecuación (6.2) es la base de la definición de la función de masa de probabilidad, la función de distribución y la función de densidad conjuntas de n dimensiones. Las probabilidades de otros sucesos se pueden expresar en términos de estas tres funciones.

6.1.2 Funciones de distribución conjuntas

La **función de distribución conjunta** de X_1, X_2, \dots, X_n se define como la probabilidad de un rectángulo semi-infinito n -dimensional asociado al punto (x_1, \dots, x_n) :

$$F_X(\mathbf{x}) \triangleq F_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) = P[X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n]. \quad (6.4)$$

La cdf conjunta está definida para variables aleatorias discretas, continuas y de tipo mixto. La probabilidad de sucesos en forma producto se puede expresar en términos de la cdf conjunta.

La cdf conjunta genera una familia de **cdf marginales** para las subcolecciones de las variables aleatorias X_1, \dots, X_n . Estas cdf marginales se calculan colocando $+\infty$ en las entradas apropiadas en la ecuación (6.4). Por ejemplo:

La cdf conjunta de X_1, \dots, X_{n-1} viene dada por

$$F_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_{n-1}, \infty)$$

y la cdf conjunta de X_1 , y X_2 viene dada por

$$F_{X_1, X_2, \dots, X_n}(x_1, x_2, \infty, \dots, \infty).$$

Ejemplo 6.4 Un transmisor de radio envía una señal a un receptor utilizando tres caminos. Sea X_1 , X_2 y X_3 las señales que llegan al receptor a través de cada camino. Calcula $P[\text{máx}(X_1, X_2, X_3) \leq 5]$.

Solución

El máximo de los tres números es menor que 5 si y sólo si cada uno de los tres números es menor que 5; entonces

$$\begin{aligned} P[A] &= P[\{X_1 \leq 5\} \cap \{X_2 \leq 5\} \cap \{X_3 \leq 5\}] \\ &= F_{X_1, X_2, X_3}(5, 5, 5). \end{aligned}$$

La **función de masa de probabilidad conjunta** de n variables aleatorias discretas se define por

$$p_{\mathbf{X}}(\mathbf{x}) \triangleq p_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) = P[X_1 = x_1, X_2 = x_2, \dots, X_n = x_n]. \quad (6.5)$$

La probabilidad de cualquier evento n -dimensional A se calcula sumando la pmf de todos los puntos del suceso

$$P[\mathbf{X} \in A] = \sum_{\mathbf{x} \in A} \cdots \sum p_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n). \quad (6.6)$$

La pdf conjunta genera una familia de **pmf marginales** que especifican las probabilidades conjuntas para subcolecciones de las n variables aleatorias. Por ejemplo, la pmf de una dimensión de X_j se calcula sumando la pmf conjunta en todas las variables excepto x_j :

$$p_{X_j}(x_j) = P[X_j = x_j] = \sum_{x_1} \cdots \sum_{x_{j-1}} \sum_{x_{j+1}} \cdots \sum_{x_n} p_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n). \quad (6.7)$$

La pmf conjunta bidimensional de cualquier par X_j y X_k se calcula sumando la pmf en todo el resto de variables y así sucesivamente. Entonces, la pmf marginal para X_1, \dots, X_{n-1} viene dada por

$$p_{X_1, \dots, X_{n-1}}(x_1, \dots, x_{n-1}) = \sum_{X_n} p_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n). \quad (6.8)$$

Se obtiene una familia de **pmf condicionadas** condicionando la pmf conjunta a diferentes subcolecciones de variables aleatorias. Por ejemplo, si $p_{X_1, \dots, X_{n-1}}(x_1, \dots, x_{n-1}) > 0$:

$$p_{X_n}(x_n | x_1, \dots, x_{n-1}) = \frac{p_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n)}{p_{X_1, \dots, X_{n-1}}(x_1, \dots, x_{n-1})}. \quad (6.9a)$$

Repetidas aplicaciones de la ecuación (6.9a) llevan a la siguiente (muy útil) expresión:

$$\begin{aligned} p_{X_1, \dots, X_n}(x_1, \dots, x_n) &= \\ &= p_{X_n}(x_n | x_1, \dots, x_{n-1}) p_{X_{n-1}}(x_{n-1} | x_1, \dots, x_{n-2}) \cdots p_{X_2}(x_2 | x_1) p_{X_1}(x_1). \end{aligned} \quad (6.9b)$$

Ejemplo 6.5 Llegadas a un conmutador de paquetes

Calcula la pmf conjunta de $\mathbf{X} = (X_1, X_2, X_3)$ del ejemplo 6.1. Calcula $P[X_1 > X_3]$.

Solución

Sea N el número total de paquetes que llegan a los tres puertos de entrada. Cada puerto de entrada tiene una llegada con probabilidad $p = 1/2$, entonces N es binomial con pmf:

$$p_N(n) = \binom{3}{n} \frac{1}{2^3} \text{ para } 0 \leq n \leq 3.$$

Dado que $N = n$, el número de paquetes que llegan a cada puerto de salida tiene una distribución multinomial:

$$p_{X_1, X_2, X_3}(i, j, k | i+j+k = n) = \begin{cases} \frac{n!}{i!j!k!} \frac{1}{3^n} & \text{para } i+j+k = n, i \geq 0, j \geq 0, k \geq 0 \\ 0 & \text{en otro caso.} \end{cases}$$

La pmf conjunta de \mathbf{X} es:

$$p_{\mathbf{X}}(0, 0, 0) = \frac{0!}{0!0!0!} \frac{1}{3^0} \binom{3}{0} \frac{1}{2^3} = \frac{1}{8}$$

$$p_{\mathbf{X}}(1, 0, 0) = p_{\mathbf{X}}(0, 1, 0) = p_{\mathbf{X}}(0, 0, 1) = \frac{1!}{0!0!1!} \frac{1}{3^1} \binom{3}{1} \frac{1}{2^3} = \frac{3}{24}$$

$$p_{\mathbf{X}}(1, 1, 0) = p_{\mathbf{X}}(1, 0, 1) = p_{\mathbf{X}}(0, 1, 1) = \frac{2!}{0!1!1!} \frac{1}{3^2} \binom{3}{2} \frac{1}{2^3} = \frac{6}{72}$$

$$p_{\mathbf{X}}(2, 0, 0) = p_{\mathbf{X}}(0, 2, 0) = p_{\mathbf{X}}(0, 0, 2) = 3/72$$

$$p_{\mathbf{X}}(1, 1, 1) = 6/216$$

$$p_{\mathbf{X}}(0, 1, 2) = p_{\mathbf{X}}(0, 2, 1) = p_{\mathbf{X}}(1, 0, 2) = p_{\mathbf{X}}(1, 2, 0) = p_{\mathbf{X}}(2, 0, 1) = p_{\mathbf{X}}(2, 1, 0) = 3/216$$

$$p_{\mathbf{X}}(3, 0, 0) = p_{\mathbf{X}}(0, 3, 0) = p_{\mathbf{X}}(0, 0, 3) = 1/216$$

Por último:

$$\begin{aligned} P[X_1 > X_3] &= p_{\mathbf{X}}(1, 0, 0) + p_{\mathbf{X}}(1, 1, 0) + p_{\mathbf{X}}(2, 0, 0) + p_{\mathbf{X}}(1, 2, 0) \\ &\quad + p_{\mathbf{X}}(2, 0, 1) + p_{\mathbf{X}}(2, 1, 0) + p_{\mathbf{X}}(3, 0, 0) \\ &= 8/27. \end{aligned}$$

Decimos que las variables aleatorias X_1, X_2, \dots, X_n son **variables aleatorias conjuntamente continuas** si la probabilidad de cualquier suceso n -dimensional A es dada por una integral de n dimensiones de una función de densidad:

$$P[\mathbf{X} \in A] = \int_{\mathbf{x} \in A} \cdots \int f_{X_1, \dots, X_n}(x'_1, \dots, x'_n) dx'_1 \cdots dx'_n, \quad (6.10)$$

donde $f_{X_1, \dots, X_n}(x_1, \dots, x_n)$ es la **función de densidad conjunta**.

La cdf conjunta de \mathbf{X} se obtiene a partir de la pdf conjunta integrando:

$$F_{\mathbf{X}}(\mathbf{x}) = F_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f_{X_1, \dots, X_n}(x'_1, \dots, x'_n) dx'_1 \cdots dx'_n. \quad (6.11)$$

La pdf conjunta (si existe la derivada) viene dada por

$$f_{\mathbf{X}}(\mathbf{x}) \triangleq f_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) = \frac{\delta^n}{\delta x_1 \cdots \delta x_n} F_{X_1, \dots, X_n}(x_1, \dots, x_n). \quad (6.12)$$

Una familia de **pdf marginales** se asocia con la pdf conjunta de la ecuación (6.12). La pdf marginal para un subconjunto de variables aleatorias se obtiene integrando en las otras variables. Por ejemplo, la pdf marginal de X_1 es

$$f_{X_1}(x_1) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1, \dots, X_n}(x_1, x'_2, \dots, x'_n) dx'_2 \cdots dx'_n. \quad (6.13)$$

Como otro ejemplo, la pdf marginal de X_1, \dots, X_{n-1} viene dada por

$$f_{X_1, \dots, X_{n-1}}(x_1, \dots, x_{n-1}) = \int_{-\infty}^{\infty} f_{X_1, \dots, X_n}(x_1, \dots, x_{n-1}, x'_n) dx'_n. \quad (6.14)$$

Una familia de **pdf condicionadas** también se asocia con la pdf conjunta. Por ejemplo, la pdf de X_n dados los valores de X_1, \dots, X_{n-1} viene dada por

$$f_{X_n}(x_n | x_1, \dots, x_{n-1}) = \frac{f_{X_1, \dots, X_n}(x_1, \dots, x_n)}{f_{X_1, \dots, X_{n-1}}(x_1, \dots, x_{n-1})} \quad (6.15a)$$

si $f_{X_1, \dots, X_{n-1}}(x_1, \dots, x_{n-1}) > 0$.

Repitiendo aplicaciones de la ecuación (6.15a) llegamos a la expresión análoga de la ecuación (6.9b):

$$f_{X_1, \dots, X_n}(x_1, \dots, x_n) = f_{X_n}(x_n | x_1, \dots, x_{n-1}) f_{X_{n-1}}(x_{n-1} | x_1, \dots, x_{n-2}) \cdots f_{X_2}(x_2 | x_1) f_{X_1}(x_1). \quad (6.15b)$$

Ejemplo 6.6 Las variables aleatorias X_1 , X_2 y X_3 tienen la pdf Normal

$$f_{X_1, X_2, X_3}(x_1, x_2, x_3) = \frac{e^{-(x_1^2 + x_2^2 - \sqrt{2}x_1x_2 + 1/2x_3^2)}}{2\pi\sqrt{\pi}}.$$

Calcula la pdf marginal de X_1 y X_3 . Calcula la pdf condicionada de X_2 dados X_1 y X_3 .

Solución

La pdf marginal para el par X_1 y X_3 se calcula integrando la pdf conjunta en x_2 :

$$f_{X_1, X_3}(x_1, x_3) = \frac{e^{-x_3^2/2}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{e^{-(x_1^2 + x_2^2 - \sqrt{2}x_1x_2)}}{2\pi/\sqrt{2}} dx_2.$$

La integral de arriba se resolvió en el ejemplo 5.18 con $\rho = -1/\sqrt{2}$. Sustituyendo el resultado de la integral, obtenemos

$$f_{X_1, X_3}(x_1, x_3) = \frac{e^{-x_3^2/2} e^{-x_1^2/2}}{\sqrt{2\pi} \sqrt{2\pi}}.$$

Entonces, X_1 y X_3 son variables aleatorias normales independientes con media cero y varianza 1.

La pdf condicionada de X_2 dadas X_1 y X_3 es:

$$\begin{aligned} f_{X_2}(x_2|x_1, x_3) &= \frac{e^{-(x_1^2+x_2^2-\sqrt{2}x_1x_2+1/2x_3^2)}}{2\pi\sqrt{\pi}} \frac{\sqrt{2\pi}\sqrt{2\pi}}{e^{-x_3^2/2}e^{-x_1^2/2}} \\ &= \frac{e^{-(1/2x_1^2+x_2^2-\sqrt{2}x_1x_2)}}{\sqrt{\pi}} = \frac{e^{-(x_2-x_1/\sqrt{2})^2}}{\sqrt{\pi}}. \end{aligned}$$

Concluimos que X_2 dadas X_1 y X_3 es una variable aleatoria Normal con media $x_1/\sqrt{2}$ y varianza $1/2$.

Ejemplo 6.7 Secuencia multiplicativa

Sea X_1 uniforme en el intervalo $[0, 1]$, X_2 uniforme en $[0, X_1]$ y X_3 uniforme en $[0, X_2]$. (Nótese que X_3 también es el producto de tres variables aleatorias uniformes.) Calcula la pdf conjunta de \mathbf{X} y la pdf marginal de X_3 .

Solución

Para $0 < z < y < x < 1$, la pdf conjunta es no nula y dada por:

$$f_{X_1, X_2, X_3}(x_1, x_2, x_3) = f_{X_3}(z|x, y) f_{X_2}(y|x) f_{X_1}(x) = \frac{1}{y} \frac{1}{x} 1 = \frac{1}{xy}.$$

La pdf conjunta de X_2 y X_3 es no nula para $0 < z < y < 1$ y se calcula integrando x entre y y 1:

$$f_{X_2, X_3}(x_2, x_3) = \int_y^1 \frac{1}{xy} dx = \frac{1}{y} \ln x \Big|_y^1 = \frac{1}{y} \ln \frac{1}{y}.$$

Obtenemos la pdf de X_3 integrando y entre z y 1:

$$f_{X_3}(x_3) = - \int_z^1 \frac{1}{y} \ln y dy = - \frac{1}{2} (\ln y)^2 \Big|_z^1 = \frac{1}{2} (\ln z)^2.$$

Nótese que la pdf de X_3 se concentra alrededor de valores cercanos a $x = 0$.

6.1.3 Independencia

La colección de variables aleatorias X_1, \dots, X_n es **independiente** si

$$P[X_1 \in A_1, X_2 \in A_2, \dots, X_n \in A_n] = P[X_1 \in A_1] P[X_2 \in A_2] \dots P[X_n \in A_n]$$

para *cualquier* conjunto de sucesos unidimensionales A_1, \dots, A_n . Se puede probar que X_1, \dots, X_n son independientes si y sólo si

$$F_{X_1, \dots, X_n}(x_1, \dots, x_n) = F_{X_1}(x_1) \dots F_{X_n}(x_n) \quad (6.16)$$

para todo x_1, \dots, x_n . Si las variables aleatorias son discretas, la ecuación (6.16) es equivalente a

$$p_{X_1, \dots, X_n}(x_1, \dots, x_n) = p_{X_1}(x_1) \dots p_{X_n}(x_n) \quad \forall x_1, \dots, x_n.$$

Si las variables aleatorias son conjuntamente continuas, la ecuación (6.16) es equivalente a

$$f_{X_1, \dots, X_n}(x_1, \dots, x_n) = f_{X_1}(x_1) \dots f_{X_n}(x_n) \quad \forall x_1, \dots, x_n.$$

Ejemplo 6.8 Las n muestras X_1, \dots, X_n de una señal ruidosa tienen una pdf conjunta dada por

$$f_{X_1, \dots, X_n}(x_1, \dots, x_n) = \frac{e^{-(x_1^2 + \dots + x_n^2)/2}}{(2\pi)^{n/2}} \quad \forall x_1, \dots, x_n.$$

Claramente, la expresión anterior es el producto de n pdf normales de una dimensión. Entonces, X_1, \dots, X_n son variables aleatorias normales independientes.

6.2 Funciones de varias variables aleatorias

Las funciones de varias variables aleatorias surgen de forma natural en experimentos aleatorios. Por ejemplo, $\mathbf{X} = (X_1, X_2, \dots, X_n)$ puede corresponder a las observaciones de n repeticiones de un experimento que genera una variable aleatoria dada. Casi siempre nos interesarán la media y la varianza muestral de las observaciones. En otro ejemplo $\mathbf{X} = (X_1, X_2, \dots, X_n)$ se puede corresponder con muestras a muestras de una señal del habla y podemos estar interesados en la extracción de características que se definen como funciones de \mathbf{X} para su uso en un sistema de reconocimiento de voz.

6.2.1 Una función de varias variables aleatorias

Sea Z la variable aleatoria definida como una función de varias variables aleatorias:

$$Z = g(X_1, X_2, \dots, X_n). \quad (6.17)$$

La cdf de Z se calcula mediante el suceso equivalente $\{Z \leq z\}$, esto es, el conjunto $R_z = \{\mathbf{x} : g(\mathbf{x}) \leq z\}$, entonces

$$F_Z(z) = P[\mathbf{X} \in R_z] = \int_{\mathbf{x} \in R_z} \dots \int f_{X_1, \dots, X_n}(x'_1, \dots, x'_n) dx'_1 \dots dx'_n. \quad (6.18)$$

La pdf de Z entonces se calcula tomando la derivada de $F_Z(z)$.

Ejemplo 6.9 Máximo y mínimo de n variables aleatorias

Sea $W = \max(X_1, X_2, \dots, X_n)$ y $Z = \min(X_1, X_2, \dots, X_n)$, donde las X_i son variables aleatorias independientes con la misma distribución. Calcula $F_W(w)$ y $F_Z(z)$.

Solución

El máximo de X_1, X_2, \dots, X_n es menor que x si y sólo si cada X_i es menor que x , por lo que:

$$\begin{aligned} F_W(w) &= P[\max(X_1, X_2, \dots, X_n) \leq w] \\ &= P[X_1 \leq w]P[X_2 \leq w] \dots P[X_n \leq w] = (F_X(w))^n. \end{aligned}$$

El mínimo de X_1, X_2, \dots, X_n es mayor que x si y sólo si cada X_i es mayor que x , por lo que:

$$\begin{aligned} 1 - F_Z(z) &= P[\min(X_1, X_2, \dots, X_n) > z] \\ &= P[X_1 > z]P[X_2 > z] \dots P[X_n > z] = (1 - F_X(z))^n \end{aligned}$$

y

$$F_Z(z) = 1 - (1 - F_X(z))^n.$$

Ejemplo 6.10 Ejemplo de fusión de llegadas independientes de Poisson

Las peticiones de páginas web llegan a un servidor desde n fuentes independientes. La fuente j genera paquetes con distribución exponencial con una tasa de tiempo entre llegadas λ_j . Calcula la distribución de los tiempos entre llegadas consecutivas al servidor.

Solución

Sea X_1, X_2, \dots, X_n los tiempos entre llegadas de las diferentes fuentes. Cada X_j satisface la propiedad de no tener memoria, por lo que el tiempo que ha transcurrido desde la última llegada de cada fuente es irrelevante. El tiempo hasta la próxima llegada del multiplexor es entonces:

$$Z = \min(X_1, X_2, \dots, X_n).$$

Por lo tanto, la pdf de Z es la siguiente:

$$\begin{aligned} 1 - F_Z(z) &= P[\min(X_1, X_2, \dots, X_n) > z] \\ &= P[X_1 > z]P[X_2 > z] \dots P[X_n > z] \\ &= (1 - F_{X_1}(z))(1 - F_{X_2}(z)) \dots (1 - F_{X_n}(z)) \\ &= e^{-\lambda_1 z} e^{-\lambda_2 z} \dots e^{-\lambda_n z} = e^{-(\lambda_1 + \lambda_2 + \dots + \lambda_n)z}. \end{aligned}$$

El tiempo entre llegadas es una variable aleatoria Exponencial con tasa $\lambda_1 + \lambda_2 + \dots + \lambda_n$.

Ejemplo 6.11 Fiabilidad de sistemas redundantes

Un cluster de ordenadores tiene n subsistemas independientes y redundantes. Cada subsistema tiene un tiempo de vida distribuido exponencialmente con parámetro λ . El grupo funcionará siempre y cuando al menos un subsistema está funcionando. Calcular la cdf del tiempo hasta el fallo del sistema.

Solución

Sea X_1, X_2, \dots, X_n el tiempo de vida de cada subsistema. El tiempo hasta el fallo del último subsistema es:

$$W = \max(X_1, X_2, \dots, X_n).$$

Entonces la cdf de W es:

$$F_W(w) = (F_X(w))^n = (1 - e^{-\lambda w})^n = 1 - \binom{n}{1}e^{-\lambda w} + \binom{n}{2}e^{-2\lambda w} + \dots$$

6.2.2 Transformaciones de vectores aleatorios

Sean X_1, X_2, \dots, X_n variables aleatorias de un experimento y sean Z_1, Z_2, \dots, Z_n variables aleatorias definidas por una transformación que consiste en n funciones de $\mathbf{X} = (X_1, \dots, X_n)$:

$$Z_1 = g_1(\mathbf{X}) \quad Z_2 = g_2(\mathbf{X}) \quad \dots \quad Z_n = g_n(\mathbf{X}).$$

La cdf conjunta de $\mathbf{Z} = (Z_1, \dots, Z_n)$ en el punto $\mathbf{z} = (z_1, \dots, z_n)$ es igual a la probabilidad de la región de \mathbf{x} donde $g_k(\mathbf{x}) \leq z_k$ para $k = 1, \dots, n$:

$$F_{Z_1, \dots, Z_n}(z_1, \dots, z_n) = P[g_1(\mathbf{X}) \leq z_1, \dots, g_n(\mathbf{X}) \leq z_n]. \quad (6.19a)$$

Si X_1, \dots, X_n tienen pdf conjunta, entonces

$$F_{Z_1, \dots, Z_n}(z_1, \dots, z_n) = \int \dots \int_{\mathbf{x}' : g_k(\mathbf{x}') \leq z_k} f_{X_1, \dots, X_n}(x'_1, \dots, x'_n) dx'_1 \dots dx'_n. \quad (6.19b)$$

Ejemplo 6.12 Dado un vector aleatorio \mathbf{X} , calcula la pdf conjunta de la siguiente transformación:

$$\begin{aligned} Z_1 &= g_1(X_1) = a_1 X_1 + b_1, \\ Z_2 &= g_2(X_2) = a_2 X_2 + b_2, \\ &\vdots \\ Z_n &= g_n(X_n) = a_n X_n + b_n. \end{aligned}$$

Solución

Nótese que $Z_k = a_k X_k + b_k \leq z_k$ si y sólo si $X_k \leq (z_k - b_k)/a_k$, si $a_k > 0$, entonces

$$\begin{aligned} F_{Z_1, Z_2, \dots, Z_n}(z_1, z_2, \dots, z_n) &= P \left[X_1 \leq \frac{z_1 - b_1}{a_1}, X_2 \leq \frac{z_2 - b_2}{a_2}, \dots, X_n \leq \frac{z_n - b_n}{a_n} \right] \\ &= F_{X_1, X_2, \dots, X_n} \left(\frac{z_1 - b_1}{a_1}, \frac{z_2 - b_2}{a_2}, \dots, \frac{z_n - b_n}{a_n} \right) \\ f_{Z_1, Z_2, \dots, Z_n}(z_1, z_2, \dots, z_n) &= \frac{\delta^n}{\delta z_1 \dots \delta z_n} F_{Z_1, Z_2, \dots, Z_n}(z_1, z_2, \dots, z_n) \\ &= \frac{1}{a_1 \dots a_n} f_{X_1, X_2, \dots, X_n} \left(\frac{z_1 - b_1}{a_1}, \frac{z_2 - b_2}{a_2}, \dots, \frac{z_n - b_n}{a_n} \right). \end{aligned}$$

6.3 Esperanza de variables aleatorias multidimensionales

En esta sección caracterizaremos una variable aleatoria multidimensional a través de las esperanzas de sus componentes y de funciones de los mismos. Nos centraremos en la caracterización de variables aleatorias multidimensionales utilizando su vector de medias y su matriz de covarianza. Presentaremos también la función característica conjunta de un vector aleatorio.

La esperanza de una función $g(\mathbf{X}) = g(X_1, \dots, X_n)$ de un vector aleatorio $\mathbf{X} = (X_1, \dots, X_n)$ viene dada por:

$$E[Z] = \begin{cases} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} g(x_1, x_2, \dots, x_n) f_{\mathbf{X}}(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n & \mathbf{X} \text{ conjuntamente continua} \\ \sum_{x_1} \dots \sum_{x_n} g(x_1, x_2, \dots, x_n) p_{\mathbf{X}}(x_1, x_2, \dots, x_n) & \mathbf{X} \text{ discreta.} \end{cases} \quad (6.20)$$

Un ejemplo importante es $g(\mathbf{X})$ igual a la suma de funciones de \mathbf{X} . El procedimiento que lleva a la ecuación (5.26) y un razonamiento de inducción demuestra que:

$$E[g_1(\mathbf{X}) + g_2(\mathbf{X}) + \dots + g_n(\mathbf{X})] = E[g_1(\mathbf{X})] + \dots + E[g_n(\mathbf{X})]. \quad (6.21)$$

Otro ejemplo importante es $g(\mathbf{X})$ igual al producto de n funciones individuales de los componentes. Si X_1, \dots, X_n son variables aleatorias *independientes*, entonces

$$E[g_1(X_1)g_2(X_2) \dots g_n(X_n)] = E[g_1(X_1)]E[g_2(X_2)] \dots E[g_n(X_n)]. \quad (6.22)$$

6.3.1 Vector de medias y matriz de covarianza

La media, varianza y covarianza proporcionan información útil sobre la distribución de una variable aleatoria y son fáciles de estimar, de modo que frecuentemente estaremos interesados en caracterizar las variables aleatorias multidimensionales en términos de sus primeros y segundos momentos. Ahora definimos el

vector de medias y la matriz de covarianza. Después exploraremos el vector de medias y la matriz de covarianza de una transformación lineal de una variable aleatoria multidimensional.

