CAPÍTULO SEIS

Distribuciones conjuntas de probabilidad

6.1 Introducción

En los capítulos anteriores se consideraron conceptos probabilísticos tomando en cuenta una variable aleatoria a la vez. Sin embargo, muchas veces resulta de interés medir más de una característica de algún fenómeno aleatorio. Por ejemplo, en un proceso de producción en el que se tiene determinado número de artículos producidos en un tiempo definido, es muy común que el interés no sólo recaiga en el número de artículos que se encuentran listos para su venta inmediatamente después de su fabricación, sino también en el número que, después de reprocesarse, cae en la categoría anterior o en el número de artículos que serán desechados. Otro ejemplo puede ser que, al estudiar la contaminación del agua en general, se mida la concentración de varios contaminantes presentes en ésta. De los ejemplos anteriores surge la necesidad de estudiar modelos de probabilidad que contengan más de una variable aleatoria. Estos modelos reciben el nombre de *modelos multivariados,* mientras que los modelos con una sola variable reciben el nombre de *univariados.* En este capítulo se examinarán conceptos generales para distribuciones de probabilidad discretas y continuas con dos variables aleatorias. La extensión de estos conceptos a un mayor número de variables aleatorias resulta directa.

6.2 Distribuciones de probabilidad bivariadas

En esta sección se considerarán las definiciones pertinentes para distribuciones, tanto discretas como continuas, de dos variables aleatorias.

Definición 6.1 Sean X y Y dos variables aleatorias discretas. La probabilidad de que $X = x$ y $Y = y$ está determinada por la función de probabilidad bivariada

$$p(x, y) = P(X = x, Y = y),$$

en donde $p(x, y) \geqslant 0$ para toda *x, y,* de *X, Y,* y $\Sigma_x \Sigma_y p(x, y) = 1$. La suma se efectúa sobre todos los valores posibles de *x* y *y*.

Con base en la definición 6.1, la función de distribución acumulativa bivariada es la probabilidad conjunta de que $X \leqslant x$ y $Y \leqslant y$, dada por

$$F(x, y) = P(X \leqslant x, Y \leqslant y) = \sum_{x_i \leqslant x} \sum_{y_i \leqslant y} p(x_i, y_i). \tag{6.1}$$

La expresión anterior es una extensión del caso univariado. La función de probabilidad conjunta de dos variables aleatorias da origen a las probabilidades puntuales conjuntas, y la función de distribución bivariada es una función escalonada creciente para cada probabilidad puntual distinta de cero, de manera tal que $X = x$ y $Y = y$.

Ejemplo 6.1 Con base en la experiencia se sabe que la proporción de unidades útiles producidas por un proceso de manufactura es p_1, y las proporciones de unidades enviadas a reprocesar y desechadas, son p_2 y P_3, respectivamente. Si se supone que el número de unidades que se produce en un lapso dado es *n* y que además éstas constituyen un conjunto de ensayos independientes de manera que $p_1 + p_2 + p_3 = 1$, desarrollar una expresión para la probabilidad de tener, de manera exacta, x_1, x_2 y x_3 unidades útiles, reprocesables y desechadas, respectivamente.

Lo que se pide es una extensión de la distribución binomial univariada. A pesar de que existen tres resultados mutuamente excluyentes (útil, reprocesable y desechado), sólo es necesario definir dos variables aleatorias dado que, para cualquier número específico de cada una, la suma de las tres es *n*. Por consiguiente, sean *X* y *Y* las variables aleatorias que representan el número de unidades útiles y reprocesables, respectivamente, del total de unidades *n*. De esta manera, si $X = x$ y $Y = y$, entonces el número de unidades que deben desecharse es $n - x - y$. Por la hipótesis de independencia, la probabilidad de tener una secuencia específica de resultados es

$$p_1^x p_2^y (1 - p_1 - p_2)^{n-x-y}.$$

Dado que existen $n!/[x!y!(n - x - y)!]$ formas igualmente probables para que ocurra una secuencia de resultados específica, la probabilidad conjunta de tener, de manera exacta, *x, y,* y $n - x - y$ unidades útiles, reprocesables y desechadas, respectivamente, es