Para $\mathbf{X} = (X_1, \dots, X_n)$ el **vector de medias** se define como el vector columna de las esperanzas de los componentes X_k :

$$\mathbf{m}_{\mathbf{X}} = E[\mathbf{X}] = E \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} \triangleq \begin{bmatrix} E[X_1] \\ E[X_2] \\ \vdots \\ E[X_n] \end{bmatrix}. \quad (6.23a)$$

La **matriz de correlación** contiene los segundos momentos de \mathbf{X} :

$$\mathbf{R}_{\mathbf{X}} = \begin{bmatrix} E[X_1^2] & E[X_1X_2] & \cdots & E[X_1X_n] \\ E[X_2X_1] & E[X_2^2] & \cdots & E[X_2X_n] \\ \vdots & \vdots & \ddots & \vdots \\ E[X_nX_1] & E[X_nX_2] & \cdots & E[X_n^2] \end{bmatrix}. \quad (6.23b)$$

La **matriz de covarianza** tiene como entradas los momentos centrales de segundo orden:

$$\mathbf{K}_{\mathbf{X}} = \begin{bmatrix} E[(X_1 - m_1)^2] & E[(X_1 - m_1)(X_2 - m_2)] & \cdots & E[(X_1 - m_1)(X_n - m_n)] \\ E[(X_2 - m_2)(X_1 - m_1)] & E[(X_2 - m_2)^2] & \cdots & E[(X_2 - m_2)(X_n - m_n)] \\ \vdots & \vdots & \ddots & \vdots \\ E[(X_n - m_n)(X_1 - m_1)] & E[(X_n - m_n)(X_2 - m_2)] & \cdots & E[(X_n - m_n)^2] \end{bmatrix}. \quad (6.23c)$$

Tanto $\mathbf{R}_{\mathbf{X}}$ como $\mathbf{K}_{\mathbf{X}}$ son matrices simétricas de dimensión $n \times n$. Los elementos de la diagonal de $\mathbf{K}_{\mathbf{X}}$ vienen dados por las varianzas $V[X_k] = E[(X_k - m_k)^2]$ de los elementos de \mathbf{X} . Si estos elementos están incorrelados, entonces $Cov(X_j, X_k) = 0$ para $j \neq k$ y $\mathbf{K}_{\mathbf{X}}$ es una matriz diagonal. Si las variables aleatorias X_1, \dots, X_n son independientes, entonces son incorreladas y $\mathbf{K}_{\mathbf{X}}$ es diagonal. Por último, si el vector de esperanzas es $\mathbf{0}$, esto es, $m_k = E[X_k] = 0$ para todo k , entonces $\mathbf{R}_{\mathbf{X}} = \mathbf{K}_{\mathbf{X}}$.

Ejemplo 6.13 Sea $\mathbf{X} = (X_1, X_2, X_3)$ la variable aleatoria Normal multivariante del ejemplo 6.6. Calcula $E[\mathbf{X}]$ y $\mathbf{K}_{\mathbf{X}}$.

Solución

Reescribimos la pdf conjunta como sigue:

$$f_{X_1, X_2, X_3}(x_1, x_2, x_3) = \frac{e^{-(x_1^2 + x_2^2 - 2\frac{1}{\sqrt{2}}x_1x_2)} e^{-x_3^2/2}}{2\pi \sqrt{1 - \left(-\frac{1}{\sqrt{2}}\right)^2} \sqrt{2\pi}}.$$

Vemos que X_3 es una variable aleatoria Normal con media 0 y varianza 1 y que es independiente de X_1 y X_2 . También observamos que X_1 y X_2 forman una

normal bivalente con media 0 y varianza 1 y con coeficiente de correlación

$$\rho_{X_1, X_2} = -\frac{1}{\sqrt{2}} = \frac{Cov(X_1, X_2)}{\sigma_{X_1} \sigma_{X_2}} = Cov(X_1, X_2).$$

Entonces el vector de medias es $\mathbf{m}_X = \mathbf{0}$ y

$$\mathbf{K}_X = \begin{bmatrix} 1 & -\frac{1}{\sqrt{2}} & 0 \\ -\frac{1}{\sqrt{2}} & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Ahora desarrollamos expresiones compactas para \mathbf{R}_X y \mathbf{K}_X . Si multiplicamos \mathbf{X} , una matriz $n \times 1$ y \mathbf{X}^T , una matriz $1 \times n$, obtenemos la siguiente matriz de dimensión $n \times n$:

$$\mathbf{X}\mathbf{X}^T = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} [X_1, X_2, \dots, X_n] = \begin{bmatrix} X_1^2 & X_1X_2 & \cdots & X_1X_n \\ X_2X_1 & X_2^2 & \cdots & X_2X_n \\ \vdots & \vdots & \cdots & \vdots \\ X_nX_1 & X_nX_2 & \cdots & X_n^2 \end{bmatrix}.$$

Si definimos la esperanza de una matriz como la matriz de las esperanzas de sus elementos, podemos escribir la matriz de correlación como:

$$\mathbf{R}_X = E[\mathbf{X}\mathbf{X}^T]. \quad (6.24a)$$

La matriz de covarianza es entonces:

$$\begin{aligned} \mathbf{K}_X &= E[(\mathbf{X} - \mathbf{m}_X)(\mathbf{X} - \mathbf{m}_X)^T] \\ &= E[\mathbf{X}\mathbf{X}^T] - \mathbf{m}_X E[\mathbf{X}^T] - E[\mathbf{X}]\mathbf{m}_X^T + \mathbf{m}_X\mathbf{m}_X^T \\ &= \mathbf{R}_X - \mathbf{m}_X\mathbf{m}_X^T. \end{aligned} \quad (6.24b)$$

6.3.2 Transformaciones lineales de vectores aleatorios

En ingeniería muchos sistemas son lineales en el sentido que será elaborado más adelante. Frecuentemente estos sistemas se pueden reducir a una transformación lineal de un vector o de variables aleatorias donde la “entrada” es \mathbf{X} y la “salida” es \mathbf{Y} :

$$\mathbf{Y} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} = \mathbf{A}\mathbf{X}.$$

La esperanza del componente k -ésimo de \mathbf{Y} es el producto escalar de la k -ésima fila de \mathbf{A} y de \mathbf{X} :

$$E[Y_k] = E \left[\sum_{j=1}^n a_{kj} X_j \right] = \sum_{j=1}^n a_{kj} E[X_j].$$

Cada componente de $E[\mathbf{Y}]$ se obtiene de la misma forma, por tanto:

$$\begin{aligned} \mathbf{m}_Y = E[\mathbf{Y}] &= \begin{bmatrix} \sum_{j=1}^n a_{1j} E[X_j] \\ \sum_{j=1}^n a_{2j} E[X_j] \\ \vdots \\ \sum_{j=1}^n a_{nj} E[X_j] \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdot & \cdot & \cdots & \cdot \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} E[X_1] \\ E[X_2] \\ \vdots \\ E[X_n] \end{bmatrix} \\ &= \mathbf{A} E[\mathbf{X}] = \mathbf{A} \mathbf{m}_X. \end{aligned} \quad (6.25a)$$

La matriz de covarianza de \mathbf{Y} es entonces:

$$\begin{aligned} \mathbf{K}_Y &= E[(\mathbf{Y} - \mathbf{m}_Y)(\mathbf{Y} - \mathbf{m}_Y)^T] = E[(\mathbf{A}\mathbf{X} - \mathbf{A}\mathbf{m}_X)(\mathbf{A}\mathbf{X} - \mathbf{A}\mathbf{m}_X)^T] \\ &= E[\mathbf{A}(\mathbf{X} - \mathbf{m}_X)(\mathbf{X} - \mathbf{m}_X)^T \mathbf{A}^T] = \mathbf{A} E[(\mathbf{X} - \mathbf{m}_X)(\mathbf{X} - \mathbf{m}_X)^T] \mathbf{A}^T \\ &= \mathbf{A} \mathbf{K}_X \mathbf{A}^T, \end{aligned} \quad (6.25b)$$

donde utilizamos el hecho de que la transpuesta de la multiplicación de matrices es el producto de las matrices transpuestas en orden inverso: $\{\mathbf{A}(\mathbf{X} - \mathbf{m}_X)\}^T = (\mathbf{X} - \mathbf{m}_X)^T \mathbf{A}^T$.

La **matriz de covarianza entre dos vectores aleatorios \mathbf{X} e \mathbf{Y}** se define como:

$$\mathbf{K}_{\mathbf{X}\mathbf{Y}} = E[(\mathbf{X} - \mathbf{m}_X)(\mathbf{Y} - \mathbf{m}_Y)^T] = E[\mathbf{X}\mathbf{Y}^T] - \mathbf{m}_X \mathbf{m}_Y^T = \mathbf{R}_{\mathbf{X}\mathbf{Y}} - \mathbf{m}_X \mathbf{m}_Y^T.$$

Estamos interesados en la covarianza entre \mathbf{X} e $\mathbf{Y} = \mathbf{A}\mathbf{X}$:

$$\begin{aligned} \mathbf{K}_{\mathbf{X}\mathbf{Y}} &= E[(\mathbf{X} - \mathbf{m}_X)(\mathbf{Y} - \mathbf{m}_Y)^T] = E[(\mathbf{X} - \mathbf{m}_X)(\mathbf{X} - \mathbf{m}_X)^T \mathbf{A}^T] \\ &= \mathbf{K}_X \mathbf{A}^T. \end{aligned} \quad (6.25c)$$

Ejemplo 6.14 Transformación de un vector aleatorio incorrelado

Supongamos que los componentes de \mathbf{X} están incorrelados y tienen varianza 1, entonces $\mathbf{K}_X = \mathbf{I}$, la matriz identidad. La matriz de covarianza para $\mathbf{Y} = \mathbf{A}\mathbf{X}$ es

$$\mathbf{K}_Y = \mathbf{A} \mathbf{K}_X \mathbf{A}^T = \mathbf{A} \mathbf{I} \mathbf{A}^T = \mathbf{A} \mathbf{A}^T. \quad (6.26)$$

En general, $\mathbf{K}_Y = \mathbf{A} \mathbf{A}^T$ no es una matriz diagonal y por tanto, los componentes de \mathbf{Y} están correlados. Siempre se puede calcular la matriz \mathbf{A} tal que la ecuación (6.26) se verifique para un \mathbf{K}_Y dado. Podemos entonces generar un vector aleatorio \mathbf{Y} con cualquier matriz de covarianza \mathbf{K}_Y que deseemos.

Supongamos que los componentes de \mathbf{X} están correlados y por tanto, \mathbf{K}_X no es una matriz diagonal. En muchas situaciones estaremos interesados en calcular una matriz de transformación tal que $\mathbf{Y} = \mathbf{A}\mathbf{X}$ tenga componentes incorrelados. Esto requiere calcular \mathbf{A} tal que $\mathbf{K}_Y = \mathbf{A}\mathbf{K}_X\mathbf{A}^T$ es una matriz diagonal. En la última parte de esta sección mostramos cómo calcular dicha matriz \mathbf{A} .

Ejemplo 6.15 Transformación a un vector aleatorio incorrelado

Supongamos que los vectores X_1 , X_2 y X_3 del ejemplo 6.13 se transforman utilizando la matriz:

$$\mathbf{A} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Calcula $E[\mathbf{Y}]$ y \mathbf{K}_Y .

Solución

Como $\mathbf{m}_X = \mathbf{0}$, entonces $E[\mathbf{Y}] = E[\mathbf{A}\mathbf{m}_X] = \mathbf{0}$. La matriz de covarianza de \mathbf{Y} es:

$$\begin{aligned} \mathbf{K}_Y &= \mathbf{A}\mathbf{K}_X\mathbf{A}^T = \frac{1}{2} \begin{bmatrix} 1 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & -\frac{1}{\sqrt{2}} & 0 \\ -\frac{1}{\sqrt{2}} & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \\ &= \frac{1}{2} \begin{bmatrix} 1 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 - \frac{1}{\sqrt{2}} & 1 + \frac{1}{\sqrt{2}} & 0 \\ 1 - \frac{1}{\sqrt{2}} & -\left(1 + \frac{1}{\sqrt{2}}\right) & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 - \frac{1}{\sqrt{2}} & 0 & 0 \\ 0 & 1 + \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \end{bmatrix}. \end{aligned}$$

La transformación lineal produce un vector de variables aleatorias $\mathbf{Y} = (Y_1, Y_2, Y_3)$ con componentes incorrelados.

6.4 Variable aleatoria normal multivariante

Las variables aleatorias X_1, X_2, \dots, X_n forman una normal multivariante si su pdf conjunta es dada por

$$f_{\mathbf{X}}(\mathbf{x}) \triangleq f_{X_1, X_2, \dots, X_n}(x_1, \dots, x_n) = \frac{\exp\left\{-\frac{1}{2}(\mathbf{x} - \mathbf{m})^T \mathbf{K}^{-1}(\mathbf{x} - \mathbf{m})\right\}}{(2\pi)^{n/2} |\mathbf{K}|^{1/2}}, \quad (6.27a)$$

donde \mathbf{x} y \mathbf{m} son vectores columna definidos por

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad \mathbf{m} = \begin{bmatrix} m_1 \\ m_2 \\ \vdots \\ m_n \end{bmatrix} = \begin{bmatrix} E[X_1] \\ E[X_2] \\ \vdots \\ E[X_n] \end{bmatrix}$$

y K es la matriz de covarianza que está definida por

$$K = \begin{bmatrix} V[X_1] & Cov(X_1, X_2) & \dots & Cov(X_1, X_n) \\ Cov(X_2, X_1) & V[X_2] & \dots & Cov(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(X_n, X_1) & \dots & \dots & V[X_n] \end{bmatrix}. \quad (6.27b)$$

La $(\cdot)^T$ en la ecuación (6.27a) denota la transpuesta de un vector o una matriz. Nótese que la matriz de covarianza es una matriz simétrica porque $Cov(X_i, X_j) = Cov(X_j, X_i)$.

La ecuación (6.27a) demuestra que *la pdf de una variable aleatoria Normal multivariante está completamente especificada por las medias y las varianzas individuales y las covarianzas de cada par*. Se puede demostrar utilizando la función característica conjunta que todas las pdf marginales asociadas con la ecuación (6.27a) también son normales y que éstas también están completamente especificadas por el mismo conjunto de medias, varianzas y covarianzas.

Ejemplo 6.16 Verifica que la pdf normal bivalente dada en la ecuación (5.61a) tiene la forma de la ecuación (6.27a).

Solución

La matriz de covarianza para el caso bidimensional es dado por

$$K = \begin{bmatrix} \sigma_1^2 & \rho_{X,Y}\sigma_1\sigma_2 \\ \rho_{X,Y}\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix},$$

donde hemos utilizado el hecho de que $Cov(X_1, X_2) = \rho_{X,Y}\sigma_1\sigma_2$. El determinante de K es $\sigma_1^2\sigma_2^2(1 - \rho_{X,Y}^2)$ de modo que el denominador de la pdf tiene la forma correcta. La inversa de la matriz de covarianza es también una matriz real simétrica:

$$K^{-1} = \frac{1}{\sigma_1^2\sigma_2^2(1 - \rho_{X,Y}^2)} \begin{bmatrix} \sigma_2^2 & -\rho_{X,Y}\sigma_1\sigma_2 \\ -\rho_{X,Y}\sigma_1\sigma_2 & \sigma_1^2 \end{bmatrix}.$$

El término del exponente es entonces:

$$\begin{aligned} & \frac{1}{\sigma_1^2\sigma_2^2(1 - \rho_{X,Y}^2)}(x - m_1, y - m_2) \begin{bmatrix} \sigma_2^2 & -\rho_{X,Y}\sigma_1\sigma_2 \\ -\rho_{X,Y}\sigma_1\sigma_2 & \sigma_1^2 \end{bmatrix} \begin{bmatrix} x - m_1 \\ y - m_2 \end{bmatrix} \\ &= \frac{1}{\sigma_1^2\sigma_2^2(1 - \rho_{X,Y}^2)}(x - m_1, y - m_2) \begin{bmatrix} \sigma_2^2(x - m_1) - \rho_{X,Y}\sigma_1\sigma_2(y - m_2) \\ -\rho_{X,Y}\sigma_1\sigma_2(x - m_1) + \sigma_1^2(y - m_2) \end{bmatrix} \\ &= \frac{((x - m_1)/\sigma_1)^2 - 2\rho_{X,Y}((x - m_1)/\sigma_1)((y - m_2)/\sigma_2) + ((y - m_2)/\sigma_2)^2}{(1 - \rho_{X,Y}^2)}. \end{aligned}$$

Entonces la pdf bidimensional tiene la forma de la ecuación (6.27a).

Ejemplo 6.17 El vector de variables aleatorias (X, Y, Z) es normal multivarian-

te con medias cero y matriz de covarianza:

$$K = \begin{bmatrix} V[X] & Cov(X, Y) & Cov(X, Z) \\ Cov(Y, X) & V[Y] & Cov(Y, Z) \\ Cov(Z, X) & Cov(Z, Y) & V[Z] \end{bmatrix} = \begin{bmatrix} 1.0 & 0.2 & 0.3 \\ 0.2 & 1.0 & 0.4 \\ 0.3 & 0.4 & 1.0 \end{bmatrix}.$$

Calcula la pdf marginal de X y Z .

Solución

Podemos resolver este problema de dos formas. La primera involucra la integración directa de la pdf para obtener la pdf marginal. La segunda implica utilizar el hecho de que la pdf marginal de X y Z también es normal y tiene el mismo conjunto de medias, varianzas y covarianzas. Vamos a utilizar el segundo método.

El par (X, Z) tiene media cero y matriz de covarianza:

$$K' = \begin{bmatrix} V[X] & Cov(X, Z) \\ Cov(Z, X) & V[Z] \end{bmatrix} = \begin{bmatrix} 1.0 & 0.3 \\ 0.3 & 1.0 \end{bmatrix}.$$

La pdf conjunta de X y Z se calcula sustituyendo un vector nulo en la media y esta matriz de covarianza en la ecuación (6.27a).

Ejemplo 6.18 Independencia de variables aleatorias normales multivariantes incorreladas

Supongamos que X_1, X_2, \dots, X_n forman una variable aleatoria Normal multivariante con $Cov(X_i, X_j) = 0$ para $i \neq j$. Demuestra que X_1, X_2, \dots, X_n son variables aleatorias independientes.

Solución

Por la ecuación (6.27b) vemos que la matriz de covarianza es una matriz diagonal:

$$K = \text{diag}[V(X_i)] = \text{diag}[\sigma_i^2].$$

Entonces

$$K^{-1} = \text{diag} \left[\frac{1}{\sigma_i^2} \right]$$

y

$$(\mathbf{x} - \mathbf{m})^T K^{-1} (\mathbf{x} - \mathbf{m}) = \sum_{i=1}^n \left(\frac{x_i - m_i}{\sigma_i} \right)^2.$$

Por tanto, por la ecuación (6.27a)

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{\exp\{-\frac{1}{2} \sum_{i=1}^n [(x_i - m_i/\sigma_i)]^2\}}{(2\pi)^{n/2}} |K|^{1/2} = \prod_{i=1}^n \frac{\exp\{-\frac{1}{2} [(x_i - m_i/\sigma_i)]^2\}}{\sqrt{2\pi\sigma_i^2}} = \prod_{i=1}^n f_{X_i}(x_i).$$

Entonces X_1, X_2, \dots, X_n son variables aleatorias normales independientes.

Ejemplo 6.19 pdf condicional de una variable aleatoria Normal

Calcula la pdf condicional de X_n dadas X_1, X_2, \dots, X_{n-1} .

Solución

Sea \mathbf{K}_n la matriz de covarianza de $\mathbf{X}_n = (X_1, X_2, \dots, X_n)$ y \mathbf{K}_{n-1} la matriz de covarianza de $\mathbf{X}_{n-1} = (X_1, X_2, \dots, X_{n-1})$. Sea $\mathbf{Q}_n = \mathbf{K}_n^{-1}$ y $\mathbf{Q}_{n-1} = \mathbf{K}_{n-1}^{-1}$, entonces estas últimas matrices son submatrices de las primeras matrices como se muestra a continuación:

$$\mathbf{K}_n = \begin{bmatrix} & & & K_{1n} \\ & \mathbf{K}_{n-1} & & K_{2n} \\ & & & \vdots \\ K_{1n} & K_{2n} & \dots & K_{nn} \end{bmatrix} \quad \mathbf{Q}_n = \begin{bmatrix} & & & Q_{1n} \\ & \mathbf{Q}_{n-1} & & Q_{2n} \\ & & & \vdots \\ Q_{1n} & Q_{2n} & \dots & Q_{nn} \end{bmatrix}.$$

Más abajo usaremos el subíndice n o $n-1$ para distinguir entre los dos vectores aleatorios y sus parámetros. La pdf marginal de X_n dadas X_1, X_2, \dots, X_{n-1} es dada por:

$$\begin{aligned} f_{X_n}(x_n|x_1, \dots, x_{n-1}) &= \frac{f_{\mathbf{X}_n}(\mathbf{x}_n)}{f_{\mathbf{X}_{n-1}}(\mathbf{x}_{n-1})} \\ &= \frac{\exp\{-\frac{1}{2}(\mathbf{x}_n - \mathbf{m}_n)^T \mathbf{Q}_n (\mathbf{x}_n - \mathbf{m}_n)\}}{(2\pi)^{n/2} |\mathbf{K}_n|^{1/2}} \frac{(2\pi)^{(n-1)/2} |\mathbf{K}_{n-1}|^{1/2}}{\exp\{-\frac{1}{2}(\mathbf{x}_{n-1} - \mathbf{m}_{n-1})^T \mathbf{Q}_{n-1} (\mathbf{x}_{n-1} - \mathbf{m}_{n-1})\}} \\ &= \frac{\exp\{-\frac{1}{2}(\mathbf{x}_n - \mathbf{m}_n)^T \mathbf{Q}_n (\mathbf{x}_n - \mathbf{m}_n) + \frac{1}{2}(\mathbf{x}_{n-1} - \mathbf{m}_{n-1})^T \mathbf{Q}_{n-1} (\mathbf{x}_{n-1} - \mathbf{m}_{n-1})\}}{\sqrt{2\pi} |\mathbf{K}_n|^{1/2} / |\mathbf{K}_{n-1}|^{1/2}}. \end{aligned}$$

Los términos de la expresión de arriba vienen dados por:

$$\begin{aligned} -\frac{1}{2}(\mathbf{x}_n - \mathbf{m}_n)^T \mathbf{Q}_n (\mathbf{x}_n - \mathbf{m}_n) + \frac{1}{2}(\mathbf{x}_{n-1} - \mathbf{m}_{n-1})^T \mathbf{Q}_{n-1} (\mathbf{x}_{n-1} - \mathbf{m}_{n-1}) \\ = Q_{nn} \{(x_n - m_n) + B\}^2 - Q_{nn} B^2 \end{aligned} \quad (6.28)$$

donde $B = \frac{1}{Q_{nn}} \sum_{j=1}^{n-1} Q_{jn}(x_j - m_j)$ y $|\mathbf{K}_n|/|\mathbf{K}_{n-1}| = 1/Q_{nn}$.

Esto implica que X_n tiene media $m_n - B$ y varianza $1/Q_{nn}$. El término $Q_{nn} B^2$ es parte de la constante de normalización. Entonces concluimos que:

$$f_{X_n}(x_n|x_1, \dots, x_{n-1}) = \frac{\exp\left\{-\frac{Q_{nn}}{2} \left(x - m_n + \frac{1}{Q_{nn}} \sum_{j=1}^{n-1} Q_{jn}(x_j - m_j)\right)^2\right\}}{\sqrt{2\pi/Q_{nn}}}.$$

Vemos que la media condicionada de X_n es una función lineal de las "observaciones" x_1, x_2, \dots, x_{n-1} .

Problemas

6.1 Sean las variables aleatorias X, Y y Z independientes. Calcula las siguientes probabilidades en términos de $F_X(x)$, $F_Y(y)$ y $F_Z(z)$.

- $P[|X| < 5, Y < 4, Z^3 > 8]$.
- $P[X = 5, Y < 0, Z > 1]$.
- $P[\text{mín}(X, Y, Z) < 2]$.
- $P[\text{máx}(X, Y, Z) > 6]$.

6.2 Un transmisor de radio envía una señal $s > 0$ a un receptor utilizando tres caminos. Las señales que llegan al receptor a lo largo de cada camino son:

$$X_1 = s + N_1, X_2 = s + N_2, X_3 = s + N_3,$$

donde N_1 , N_2 y N_3 son variables aleatorias Normales independientes con media 0 y varianza 1.

- Calcula la pdf conjunta de $\mathbf{X} = (X_1, X_2, X_3)$. ¿Son X_1 , X_2 y X_3 variables aleatorias independientes?
- Calcula la probabilidad de que el mínimo de las tres señales sea positivo.
- Calcula la probabilidad de que la mayoría de las señales sean positivas.

6.3 Sea la pdf conjunta de X , Y y Z :

$$f_{X,Y,Z}(x, y, z) = k(x + y + z) \text{ para } 0 \leq x \leq 1, 0 \leq y \leq 1, 0 \leq z \leq 1.$$

- Calcula k .
- Calcula $f_X(x|y, z)$ y $F_Z(z|x, y)$.
- Calcula $f_X(x)$, $f_Y(y)$ y $F_Z(z)$.

6.4 Sea X_1, X_2, \dots, X_n variables aleatorias binarias que toman valores 0 o 1 para denotar si un altavoz está en silencio (0) o activo (1). Un altavoz en silencio continua en ese estado durante el siguiente bloque de tiempo con probabilidad $3/4$ y un altavoz activo sigue activo con probabilidad $1/2$. Calcula la pmf conjunta de X_1, X_2, X_3 y la pmf marginal de X_3 . Asumiendo que el altavoz comienza en silencio.

6.5 Un experimento aleatorio tiene cuatro posibles resultados. Supongamos que el experimento se repite n veces de forma independiente y sea X_k el número de veces que ocurre el resultado k . La pmf conjunta de (X_1, X_2, X_3) viene dada por

$$p(k_1, k_2, k_3) = \frac{n!3!}{(n+3)!} = \binom{n+3}{3}^{-1} \text{ para } 0 \leq k_i \text{ y } k_1 + k_2 + k_3 \leq n.$$

- Calcula la pmf marginal de (X_1, X_2) .
- Calcula la pmf marginal de X_1 .
- Calcula la pmf condicionada conjunta de (X_2, X_3) dado que $X_1 = m$, donde $0 \leq m \leq n$.

6.6 El número de peticiones de tipo 1, 2 y 3, respectivamente, que llegan a una estación de servicio en t segundos son variables aleatorias de Poisson independientes con medias $\lambda_1 t$, $\lambda_2 t$ y $\lambda_3 t$. Sea N_1 , N_2 y N_3 el número de peticiones que llegan durante un tiempo T distribuido exponencialmente con media αt .

- Calcula la pmf conjunta de N_1 , N_2 y N_3 .
- Calcula la pmf marginal de N_1 .
- Calcula la pmf condicionada de N_1 y N_2 dado que N_3 .

6.7 Se instala N dispositivos al mismo tiempo. Sea Y el tiempo hasta el fallo del primer dispositivo.

- Calcula la pdf de Y si los tiempos de vida de los dispositivos son independientes y tienen la misma distribución de Pareto.
- Repite la parte a) si los tiempos de vida tienen distribución Weibull.
- Sea $I_k(t)$ la función indicadora del suceso “el dispositivo k -ésimo está en funcionamiento en el tiempo t ”. Sea $N(t)$ el número de dispositivos que están en funcionamiento en el tiempo t : $N(t) = I_1(t) + I_2(t) + \dots + I_N(t)$. Calcula la pmf de $N(t)$, su media y su varianza.

6.8 Un receptor de un sistema de comunicación multiusuario acepta K señales binarias de K transmisores independientes: $\mathbf{Y} = (Y_1, Y_2, \dots, Y_k)$, donde Y_k es la señal recibida desde el transmisor k -ésimo. Un sistema ideal para el vector recibido viene dado por:

$$\mathbf{Y} = \mathbf{A}\mathbf{b} + \mathbf{N}$$

donde $\mathbf{A} = [\alpha_k]$ es una matriz diagonal de ganancias positivas del canal, $\mathbf{b} = (b_1, b_2, \dots, b_K)$ es el vector de bits de cada transmisor donde $b_k = \pm 1$ y \mathbf{N} es un vector de K variables aleatorias independientes Normales con media 0 y varianza 1.

- Calcula la pdf conjunta de \mathbf{Y} .
- Supongamos que $\mathbf{b} = (1, 1, \dots, 1)$, calcula la probabilidad de que todos los componentes de \mathbf{Y} sean positivos.

6.9 (a) Calcula la pdf conjunta de la media y la varianza muestral de dos variables aleatorias:

$$M = \frac{X_1 + X_2}{2} \quad V = \frac{(X_1 - M)^2 + (X_2 - M)^2}{2}$$

en términos de la pdf conjunta de X_1 y X_2 .

- Evalúa la pdf conjunta si X_1 y X_2 son variables aleatorias Normales independientes con la misma media y varianza 1.
- Evaluar la pdf conjunta si X_1 y X_2 son variables aleatorias independientes Exponenciales con parámetro 1.

6.10 Calcula $E[M]$, $E[V]$ y $E[MV]$ en el problema anterior.

6.11 Calcula el vector de medias y la matriz de covarianza de las tres señales $\mathbf{X} = (X_1, X_2, X_3)$ en el problema 6.2

6.12 Calcula el vector media y la matriz de covarianza del número de llegadas de paquetes en un tiempo aleatorio N_1 , N_2 y N_3 en el problema 6.6. Sugerencia: Utiliza la esperanza condicionada.

6.13 Sea $\mathbf{X} = (X_1, X_2, X_3, X_4)$ variables aleatorias independientes con media igual y varianza 1. Calcula el vector de medias y las matrices de covarianza de $\mathbf{Y} = \mathbf{A}\mathbf{X}$:

$$(a) \mathbf{A} = \begin{bmatrix} 1 & 1/2 & 1/4 & 1/8 \\ 0 & 1 & 1/2 & 1/4 \\ 0 & 0 & 1 & 1/2 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$(b) \mathbf{A} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}.$$

6.14 Sea $X = aU + bV$ y $y = cU + dV$, donde $|ad - bc| \neq 0$. Calcula una expresión de $E[XY]$ en términos de los momentos conjuntos de U y V .

6.15 Sea $\mathbf{X} = (X_1, X_2, X_3)$ una variable aleatoria Normal multivariante con media y covarianza dadas por:

$$\mathbf{m}_{\mathbf{X}} = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} \quad \mathbf{K}_{\mathbf{X}} = \begin{bmatrix} 3/2 & 0 & 1/2 \\ 0 & 1 & 0 \\ 1/2 & 0 & 3/2 \end{bmatrix}.$$

- Calcula la pdf de \mathbf{X} en notación matricial.
- Calcula la pdf de \mathbf{X} utilizando la expresión cuadrática en la exponencial.
- Calcula las pdf marginales de X_1 , X_2 y X_3 .

6.16 Sea U_1 , U_2 y U_3 variables aleatorias Normales independientes de media 0 y varianza 1 y sea $X = U_1$, $Y = U_1 + U_2$ y $Z = U_1 + U_2 + U_3$.

- Calcula la matriz de covarianza de (X, Y, Z) .
- Calcula la pdf conjunta de (X, Y, Z) .
- Calcula la pdf condicionada de Y y Z dada X .
- Calcula la pdf condicionada de Z dadas X e Y .

6.17 Un modelo más realista para el receptor del sistema de comunicación multiusuario del problema 6.8 consiste en que la K señales recibidas $\mathbf{Y} = (Y_1, Y_2, \dots, Y_K)$ vienen dadas por:

$$\mathbf{Y} = \mathbf{A}\mathbf{R}\mathbf{b} + \mathbf{N}$$

donde $\mathbf{A} = [\alpha_k]$ es una matriz diagonal de ganancias positivas del canal, \mathbf{R} es una matriz simétrica que tiene en cuenta las interferencias entre usuarios y $\mathbf{b} = (b_1, b_2, \dots, b_K)$ es el vector de bits de cada transmisor. \mathbf{N} es el vector de K variables aleatorias Normales independientes con media 0 y varianza 1.

- Calcula la pdf conjunta de \mathbf{Y} .
- Supongamos que con el objetivo de recuperar \mathbf{b} , el receptor calcula $\mathbf{Z} = (\mathbf{A}\mathbf{R})^{-1}\mathbf{Y}$. Calcula la pdf conjunta de \mathbf{Z} .

6.18 Sea U_0, U_1, \dots una serie de variables aleatorias Normales independientes de media 0 y varianza 1. Un “filtro de paso bajo” toma la secuencia U_i y produce la serie de salida $X_n = (U_n + U_{n-1})/2$ y un “filtro de paso alto” produce la serie de salida $Y_n = (U_n - U_{n-1})/2$.

- Calcula la pdf conjunta de X_{n+1} , X_n y X_{n-1} y también de X_n , X_{n+m} y X_{n+2m} , $m > 1$.
- Repite la parte a) para Y_n .
- Calcula la pdf conjunta de X_n , X_m , Y_n e Y_m .
- Calcula la función característica correspondiente conjunta correspondiente a los apartados a), b) y c).

6.19 Sea X_1, X_2, \dots, X_n muestras de una señal. Supongamos que las muestras son variables aleatorias Normales con covarianza

$$\text{Cov}(X_i, X_j) = \begin{cases} \sigma^2 & i = j \\ \rho\sigma^2 & |i - j| = 1 \\ 0 & \text{en otro caso.} \end{cases}$$

Supongamos que tomamos bloques de dos muestras consecutivas para formar un vector \mathbf{X} que es transformado para formar $\mathbf{Y} = \mathbf{A}\mathbf{X}$.

- Calcula la matriz \mathbf{A} tal que los componentes de \mathbf{Y} son variables aleatorias independientes.
- Sea \mathbf{X}_i y \mathbf{X}_{i+1} dos bloques consecutivos y sea \mathbf{Y}_i e \mathbf{Y}_{i+1} las variables transformadas correspondientes. ¿Son independientes los componentes de \mathbf{Y}_i e \mathbf{Y}_{i+1} ?

6.20 Un multiplexor combina N señales digitales de TV en una línea de comunicaciones común. La señal de TV n genera X_n bits cada 33 ms, donde X_n es una variable aleatoria Normal con media m y varianza σ^2 . Supongamos que el multiplexor acepta un total máximo de T bits de las fuentes combinadas cada 33 ms y el resto son descartados. Asumimos que las N señales son independientes.

- Calcula la probabilidad de que los bits sean descartados en un periodo dado de 33 ms, si $T = m_\alpha + t\sigma$, donde m_α es la media total de bits generados por las fuentes combinadas y σ es la desviación típica del número total de bits producidos por las fuentes combinadas.
- Calcula el número medio de bits descartados por periodo.
- Calcula la fracción límite de bits perdidos por el multiplexor.
- Calcula el número medio de bits por fuente en el apartado a) y calcula el número medio de bits perdidos por fuente. ¿Qué pasa cuando N crece (tiende a infinito)?
- Supongamos que se necesita que t se ajuste con N tal que la fracción de bits perdidos por fuente se mantenga constante. Calcula una ecuación cuya solución lleve al valor deseado de t .

- (f) ¿Cambian los resultados anteriores si las señales tienen una covarianza dos a dos igual a ρ ?

7 Suma de variables aleatorias

Muchos problemas implican contar las veces que ocurren algunos sucesos, medir los efectos acumulados o calcular medias aritméticas de series de medidas. Normalmente, estos problemas se reducen al problema de calcular, de forma exacta o aproximada, la distribución de una variable aleatoria que consiste en la suma de n variables aleatorias independientes e idénticamente distribuidas. En este capítulo, investigamos la suma de variables aleatorias y sus propiedades cuando n crece.

En la sección 7.1 mostramos que la función característica se utiliza para calcular la pdf de la suma de variables aleatorias independientes. En la sección 7.2 discutimos el estimador de la media muestral para la esperanza de una variable aleatoria y el estimador de la frecuencia relativa para la probabilidad de un suceso. Definiremos medidas para comprobar la bondad de estos estimadores. Después estudiaremos las leyes de los grandes números, que son teoremas que establecen que los estimadores de la media muestral y la frecuencia relativa convergen a sus correspondientes esperanzas y probabilidades al aumentar el número de muestras. Estos resultados teóricos demuestran la remarcable consistencia entre la teoría de la probabilidad y el comportamiento observado, reforzando la idea de la interpretación de la probabilidad como la frecuencia relativa.

En la sección 7.3, presentaremos el teorema central del límite, que establece que, bajo condiciones muy generales, la cdf de la suma de variables aleatorias se aproxima a la de la variable aleatoria Normal incluso cuando la cdf de las variables aleatorias individuales sea muy diferente a la de la normal. Este resultado permite aproximar la pdf de la suma de variables aleatorias por la pdf de la variable aleatoria Normal. Este resultado también explica por qué la variable aleatoria Normal aparece en muchas y muy diversas aplicaciones.

7.1 Suma de variables aleatorias

Sea X_1, X_2, \dots, X_n una serie de variables aleatorias y sea S_n su suma:

$$S_n = X_1 + X_2 + \dots + X_n. \quad (7.1)$$

En esta sección calcularemos la media y la varianza de S_n , así como la pdf de S_n en el caso especial en el que las X_j son variables aleatorias independientes.

7.1.1 Media y varianza de la suma de variables aleatorias

En la sección 6.3 se demostró que *independientemente de la dependencia estadística, la esperanza de la suma de n variables aleatorias es igual a la suma de las esperanzas*:

$$E[X_1 + X_2 + \cdots + X_n] = E[X_1] + \cdots + E[X_n]. \quad (7.2)$$

Entonces, es suficiente con conocer las medias de las X_j para calcular la media de S_n .

El siguiente ejemplo muestra que para calcular la varianza de la suma de variables aleatorias, necesitamos conocer las varianzas y las covarianzas de las X_j .

Ejemplo 7.1 Calcula la varianza de $Z = X + Y$.

Solución

Según la ecuación (7.2), $E[Z] = E[X + Y] = E[X] + E[Y]$. La varianza de Z es por tanto

$$\begin{aligned} V[Z] &= E[(Z - E[Z])^2] = E[(X + Y - E[X] - E[Y])^2] \\ &= E[\{(X - E[X]) + (Y - E[Y])\}^2] \\ &= E[(X - E[X])^2 + (Y - E[Y])^2 + (X - E[X])(Y - E[Y]) \\ &\quad + (Y - E[Y])(X - E[X])] \\ &= V[X] + V[Y] + COV(X, Y) + COV(Y, X) \\ &= V[X] + V[Y] + 2COV(X, Y). \end{aligned}$$

En general, la covarianza $COV(X, Y)$ no es igual a cero, de modo que la varianza de la suma no es necesariamente igual a la suma de las varianzas individuales.

El resultado del ejemplo 7.1 se puede generalizar al caso de n variables aleatorias:

$$\begin{aligned} V(X_1 + X_2 + \cdots + X_n) &= E \left\{ \sum_{j=1}^n (X_j - E[X_j]) \sum_{k=1}^n (X_k - E[X_k]) \right\} \\ &= \sum_{j=1}^n \sum_{k=1}^n E[(X_j - E[X_j])(X_k - E[X_k])] \\ &= \sum_{k=1}^n V[X_k] + \sum_{j=1}^n \sum_{k=1}^n COV(X_j, X_k). \quad (7.3) \end{aligned}$$

Entonces, *en general, la varianza de la suma de variables aleatorias no es igual a la suma de las varianzas individuales.*

Un caso especial importante es cuando las X_j son variables aleatorias independientes. Si X_1, X_2, \dots, X_n son variables aleatorias independientes, entonces la $COV(X_j, X_k) = 0$ para $j \neq k$ y

$$V(X_1 + X_2 + \dots + X_n) = V(X_1) + \dots + V(X_n). \quad (7.4)$$

Ejemplo 7.2 Suma de variables aleatorias iid

Calcula la media y la varianza de la suma de n variables aleatorias **independientes e idénticamente distribuidas** (iid), cada una de ellas con media μ y varianza σ^2 .

Solución

La media de S_n se obtiene con la ecuación (7.2):

$$E[S_n] = E[X_1] + \dots + E[X_n] = n\mu.$$

La covarianza de los pares de variables aleatorias independientes es cero, de modo que por la ecuación (7.4),

$$V[S_n] = nV[X_j] = n\sigma^2,$$

ya que $V[X_j] = \sigma^2$ para $j = 1, \dots, n$.

7.1.2 pdf de la suma de variables aleatorias independientes

Sea X_1, X_2, \dots, X_n n variables aleatorias independientes. En esta sección mostramos cómo los métodos de la transformada se pueden usar para hallar la pdf de $S_n = X_1 + X_2 + \dots + X_n$.

Primero consideremos el caso para $n = 2$, $Z = X + Y$ donde X e Y son variables aleatorias independientes. La función característica de Z viene dada por

$$\begin{aligned} \Phi_Z(\omega) &= E[e^{j\omega Z}] \\ &= E[e^{j\omega(X+Y)}] \\ &= E[e^{j\omega X} e^{j\omega Y}] \\ &= E[e^{j\omega X}] E[e^{j\omega Y}] \\ &= \Phi_X(\omega) \Phi_Y(\omega), \end{aligned} \quad (7.5)$$

donde la cuarta igualdad se deduce por el hecho de que las funciones de variables aleatorias independientes también son variables aleatorias independientes, como discutimos en el ejemplo 5.25. Entonces, la función característica de Z es el producto de las funciones características de X e Y .

En el ejemplo 5.39 vimos que la pdf de $Z = X + Y$ viene dada por la convolución de las pdf de X y de Y :

$$f_Z(z) = f_X(x) * f_Y(y). \quad (7.6)$$

Recordemos que $\Phi_Z(\omega)$ también puede verse como la transformada de Fourier de la pdf de Z :

$$\Phi_Z(\omega) = \mathcal{F}\{f_Z(z)\}.$$

Calculando la transformada de la ecuación (7.6) a la ecuación (7.5) obtenemos

$$\Phi_Z(\omega) = \mathcal{F}\{f_Z(z)\} = \mathcal{F}\{f_X(x) * f_Y(y)\} = \Phi_X(\omega)\Phi_Y(\omega). \quad (7.7)$$

La ecuación (7.7) establece el conocido resultado de que la transformada de Fourier de la convolución de dos funciones es igual al producto de transformadas de Fourier individuales.

Ahora consideremos la suma de n variables aleatorias independientes:

$$S_n = X_1 + X_2 + \cdots + X_n.$$

La función característica de S_n es

$$\begin{aligned} \Phi_{S_n}(\omega) &= E[e^{j\omega S_n}] = E[e^{j\omega(X_1+X_2+\cdots+X_n)}] \\ &= E[e^{j\omega X_1}] \cdots E[e^{j\omega X_n}] \\ &= \Phi_{X_1}(\omega) \cdots \Phi_{X_n}(\omega). \end{aligned} \quad (7.8)$$

Entonces la pdf de S_n se puede calcular por medio de la transformada inversa de Fourier del producto de las funciones características individuales de las X_j .

$$f_{S_n}(X) = \mathcal{F}^{-1}\{\Phi_{X_1}(\omega) \cdots \Phi_{X_n}(\omega)\}. \quad (7.9)$$

Ejemplo 7.3 Suma de variables aleatorias normales independientes

Sea S_n la suma de n variables aleatorias normales independientes con medias m_1, \dots, m_n y varianzas $\sigma_1^2, \dots, \sigma_n^2$. Calcula la pdf de S_n .

Solución

La función característica de X_k es

$$\Phi_{X_k} = e^{+j\omega m_k - \omega^2 \sigma_k^2 / 2}$$

entonces, por la ecuación (7.8),

$$\begin{aligned} \Phi_{S_n}(\omega) &= \prod_{k=1}^n e^{+j\omega m_k - \omega^2 \sigma_k^2 / 2} \\ &= \exp\{+j\omega(m_1 + \cdots + m_n) - \omega^2(\sigma_1^2 + \cdots + \sigma_n^2)/2\} \end{aligned}$$

Esto es la función característica de una variable aleatoria Normal. Entonces S_n es una variable aleatoria Normal con media $m_1 + \cdots + m_n$ y varianza $\sigma_1^2 + \cdots + \sigma_n^2$.

Ejemplo 7.4 Suma de variables aleatorias iid

Calcula la pdf de la suma de n variables aleatorias independientes e idénticamente distribuidas con funciones características

$$\Phi_{X_k}(\omega) = \Phi_X(\omega) \text{ para } k = 1, \dots, n.$$

Solución

La ecuación (7.8) implica inmediatamente que la función característica de S_n es

$$\Phi_{S_n}(\omega) = \{\Phi_X(\omega)\}^n. \quad (7.10)$$

La pdf de S_n se calcula tomando la transformada inversa de esta expresión.

Ejemplo 7.5 Suma de variables aleatorias exponenciales iid

Calcula la pdf de la suma de n variables aleatorias independientes y distribuidas exponencialmente, todas con parámetro α .

Solución

La función característica de una variable aleatoria Exponencial es

$$\Phi_X(\omega) = \frac{\alpha}{\alpha - j\omega}.$$

Por el ejemplo anterior tenemos entonces que

$$\Phi_{S_n}(\omega) = \left\{ \frac{\alpha}{\alpha - j\omega} \right\}^n.$$

Vemos que S_n es una variable aleatoria Erlang de parámetro m .

Cuando trabajamos con variables aleatorias enteras, por lo general, es preferible trabajar con la función generadora de probabilidad

$$G_N(z) = E[z^N].$$

La función generadora para la suma de variables aleatorias discretas independientes, $N = X_1 + \dots + X_n$, es

$$\begin{aligned} G_N(z) &= E[z^{X_1 + \dots + X_n}] = E[z^{X_1}] \dots E[z^{X_n}] \\ &= G_{X_1}(z) \dots G_{X_n}(z). \end{aligned} \quad (7.11)$$

Ejemplo 7.6 Calcula la función generatriz de una suma de n variables aleatorias independientes e idénticamente distribuidas geométricas.

Solución

La función generadora de una variable aleatoria Geométrica está dada por

$$G_X(z) = \frac{pz}{1 - qz}.$$

Por lo tanto, la función generadora de la suma de n de estas variables aleatorias independientes es

$$G_N(z) = \left\{ \frac{pz}{1 - qz} \right\}^n.$$

Vimos que ésta es la función generatriz de una variable aleatoria Binomial negativa de parámetros p y n .

7.2 La media muestral y las leyes de los grandes números

Sea X una variable aleatoria cuya media, $E[X] = \mu$, es desconocida. Sean X_1, \dots, X_n n medidas repetidas e independientes de X ; esto es, las X_j son variables aleatorias **independientes e idénticamente distribuidas** (iid) con la misma pdf que X . La **media muestral** de la serie se utiliza para estimar $E[X]$:

$$M_n = \frac{1}{n} \sum_{j=1}^n X_j. \quad (7.12)$$

En esta sección calcularemos la esperanza y la varianza de M_n con el objetivo de comprobar la efectividad de M_n como estimador para $E[X]$. También investigaremos el comportamiento de M_n cuando n crece.

El siguiente ejemplo muestra que el estimador de frecuencia relativa para la probabilidad de un suceso es un caso especial de la media muestral. Entonces, los resultados derivados más abajo para la media muestral son también aplicables al estimador de la frecuencia relativa.

Ejemplo 7.7 Frecuencia relativa

Consideremos una secuencia de repeticiones independientes de un experimento aleatorio y sea I_j la variable aleatoria función indicatriz de la ocurrencia del suceso A en la prueba j -ésima. El número total de ocurrencias de A en los primeros n ensayos es entonces

$$N_n = I_1 + I_2 + \dots + I_n.$$

La **frecuencia relativa** del suceso A en las n primeras repeticiones del experimento es entonces

$$f_A(n) = \frac{1}{n} \sum_{j=1}^n I_j. \quad (7.13)$$

Entonces, la frecuencia relativa $f_A(n)$ es simplemente la media muestral de las variables aleatorias I_j .

La media muestral en sí es una variable aleatoria, por lo que presenta una variación aleatoria. Un buen estimador debe tener las siguientes dos propiedades: (1) En el promedio, debe dar el valor correcto del parámetro que se calcula,

es decir, $E[M_n] = \mu$; y (2) no debe variar demasiado del valor correcto del parámetro, es decir, $E[(M_n - \mu)^2]$ es pequeño.

La esperanza de la media muestral viene dada por

$$E[M_n] = E\left[\frac{1}{n} \sum_{j=1}^n X_j\right] = \frac{1}{n} \sum_{j=1}^n E[X_j] = \mu, \quad (7.14)$$

dado que $E[X_j] = E[X] = \mu$ para todo j . Así, la media muestral es igual a $E[X] = \mu$, en promedio. Por esta razón, se dice que la media muestral es un **estimador insesgado** de μ . La ecuación (7.14) implica que el error medio cuadrático de la media muestral alrededor de μ es igual a la varianza de M_n , es decir,

$$E[(M_n - \mu)^2] = E[(M_n - E[M_n])^2].$$

Nótese que $M_n = S_n/n$, donde $S_n = X_1 + X_2 + \dots + X_n$. De la ecuación (7.4), $V[S_n] = n V[X_j] = n\sigma^2$, ya que las X_j son variables aleatorias iid. Así

$$V[M_n] = \frac{1}{n^2} V[S_n] = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}. \quad (7.15)$$

La ecuación (7.15) indica que la varianza de la media muestral se aproxima a cero, cuando el número de muestras crece. Esto implica que la probabilidad de que la media muestral esté cerca de la media real se acerca a uno a medida que n se hace muy grande. Podemos formalizar esta afirmación mediante el uso de la desigualdad de Chebyshev, la ecuación (4.76):

$$P[|M_n - E[M_n]| \geq \varepsilon] \leq \frac{V[M_n]}{\varepsilon^2}.$$

Sustituyendo $E[M_n]$ y $V[M_n]$, obtenemos

$$P[|M_n - \mu| \geq \varepsilon] \leq \frac{\sigma^2}{n\varepsilon^2} \quad (7.16)$$

Si tenemos en cuenta el complementario del suceso considerado en la ecuación (7.16), obtenemos

$$P[|M_n - \mu| < \varepsilon] \geq 1 - \frac{\sigma^2}{n\varepsilon^2}. \quad (7.17)$$

Así, para cualquier elección del error ε y probabilidad $1 - \delta$, podemos seleccionar el número de muestras n para que M_n se encuentre dentro de un intervalo de ancho ε alrededor de la media verdadera con probabilidad $1 - \delta$ o superior. Los siguientes ejemplos ilustran esto.

Ejemplo 7.8 Se mide un voltaje de valor constante pero desconocido. Cada medida X_j es en realidad la suma de las v tensiones deseadas y un ruido N_j de media cero y desviación típica de 1 microvoltio (μV):

$$X_j = v + N_j.$$

Supongamos que las tensiones de ruido son variables aleatorias independientes. ¿Cuántas mediciones son necesarias para que la probabilidad de que M_n se encuentra a menos de $\varepsilon = 1\mu V$ de la media real sea por lo menos 0.99?

Solución

Cada medición X_j tiene media v y varianza 1, así que a partir de la ecuación (7.17) se requiere que n satisfaga

$$1 - \frac{\sigma^2}{n\varepsilon^2} = 1 - \frac{1}{n} = 0.99.$$

Esto implica que $n = 100$.

Por tanto, si tuviéramos que repetir la medición 100 veces y calcular la media de la muestra, en promedio, por lo menos 99 de cada 100 veces, la media muestral resultante estará a menos de $1\mu V$ del verdadero valor.

Notese que si n tiende a infinito en la ecuación (5.20) obtenemos

$$\lim_{n \rightarrow \infty} P[|M_n - \mu| < \varepsilon] = 1.$$

La ecuación (7.17) requiere que las X_j tengan varianza finita. Se puede demostrar que este límite se mantiene incluso si la varianza de las X_j no existe. Presentamos este resultado más general:

TEOREMA 7.1 (Ley débil de los grandes números) *Sea X_1, X_2, \dots una secuencia de variables aleatorias iid con media finita $E[X] = \mu$, entonces para $\varepsilon > 0$,*

$$\lim_{n \rightarrow \infty} P[|M_n - \mu| < \varepsilon] = 1. \quad (7.18)$$

La ley débil de los grandes números establece que para un valor fijo suficientemente grande de n , la media muestral utilizando n muestras estará cerca de la media real con una probabilidad alta. La ley débil de los grandes números no responde a la pregunta sobre qué pasa con la media muestral como función de n al realizar mediciones adicionales. Esta pregunta se responde con la ley fuerte de los grandes números, que discutiremos a continuación.

Supongamos que realizamos una serie de mediciones independientes de la misma variable aleatoria. Sea X_1, X_2, \dots la secuencia resultante de variables aleatorias iid con media μ . Consideremos ahora la *secuencia de medias muestrales* resultante de las mediciones anteriores: M_1, M_2, \dots , donde M_j es la media muestral calculada usando desde la muestra X_1 hasta X_j . La noción de regularidad estadística discutida en el capítulo 1 nos lleva a esperar que esta secuencia de medias muestrales converja a μ , es decir, que esperamos que con probabilidad alta, *cada secuencia de medias muestrales se acerca a μ y se queda allí*, como se muestra en la figura 7.1. En términos de probabilidades, esperamos lo siguiente:

$$P\left[\lim_{n \rightarrow \infty} M_n = \mu\right] = 1;$$

esto es, con certeza virtual, cualquier secuencia de cálculo de medias muestrales

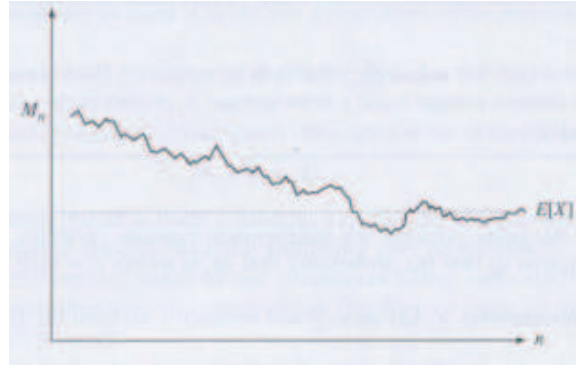


Figura 7.1 Convergencia de la secuencia de medias muestrales a $E[X]$.

converge a la media verdadera de la cantidad. La demostración de este resultado está muy por encima del nivel de este curso, pero tendremos la oportunidad de aplicar el resultado en situaciones diversas en secciones posteriores.

TEOREMA 7.2 (Ley fuerte de los grandes números) *Sea X_1, X_2, \dots una secuencia de variables aleatorias iid con media finita $E[X] = \mu$ y varianza finita, entonces*

$$P\left[\lim_{n \rightarrow \infty} M_n = \mu\right] = 1. \quad (7.19)$$

La ecuación (7.19) parece similar a la ecuación (7.18) pero en realidad son afirmaciones drásticamente diferentes. Establece que *con probabilidad 1, cualquier secuencia de cálculos de medias muestrales se aproximarán eventualmente y permanecerán cerca de $E[X] = \mu$* . Este es el tipo de convergencia que esperamos en situaciones físicas donde la regularidad estadística se verifica.

Con la ley fuerte de los grandes números cerramos el círculo del proceso de modelización. Comenzamos en el capítulo 1 afirmando que la regularidad estadística se observa en muchos fenómenos físicos y de esto dedujimos un número de propiedades de la frecuencia relativa. Estas propiedades se usaron para formular una serie de axiomas a partir de los cuales desarrollamos la teoría matemática de la probabilidad. Ahora hemos cerrado el círculo y demostrado que, bajo ciertas condiciones, la *teoría* predice la convergencia de las medias muestrales a los valores esperados. Hay aún puntos no unidos entre la teoría matemática y el mundo real (es decir, nunca vamos a poder llevar a cabo un número infinito de mediciones y calcular un número infinito de medias muestrales). Sin embargo, la ley fuerte de los grandes números demuestra la notable coherencia entre la teoría y el comportamiento físico observado.

Ya hemos indicado que las frecuencias relativas son casos especiales de la media muestral. Si aplicamos la ley débil de los grandes números a la frecuencia relativa de un suceso A , $f_A(n)$, en una secuencia de repeticiones independientes de un experimento aleatorio, obtenemos

$$\lim_{n \rightarrow \infty} P[|f_A(n) - P[A]| < \varepsilon] = 1. \quad (7.20)$$

Si aplicamos la ley fuerte de los grandes números, obtenemos

$$P\left[\lim_{n \rightarrow \infty} f_A(n) = P[A]\right] = 1. \quad (7.21)$$

Ejemplo 7.9 Con el fin de estimar la probabilidad de un suceso A , se lleva a cabo una secuencia de ensayos de Bernoulli y se observa la frecuencia relativa de A . ¿Cuánto ha de ser n para tener una probabilidad 0.95 de que la frecuencia relativa esté a menos de 0.01 de $p = P[A]$?

Solución

Sea $X = I_A$ la función indicatriz de A . La media de I_A es $\mu = p$ y la varianza es $\sigma^2 = p(1-p)$. Como p es desconocido, σ^2 también es desconocida. Sin embargo, es fácil demostrar que $p(1-p)$ es a lo sumo $1/4$ para $0 \leq p \leq 1$. Por lo tanto, por la ecuación (7.16),

$$P[|f_A(n) - p| \geq \varepsilon] \leq \frac{\sigma^2}{n\varepsilon^2} \leq \frac{1}{4n\varepsilon^2}.$$

La precisión deseada es $\varepsilon = 0.01$ y la probabilidad deseada es

$$1 - 0.95 = \frac{1}{4n\varepsilon^2}.$$

Entonces, resolviendo para n obtenemos $n = 50000$. Ya se ha señalado que la desigualdad de Chebyshev proporciona límites muy poco ajustados, así que esperamos que este valor de n sea probablemente demasiado conservador. En la siguiente sección, se presenta una estimación mejor para el valor requerido de n .

7.3 El teorema central del límite

Sea X_1, X_2, \dots una sucesión de variables aleatorias iid con media finita μ y varianza finita σ^2 y sea S_n la suma de las n primeras variables aleatorias en la secuencia:

$$S_n = X_1 + X_2 + \dots + X_n. \quad (7.22)$$

En la sección 7.1 hemos desarrollado métodos para determinar la pdf exacta de S_n . A continuación se presenta el teorema central del límite, que establece que, a medida que n crece, la cdf de S_n adecuadamente estandarizada se aproxima a la de una variable aleatoria Normal. Esto nos permite aproximar la cdf de S_n con la de una variable aleatoria Normal.

El teorema central del límite explica por qué la variable aleatoria Normal aparece en tantas aplicaciones y tan diversas. En la naturaleza, muchos de los fenómenos macroscópicos resultan de la suma de numerosos procesos independientes, microscópicos, lo que da lugar a la variable aleatoria Normal. En muchas

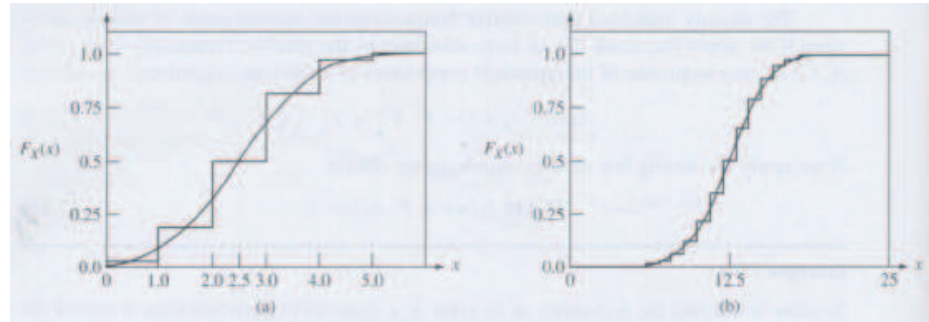


Figura 7.2 (a) La cdf de la suma de cinco variables aleatorias independientes de Bernoulli con $p = 1/2$ y la cdf de una variable aleatoria Normal con la misma media y varianza. (b) La cdf de la suma de 25 variables aleatorias de Bernoulli con $p = 1/2$ y la cdf de una variable aleatoria Normal con la misma media y varianza.

aplicaciones artificiales, estamos interesados en promedios que consisten a menudo en la suma de variables aleatorias independientes. De nuevo, esto da lugar a la variable aleatoria Normal.

A partir del ejemplo 7.2, sabemos que si las X_j son iid, entonces S_n tiene media $n\mu$ y varianza $n\sigma^2$. El teorema central del límite establece que la cdf de la versión estandarizada de S_n se aproxima a la de la variable aleatoria Normal.

TEOREMA 7.3 (Teorema Central del Límite) *Sea S_n la suma de n variables aleatorias iid con media finita $E[X] = \mu$ y varianza finita σ^2 y sea Z_n una variable aleatoria de media cero y varianza uno definida como*

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}, \quad (7.23a)$$

entonces

$$\lim_{n \rightarrow \infty} P[Z_n \leq z] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-x^2/2} dx. \quad (7.23b)$$

Nótese que Z_n se escribe a veces en términos de la media muestral:

$$Z_n = \sqrt{n} \frac{M_n - \mu}{\sigma}. \quad (7.24)$$

Lo fascinante de el teorema central del límite es que los sumandos X_j pueden tener *cualquier* distribución siempre que tengan media y varianza finitas. Esto le proporciona una gran aplicabilidad al resultado.

Las figuras 7.2 a 7.4 comparan la cdf exacta y la aproximación normal para la suma de variables aleatorias de Bernoulli, uniformes y exponenciales, respectivamente. En estos tres casos, se puede ver que la aproximación mejora con el aumento de términos en la suma. La prueba del teorema central del límite se discute en la última parte de esta sección.

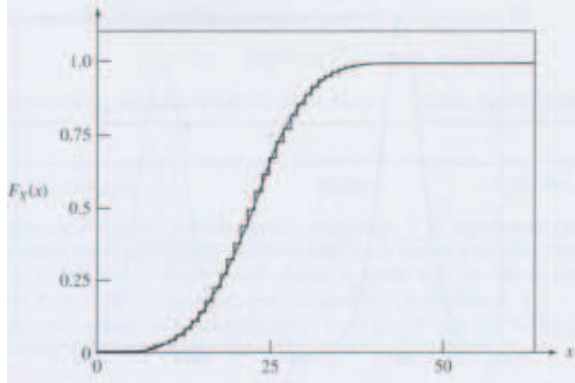


Figura 7.3 La cdf de la suma de cinco variables aleatorias independientes discretas uniformes en el conjunto $\{0, 1, \dots, 9\}$ y la cdf de una variable aleatoria Normal con la misma media y varianza.

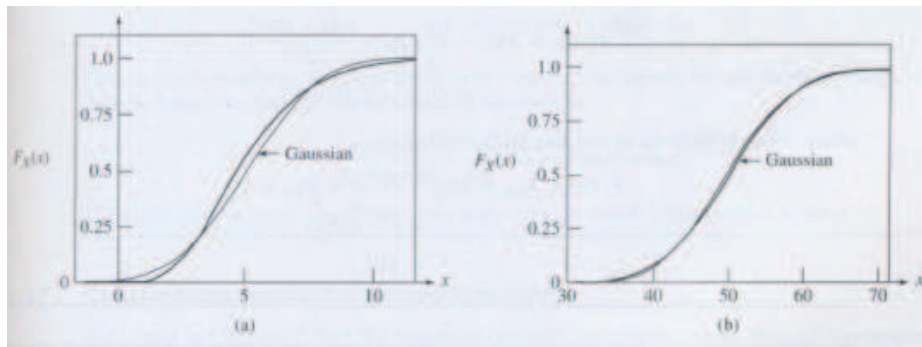


Figura 7.4 (a) La cdf de la suma de cinco variables aleatorias independientes exponenciales con media 1 y la cdf de una variable aleatoria Normal con la misma media y varianza. (b) La cdf de la suma de 50 variables aleatorias exponenciales de media 1 y la cdf de una variable aleatoria Normal con la misma media y varianza.

Ejemplo 7.10 Supongamos que los pedidos en un restaurante son variables aleatorias iid con media $\mu = 8$ euros y desviación típica $\sigma = 2$ euros. Estima la probabilidad de que los primeros 100 clientes se gasten un total de más de 840 euros. Estima la probabilidad de que los primeros 100 clientes se gasten un total de entre 780 y 820 euros.

Solución

Sea X_k el gasto del cliente k -ésimo, entonces el gasto total de los primeros 100 clientes es

$$S_{100} = X_1 + X_2 + \dots + X_{100}.$$

La media de S_{100} es $n\mu = 800$ y la varianza es $n\sigma^2 = 400$. La figura 7.5 muestra la pdf de S_{100} donde se puede ver que la pdf se concentra mucho alrededor de la media. La forma estandarizada de S_{100} es

$$Z_{100} = \frac{S_{100} - 800}{20}.$$

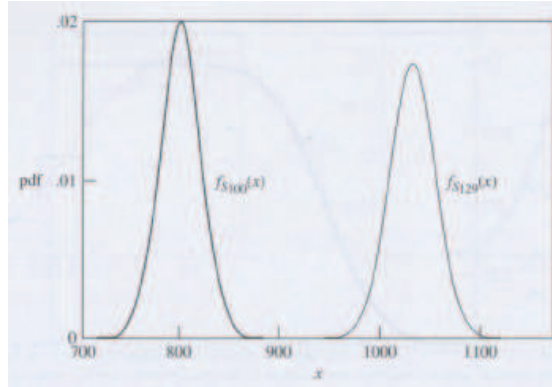


Figura 7.5 Aproximación normal de la pdf de S_{100} y S_{129} en los ejemplos 7.10 y 7.11.

Entonces

$$P[S_{100} > 840] = P\left[Z_{100} > \frac{840 - 800}{20}\right] \\ \simeq Q(2) = 2.28(10^{-2}),$$

donde utilizamos la tabla de la Normal para evaluar $Q(2)$. De forma similar,

$$P[780 \leq S_{100} \leq 820] = P[-1 \leq Z_{100} \leq 1] \\ \simeq 1 - 2Q(1) \\ = 0.682.$$

Ejemplo 7.11 En el ejemplo 7.10, ¿después de cuántos pedidos podemos estar seguros al 90 % de que el gasto total de todos los clientes es mayor a 1000 euros?

Solución

El problema aquí es calcular el valor de n para el cual

$$P[S_n > 1000] = 0.90.$$

S_n tiene media $8n$ y varianza $4n$. Procediendo como en el ejemplo anterior, tenemos

$$P[S_n > 1000] = P\left[Z_n > \frac{1000 - 8n}{2\sqrt{n}}\right] = 0.90.$$

Utilizando el hecho de que $Q(-x) = 1 - Q(x)$, la tabla 4.2 implica que n debe satisfacer

$$\frac{1000 - 8n}{2\sqrt{n}} = -1.2815,$$

que arroja la siguiente ecuación cuadrática para \sqrt{n} :

$$8n - 1.2815(2)\sqrt{n} - 1000 = 0.$$

La raíz positiva de la ecuación lleva a $\sqrt{n} = 11.34$ o $n = 128.6$. La figura 7.5 muestra la pdf para S_{129} .

Ejemplo 7.12 Los tiempos entre sucesos en cierto experimento aleatorio son variables aleatorias exponenciales iid con media m segundos. Calcula la probabilidad de que el 1000-ésimo suceso ocurra en el intervalo $(1000 \pm 50)m$.

Solución

Sea X_j el tiempo entre sucesos y sea S_n el tiempo del n -ésimo suceso, entonces S_n viene dado por la ecuación (7.22). Por la tabla 4.1 la media y la varianza de X_j viene dada por $E[X_j] = m$ y $V[X_j] = m^2$. La media y la varianza de S_n son entonces $E[S_n] = nE[X_j] = nm$ y $V[S_n] = nV[X_j] = nm^2$. El teorema central del límite entonces da

$$\begin{aligned} P[950m \leq S_{1000} \leq 1050m] &= P\left[\frac{950m - 1000m}{m\sqrt{1000}} \leq Z_n \leq \frac{1050m - 1000m}{m\sqrt{1000}}\right] \\ &\simeq Q(1.58) - Q(-1.58) \\ &= 1 - 2Q(1.58) \\ &= 1 - 2(0.0567) = 0.8866. \end{aligned}$$

Entonces, a medida que n crece, es muy probable que S_n esté cerca de su media nm . Podemos por tanto conjeturar que la tasa media asintótica a la que suceden los sucesos es

$$\frac{n \text{ sucesos}}{S_n \text{ segundos}} = \frac{n}{nm} = \frac{1}{m} \text{ sucesos/segundo.} \quad (7.25)$$

7.3.1 Aproximación normal de probabilidades binomiales

En el capítulo 2 vimos que la variable aleatoria Binomial es difícil de calcular de forma directa para n grande por la necesidad de calcular términos factoriales. Una aplicación particularmente importante del teorema central del límite es la aproximación de probabilidades binomiales. Como la variable aleatoria Binomial es una suma de variables aleatorias de Bernoulli iid (que tienen media y varianza finitas), su cdf se aproxima a aquella de la variable aleatoria Normal. Sea X una variable aleatoria Binomial con media np y varianza $np(1-p)$ y sea Y una variable aleatoria Normal con la misma media y varianza, entonces por el teorema central del límite para n grande la probabilidad de que $X = k$ es aproximadamente igual a la integral de la pdf normal en un intervalo de longitud k , como se muestra en la figura 7.6:

$$\begin{aligned} P[X = k] &\simeq P\left[k - \frac{1}{2} < Y < k + \frac{1}{2}\right] \\ &= \frac{1}{\sqrt{2\pi np(1-p)}} \int_{k-1/2}^{k+1/2} e^{-(x-np)^2/2np(1-p)} dx. \end{aligned} \quad (7.26)$$

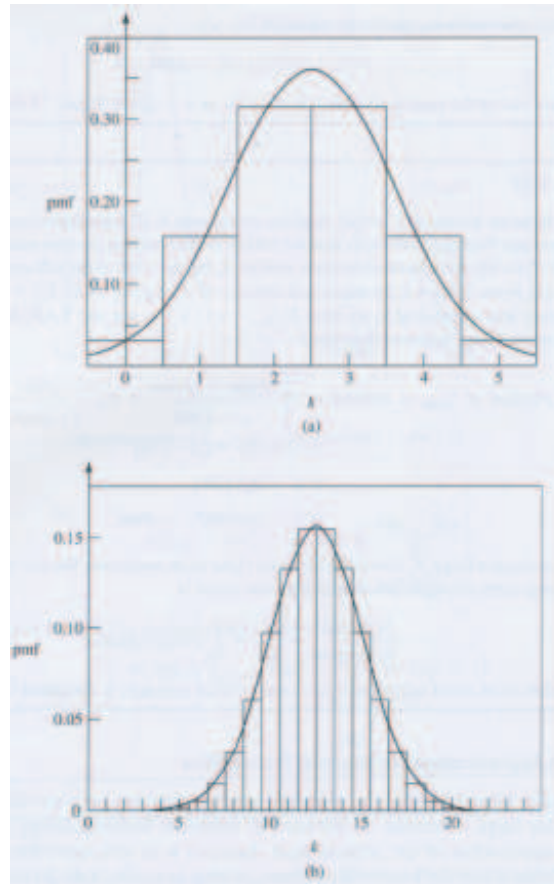


Figura 7.6 (a) Aproximación normal de probabilidades binomiales con $n = 5$ y $p = 1/2$. (b) Aproximación normal de la binomial con $n = 25$ y $p = 1/2$.

La aproximación de arriba se puede simplificar aproximando la integral con el producto del integrando en el centro del intervalo de integración (esto es, $x = k$) y la longitud del intervalo de integración (uno):

$$P[X = k] \simeq \frac{1}{\sqrt{2\pi np(1-p)}} e^{-(k-np)^2/2np(1-p)}. \quad (7.27)$$

Las figuras 7.6(a) y 7.6(b) comparan las probabilidades binomiales y las aproximaciones normales utilizando la ecuación (7.27).

Ejemplo 7.13 En el ejemplo 7.9 en la sección 7.2, utilizamos la desigualdad de Chebyshev para estimar el número de muestras requeridas para que haya una probabilidad de 0.95 de que la frecuencia relativa estime que la probabilidad de un

suceso A estará a menos de 0.01 de $P[A]$. Ahora estimamos el número necesario de muestras utilizando la aproximación normal de la distribución binomial.

Solución

Sea $f_A(n)$ la frecuencia relativa de A en n intentos de Bernoulli. Como $f_A(n)$ tiene media p y varianza $p(1-p)/n$, entonces

$$Z_n = \frac{f_A(n) - p}{\sqrt{p(1-p)/n}}$$

tiene media cero y varianza uno y es aproximadamente normal para n suficientemente grande. La probabilidad de interés es

$$P[|f_A(n) - p| < \varepsilon] \simeq P\left[|Z_n| < \frac{\varepsilon\sqrt{n}}{\sqrt{p(1-p)}}\right] = 1 - 2Q\left(\frac{\varepsilon\sqrt{n}}{\sqrt{p(1-p)}}\right).$$

La probabilidad de arriba no se puede calcular porque p es desconocida. Sin embargo, se demuestra fácilmente que $p(1-p) \leq 1/4$ para p en el intervalo unidad. Entonces se sigue que para dicho p , $\sqrt{p(1-p)} \leq 1/2$ y como $Q(x)$ decrece con un argumento creciente

$$P[|f_A(n) - p| < \varepsilon] > 1 - 2Q(2\varepsilon\sqrt{n}).$$

Queremos que la probabilidad de arriba sea igual a 0.95. Esto implica que $Q(2\varepsilon\sqrt{n}) = (1 - 0.95)/2 = 0.025$. Por la tabla de la Normal, vemos que el argumento de $Q(x)$ debe ser aproximadamente 1.95, entonces

$$2\varepsilon\sqrt{n} = 1.95.$$

Despejando la n , obtenemos

$$n = (0.98)^2/\varepsilon^2 = 9506.$$

Problemas

7.1 Sea X_1, \dots, X_n variables aleatorias con la misma media y con función de covarianza:

$$\text{Cov}(X_i, X_j) = \begin{cases} \sigma^2 & i = j \\ \rho\sigma^2 & |i - j| = 1, \\ 0 & \text{en otro caso.} \end{cases}$$

donde $|\rho| < 1$. Calcula la media y la varianza de $S_n = X_1 + \dots + X_n$.

7.2 Sea $S_k = X_1 + \dots + X_k$, donde las X_j son variables aleatorias independientes donde X_i es una variable aleatoria Chi cuadrado con n_i grados de libertad. Muestra que S_k es una variable aleatoria Chi-cuadrado con $n = n_1 + \dots + n_k$ grados de libertad.

7.3 El número X_j de ganadores de un sorteo en la clase j es una variable aleatoria binomial con parámetros n_j y p . Supongamos que la escuela tiene K clases. Calcula la pmf del número total de ganadores en la escuela, asumiendo que las X_j son las variables aleatorias independientes.

7.4 El número de llegadas de paquetes X_i al puerto i de un router es una variable aleatoria de Poisson con media α_i . Dado que el router tiene k puertos, calcula la pmf del número total de llegadas de paquetes al router. Asumamos que los X_i son variables aleatorias independientes.

7.5 Supongamos que el 20% de los votantes están a favor de cierta legislación. Se pregunta a un número alto n de votantes y se obtiene un estimador de frecuencia relativa $f_A(n)$ para la proporción anterior. Utiliza la ecuación (7.17) para determinar cuántos votantes hay que preguntar para que $f_A(n)$ difiera de 0.2 menos de 0.02 con probabilidad 0.95.

7.6 El tiempo de vida de una bombilla barata es una variable aleatoria exponencial con media 1 semana. Suponiendo que se prueban 16 bombillas, midiendo sus tiempos de vida. Usa el teorema central del límite para estimar la probabilidad de que la suma de los tiempos de vida sea menor que 600 horas.

7.7 Un canal de transmisiones binario introduce errores en un bit con probabilidad 0.15. Estima la probabilidad de que haya 20 o menos errores en la transmisión de 100 bits.

Parte III

Procesos estocásticos

8 Procesos estocásticos

En algunos experimentos aleatorios, el resultado es una función del tiempo o del espacio. Por ejemplo, en los sistemas de reconocimiento de voz, las decisiones se toman en base a una onda de tensión que corresponde a una expresión verbal. En un sistema de procesamiento de imágenes, la intensidad y el color de la imagen varían a lo largo de una región rectangular. En una red peer-to-peer, el número de pares en el sistema varía con el tiempo. En algunos casos, puede ser de interés considerar dos o más funciones de tiempo. Por ejemplo, la temperatura en una ciudad y la demanda de energía eléctrica locales varían en el tiempo. Las funciones aleatorias del tiempo en los ejemplos anteriores se pueden ver como cantidades numéricas que evolucionan aleatoriamente en el tiempo o en el espacio. Por lo tanto lo que realmente tenemos es una familia de variables aleatorias indexadas por el tiempo o la variable espacial. En este capítulo comenzamos el estudio de los procesos estocásticos. Se procederá de la siguiente manera:

- En la sección 8.1 se introduce la noción de proceso estocástico (o proceso aleatorio) que se define como una familia indexada de variables aleatorias.
- Estamos interesados en la especificación del comportamiento conjunto de las variables aleatorias en una familia (por ejemplo, la temperatura en dos instantes de tiempo). En la sección 8.2, vemos que ésta se realiza mediante la especificación de funciones de distribución conjunta, así como las funciones media y covarianza.
- En la sección 8.3 se presentan ejemplos de procesos estocásticos y se muestra cómo modelos de procesos complejos se pueden desarrollar a partir de modelos sencillos.
- En la sección 8.4, se introduce la clase de procesos estocásticos estacionarios que pueden ser vistos como procesos estocásticos en “estado estable”.

8.1 Definición de proceso estocástico

Consideremos un experimento aleatorio especificado por los resultados ζ de un cierto espacio muestral S , por los sucesos definidos en S , y por las probabilidades de estos sucesos. Supongamos que a todos los resultados $\zeta \in S$, se les asigna una función del tiempo:

$$X(t, \zeta), \quad t \in I.$$

La gráfica de la función $X(t, \zeta)$ en función de t con ζ fijo, se denomina realización, camino muestral o función muestral del proceso estocástico.

Así, podemos ver el resultado del experimento aleatorio como la generación de una función completa del tiempo, como se muestra en la Figura 8.1. Por otro lado, si fijamos un tiempo t_k perteneciente al conjunto de índices I , entonces $X(t_k, \zeta)$ es una variable aleatoria (ver Figura 8.1), ya que estamos asociando números reales a valores de ζ .

Así, hemos creado una familia (o conjunto) de variables aleatorias indexadas por el parámetro t , $\{X(t, \zeta), t \in I\}$. Esta familia se llama proceso aleatorio. Los procesos aleatorios se llaman también procesos estocásticos. En general, se elimina el símbolo ζ y se usa $X(t)$ para denotar un proceso estocástico. Un proceso estocástico se dice que es en tiempo discreto si el conjunto de índices es un conjunto numerable (es decir, el conjunto de los números enteros o el conjunto de números enteros no negativos).

Cuando se tratan procesos en tiempo discreto, se suele utilizar n para denotar el índice de tiempo y X_n para indicar el proceso estocástico. Un proceso estocástico en tiempo continuo es uno en el que I es continuo (es decir, la recta real o la recta real no negativa). El siguiente ejemplo muestra que podemos imaginar un proceso estocástico como el resultado de seleccionar ζ en el instante de tiempo inicial y revelarla poco a poco en el tiempo a través de $X(t, \zeta)$.

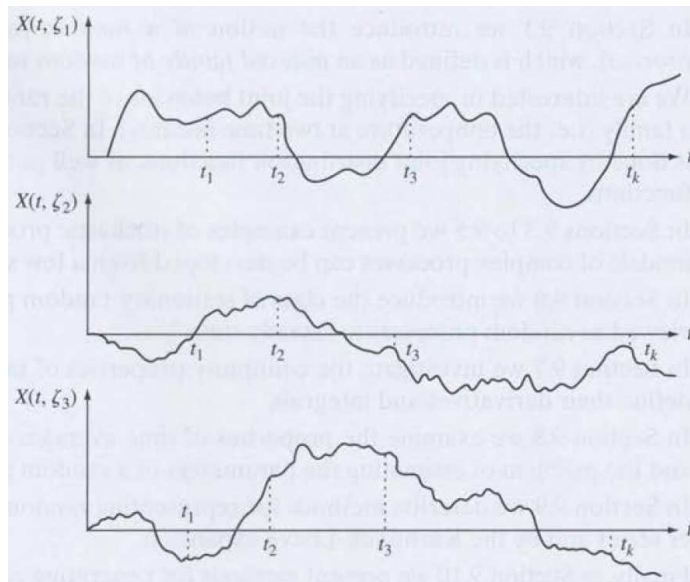


Figura 8.1 Distintas realizaciones de un proceso estocástico.

Ejemplo 8.1 Secuencia binaria aleatoria

Sea ζ un número aleatorio perteneciente al intervalo $[0, 1]$, y sea b_1, b_2, \dots la expansión binaria de ζ :

$$\zeta = \sum_{i=1}^{\infty} b_i 2^{-i}, \text{ donde } b_i \in \{0, 1\}.$$

Si se define el proceso estocástico en tiempo discreto $X(n, \zeta)$ como

$$X(n, \zeta) = b_n, \quad n = 1, 2, \dots$$

el proceso resultante es una secuencia de números binarios, con $X(n, \zeta)$ igual al n -ésimo número de la expansión binaria de ζ .

Ejemplo 8.2 Sinusoides aleatorias

Sea ζ un número aleatorio perteneciente al intervalo $[-1, 1]$. Definimos el proceso estocástico en tiempo continuo $X(t, \zeta)$ por

$$X(t, \zeta) = \zeta \cos(2\pi t) \quad -\infty < t < \infty.$$

Las realizaciones de este proceso estocástico son sinusoides con amplitud ζ (como se muestra en la Figura 8.2 (a)). Sea ζ un número aleatorio perteneciente al intervalo $(-\pi, \pi)$ y sea $Y(t, \zeta) = \cos(2\pi t + \zeta)$. Las realizaciones de $Y(t, \zeta)$ son versiones desfasadas de $\cos 2\pi t$ como se muestra en la Figura 8.2 (b).

La aleatoriedad en ζ induce aleatoriedad en la función observada $X(t, \zeta)$. En principio, se puede deducir la probabilidad de eventos que involucran un proceso estocástico en distintos instantes de tiempo de las probabilidades que involucran ζ mediante el método del suceso equivalente introducido en el Capítulo 4.

Ejemplo 8.3 Encuentra las siguientes probabilidades para el proceso estocástico presentado en el Ejemplo 8.1: $P[X(1, \zeta) = 0]$ y $P[X(1, \zeta) = 0 \text{ y } X(2, \zeta) = 1]$.

Las probabilidades se obtienen mediante la búsqueda de los sucesos equivalentes en términos de ζ :

$$P[X(1, \zeta) = 0] = P\left[0 \leq \zeta < \frac{1}{2}\right] = \frac{1}{2}$$

$$P[X(1, \zeta) = 0 \text{ y } X(2, \zeta) = 1] = P\left[\frac{1}{4} \leq \zeta < \frac{1}{2}\right] = \frac{1}{4},$$

porque todos los puntos en el intervalo $[0, 1/2)$ comienzan con $b_1 = 0$ y todos

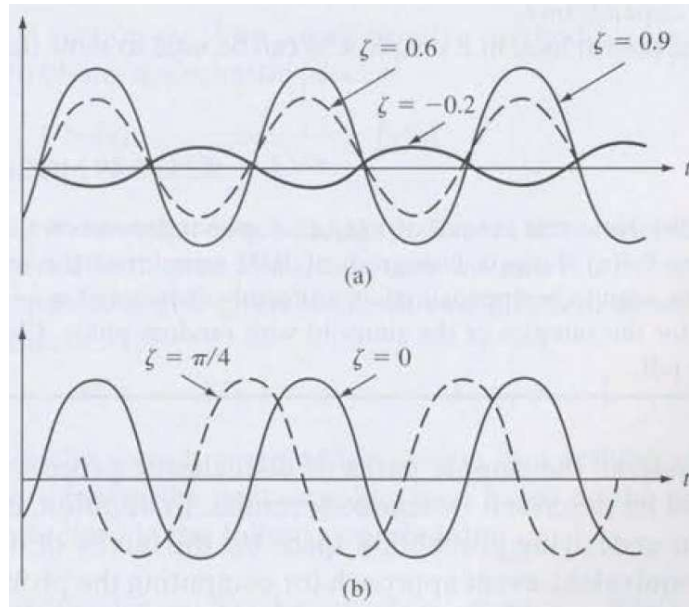


Figura 8.2 (a) Sinusoide amplitud aleatoria, (b) Sinusoide con fase aleatoria.

los puntos en $[1/4, 1/2)$ comienzan con $b_1 = 0$ y $b_2 = 1$. Claramente, cualquier secuencia de k bits tiene un subintervalo correspondiente de amplitud (y por tanto de probabilidad) 2^{-k} .

Ejemplo 8.4 Encontrar la pdf de $X_0 = X(t_0, \zeta)$ e $Y(t_0, \zeta)$, en el Ejemplo 8.2.

Si t_0 es tal que $\cos(2\pi t_0) = 0$, entonces $X(t_0, \zeta) = 0$ para todo ζ y la pdf de la $X(t_0)$ es una función delta en 0. De lo contrario, $X(t_0, \zeta)$ se distribuye uniformemente en el intervalo $(-\cos 2\pi t_0, \cos 2\pi t_0)$, ya que ζ se distribuye uniformemente en $[-1, 1]$ (ver Figura 8.3(a)). Notese que la pdf de $X(t_0, \zeta)$ depende de t_0 . El enfoque utilizado en el Ejemplo 4.33 se puede utilizar para demostrar que $Y(t_0, \zeta)$ tiene una distribución arcoseno:

$$f_Y(y) = \frac{1}{\pi\sqrt{1-y^2}}, \quad |y| < 1$$

(ver Figura 8.3(b)). Tengamos en cuenta que la pdf de la $Y(t_0, \zeta)$ no depende de t_0 . La Figura 8.3(c) muestra un histograma de 1000 muestras de las amplitudes $X(t_0, \zeta)$ en $t_0 = 0$, que como se puede ver tiene distribución aproximadamente uniforme en $[-1, 1]$. La Figura 8.3(d) muestra el histograma para las muestras de la sinusoide con fase aleatoria. Es evidente que es consistente con una pdf

arcoseno.

En general, los caminos muestrales de un proceso estocástico pueden ser bastante complicados y no se pueden describir mediante fórmulas simples. Además, normalmente no es posible identificar un espacio de probabilidad subyacente para la familia de funciones del tiempo observadas. Así, el enfoque del suceso equivalente para el cálculo de la probabilidad de sucesos que involucran $X(t, \zeta)$ en términos de las probabilidades de sucesos relacionados con ζ , no es útil en la práctica. En la siguiente sección se describe un método alternativo para especificar las probabilidades de sucesos que involucran un proceso estocástico.

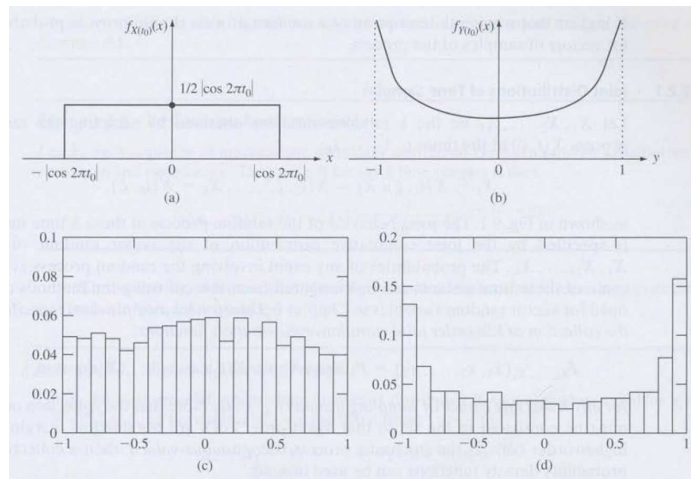


Figura 8.3 (a) pdf de una senoide con amplitud aleatoria. (b) pdf de una senoide con fase aleatoria. (c) Histograma de las muestras de una senoide con amplitud aleatoria uniformemente distribuida en $t = 0$. (d) Histograma de las muestras de una senoide con fase aleatoria en $t = 0$.

8.2 Especificación de un proceso estocástico

Hay muchas preguntas acerca de los procesos estocásticos que no se pueden responder únicamente con el conocimiento de la distribución en un solo instante de tiempo. Por ejemplo, nos puede interesar la temperatura en un lugar determinado en dos momentos diferentes. Esto requiere la siguiente información:

$$P[a < X(t_1) \leq b, c < X(t_2) \leq d].$$

En otro ejemplo, el sistema de compresión de voz en un teléfono celular predice el valor de la señal de voz en el siguiente instante de muestreo en base a las k

muestras anteriores. Así, podemos estar interesados en la siguiente probabilidad:

$$P[a < X(t_{k+1}) \leq b | X(t_1) = x_1, X(t_2) = x_2, \dots, X(t_k) = x_k].$$

Está claro que una descripción general de un proceso estocástico debe proporcionar probabilidades para los vectores de las muestras del proceso.

8.2.1 Distribuciones conjuntas de las muestras en el tiempo

Sean X_1, X_2, \dots, X_k las k variables aleatorias obtenidas mediante muestreo del proceso $X(t, \zeta)$ en los instantes t_1, t_2, \dots, t_k :

$$X_1 = X(t_1, \zeta), X_2 = X(t_2, \zeta), \dots, X_k = X(t_k, \zeta),$$

como se muestra en la Figura 8.1. El comportamiento conjunto del proceso estocástico en estos k instantes se especifica mediante la función de distribución conjunta del vector de variables aleatorias X_1, X_2, \dots, X_k . Las probabilidades de cualquier suceso relacionado con el proceso estocástico en todo o algunos de estos instantes de tiempo se puede calcular a partir de esta cdf usando los métodos desarrollados en el Capítulo 6 para variables aleatorias vectoriales. Por lo tanto, un proceso estocástico se especifica mediante la familia de funciones de distribución conjunta de orden k :

$$F_{X_1, \dots, X_k}(x_1, x_2, \dots, x_k) = P[X(t_1) \leq x_1, X(t_2) \leq x_2, \dots, X(t_k) \leq x_k],$$

para cualquier k , y cualquier conjunto de instantes de muestreo t_1, \dots, t_k . Nótese que la colección de cdfs deben ser consistentes en el sentido de que las cdfs de orden más bajo se obtienen como marginales de las cdfs de orden más alto.

Si el proceso estocástico toma valores continuos, entonces una colección de funciones de densidad de probabilidad se puede utilizar en su lugar:

$$f_{X_1, \dots, X_k}(x_1, x_2, \dots, x_k) dx_1 \dots dx_k = P[x_1 < X(t_1) \leq x_1 + dx_1, \dots, x_k < X(t_k) \leq x_k + dx_k].$$

Si el proceso estocástico toma valores discretos, entonces una colección de funciones de masa de probabilidad se puede utilizar para especificar el proceso estocástico:

$$f_{X_1, \dots, X_k}(x_1, x_2, \dots, x_k) = P[X(t_1) = x_1, X(t_2) = x_2, \dots, X(t_k) = x_k].$$

para cualquier k , y cualquier conjunto de instantes de muestreo n_1, \dots, n_k .

A primera vista no parece que hayamos avanzado mucho en la especificación de procesos estocásticos, porque ahora estamos ante la tarea de especificar una gran colección de conjuntos de cdfs. Sin embargo, este método funciona porque la mayoría de los modelos más útiles de procesos estocásticos se obtienen mediante la elaboración de unos pocos modelos simples, por lo que los métodos desarrollados en los capítulos 5 y 6 de este libro se pueden utilizar para obtener las cdfs necesarias. Los siguientes ejemplos dan una vista previa de cómo construir

modelos complejos a partir de modelos simples. Desarrollamos estos importantes ejemplos con más detalle en la sección 8.3.

Ejemplo 8.5 Variables aleatorias iid de Bernoulli

Sea X_n una secuencia de variables aleatorias de Bernoulli independientes e idénticamente distribuidas con $p = 1/2$. La pmf conjunta de las muestras en k instantes de tiempo es entonces

$$P[X_1 = x_1, X_2 = x_2, \dots, X_k = x_k] = P[X_1 = x_1] \cdots P[X_k = x_k] = \left(\frac{1}{2}\right)^k$$

donde $x_i \in \{0, 1\}$ para todo i . Este proceso estocástico binario es equivalente a el que se ha descrito en el Ejemplo 8.1.

Ejemplo 8.6 Variables aleatorias iid gaussianas

Sea X_n una secuencia de variables aleatorias independientes e idénticamente distribuidas gaussianas con media cero y varianza σ_X^2 . La pdf conjunta de las muestras en k instantes de tiempo es entonces

$$f_{X_1, X_2, \dots, X_k}(x_1, x_2, \dots, x_k) = \frac{1}{(2\pi\sigma^2)^{k/2}} e^{-(x_1^2 + x_2^2 + \dots + x_k^2)/2\sigma^2}$$

Los dos ejemplos siguientes muestran cómo los procesos más complejos e interesantes se puede construir a partir de secuencias de variables aleatorias iid.

Ejemplo 8.7 Proceso de conteo binomial

Sea X_n una secuencia de variables aleatorias independientes e idénticamente distribuidas de Bernoulli con $p = 1/2$. Sea S_n el número de unos en los primeros n ensayos:

$$S_n = X_1 + X_2 + \dots + X_n \text{ para } n = 0, 1, \dots$$

S_n es una función no decreciente con valores enteros de n que crece a pasos unitarios después de un número aleatorio de instantes de tiempo. De los capítulos anteriores, sabemos que S_n es una variable aleatoria binomial con parámetros n y $p = 1/2$. En la siguiente sección se muestra cómo encontrar las pmf conjuntas de S_n usando probabilidades condicionadas.

Ejemplo 8.8 Señal con ruido filtrada

Sea X_j una secuencia de observaciones independientes e idénticamente distribuidas de una señal de tensión μ corrompida por un ruido gaussiano N_j con media cero y varianza σ^2

$$X_j = \mu + N_j \text{ para } j = 0, 1, \dots$$

Considere la señal que resulta de calcular el promedio de la secuencia de observaciones:

$$S_n = (X_1 + X_2 + \dots + X_n)/n \text{ para } n = 0, 1, \dots$$

De los capítulos anteriores se sabe que S_n es la media muestral de una secuencia de variables aleatorias iid gaussianas. Sabemos que S_n es una variable aleatoria gaussiana con media μ y varianza σ^2/n , por lo que tiende hacia el valor de μ a medida que aumenta n . En una sección posterior, se muestra que S_n es un ejemplo de la clase de los procesos estocásticos gaussianos.

8.2.2 Las funciones media, autocorrelación y autocovarianza

Los momentos de las muestras en el tiempo de un proceso estocástico se pueden utilizar para especificar parcialmente el proceso estocástico, ya que resumen la información contenida en las cdf conjuntas.

La función media $m_X(t)$ y la función varianza $\text{VAR}[X(t)]$ del proceso estocástico de tiempo continuo $X(t)$ se definen como

$$m_X(t) = E[X(t)] = \int_{-\infty}^{\infty} x f_{X(t)}(x) dx \quad (8.1)$$

y

$$\text{VAR}[X(t)] = \int_{-\infty}^{\infty} (x - m_X(t))^2 f_{X(t)}(x) dx \quad (8.2)$$

donde $f_{X(t)}(x)$ es la pdf de $X(t)$. Nótese que $m_X(t)$ y $\text{VAR}[X(t)]$ son funciones deterministas del tiempo. La tendencia en el comportamiento de $X(t)$ se refleja en la variación de $m_X(t)$ con el tiempo. La variancia da una idea de la dispersión de los valores tomados por $X(t)$ en instantes de tiempo distintos.

La autocorrelación $R_X(t_1, t_2)$ de un proceso estocástico $X(t)$ se define como el momento conjunto de $X(t_1)$ y $X(t_2)$:

$$R_X(t_1, t_2) = E[X(t_1)X(t_2)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f_{X(t_1), X(t_2)}(x, y) dx dy, \quad (8.3)$$

donde $xy f_{X(t_1), X(t_2)}(x, y)$ es la pdf de segundo orden de $X(t)$. En general, la autocorrelación es una función de t_1 y t_2 . Nótese que $R_X(t, t) = E[X^2(t)]$. La

autocovarianza $C_X(t_1, t_2)$ de un proceso estocástico $X(t)$ se define como la covarianza de $X(t_1)$ y $X(t_2)$:

$$C_X(t_1, t_2) = E[\{X(t_1) - m_X(t_1)\}\{X(t_2) - m_X(t_2)\}] \quad (8.4)$$

De la Ecuación (5.30), la autocovarianza se puede expresar en términos de la autocorrelación y de la media:

$$C_X(t_1, t_2) = R_X(t_1, t_2) - m_X(t_1)m_X(t_2).$$

Se puede observar que la varianza de $X(t)$ se puede obtener de $C_X(t_1, t_2)$:

$$\text{VAR}[X(t)] = E[(X(t) - m_X(t))^2] = C_X(t, t).$$

El coeficiente de correlación de $X(t)$ se define como el coeficiente de correlación de $X(t_1)$ y $X(t_2)$ (véase la Ecuación (5.31)):

$$\rho_X(t_1, t_2) = \frac{C_X(t_1, t_2)}{\sqrt{C_X(t_1, t_1)}\sqrt{C_X(t_2, t_2)}}.$$

De la Ecuación (5.32) se tiene que $|\rho_X(t_1, t_2)| \leq 1$. Recordemos que el coeficiente de correlación es una medida del grado en que una variable aleatoria se puede predecir como una función lineal de otra.

Las funciones media, varianza, autocorrelación y autocovarianza de los procesos estocásticos en tiempo discreto se definen de la misma manera que la anterior. Usamos una notación un poco diferente para el índice de tiempo. La media y la varianza de un proceso estocástico de tiempo discreto X_n se definen como:

$$m_X(n) = E[X_n]$$

y

$$\text{VAR}[X_n] = E[(X_n - m_X(n))^2]$$

Las funciones de autocorrelación y autocovarianza de un proceso estocástico de tiempo discreto X_n se definen como sigue:

$$R_X(n_1, n_2) = E[X(n_1), X(n_2)]$$

y

$$C_X(n_1, n_2) = E[\{X(n_1) - m_X(n_1)\}\{X(n_2) - m_X(n_2)\}] = R_X(n_1, n_2) - m_X(n_1)m_X(n_2).$$

Antes de proceder con los ejemplos, reiteramos que las funciones media, autocorrelación y autocovarianza sólo son descripciones parciales de un proceso estocástico. Así, veremos más adelante en este capítulo que es posible que dos procesos estocásticos muy diferentes tengan las mismas funciones media, autocorrelación y autocovarianza.

Ejemplo 8.9 Sinusoides con amplitud aleatoria

Sea $X(t) = A\cos 2\pi t$, donde A es una variable aleatoria (ver Figura 8.2a). La media de $X(t)$ se encuentra utilizando la Ecuación (4.28):

$$m_X(t) = E[A\cos 2\pi t] = E[A]\cos 2\pi t.$$

Hay que tener en cuenta que la media varía con t . En particular, el proceso es siempre igual a cero para los valores de t cuando $\cos 2\pi t = 0$.

La autocorrelación es

$$R_X(t_1, t_2) = E[A\cos 2\pi t_1 A\cos 2\pi t_2] = E[A^2]\cos 2\pi t_1 \cos 2\pi t_2,$$

y la autocovarianza es entonces

$$\begin{aligned} C_X(t_1, t_2) &= R_X(t_1, t_2) - m_X(t_1)m_X(t_2) \\ &= \{E[A^2] - E[A]^2\}\cos 2\pi t_1 \cos 2\pi t_2 \\ &= \text{VAR}[A]\cos 2\pi t_1 \cos 2\pi t_2. \end{aligned}$$

Ejemplo 8.10 Sinusoides con fase aleatoria Sea $X(t) = \cos(\omega t + \Theta)$, donde Θ tiene distribución uniforme en el intervalo $(-\pi, \pi)$ (ver Figura 8.2b). La media de $X(t)$ se encuentra mediante la Ecuación (4.28):

$$m_X = E[\cos(\omega t + \Theta)] = \frac{1}{2\pi} \int_{-\pi}^{\pi} \cos(\omega t + \theta) d\theta = 0$$

La autocorrelación y autocovarianza son entonces

$$\begin{aligned} C_X(t_1, t_2) &= R_X(t_1, t_2) = E[\cos(\omega t_1 + \Theta)\cos(\omega t_2 + \Theta)] \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{2} \{ \cos[\omega(t_1 - t_2)] + \cos[\omega(t_1 + t_2) + 2\theta] \} d\theta \\ &= \frac{1}{2} \cos[\omega(t_1 - t_2)]. \end{aligned}$$

donde se ha utilizado la identidad $\cos(a)\cos(b) = 1/2\cos(a+b) + 1/2\cos(a-b)$. Tengamos en cuenta que $m_X(t)$ es una constante y que $C_X(t_1 - t_2)$ depende sólo de $|t_1 - t_2|$. Tengamos en cuenta también que las muestras en los instantes t_1 y t_2 no están correlacionadas si $\omega(t_1 - t_2) = k\pi$ donde k es cualquier número entero.

8.2.3 Procesos estocásticos múltiples

En la mayoría de las situaciones hay que lidiar con más de un proceso estocástico a la vez. Por ejemplo, podemos estar interesados en las temperaturas

en una ciudad A , $X(t)$, y una ciudad B , $Y(t)$. Otro ejemplo muy común consiste en un proceso estocástico $X(t)$ que es la “entrada” de un sistema y otro proceso estocástico $Y(t)$ que es la “salida” del sistema. Naturalmente, estamos interesados en la relación entre $X(t)$ e $Y(t)$. El comportamiento conjunto de dos o más procesos estocásticos se especifica mediante la colección de distribuciones conjuntas de todas las opciones posibles de las muestras en el tiempo de los procesos. Así, para dos procesos estocásticos continuos $X(t)$ e $Y(t)$ se deben especificar todas las posibles funciones de densidad conjunta de $X(t_1), \dots, X(t_k)$ e $Y(t'_1), \dots, Y(t'_j)$ para todos los posibles valores de k y j y todas las opciones de t_1, \dots, t_k y t'_1, \dots, t'_j . Por ejemplo, la pdf conjunta más simple sería la siguiente:

$$f_{X(t_1), Y(t_2)}(x, y) dx dy = P\{x < X(t_1) \leq x + dx, y < Y(t_2) \leq y + dy\}.$$

Nótese que los índices de tiempo de $X(t)$ e $Y(t)$ no tienen por qué ser los mismos. Por ejemplo, podemos estar interesados en la entrada en el tiempo t_1 y la salida en un tiempo posterior t_2 . Los procesos estocásticos $X(t)$ e $Y(t)$ se definen como procesos estocásticos independientes si las variables aleatorias vectoriales $\mathbf{X} = (X(t_1), \dots, X(t_k))$ e $\mathbf{Y} = (Y(t'_1), \dots, Y(t'_j))$ son independientes para todos los posibles valores de k y j , y todas las opciones de t_1, \dots, t_k y t'_1, \dots, t'_j :

$$F_{\mathbf{X}, \mathbf{Y}}(x_1, \dots, x_k, y_1, \dots, y_j) = F_{\mathbf{X}}(x_1, \dots, x_k) F_{\mathbf{Y}}(y_1, \dots, y_j).$$

La correlación cruzada $R_{X,Y}(t_1, t_2)$ de $X(t)$ y $Y(t)$ se define como

$$R_{X,Y}(t_1, t_2) = E[X(t_1)Y(t_2)].$$

Los procesos $X(t)$ e $Y(t)$ se definen procesos estocásticos ortogonales si

$$R_{X,Y}(t_1, t_2) = 0 \text{ para todo } t_1 \text{ y } t_2.$$

La covarianza cruzada $C_{X,Y}(t_1, t_2)$ de $X(t)$ e $Y(t)$ se define como

$$\begin{aligned} C_{X,Y}(t_1, t_2) &= E[\{X(t_1) - m_X(t_1)\}\{Y(t_2) - m_X(t_2)\}] \\ &= R_{X,Y}(t_1, t_2) - m_X(t_1)m_X(t_2). \end{aligned} \quad (8.5)$$

Los procesos $X(t)$ e $Y(t)$ se definen procesos estocásticos no correlacionados si

$$C_{X,Y}(t_1, t_2) = 0 \text{ para todo } t_1 \text{ y } t_2.$$

Ejemplo 8.11 Sean $X(t) = \cos(\omega t + \Theta)$ e $Y(t) = \sin(\omega t + \Theta)$, donde Θ es una variable aleatoria con distribución uniforme en $[-\pi, \pi]$. Encontrar la covarianza cruzada de $X(t)$ e $Y(t)$.

A partir del Ejemplo 8.10, sabemos que $X(t)$ e $Y(t)$ tienen media cero. De la

ecuación (8.5), la covarianza cruzada es entonces igual a la correlación cruzada:

$$\begin{aligned} C_{X,Y}(t_1, t_2) &= R_{X,Y}(t_1, t_2) = E[\cos(\omega t_1 + \Theta)\sin(\omega t_2 + \Theta)] \\ &= E\left[-\frac{1}{2}\sin(\omega(t_1 - t_2)) + \frac{1}{2}\sin(\omega(t_1 + t_2) + 2\Theta)\right] \\ &= -\frac{1}{2}\sin(\omega(t_1 - t_2)), \end{aligned}$$

ya que $E[\sin(\omega(t_1 + t_2) + 2\Theta)] = 0$. Se ha utilizado la identidad $\cos(a)\cos(b) = 1/2\sin(a+b) - 1/2\sin(a-b)$. $X(t)$ e $Y(t)$ son procesos estocásticos correlacionados, porque la covarianza cruzada no es igual a cero para todas las posibles muestras de tiempo. Nótese que, sin embargo, $X(t_1)$ e $Y(t_2)$ son variables aleatorias no correlacionadas para t_1 y t_2 tales que $\omega(t_1 - t_2) = k\pi$, donde k es un número entero arbitrario.

Ejemplo 8.12 Señal más ruido

Supongamos que el proceso $Y(t)$ consiste en una señal $X(t)$ deseada, más un ruido $N(t)$:

$$Y(t) = X(t) + N(t).$$

Encontrar la correlación cruzada entre la señal observada y la señal deseada suponiendo que $X(t)$ y $N(t)$ son procesos estocásticos independientes.

De la ecuación (8.14), tenemos que

$$\begin{aligned} R_{XY}(t_1, t_2) &= E[X(t_1)Y(t_2)] \\ &= E[X(t_1)\{X(t_2) + N(t_2)\}] \\ &= R_X(t_1, t_2) + E[X(t_1)]E[N(t_2)] \\ &= R_X(t_1, t_2) + m_X(t_1)m_N(t_2), \end{aligned}$$

donde la tercera igualdad es debida al hecho de que $X(t)$ y $N(t)$ son independientes.

8.3 Procesos estocásticos en tiempo discreto: proceso de suma, proceso de conteo binomial y paseo aleatorio

En esta sección presentamos varios importantes procesos estocásticos en tiempo discreto. Comenzamos con la clase más simple de procesos estocásticos - las secuencias independientes e idénticamente distribuidas- y luego consideramos el proceso de suma que resulta de la adición de una secuencia iid.

Se demuestra que el proceso de suma satisface la propiedad de incrementos independientes así como la propiedad de Markov. Estas dos propiedades facilitan

enormemente el cálculo de las probabilidades conjuntas. También presentamos el proceso de conteo binomial y el paseo aleatorio como casos especiales de los procesos de suma.

8.3.1 Procesos estocásticos iid

Sea X_n es un proceso estocástico de tiempo discreto que consta de una serie de variables aleatorias independientes, idénticamente distribuidas (iid) con función de distribución común $F_X(x)$, con media m , y varianza σ^2 . La secuencia X_n se llama proceso estocástico iid. La cdf conjunta para cualquier conjunto de instantes de tiempo n_1, \dots, n_k está dada por

$$\begin{aligned} F_{X_1, \dots, X_k}(x_1, x_2, \dots, x_k) &= P[X_1 \leq x_1, X_2 \leq x_2, \dots, X_k \leq x_k] \\ &= F_X(x_1)F_X(x_2) \dots F_X(x_k), \end{aligned} \quad (8.6)$$

donde, por simplicidad, X_k denota X_{n_k} . La ecuación (8.6) implica que, si X_n toma valores discretos, la pmf conjunta queda factorizada en el producto de las pmf individuales, y si X_n toma valores continuos, la pdf conjunta queda factorizada en el producto de las pdf individuales.

La media de un proceso iid se obtiene de la ecuación (8.1):

$$m_X(n) = E[X_N] = m \text{ para todo } n.$$

Así, la media es constante.

La función de autocovarianza se obtiene de la ecuación (8.3) de la siguiente manera. Si $n_1 \neq n_2$, entonces

$$\begin{aligned} C_X(n_1, n_2) &= E[(X_{n_1} - m)(X_{n_2} - m)] \\ &= E[(X_{n_1} - m)]E[(X_{n_2} - m)] = 0, \end{aligned}$$

dado que X_{n_1} y X_{n_2} son variables aleatorias independientes. Si $n_1 = n_2 = n$, entonces

$$C_X(n_1, n_2) = E[(X_n - m)^2] = \sigma^2.$$

Podemos expresar la autocovarianza del proceso iid de forma compacta de la siguiente manera:

$$C_X(n_1, n_2) = \sigma^2 \delta_{n_1 n_2}.$$

donde $\delta_{n_1 n_2} = 1$ si $n_1 = n_2$, y 0 en caso contrario. Por lo tanto la función de autocovarianza es cero siempre excepto cuando $n_1 = n_2$. La función de autocorrelación del proceso iid se puede calcular a partir de la ecuación (8.4):

$$R_X(n_1, n_2) = C_X(n_1, n_2) + m^2.$$

Ejemplo 8.13 Proceso estocástico de Bernoulli

Sea I_n una sucesión de variables aleatorias independientes de Bernoulli. I_n es

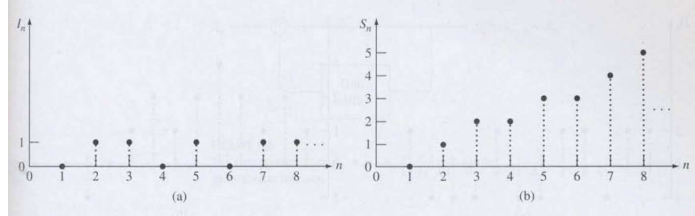


Figura 8.4 (a) Realización de un proceso de Bernoulli. $I_n = 1$ indica que una bombilla se estropea y se sustituye en el día n . (b) Realización de un proceso binomial. S_n denota de número de bombillas que se han estropeado hasta el día n .

entonces un proceso estocástico iid que toma valores del conjunto $\{0, 1\}$. Una realización de este proceso se muestra en la Figura 8.4a. Por ejemplo, I_n podría ser una función indicatriz para el evento “una bombilla se estropea y se sustituye en el día n ”. Como I_n es una variable aleatoria de Bernoulli, tiene media y varianza

$$m_I = p, \quad \text{VAR}[I_n] = p(1 - p).$$

La independencia de las I_n facilita el cálculo de las probabilidades. Por ejemplo, la probabilidad de que los primeros cuatro bits en la secuencia sean 1001 es

$$\begin{aligned} P[I_1 = 1, I_2 = 0, I_3 = 0, I_4 = 1] \\ &= P[I_1 = 1]P[I_2 = 0]P[I_3 = 0]P[I_4 = 1] \\ &= p^2(1 - p)^2. \end{aligned}$$

Del mismo modo, la probabilidad de que el segundo bit sea 0 y el séptimo sea un 1

$$P[I_2 = 0, I_7 = 1] = P[I_2 = 0]P[I_7 = 1] = p(1 - p).$$

Ejemplo 8.14 Escalón aleatorio

Supongamos que una serie de pulsos $+1$ o -1 activan un contador arriba-abajo y que la entrada del contador es $D_n = 2I_n - 1$, donde I_n es el proceso estocástico de Bernoulli. Entonces será

$$D_n = \begin{cases} +1 & \text{si } I_n = 1 \\ -1 & \text{si } I_n = 0. \end{cases}$$

Por ejemplo, D_n podría representar el cambio en la posición de una partícula que se mueve a lo largo de una línea recta efectuando saltos de ± 1 por cada unidad de tiempo. Una realización de la D_n se muestra en la Figura 8.5(a). La

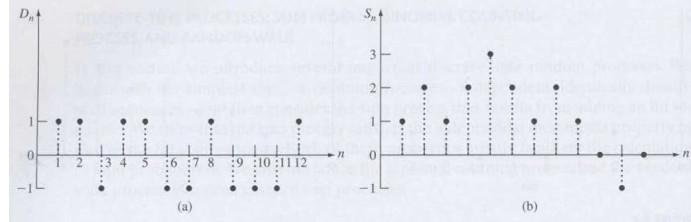


Figura 8.5 (a) Realización de un proceso escalón aleatorio. (b) Realización de un proceso paseo aleatorio. S_n denota la posición de una partícula en el tiempo n .

media de la D_n es

$$m_D(n) = E[D_n] = E[2I_n - 1] = 2E[I_n] - 1 = 2p - 1.$$

La varianza de D_n se encuentra mediante las ecuaciones (4.36) y (4.37):

$$\text{VAR}[D_n] = \text{VAR}[2I_n - 1] = 2^2 \text{VAR}[I_n] = 4p(1 - p).$$

Las probabilidades de sucesos relacionados con D_n , se calculan como en el Ejemplo 8.13.

8.3.2 Incrementos independientes y propiedad de Markov de procesos estocásticos

Antes de proceder a la construcción de procesos estocásticos a partir de procesos iid, se presentan dos propiedades muy útiles de los procesos estocásticos. Sea $X(t)$ un proceso estocástico y consideremos dos instantes de tiempo, $t_1 < t_2$.

El incremento del proceso estocástico en el intervalo $t_1 < t \leq t_2$ se define como $X(t_2) - X(t_1)$. Se dice que un proceso estocástico $X(t)$ tiene incrementos independientes si los incrementos en intervalos disjuntos son variables aleatorias independientes, es decir, para cualquier k y cualquier elección de los instantes de muestreo $t_1 < t_2 < \dots < t_k$ los incrementos asociados

$$X(t_2) - X(t_1), X(t_3) - X(t_2), \dots, X(t_k) - X(t_{k-1}),$$

son variables aleatorias independientes. En la subsección siguiente, se muestra que la pdf (pmf) conjunta de $X(t_1), X(t_2), \dots, X(t_k)$ está dada por el producto de la pdf (pmf) de $X(t_1)$ y las pdfs (pmfs) marginales de los incrementos individuales.

Otra característica útil de los procesos estocásticos que nos permiten obtener fácilmente la probabilidades conjuntas es la propiedad de Markov. Se dice que un proceso estocástico $X(t)$ es de Markov si el futuro del proceso dado el presente es independiente del pasado, es decir, para cualquier k y cualquier elección de los instantes de muestreo $t_1 < t_2 < \dots < t_k$ y para cualquier x_1, x_2, \dots, x_k

$$f_{X(t_k)}(x_k | X(t_{k-1}) = x_{k-1}, \dots, X(t_1) = x_1) = f_{X(t_k)}(x_k | X(t_{k-1}) = x_{k-1})$$

si $X(t)$ toma valores continuos, y

$$P[X(t_k) = x_k | X(t_{k-1}) = x_{k-1}, \dots, X(t_1) = x_1] = P[X(t_k) = x_k | X(t_{k-1}) = x_{k-1}]$$

si $X(t)$ toma valores discretos.

Las expresiones en el lado derecho de las dos anteriores ecuaciones se llaman pdf de transición y pmf de transición, respectivamente. En las siguientes secciones nos encontramos con varios procesos que satisfacen la propiedad de Markov. El Capítulo ?? se dedica por completo a los procesos estocásticos que satisfacen esta propiedad. Es fácil demostrar que un proceso estocástico que tiene incrementos independientes es también un proceso de Markov. Lo contrario no es verdad, es decir, la propiedad de Markov no implica incrementos independientes.

8.3.3 Procesos de suma: los procesos conteo binomial y paseo aleatorio

Muchos procesos estocásticos interesantes se obtienen como la suma de una secuencia de variables aleatorias iid, X_1, X_2, \dots :

$$\begin{aligned} S_n &= X_1 + X_2 + \dots + X_n \\ &= S_{n-1} + X_n, \quad n = 1, 2, \dots \end{aligned}$$

donde $S_0 = 0$. El proceso S_n se llama proceso de suma.

La pdf o pmf de S_n se encuentra usando el método de la convolución o el método de la ecuación característica presentados en la Sección 7.1. Nótese que S_n depende del “pasado”, S_1, \dots, S_{n-1} , sólo a través de S_{n-1} , es decir, S_n es independiente del pasado cuando S_{n-1} es conocida.

Esto se puede ver claramente en la Figura 8.6, que muestra un procedimiento recursivo para el cálculo de S_n en términos de S_{n-1} y del incremento X_n . Así S_n es un proceso de Markov.

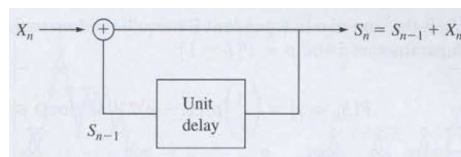


Figura 8.6 El proceso de suma $S_n = X_1 + \dots + X_n$, $S_0 = 0$ puede ser generado de esta forma.

Ejemplo 8.15 Proceso de conteo binomial

Sea I_i la secuencia de variables aleatorias independientes de Bernoulli en el Ejemplo 8.13, y sea S_n el correspondiente proceso de suma. S_n es entonces el proceso de conteo que da el número de éxitos en los primeros n ensayos de Bernoulli. La función muestral de S_n que corresponde a una determinada secuencia de I_i s

se muestra en la Figura 8.4(b). Nótese que el proceso de conteo sólo puede aumentar con el tiempo. Nótese también que el proceso binomial puede aumentar como mucho de una unidad a la vez. Si I_n indica que una bombilla se estropea y se sustituye el día n , entonces S_n denota el número de bombillas que han fallado hasta el día n .

Puesto que S_n es la suma de n variables aleatorias independientes de Bernoulli, S_n es una variable aleatoria binomial con parámetros n y $p = P[I = 1]$:

$$P[S_n = j] = \binom{n}{j} p^j (1-p)^{n-j} \quad \text{para } 0 \leq j \leq n,$$

y cero en caso contrario. Así pues, S_n tiene media np y varianza $np(1-p)$. Observemos que la media y la varianza de este proceso crecen linealmente con el tiempo. Esto refleja el hecho de que a medida que avanza el tiempo, es decir, a medida que n crece, el rango de valores que pueden ser asumidos por el proceso aumenta. Si $p > 0$ entonces también sabemos que S_n tiene tendencia a crecer sin límite en el tiempo. La propiedad de Markov del proceso de conteo binomial es fácil de deducir. Dado que el valor actual del proceso en el tiempo $n-1$ es $S_{n-1} = k$, el proceso en el instante siguiente será k con probabilidad $1-p$ o $k+1$ con probabilidad p . Una vez que conocemos el valor del proceso en el tiempo $n-1$, los valores del proceso estocástico antes del tiempo $n-1$ son irrelevantes.

Ejemplo 8.16 Paseo aleatorio unidimensional

Sea D_n el proceso iid de variables aleatorias ± 1 en el Ejemplo 8.14, y sea S_n el correspondiente proceso de suma. S_n puede representar la posición de una partícula en el instante n . El proceso estocástico S_n es un ejemplo de un paseo aleatorio unidimensional. Una función muestral de S_n se muestra en la Figura 8.5(b). A diferencia de el proceso binomial, el paseo aleatorio puede aumentar o disminuir con el tiempo. El proceso paseo aleatorio cambia en una unidad a la vez. La pmf de S_n se encuentra de la siguiente manera. Si hay k “+1”s en los primeros n ensayos, entonces hay $n-k$ “-1”s, y $S_n = k - (n-k) = 2k - n$. Por el contrario, $S_n = j$ si el número de +1s es $k = (j+n)/2$. Si $(j+n)/2$ no es un número entero, S_n no puede ser igual a j . Así

$$P[S_n = 2k - n] = \binom{n}{k} p^k (1-p)^{n-k} \quad \text{para } k \in \{0, 1, \dots, n\}.$$

Como k es el número de éxitos n ensayos de Bernoulli, la media del paseo aleatorio es:

$$E[S_n] = 2np - n = n(2p - 1),$$

y la varianza es:

$$\text{VAR}[S_n] = 4np(1-p).$$

Con el tiempo, el paseo aleatorio puede fluctuar en un rango cada vez más amplio de valores positivos y negativos. S_n tiene una tendencia a crecer si $p > 1/2$, o a disminuir si $p < 1/2$. El caso $p = 1/2$ proporciona un equilibrio precario, y como se verá más adelante, en el capítulo 12, una dinámica muy interesante. La Figura 8.7(a) muestra los primeros 100 pasos de una función muestral del paseo aleatorio con $p = 1/2$. La Figura 8.7(b) muestra cuatro funciones muestrales del proceso paseo aleatorio con $p = 1/2$ y 1000 pasos. La Figura 8.7(c) muestra cuatro funciones muestrales en el caso asimétrico, donde $p = 3/4$. Nótese la fuerte tendencia de crecimiento lineal en el proceso.

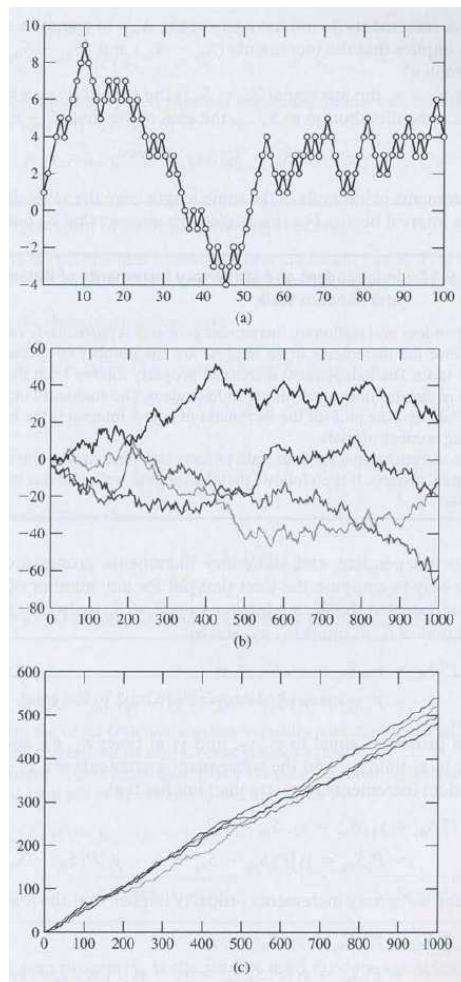


Figura 8.7 (a) Proceso paseo aleatorio con $p = 1/2$. (b) Cuatro funciones muestrales de un proceso paseo aleatorio simétrico con $p = 1/2$. (c) Cuatro funciones muestrales de un proceso paseo aleatorio asimétrico con $p = 3/4$.

El proceso de suma S , tiene incrementos independientes en intervalos de tiempo que no se superponen. Para ver esto consideremos dos intervalos de tiempo: $n_0 < n \leq n_1$ y $n_2 < n \leq n_3$, donde $n_1 \leq n_2$. Los incrementos de S_n en estos intervalos de tiempo disjuntos son dados por

$$\begin{aligned} S_{n_1} - S_{n_0} &= X_{n_0+1} + \cdots + X_{n_1} \\ S_{n_3} - S_{n_2} &= X_{n_2+1} + \cdots + X_{n_3}. \end{aligned}$$

Los incrementos anteriores no tienen ninguno de los de X_n en común, por lo que la independencia de las X_n implica que los incrementos $(S_{n_1} - S_{n_0})$ y $(S_{n_3} - S_{n_2})$ son variables aleatorias independientes.

Para $n' > n$, el incremento $S_{n'} - S_n$ es la suma de $n' - n$ variables aleatorias iid, por lo que tiene la misma distribución que $S_{n'-n}$, la suma de los primeros $n' - n$ X s, es decir,

$$P[S_{n'} - S_n = y] = P[S_{n'-n} = y].$$

Así, los incrementos en los intervalos de la misma longitud tienen la misma distribución, independientemente de cuando comienza el intervalo. Por esta razón, también se dice que S_n tiene incrementos estacionarios.

Ejemplo 8.17 Incrementos independientes y estacionarios del proceso binomial y del proceso paseo aleatorio

La propiedad de incrementos independientes y estacionarios es particularmente fácil de ver en el proceso binomial, ya que los incrementos en un intervalo son el número de éxitos en los correspondientes experimentos de Bernoulli. La propiedad de incrementos independientes se deduce del hecho de que el número de éxitos en intervalos de tiempo disjuntos son independientes. La propiedad de incrementos estacionarios se debe al hecho de que la pmf para el incremento en un intervalo de tiempo es la pmf binomial con el correspondiente número de experimentos.

El incremento en un proceso paseo aleatorio es determinado por el mismo número de éxitos de un proceso binomial. De esto se deduce que el paseo aleatorio tiene también incrementos independientes y estacionarios.

La propiedad de incrementos independientes y estacionarios del proceso de suma S_n hace que sea fácil de calcular la pmf/pdf conjunta para cualquier número de instantes de tiempo. Para simplificar, supongamos que la X_n tome valores enteros, por lo que S_n también toma valores enteros. Calculamos la pmf conjunta de S_n en los instantes n_1, n_2 y n_3 .

$$\begin{aligned} P[S_{n_1} = y_1, S_{n_2} = y_2, S_{n_3} = y_3] &= \\ P[S_{n_1} = y_1, S_{n_2} - S_{n_1} = y_2 - y_1, S_{n_3} - S_{n_2} = y_3 - y_2], & \end{aligned}$$

ya que el proceso es igual a y_1, y_2, y_3 , y en los instantes n_1, n_2 y n_3 , si y sólo si es igual a y_1 en el instante n_1 , y los incrementos posteriores son $y_2 - y_1$ e $y_3 - y_2$. La propiedad de incrementos independientes implica entonces que

$$\begin{aligned} P[S_{n_1} = y_1, S_{n_2} = y_2, S_{n_3} = y_3] &= \\ P[S_{n_1} = y_1]P[S_{n_2} - S_{n_1} = y_2 - y_1]P[S_{n_3} - S_{n_2} = y_3 - y_2], \end{aligned}$$

Finalmente, la propiedad incrementos estacionarios implica que la pmf conjunta de S_n , viene dada por

$$\begin{aligned} P[S_{n_1} = y_1, S_{n_2} = y_2, S_{n_3} = y_3] &= \\ P[S_{n_1} = y_1]P[S_{n_2 - n_1} = y_2 - y_1]P[S_{n_3 - n_2} = y_3 - y_2], \end{aligned}$$

Claramente, podemos usar este procedimiento para expresar la pmf conjunta de S_n en cualquier instante de tiempo $n_1 < n_2 < \dots < n_k$ en términos de la pmf en el instante inicial y las pmf de los posteriores incrementos:

$$\begin{aligned} P[S_{n_1} = y_1, S_{n_2} = y_2, \dots, S_{n_k} = y_k] &= \\ P[S_{n_1} = y_1]P[S_{n_2 - n_1} = y_2 - y_1] \cdots P[S_{n_k - n_{k-1}} = y_k - y_{k-1}], \end{aligned}$$

Si las X_n son variables aleatorias con valores continuos, entonces se puede demostrar que la densidad de probabilidad conjunta de S_n en los instantes n_1, n_2, \dots, n_k es:

$$\begin{aligned} f_{S_{n_1}, S_{n_2}, \dots, S_{n_k}}(y_1, y_2, \dots, y_k) &= \\ f_{S_{n_1}}(y_1)f_{S_{n_2 - n_1}}(y_2 - y_1) \cdots f_{S_{n_k - n_{k-1}}}(y_k - y_{k-1}). \end{aligned}$$

Ejemplo 8.18 pmf conjunta del proceso de conteo binomial

Calcular la pmf conjunta para el proceso de conteo binomial en los instantes n_1 y n_2 . Calcular la probabilidad de que $P[S_{n_1} = 0, S_{n_2} = n_2 - n_1]$, es decir, los primeros n_1 experimentos son fracasos y los restantes experimentos son éxitos.

Siguiendo el enfoque de arriba se obtiene

$$\begin{aligned} P[S_{n_1} = y_1, S_{n_2} = y_2] &= P[S_{n_1} = y_1]P[S_{n_2} - S_{n_1} = y_2 - y_1] \\ &= \binom{n_2 - n_1}{y_2 - y_1} p^{y_2 - y_1} (1 - p)^{n_2 - n_1 - y_2 + y_1} \binom{n_1}{y_1} p^{y_1} (1 - p)^{n_1 - y_1} \\ &= \binom{n_2 - n_1}{y_2 - y_1} \binom{n_1}{y_1} p^{y_2} (1 - p)^{n_2 - y_2} \end{aligned}$$

La probabilidad requerida es entonces:

$$P[S_{n_1} = 0, S_{n_2} = n_2 - n_1] = \binom{n_2 - n_1}{n_2 - n_1} \binom{n_1}{0} p^{n_2 - n_1} (1 - p)^{n_1} = p^{n_2 - n_1} (1 - p)^{n_1}$$

que es lo que se obtendría de un cálculo directo para experimentos de Bernoulli.

Ejemplo 8.19 pdf conjunta de la suma de la secuencia de gaussianas iid

Sea X_n una secuencia de las variables aleatorias iid gaussianas con media cero y varianza σ^2 . Encontrar la pdf conjunta del correspondiente proceso de suma en los instantes n_1 y n_2 .

A partir del Ejemplo 7.3, se sabe que S_n es una variable aleatoria gaussiana de media cero y varianza $n\sigma^2$. La pdf conjunta de S_n en los instantes n_1 y n_2 está dada por

$$\begin{aligned} f_{S_{n_1}, S_{n_2}}(y_1, y_2) &= f_{S_{n_2-n_1}}(y_2 - y_1) f_{S_{n_1}}(y_1) \\ &= \frac{1}{\sqrt{2\pi(n_2 - n_1)\sigma^2}} e^{-(y_2 - y_1)^2 / [2(n_2 - n_1)\sigma^2]} \frac{1}{\sqrt{2\pi n_1 \sigma^2}} e^{-y_1^2 / 2n_1 \sigma^2}. \end{aligned}$$

Dado que el proceso suma S_n es la suma de n variables aleatorias iid, tiene media y varianza:

$$\begin{aligned} m_S(n) &= E[S_n] = nE[X] = nm \\ \text{VAR}[S_n] &= n\text{VAR}[X] = n\sigma^2. \end{aligned} \quad (8.7)$$

La propiedad de incrementos independientes nos permite calcular la autocovarianza en una manera interesante. Supongamos que $n \leq k$ de manera que $n = \min(n, k)$, entonces

$$\begin{aligned} C_S(n, k) &= E[(S_n - nm)(S_k - km)] \\ &= E[(S_n - nm)\{(S_n - nm) + (S_k - km) - (S_n - nm)\}] \\ &= E[(S_n - nm)^2] + E[(S_n - nm)(S_k - S_n - (k - n)m)]. \end{aligned}$$

Dado que S_n y el incremento $S_k - S_n$ son independientes,

$$\begin{aligned} C_S(n, k) &= E[(S_n - nm)^2] + E[(S_n - nm)]E[(S_k - S_n - (k - n)m)] \\ &= E[(S_n - nm)^2] \\ &= \text{VAR}[S_n] = n\sigma^2, \end{aligned}$$

ya que $E[S_n - nm] = 0$. Del mismo modo, si $k = \min(n, k)$, se obtendría $k\sigma^2$. Por lo tanto, la autocovarianza del proceso de suma es

$$C_S(n, k) = \min(n, k)\sigma^2. \quad (8.8)$$

Ejemplo 8.20 Encontrar la autocovarianza del paseo aleatorio unidimensional.

A partir del Ejemplo 8.14 y de las ecuaciones (8.7) y (8.8), S_n tiene media $n(2p - 1)$ y varianza $4np(1 - p)$. Así pues, su autocovarianza está dada por

$$C_S(n, k) = \min(n, k)4p(1 - p).$$

El proceso de suma se puede generalizar en varios modos. Por ejemplo, la estructura recursiva de la Figura 8.6 se puede modificar como se muestra en la Figura 8.8(a). Se obtienen entonces procesos estocásticos de primer orden autorregresivos, que son de interés en el análisis de series temporales y en el procesamiento digital de la señal. Si por el contrario utilizamos la estructura que se muestra en la Figura 8.8(b), se obtiene un ejemplo de proceso de media móvil.

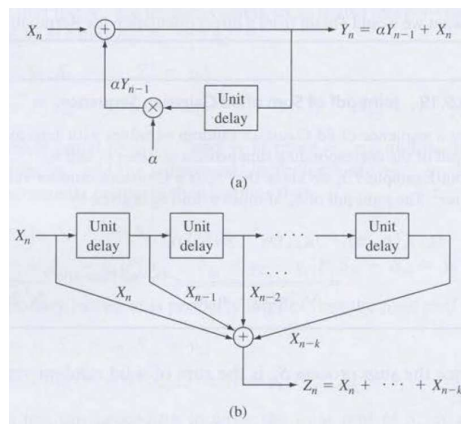


Figura 8.8 (a) Proceso autorregresivo de primer orden. (b) Proceso de media móvil.

8.4 Procesos estocásticos estacionarios

Muchos procesos estocásticos tienen la propiedad de que la naturaleza de la aleatoriedad en el proceso no cambia con el tiempo. Una observación del proceso en el intervalo de tiempo (t_0, t_1) presenta el mismo tipo de comportamiento aleatorio de una observación en algún otro intervalo de tiempo $(t_0 + \tau, t_1 + \tau)$. Esto nos lleva a postular que las probabilidades de las muestras del proceso no dependen del instante en que comienzan a tomarse las observaciones, es decir, las probabilidades que involucran muestras tomadas en los instantes t_1, \dots, t_k no difieren de las probabilidades de las muestras tomadas en los instantes $t_1 + \tau, \dots, t_k + \tau$.

Ejemplo 8.21 Estacionariedad y transitoriedad

Una urna tiene 6 bolas blancas, cada una con la etiqueta “0” y 5 bolas negras con la etiqueta “1”. Se lleva a cabo la siguiente secuencia de experimentos: Se selecciona una bola y se anota el número; la primera vez que se selecciona una

bola blanca no se vuelve a poner en la urna, pero en los otros casos siempre se ponen de nuevo la bolas en la urna. El proceso estocástico que resulta de esta secuencia de experimentos claramente tiene una fase transitoria y una fase estacionaria. La fase transitoria consiste en una cadena de n “1” consecutivos y termina con la primera aparición de un “0”. Durante la fase transitoria $P[I_n = 0] = 6/11$ y la duración media de la fase transitoria tiene distribución geométrica con media $11/6$. Después de la primera ocurrencia de un “0”, el proceso entra en la fase estacionaria en que el proceso es una secuencia binaria equiprobable iid. El comportamiento estadístico del proceso no cambia una vez que se alcanza la fase estacionaria.

Si consideramos procesos estocásticos que comienzan en el instante $t = -\infty$, entonces la condición anterior se puede formular con precisión de la siguiente manera. Un proceso estocástico en tiempo discreto o en tiempo continuo $X(t)$ es estacionario si la distribución conjunta de cualquier conjunto de muestras no depende de la posición del origen del tiempo. Esto significa que la función de distribución conjunta de $X(t_1), X(t_2), \dots, X(t_k)$ es la misma de la de $X(t_1 + \tau), X(t_2 + \tau), \dots, X(t_k + \tau)$:

$$F_{X(t_1), \dots, X(t_k)}(x_1, \dots, x_k) = F_{X(t_1 + \tau), \dots, X(t_k + \tau)}(x_1, \dots, x_k), \quad (8.9)$$

para todo desplazamiento en tiempo τ , todo k , y todas las posibles opciones de instantes de muestreo t_1, \dots, t_k . Si un proceso comienza en algún instante determinado (es decir, $n = 0$ o $t = 0$), entonces se dice que es estacionario si su distribución conjunta no cambia en presencia de desplazamientos a la derecha del tiempo. Se dice que dos procesos $X(t)$ e $Y(t)$ son conjuntamente estacionarios si las cdf conjuntas de $X(t_1), \dots, X(t_k)$ e $Y(t'_1), \dots, Y(t'_j)$ no dependen de la ubicación del origen de tiempo para todo k y j y todas las opciones de instantes de muestreo t_1, \dots, t_k y t'_1, \dots, t'_j . La cdf de primer orden de un proceso estocástico estacionario debe ser independiente del tiempo, ya que por la Ecuación 8.9,

$$F_{X(t)}(x) = F_{X(t + \tau)}(x) = F_X(x) \quad \text{para todo } t, \tau.$$

Esto implica que la media y la varianza de $X(t)$ son constantes e independientes del tiempo:

$$\begin{aligned} m_X(t) &= E[X(t)] = m \quad \text{para todo } t \\ \text{VAR}[X(t)] &= E[(X(t) - m)^2] = \sigma^2 \quad \text{para todo } t. \end{aligned}$$

La cdf de segundo orden de un proceso estocástico estacionario sólo puede depender de la diferencia temporal entre las muestras y no del instante concreto de las muestras, ya que por la Ecuación 8.9,

$$F_{X(t_1), X(t_2)}(x_1, x_2) = F_{X(0), X(t_2 - t_1)}(x_1, x_2) \quad \text{para todo } t_1, t_2.$$

Esto implica que la autocorrelación y la autocovarianza de $X(t)$ pueden depender sólo de $t_2 - t_1$:

$$\begin{aligned} R_X(t_1, t_2) &= R_X(t_2 - t_1) && \text{para todo } t_1, t_2 \\ C_X(t_1, t_2) &= C_X(t_2 - t_1) && \text{para todo } t_1, t_2. \end{aligned}$$

Ejemplo 8.22 Proceso estocástico iid

Demostrar que el proceso estocástico iid es estacionario.

La cdf conjunta de las muestras en cualquier conjunto de k instantes de tiempo t_1, \dots, t_k es

$$\begin{aligned} F_{X(t_1), \dots, X(t_k)}(x_1, x_2, \dots, x_k) &= F_X(x_1)F_X(x_2) \cdots F_X(x_k) \\ &= F_{X(t_1+\tau), \dots, X(t_k+\tau)}(x_1, \dots, x_k) \end{aligned}$$

para todo k, t_1, \dots, t_k . Así la Ecuación 8.9 se satisface y el proceso estocástico iid es estacionario.

Ejemplo 8.23 ¿Es el proceso de suma un proceso en tiempo discreto estacionario?

El proceso de suma se define como $S_n = X_1 + X_2 + \cdots + X_n$, donde las X_i representan una secuencia iid. El proceso tiene media y varianza

$$m_S(n) = nm \quad \text{VAR}[S_n] = n\sigma^2,$$

donde m y σ^2 son la media y la varianza de las X_n . Se puede observar que la media y la varianza no son constantes sino que crecen linealmente con el índice temporal n . Por lo tanto el proceso de suma no puede ser un proceso estacionario.

8.4.1 Procesos estocásticos estacionarios en sentido amplio

En muchas situaciones no podemos determinar si un proceso estocástico es estacionario, pero podemos determinar si la media es una constante:

$$m_X(t) = m \text{ para todo } t, \tag{8.10}$$

y si la autocovarianza (o, equivalentemente, la autocorrelación) es una función sólo de $t_1 - t_2$:

$$C_X(t_1, t_2) = C_X(t_1 - t_2) \text{ para todo } t_1, t_2, \tag{8.11}$$

Un proceso estocástico $X(t)$ de tiempo discreto o de tiempo continuo es estacionario en sentido amplio (WSS) si se cumplen las ecuaciones (8.10) y (8.11). Del

mismo modo, podemos decir que los procesos $X(t)$ e $Y(t)$ son conjuntamente estacionarios en sentido amplio cuando ambos son estacionarios en sentido amplio y su covarianza cruzada sólo depende de $t_1 - t_2$. Cuando $X(t)$ es estacionario en sentido amplio, podemos escribir

$$C_X(t_1, t_2) = C_X(\tau) \text{ y } R_X(t_1, t_2) = R_X(\tau)$$

donde $\tau = t_1 - t_2$. Todos los procesos estocásticos estacionarios son estacionarios en sentido amplio ya que cumplen las ecuaciones (8.10) y (8.11). El siguiente ejemplo muestra que un algunos procesos estacionarios en sentido amplio no son estacionarios.

Ejemplo 8.24 Supongamos que X_n consta de dos secuencias intercaladas de variables aleatorias independientes. En particular, supongamos que para n par, X_n asume los valores ± 1 con probabilidad $1/2$ y que para n impar, X_n asume los valores de $1/3$ y -3 con probabilidades de $9/10$ y $1/10$, respectivamente.

X_n no es estacionario porque su pmf varía con n . Es fácil demostrar que X_n tiene una media de

$$m_X(n) = 0 \text{ para todo } n$$

y función de covarianza

$$C_X(i, j) = \begin{cases} E[X_i]E[X_j] = 0 & \text{para } i \neq j \\ E[X_i^2] = 1 & \text{para } i = j. \end{cases}$$

X_n es, por lo tanto, estacionario en sentido amplio.

Veremos que la función de autocorrelación de procesos estacionarios en sentido amplios juega un papel crucial en el diseño de algoritmos lineales de procesamiento de señales. Ahora desarrollamos varios resultados que nos permiten deducir las propiedades de un proceso WSS a partir de las propiedades de su función de autocorrelación.

En primer lugar, la función de autocorrelación en $\tau = 0$ da la potencia media (segundo momento) del proceso:

$$R_X(0) = E[X(t)^2] \text{ para todo } t.$$

En segundo lugar, la función de autocorrelación es una función par de τ ya que

$$R_X(\tau) = E[X(t+T)X(t)] = E[X(t)X(t+T)] = R_X(-\tau).$$

En tercer lugar, la función de autocorrelación es una medida de la tasa de variación del proceso estocástico en el siguiente sentido. Consideremos el cambio

en el proceso del instante de tiempo t al instante $t + \tau$:

$$\begin{aligned} P[|X(t + \tau) - X(t)| > \varepsilon] &= P[(X(t + \tau) - X(t))^2 > \varepsilon^2] \\ &\leq \frac{E[(X(t + \tau) - X(t))^2]}{\varepsilon^2} = \frac{2\{R_X(0) - R_X(\tau)\}}{\varepsilon^2} \end{aligned} \quad (8.12)$$

donde hemos utilizado la desigualdad de Markov, ecuación (4.75), para obtener el límite superior. La ecuación (8.12) establece que si $R_X(0) - R_X(\tau)$ es pequeño, es decir, $R_X(\tau)$ disminuye lentamente, la probabilidad de que un gran cambio en $X(t)$ en τ segundos es pequeña.

En cuarto lugar, la función de autocorrelación es máxima en $\tau = 0$. Usamos la desigualdad de Cauchy-Schwarz:

$$E[XY]^2 \leq E[X^2]E[Y^2] \quad (8.13)$$

para cualquier par de variables aleatorias X e Y . Si aplicamos esta ecuación a $X(t + \tau)$ y $X(t)$, obtenemos

$$R_X(\tau)^2 = E[X(t + \tau)X(t)]^2 \leq E[X^2(t + \tau)]E[X^2(t)] = R_X(0)^2$$

Así

$$|R_X(\tau)| \leq R_X(0).$$

En quinto lugar, si $R_X(0) = R_X(d)$, entonces $R_X(\tau)$ es periódica con periodo d y $X(t)$ es decir es periódica en valor cuadrático medio, es decir, $E[(X(t + d) - X(t))^2] = 0$. Si aplicamos la ecuación (8.13) a $X(t + \tau + d) - X(t + \tau)$ y $X(t)$, obtenemos

$$\begin{aligned} E[(X(t + \tau + d) - X(t + \tau))X(t)]^2 \\ \leq E[(X(t + \tau + d) - X(t + \tau))^2]E[X^2(t)], \end{aligned}$$

lo que implica que

$$\{R_X(\tau + d) - R_X(\tau)\}^2 \leq 2\{R_X(0) - R_X(d)\}R_X(0).$$

Así $R_X(d) = R_X(0)$ implica que el lado derecho de la ecuación es igual a cero, y por lo tanto que $R_X(\tau + d) = R_X(\tau)$ para todo τ . Aplicando iterativamente este resultado se puede concluir que $R_X(\tau)$ es periódica con periodo d . El hecho de que $X(t)$ es periódica en valor cuadrático medio se deduce de

$$E[(X(t + d) - X(t))^2] = 2\{R_X(0) - R_X(d)\} = 0.$$

En sexto lugar, sea $X(t) = m + N(t)$, donde $N(t)$ es un proceso de media cero para el que $R_N(\tau) \rightarrow 0$ por $\tau \rightarrow \infty$, entonces

$$\begin{aligned} R_X(\tau) &= E[(m + N(t + \tau))(m + N(t))] = m^2 + 2mE[N(t)] + R_N(\tau) \\ &= m^2 + R_N(\tau) \rightarrow m^2 \text{ por } \tau \rightarrow \infty. \end{aligned}$$

En otras palabras, $R_X(\tau)$ se acerca al valor cuadrático medio de $X(t)$ por $\tau \rightarrow \infty$. En resumen, la función de autocorrelación puede tener tres tipos de componentes:

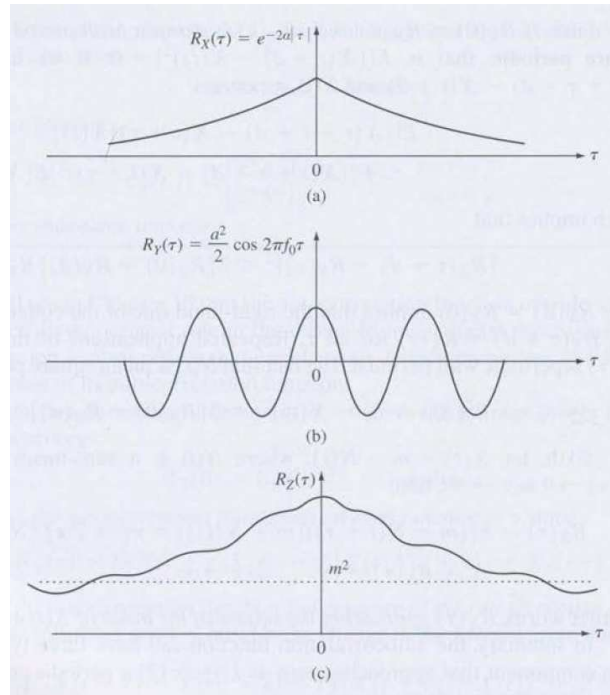


Figura 8.9 (a) Función de autocorrelación de una señal telégrafica aleatoria. (b) Función de autocorrelación de una senoide con fase aleatoria. (c) función de autocorrelación de un proceso estocástico que tiene media distinta de cero, una componente periódica, y una componente “aleatoria”.

(1) una componente que se aproxima a cero cuando $\tau \rightarrow \infty$, (2) una componente periódica, y (3) una componente debida a una media distinta de cero.

Ejemplo 8.25 La Figura 8.9 muestra varias funciones de autocorrelación típicas. La Figura 8.9(a) muestra la función de autocorrelación de una señal telégrafica aleatoria $X(t)$

$$R_X(\tau) = e^{-2\alpha|\tau|} \text{ para todo } \tau.$$

$X(t)$ es de media cero y $R_X(\tau) \rightarrow 0$ por $|\tau| \rightarrow \infty$. Figura 8.9(b) muestra la función de autocorrelación de una senoide $Y(t)$ con una amplitud a y fase aleatoria (ver Ejemplo 8.10):

$$R_Y(\tau) = \frac{a^2}{2} \cos(2\pi f_0 \tau) \text{ para todo } \tau.$$

$Y(t)$ es de media cero y $R_Y(t)$ es periódica con periodo de $1/f_0$. Figura 8.9(c) muestra la función de autocorrelación del proceso $Z(t) = X(t) + Y(t) + m$, donde $X(t)$ es el proceso telégrafo aleatorio, $Y(t)$ es una senoide con fase aleatoria, y

m es una constante. Si asumimos que $X(t)$ e $Y(t)$ son procesos independientes,

$$R_Z(\tau) = E[\{X(t + \tau) + Y(t + \tau) + m\}\{X(t) + Y(t) + m\}] \\ R_X(\tau) + R_Y(\tau) + m^2.$$

8.4.2 Procesos estocásticos gaussianos estacionarios en sentido amplio

Si un proceso estocástico gaussiano es estacionario en sentido amplio, entonces también es estacionario. La pdf conjunta de un proceso estocástico gaussiano es completamente determinada por la media $m_X(t)$ y la autocovarianza $C_X(t_1, t_2)$. Si $X(t)$ es estacionario en sentido amplio, entonces su media es una constante m y su autocovarianza depende sólo de la diferencia de los tiempos de muestreo, $t_i - t_j$. De esto se deduce que la pdf conjunta de $X(t)$ depende sólo de este conjunto de diferencias, y por lo tanto es invariante con respecto a desplazamientos en el tiempo. Así, el proceso es también estacionario. El resultado anterior hace que los procesos estocásticos gaussianos WSS sean fáciles de tratar ya que toda la información necesaria para especificar la pdf conjunta está contenida en m y $C_X(\tau)$.

8.5 Promedios temporales de procesos estocásticos y teoremas ergódicos

En algunos casos, los parámetros de un proceso estocástico deben ser obtenidos a través de medición. Los resultados del Capítulo 7 sugieren que se repita el experimento aleatorio que da lugar al proceso estocástico un gran número de veces y se tome la media aritmética de las magnitudes de interés. Por ejemplo, para estimar la media $m_X(t)$ de un proceso estocástico $X(t, \zeta)$ se puede repetir el experimento aleatorio y tomar el promedio siguiente:

$$\hat{m}_X(t) = \frac{1}{N} \sum_{i=1}^N X(t, \zeta_i),$$

donde N es el número de repeticiones del experimento, y $X(t, \zeta_i)$ es la realización observada en la i -ésima repetición.

En algunas situaciones, estamos interesados en estimar las funciones media o autocorrelación de la media temporal de una realización única, es decir,

$$\langle X(t) \rangle_T = \frac{1}{2T} \int_{-T}^T X(t, \zeta) dt. \quad (8.14)$$

Un teorema ergódico establece las condiciones bajo las cuales una media temporal converge cuando el intervalo de observación se vuelve grande. En esta

sección, estamos interesados en los teoremas ergódicos que establecen cuando los promedios temporales convergen a los promedios del conjunto (valor esperado).

La ley fuerte de los grandes números, presentada en el capítulo 7, es uno de los más importantes teoremas ergódicos. Establece que, si X_n es un proceso estocástico en tiempo discreto iid con media finita $E[X_n] = m$, entonces la media temporal de las muestras converge a la media del conjunto con probabilidad uno:

$$P \left[\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n X_i = m \right] = 1.$$

Este resultado nos permite estimar m tomando el promedio temporal de una sola realización del proceso. Estamos interesados en obtener resultados de este tipo para una clase más amplia de procesos estocásticos, es decir, para procesos estocásticos en tiempo discreto no iid, y para procesos estocásticos en tiempo continuo.

El siguiente ejemplo muestra que, en general, los promedios temporales no convergen a promedios de conjuntos.

Ejemplo 8.26 Sea $X(t) = A$ para todo t , donde A es una variable aleatoria de media cero y varianza unitaria. Hallar el valor límite del promedio temporal. La media del proceso es $m_X(t) = E[X(t)] = E[A] = 0$. sin embargo, la Ecuación 8.15 da

$$\langle X(t) \rangle_T = \frac{1}{2T} \int_{-T}^T A dt = A.$$

Así, el promedio temporal no siempre converge a $m_X = 0$. Nótese que este proceso es estacionario. Así este ejemplo muestra que los procesos estacionarios no siempre son también ergódicos.

Consideremos la estimación dada por la Ecuación 8.15 para $E[X(t)] = m_X(t)$. La estimación devuelve un solo número, así que obviamente tiene sentido considerar procesos para los cuales $m_X(t) = m$, con m una constante. Se presentará ahora un teorema ergódico para el promedio temporal de procesos estacionarios en sentido amplio.

Sea $X(t)$ un proceso WSS. El valor esperado de $\langle X(t) \rangle_T$ es

$$E[\langle X(t) \rangle_T] = E \left[\frac{1}{2T} \int_{-T}^T X(t) dt \right] = \frac{1}{2T} \int_{-T}^T E[X(t)] dt = m.$$

Esta ecuación establece que $\langle X(t) \rangle_T$ es un estimador insesgado de m .

TEOREMA 8.1 Sea $X(t)$ un proceso WSS con media $m_X(t) = m$ y autocovarianza $C_X(\tau)$ con $\tau = t_2 - t_1$, entonces

$$\lim_{T \rightarrow \infty} \langle X(t) \rangle_T = m$$

en valor cuadrático medio, si y solo si

$$\lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-2T}^{2T} \left(1 - \frac{|u|}{2T}\right) C_X(u) du = 0.$$

Por lo tanto, $\langle X(t) \rangle_T$ se acercará a m en valor cuadrático medio, es decir,

$$\lim_{T \rightarrow \infty} E[(\langle X(t) \rangle_T - m)^2] = 0$$

si y solo si

$$VAR[\langle X(t) \rangle_T] = \frac{1}{2T} \int_{-2T}^{2T} \left(1 - \frac{|u|}{2T}\right) C_X(u) du.$$

se tiende a cero por T creciente.

Según la terminología usada en la ingeniería, se dice que un proceso WSS es ergódico respecto a la media si satisface las condiciones del teorema anterior.

El teorema anterior se puede utilizar para obtener teoremas ergódicos para el promedio temporal de otras magnitudes. Por ejemplo, si se sustituye $X(t)$ con $Y(t + \tau)Y(t)$ en la Ecuación 8.15, se obtiene una estimación del promedio temporal de la función de autocorrelación del proceso $Y(t)$:

$$\langle Y(t + \tau)Y(t) \rangle_T = \frac{1}{2T} \int_{-T}^T Y(t + \tau)Y(t) dt. \quad (8.15)$$

Se demuestra fácilmente que $E[\langle Y(t + \tau)Y(t) \rangle_T] = R_Y(\tau)$ si $Y(t)$ es WSS. Sea $Y(t)$ un proceso WSS. El valor esperado de $\langle Y(t + \tau)Y(t) \rangle_T$ es

$$\begin{aligned} E[\langle Y(t + \tau)Y(t) \rangle_T] &= E \left[\frac{1}{2T} \int_{-T}^T Y(t + \tau)Y(t) dt \right] \\ &= \frac{1}{2T} \int_{-T}^T E[Y(t + \tau)Y(t)] dt = R_Y(\tau). \end{aligned}$$

El teorema ergódico anterior implica entonces que el promedio temporal de la autocorrelación converge a $R_Y(\tau)$ en valor cuadrático medio, es decir

$$\lim_{T \rightarrow \infty} E[(\langle Y(t + \tau)Y(t) \rangle_T - R_Y(\tau))^2] = 0$$

si $VAR[\langle Y(t + \tau)Y(t) \rangle_T]$ converge a cero por T creciente.

Ejemplo 8.27 El proceso telégrafo aleatorio, $X(t)$, tiene media $m_X(t) = 1$, varianza $VAR[X(t)] = 1$ y autocovarianza $C_X(t_1, t_2) = e^{-2\alpha|t_2 - t_1|}$.

¿Es el proceso telégrafo aleatorio ergódico respecto a la media?

La función de autocovarianza para el proceso telégrafo aleatorio se puede expresar como $C_X(\tau) = e^{-2\alpha|\tau|}$, con $\tau = t_2 - t_1$, por lo que la varianza de $\langle X(t) \rangle_T$

es

$$\begin{aligned}
 VAR[\langle X(t) \rangle_T] &= \frac{1}{2T} \int_{-2T}^{2T} \left(1 - \frac{|u|}{2T}\right) C_X(u) du \\
 &= \frac{1}{2T} \int_{-2T}^{2T} \left(1 - \frac{|u|}{2T}\right) e^{-2\alpha|u|} du \\
 &= \frac{2}{2T} \int_0^{2T} \left(1 - \frac{u}{2T}\right) e^{-2\alpha u} du \\
 &< \frac{1}{T} \int_0^{2T} e^{-2\alpha u} du \\
 &= \frac{1 - e^{-4\alpha T}}{2\alpha T}.
 \end{aligned}$$

La cota superior tiende a cero por $T \rightarrow \infty$ y por consiguiente $VAR[\langle X(t) \rangle_T] \rightarrow 0$ por $T \rightarrow \infty$. Entonces el proceso es ergódico respecto a la media.

Si el proceso aleatorio en consideración es en tiempo discreto, entonces las estimaciones mediante promedio temporal de las funciones media y autocorrelación de X_n están dadas por

$$\begin{aligned}
 \langle X_n \rangle_T &= \frac{1}{2T+1} \sum_{n=-T}^T X_n, \\
 \langle X_{n+k} X_n \rangle_T &= \frac{1}{2T+1} \sum_{n=-T}^T X_{n+k} X_n.
 \end{aligned}$$

Si X_n es un proceso aleatorio WSS, entonces $E[\langle X_n \rangle_T] = m$, por lo que $\langle X_n \rangle_T$ es una estimación no sesgada de m . Es fácil demostrar que la varianza de $\langle X_n \rangle_T$ es

$$VAR[\langle X_n \rangle_T] = \frac{1}{2T+1} \sum_{k=-2T}^{2T} \left(1 - \frac{|k|}{2T+1}\right) C_X(k)$$

Por lo tanto, $\langle X_n \rangle_T$ tiende a m en valor cuadrático medio y es ergódico respecto a la media si $VAR[\langle X_n \rangle_T]$ tiende a cero por valores crecientes de T .

Problemas

Secciones 8.1 y 8.2: Definición y especificación de un proceso estocástico

8.1 En el Ejemplo 8.1 hallar la pmf conjunta de X_1 y X_2 . Por qué X_1 y X_2 son independientes?

8.2 Un proceso estocástico en tiempo discreto X_n , se define de la siguiente manera. Se lanza un dado se observa el resultado que denotamos con k . El proceso, entonces, está definido por $X_n = k$ para todo n .

- Dibujar algunos caminos muestrales del proceso.
- Hallar la pmf de X_n .
- Calcular la pmf conjunta de X_n y X_{n+k} .
- Hallar las funciones media y autocovarianza de X_n .

8.3 Un proceso estocástico en tiempo discreto X_n , se define de la siguiente manera. Se lanza una moneda. Si el resultado es cara, $X_n = (-1)^n$, para todo n ; si el resultado es cruz, $X_n = (-1)^{n+1}$ para todo n .

- Dibujar algunos caminos muestrales del proceso.
- Hallar la pfm de X_n .
- Calcular la pmf conjunta de X_n y X_{n+k} .
- Hallar las funciones media y autocovarianza de X_n .

8.4 Un proceso estocástico de tiempo discreto está definido por $X_n = s^n$, para $n \geq 0$, donde s se selecciona al azar en el intervalo $(0, 1)$.

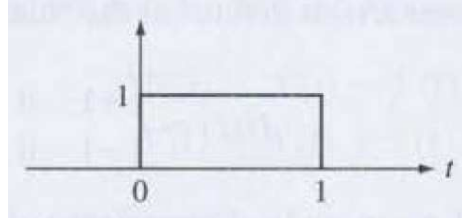
- Dibujar algunos caminos muestrales del proceso.
- Hallar la cdf de X_n .
- Calcular la cdf conjunta de X_n y X_{n+1} .
- Hallar las funciones media y autocovarianza de X_n .
- Repetir los puntos (a), (b), (c) y (d), si s es uniforme en $(1, 2)$.

8.5 Sea $g(t)$ el pulso rectangular que se muestra en la Figura 8.5. El proceso estocástico $X(t)$ se define como $X(t) = Ag(t)$, donde A toma los valores ± 1 con igual probabilidad.

- Hallar la pmf de $X(t)$.
- Hallar $m_X(t)$.
- Calcular la pmf conjunta de $X(t)$ y $X(t+d)$.
- Calcular $C_X(t, t+d)$, $d > 0$.

8.6 Un proceso estocástico está definido por $Y(t) = g(t-T)$, donde $g(t)$ es el pulso rectangular de la Figura 8.5 y T es una variable aleatoria uniformemente distribuida en el intervalo $(0, 1)$.

- Hallar la pmf de $Y(t)$.



(b) Hallar $m_Y(t)$ y $C_Y(t_1, t_2), d$.

8.7 Un proceso estocástico está definido por

$$X(t) = g(t - T),$$

donde T es una variable aleatoria uniforme en el intervalo $(0, 1)$ y $g(t)$ es la forma de onda triangular periódica que se muestra en la Figura 8.10.

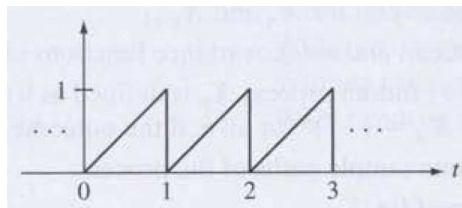


Figura 8.10

(a) Hallar la cdf de $X(t)$ para $0 < t < 1$.

(b) Calcular $m_X(t)$ y $C_X(t_1, t_2)$.

8.8 Sea $Y(t) = g(t - T)$, donde $g(t)$ es el pulso rectangular de la Figura 8.10 y T es una variable aleatoria con distribución exponencial con parámetro α .

(a) Hallar la pmf de $Y(t)$.

(b) Calcular la pmf conjunta de $Y(t)$ e $Y(t + d)$. Considerar los dos casos: $d > 1$ y $0 < d < 1$.

(c) Hallar $m_Y(t)$ y $C_Y(t, t + d)$ para $d > 1$ y $0 < d < 1$.

8.9 Sea $Z(t) = At^3 + B$, donde A y B son variables aleatorias independientes.

(a) Hallar la pdf de $Z(t)$.

(b) Calcular $m_Z(t)$ y $C_Z(t_1, t_2)$.

8.10 Hallar una expresión para $E[|X_{t_2} - X_{t_1}|^2]$ en términos de la función de autocorrelación.

8.11 El proceso estocástico $H(t)$ se define como la versión “duramente limitada” de $X(t)$:

$$H(t) = \begin{cases} +1 & \text{si } X(t) \geq 0, \\ -1 & \text{si } X(t) < 0. \end{cases}$$

- (a) Hallar la pdf, la media y la autocovarianza de $H(t)$ si $X(t)$ es la senoide con amplitud aleatoria presentada en el Ejemplo 8.2.
- (b) Hallar la pdf, la media y la autocovarianza de $H(t)$ si $X(t)$ es la senoide con fase aleatoria presentada en el Ejemplo 8.9.
- (c) Hallar una expresión general para la media de $H(t)$ en términos de la cdf de $X(t)$.

- 8.12** (a) ¿Son ortogonales los procesos estocásticos independientes? Explicar.
 (b) ¿Son incorrelacionados los procesos estocásticos ortogonales? Explicar.
 (c) ¿Son independientes los procesos estocásticos incorrelacionados?
 (d) ¿Son ortogonales los procesos estocásticos incorrelacionados?

8.13 El proceso estocástico $Z(t)$ se define como

$$Z(t) = 2Xt - Y,$$

donde X e Y son dos variables aleatorias con medias m_X, m_Y , varianzas σ_X^2, σ_Y^2 , y coeficiente de correlación $\rho_{X,Y}$. Hallar la media y la autocovarianza de $Z(t)$.

8.14 Sea $H(t)$ la salida del limitador duro considerado en el problema 8.11.

- (a) Hallar la cross-correlación y la cross-covarianza entre $H(t)$ y $X(t)$ cuando la de entrada es una senoide de amplitud aleatoria como en el problema 8.11a.
- (b) Repetir el cálculo suponiendo que la entrada sea una senoide con fase aleatoria, como en el problema 8.11b.
- (b) ¿Son incorrelacionados los procesos de entrada y de salida? ¿Son ortogonales?

8.15 Sea $Y_n = X_n + g(n)$, donde X_n es un proceso estocástico en tiempo discreto con media cero y $g(n)$ es una función determinista de n .

- (a) Hallar la media y la varianza de Y_n .
- (b) Hallar la cdf conjunta de Y_n e Y_{n+1} .
- (c) Hallar la función de autocovarianza de Y_n .
- (d) Representar funciones muestrales típicas para X_n e Y_n si: $g(n) = n$, $g(n) = 1/n^2$ y $g(n) = 1/n$.

8.16 Sea $Y_n = c(n)X_n$ donde X_n es un proceso estocástico en tiempo discreto con media cero y varianza uno y $c(n)$ es una función determinista de n .

- (a) Hallar la media y la varianza de Y_n .
- (b) Hallar la cdf conjunta de Y_n e Y_{n+1} .

- (c) Hallar la función de autocovarianza de Y_n .
 (d) Representar funciones muestrales típicas para X_n e Y_n si: $c(n) = n$, $c(n) = 1/n^2$ y $c(n) = 1/n$.

8.17 (a) Hallar la cross-correlación y la cross-covarianza de X_n y Y_n considerados en el problema 8.15.

- (b) Hallar la pdf conjunta de X_n e Y_{n+1} .
 (c) Determinar si X_n e Y_n son procesos estocásticos incorrelacionados, independientes, u ortogonales.

8.18 (a) Hallar la cross-correlación y la cross-covarianza de X_n y Y_n considerados en el problema 8.16.

- (b) Hallar la pdf conjunta de X_n e Y_{n+1} .
 (c) Determinar si X_n e Y_n son procesos estocásticos incorrelacionados, independientes, u ortogonales.

8.19 Supongamos que $X(t)$ e $Y(t)$ son procesos estocásticos independientes y sean

$$U(t) = X(t) - Y(t)$$

$$V(t) = X(t) + Y(t).$$

- (a) Hallar $C_{UX}(t_1, t_2)$, $C_{UY}(t_1, t_2)$ y $C_{UV}(t_1, t_2)$.
 (b) Hallar $f_{U(t_1)X(t_2)}(u, x)$ y $f_{U(t_1), V(t_2)}(u, v)$. Sugerencia: Utilice variables auxiliares.

8.20 Repetir el problema 8.19, si $X(t)$ e $Y(t)$ son procesos estocásticos en tiempo discreto independientes y $X(t)$ e $Y(t)$ son diferentes procesos estocásticos iid.

Sección 8.3: Proceso de suma, proceso de conteo binomial y paseo aleatorio

8.21 (a) Sea Y_n ser el proceso que se produce cuando en un proceso de Bernoulli 1 aislados se borran con probabilidad α . Encontrar las pmf de S'_n , el proceso de conteo para Y_n ¿Tiene Y_n incrementos independientes y estacionarios?

- (b) Repetir el punto anterior si, además del borrado, 0 aislados en el proceso de Bernoulli se cambian a 1 con probabilidad β .

8.22 Supongamos que S_n denota un proceso de conteo binomial.

- (a) Demostrar que $P[S_n = j, S_{n'} = i] \neq P[S_n = j]P[S_{n'} = i]$.
 (b) Hallar $P[S_{n_2} = j | S_{n_1} = i]$, donde $n_2 > n_1$.

- (c) Demostrar que $P[S_{n_2} = j | S_{n_1} = i, S_{n_0} = k] = P[S_{n_2} = j | S_{n_1} = i]$, donde $n_2 > n_1 > n_0$.

- 8.23** (a) Hallar $P[S_n = 0]$ para el proceso paseo aleatorio.
 (b) ¿Cuál es la respuesta en el punto anterior si $p = 1/2$?

- 8.24** Consideremos los siguientes procesos de media móvil:

$$Y_n = 1/2(X_n + X_{n-1}), X_0 = 0$$

$$Z_n = 2/3X_n + 1/3X_{n-1}, X_0 = 0.$$

- (a) Hallar la varianza media, y covarianza de Y_n y Z_n , si X_n es un proceso estocástico de Bernoulli.
 (b) Repetir el punto (a) si X_n es el proceso escalón aleatorio.
 (c) Generar 100 resultados de un proceso estocástico de Bernoulli X_n , y encontrar las Y_n y Z_n resultantes. ¿Son las medias muestrales de Y_n y Z_n del punto (a) próximas a sus respectivas medias?
 (d) Repetir el punto (c) con X_n producida por el proceso escalón aleatorio.

- 8.25** Considerar los procesos autorregresivos siguientes:

$$W_n = 2W_{n-1} + X_n, W_0 = 0$$

$$Z_n = 3/4Z_{n-1} + X_n, Z_0 = 0.$$

- (a) Supongamos que X_n es un proceso de Bernoulli. ¿Qué tendencias de la manifiestan los procesos?
 (b) Expresar W_n y Z_n en términos de X_n, X_{n-1}, \dots, X_1 y luego encontrar $E[W_n]$ y $E[Z_n]$. ¿Estos resultados concuerdan con las tendencias esperadas?
 (c) ¿ W_n o Z_n tienen incrementos independientes? ¿Tienen incrementos estacionarios?
 (d) Generar 100 resultados de un proceso de Bernoulli. Encontrar las realizaciones resultantes de W_n y Z_n . ¿Es la media muestral significativa para estos procesos?
 (e) Repetir el punto (a) suponiendo que X_n es el proceso escalón aleatorio.

- 8.26** Sea M_n es el proceso en tiempo discreto definidos como la secuencia de medias muestrales de una secuencia iid:

$$M_n = \frac{X_1 + X_2 + \dots + X_n}{n}.$$

- (a) Hallar la media, la varianza y la covarianza de M_n .
 (b) ¿Tiene M_n incrementos independientes? ¿Tiene incrementos estacionarios?

8.27 Hallar la pdf de los procesos definidos en el problema 8.24 si las X_n proceden de una secuencia iid de variables aleatorias gaussianas de media cero y varianza unitaria.

8.28 Supongamos que las X_n consisten en una secuencia iid de variables aleatorias de Cauchy.

- (a) Hallar la pdf del proceso de suma S_n . Sugerencia: Usar el método de la función característica.
- (b) Hallar la pdf conjunta de S_n y S_{n+k} .

8.29 Supongamos que las X_n consisten en una secuencia iid de variables aleatorias de Poisson con media α .

- (a) Hallar la pmf del proceso suma S_n .
- (b) Hallar la pdf conjunta de S_n y S_{n+k} .

8.30 Sea X_n una secuencia iid de variables aleatorias gaussianas con media cero y varianza unitaria.

- (a) Hallar la pdf de M_n definida en el problema 8.26.
- (b) Hallar la pdf conjunta de M_n y M_{n+k} . Sugerencia: Utilizar la propiedad de incrementos independientes de S_n .

8.31 Repetir el problema 8.26 con $X_n = 1/2(Y_n + Y_{n-1})$, donde Y_n es un proceso estocástico iid. ¿Qué ocurre a la varianza de M_n a medida que aumenta n ?

8.32 Repetir el problema 8.26 con $X_n = 3/4X_{n-1} + Y_n$, donde Y_n es un proceso estocástico iid. ¿Qué ocurre a la varianza de M_n a medida que aumenta n ?

8.33 Supongamos que un experimento aleatorio tiene tres resultados posibles, por ejemplo 0, 1 y 2, y supongamos que ocurren con probabilidades p_0, p_1 , y p_2 , respectivamente.

Consideremos una secuencia de repeticiones independientes del experimento, y sea $X_j(n)$ la función indicatriz del resultado j . El vector

$$\mathbf{X}(n) = (X_0(n), X_1(n), X_2(n))$$

constituye entonces un proceso estocástico vectorial de Bernoulli. Consideremos el proceso de conteo para $\mathbf{X}(n)$:

$$\mathbf{S}(n) = \mathbf{X}(n) + \mathbf{X}(n-1) + \cdots + \mathbf{X}(1), \mathbf{S}(0) = \mathbf{0}.$$

- (a) Demostrar que $\mathbf{S}(n)$ tiene una distribución multinomial.
- (b) Demostrar que $\mathbf{S}(n)$ tiene incrementos independientes, y luego hallar la pmf conjunta de $\mathbf{S}(n)$ y $\mathbf{S}(n+k)$.
- (c) Demostrar que las componentes $S_j(n)$ del proceso vectorial constituyen un proceso de conteo binomial.

Sección 8.4: Procesos estacionarios

- 8.34** (a) ¿Es la senoide de amplitud aleatoria en el Ejemplo 8.9 un proceso estocástico estacionario? ¿Es estacionario en sentido amplio?
 (b) Repita la parte (a) para la senoide con fase aleatoria en el Ejemplo 8.10.

- 8.35** Un proceso estocástico de tiempo discreto X_n se define como sigue. Se lanza una moneda equilibrada, y si el resultado es cara entonces $X_n = 1$ para todo n , y en caso contrario $X_n = -1$ para todo n .
 (a) ¿Es X_n un proceso estocástico WSS?
 (b) ¿Es X_n un proceso estocástico estacionario?
 (c) ¿Cambian las respuestas dadas en las partes (a) y (b) si la moneda es sesgada?

- 8.36** Let X_n be proceso estocástico del Problema 8.3.
 (a) ¿Es X_n un proceso estocástico WSS?
 (b) ¿Es X_n un proceso estocástico estacionario?
 (c) ¿Es X_n un proceso estocástico cicloestacionario?

- 8.37** Sea $X(t) = g(t-T)$, donde $g(t)$ es la forma de onda periódica introducida en el Problema 8.7 y T es una variable aleatoria uniformemente distribuida en el intervalo $(0, 1)$. ¿Es $X(t)$ un proceso estocástico estacionario? Es $X(t)$ un proceso estocástico estacionario en sentido amplio?

- 8.38** Sea $X(t)$ definido por

$$X(t) = A\cos(\omega t) + B\sin(\omega t),$$

donde A y B son variables aleatorias iid.

- (a) ¿Bajo qué condiciones $X(t)$ es estacionario en sentido amplio?
 (b) Demostrar que $X(t)$ es no estacionario. Sugerencia: Considere $E[X^3(t)]$.

- 8.39** Considere el siguiente proceso de media móvil:

$$Y_n = 1/2(X_n + X_{n-1}), X_0 = 0.$$

- (a) ¿Es Y_n un proceso estocástico estacionario si X_n es un proceso iid con valores enteros?
 (b) ¿Es Y_n un proceso estocástico estacionario si X_n es un proceso estacionario?
 (c) ¿Son Y_n y X_n procesos estocásticos conjuntamente estacionarios si X_n es un proceso iid?

- 8.40** Sea X_n es un proceso iid de media cero, y sea Z_n un proceso estocástico autorregresivo

$$Z_n = 3/4Z_{n-1} + X_n, Z_0 = 0.$$

- (a) Hallar la autocovarianza de Z_n y determinar si Z_n es estacionario en sentido amplio. Sugerencia: Expresar Z_n en términos de X_n, X_{n-1}, \dots, X_1 .
- (b) ¿Tiende Z_n a un comportamiento estacionario?
- (c) Hallar la pdf de Z_n , si X_n es una secuencia de variables aleatorias iid gaussianas con media cero y varianza uno. ¿Cuál es la pdf de Z_n por $n \rightarrow \infty$?

8.41 Sea $Y(t) = X(t+s) - \beta X(t)$, donde $X(t)$ es un proceso estocástico estacionario en sentido amplio.

- (a) Determinar si $Y(t)$ es también un proceso estocástico estacionario en sentido amplio.
- (b) Hallar la función de covarianza cruzada de $Y(t)$ y $X(t)$. ¿Son estos procesos conjuntamente estacionarios en sentido amplio?
- (c) Hallar la pdf de $Y(t)$ si $X(t)$ es un proceso estocástico gaussiano.
- (d) Hallar la pdf conjunta de $Y(t_1)$ e $Y(t_2)$ en la parte (c).
- (e) Hallar la pdf conjunta de $Y(t_1)$ y $X(t_2)$ en la parte (c).

8.42 Sean $X(t)$ e $Y(t)$ procesos estocásticos estacionarios en sentido amplio independientes con medias cero y con la misma función de covarianza $C_X(\tau)$. Sea $Z(t)$ definido por

$$Z(t) = 3X(t) - 5Y(t).$$

- (a) Determinar si $Z(t)$ es también estacionario en sentido amplio.
- (b) Hallar la pdf de $Z(t)$ si $X(t)$ e $Y(t)$ son también procesos estocásticos conjuntamente gaussianos con media cero y con $C_X(\tau) = 4e^{-|\tau|}$.
- (c) Hallar la pdf conjunta de $Z(t_1)$ y $Z(t_2)$ en la parte (b).
- (d) Hallar la covarianza cruzada entre $Z(t)$ y $X(t)$. ¿Son $Z(t)$ y $X(t)$ procesos estocásticos conjuntamente estacionarios?
- (e) Hallar la pdf conjunta de $Z(t_1)$ y $X(t_2)$ en la parte (b). Sugerencia: Utilizar variables auxiliares.

8.43 Sean $X(t)$ e $Y(t)$ procesos estocásticos estacionarios en sentido amplio independientes con medias cero y con la misma función de covarianza $C_X(\tau)$. Sea $Z(t)$ definido por

$$Z(t) = X(t)\cos\omega t + Y(t)\sin\omega t.$$

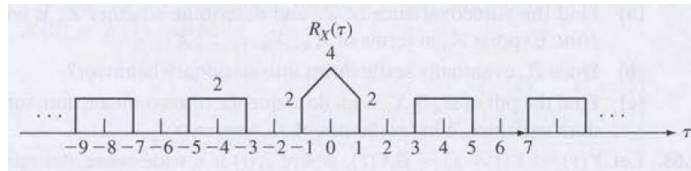
- (a) Determinar si $Z(t)$ es un proceso estocástico estacionario en sentido amplio.
- (b) Hallar la pdf de $Z(t)$ si $X(t)$ e $Y(t)$ son también procesos estocásticos conjuntamente gaussianos con media cero y con $C_X(\tau) = 4e^{-|\tau|}$.
- (c) Hallar la pdf conjunta de $Z(t_1)$ y $Z(t_2)$ en la parte (b).
- (d) Hallar la covarianza cruzada entre $Z(t)$ y $X(t)$. ¿Son $Z(t)$ y $X(t)$ procesos estocásticos conjuntamente estacionarios?
- (e) Hallar la pdf conjunta de $Z(t_1)$ y $X(t_2)$ en la parte (b).

8.44 Sea $X(t)$ un proceso estocástico estacionario en sentido amplio gaussiano de media cero con función de autocorrelación $R_X(\tau)$. La salida de un "detector de ley del cuadrado" es

$$Y(t) = X(t)^2.$$

Demostrar que $R_Y(\tau) = R_X(0)^2 + 2R_X^2(\tau)$. Sugerencia: Para variables aleatorias conjuntamente gaussianas con media cero, $E[X^2Z^2] = E[X^2]E[Z^2] + 2E[XZ]^2$.

8.45 Un proceso $X(t)$ WSS tiene media 1 y función de autocorrelación dada en la figura siguiente



- Hallar la componente media de $R_X(\tau)$.
- Hallar la componente periódica de $R_X(\tau)$.
- Hallar la componente restante de $R_X(\tau)$.

Sección 8.5: Promedios temporales de procesos estocásticos y teoremas ergódicos

8.46 Hallar la varianza del promedio temporal del Ejemplo 8.26.

8.47 Establecer si los procesos siguientes son WSS y ergódicos respecto a la media.

- El proceso en tiempo discreto del Problema 8.2.
- El proceso en tiempo discreto del Problema 8.3.
- El proceso $X_n = s^n$, $n \geq 0$ del Problema 8.4.

8.48 ¿Es el proceso estocástico WSS $X(t)$ ergódico respecto a la media?

$$R_X(\tau) = \begin{cases} 0 & |\tau| > 1 \\ 5(1 - |\tau|) & |\tau| \leq 1 \end{cases}$$

8.49 Sea $X(t) = A \cos(2\pi ft)$, donde A es una variable aleatoria con media m y varianza σ^2 .

- Calcular $\langle X(t) \rangle_T$, hallar su límite por $T \rightarrow \infty$ y compararlo con $m_X(t)$.
- Calcular $\langle X(t + \tau)X(t) \rangle_T$, hallar su límite por $T \rightarrow \infty$ y compararlo con $R_X(t + \tau, t)$.

8.50 Repetir el Problema 8.49 con $X(t) = A\cos(2\pi ft + \Theta)$, donde A es la misma del Problema 8.49, Θ es una variable aleatoria con distribución uniforme en $(0, 2\pi)$ y A y Θ son variables aleatorias independientes.

8.51 Hallar una expresión exacta para $VAR[\langle X(t) \rangle_T]$ en el Ejemplo 8.27. Hallar el límite por $T \rightarrow \infty$.

8.52 El proceso estocástico WSS X_n tiene media m y autocovarianza $C_X(k) = (1/2)^{|k|}$. ¿Es X_n ergódico respecto a la media?