$$p(x, y; n, p_1, p_2) = \frac{n!}{x!y!(n-x-y)!} p_1^x p_2^y (1 - p_1 - p_2)^{n-x-y},$$
$$x, y = 0, 1, 2, \ldots, n, \tag{6.2}$$

en donde $p_3 = 1 - p_1 - p_2$. La expresión (6.2) es la función de probabilidad conjunta de lo que se conoce como la distribución trinomial. Los parámetros de esta distribución son *n*, p_1 y p_2, dado que p_3 se determina de manera exacta si se conocen

p_1 y p_2. La distribución trinomial se ha aplicado, de manera extensa, a situaciones en que existen tres resultados distintos, como en las encuestas sobre la preferencia del consumidor en relación a tres marcas comerciales o en encuestas de tipo político en que se pide la opinión con respecto a tres candidatos.

Si existen k resultados distintos excluyentes con probabilidades $p_1, p_2, \ldots p_k$, respectivamente, entonces para n ensayos independientes, la distribución trinomial se generaliza para originar la distribución multinomial cuya función de probabilidad es:

$$p(x_1, x_2, \ldots, x_{k-1}; n, p_1, p_2, \ldots, p_{k-1}) = \frac{n!}{x_1!x_2! \ldots x_k!} p_1^{x_1} p_2^{x_2} \ldots p_k^{x_k}$$

$$x_i = 0, 1, 2, \ldots, n \quad \text{for } i = 1, 2, \ldots, k, \tag{6.3}$$

en donde $x_k = n - x_1 - x_2 - \cdots - x_{k-1}$ y $p_k = 1 - p_1 - p_2 - \cdots - p_{k-1}$.

Definición 6.2 Sean X y Y dos variables aleatorias continuas. Si existe una función $f(x, y)$ tal que la probabilidad conjunta:

$$P(a < X < b, c < Y < d) = \int_a^b \int_c^d f(x, y) dy dx$$

para cualquier valor de a, b, c, y d en donde $f(x, y) \geqslant 0$, $-\infty < x, y < \infty$, y $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y) dy dx = 1$, entonces $f(x, y)$ es la función de densidad de probabilidad bivariada de X y Y.

La función de densidad de probabilidad de dos variables aleatorias continuas X y Y es una superficie en el espacio de tres dimensiones donde el volumen por debajo de ésta y por encima de un rectángulo específico $a < X < b$ y $c < Y < d$ es igual a la probabilidad de que las variables aleatorias tomen valores iguales a los puntos que se encuentren dentro del rectángulo.

La función de distribución bivariada acumulativa de X y Y es la probabilidad conjunta de que $X \leqslant x$ y $Y \leqslant y$, dada por:

$$P(X \leqslant x, Y \leqslant y) = F(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u, v) dv du. \tag{6.4}$$

Por lo tanto, la función de densidad bivariada se encuentra diferenciando $F(x, y)$ con respecto a x y y; es decir,

$$f(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y}. \tag{6.5}$$

Ejemplo 6.2 Sean X y Y dos variables aleatorias continuas con función de densidad de probabilidad conjunta dada por:

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

Graficar la función de densidad de probabilidad conjunta, determinar la función de distribución acumulativa conjunta y obtener la probabilidad conjunta de que $X \leq 1/2$ y $Y \leq 3/4$.

La gráfica de la función de densidad conjunta se ilustra en la figura 6.1. Nótese que $f(x, y)$ es una función de densidad de probabilidad conjunta, dado que

$$\int_0^1\int_0^1 (x + y)dydx = \int_0^1 \left(xy + \frac{y^2}{2}\right)\Bigg|_0^1 dx = \int_0^1 \left(x + \frac{1}{2}\right)dx = 1.$$

Entonces

$$F(x, y) = \int_0^x\int_0^y (u + v)dvdu = \int_0^x \left(uy + \frac{y^2}{2}\right)du = xy(x + y)/2, \qquad 0 \leq x, y \leq 1.$$

De esta forma se tiene

$$F(1/2, 3/4) = \left(\frac{1}{2}\right)\left(\frac{1}{2}\right)\left(\frac{3}{4}\right)\left(\frac{1}{2} + \frac{3}{4}\right) = 15/64.$$

Además

$$\frac{\partial F(x, y)}{\partial x} = xy + \frac{y^2}{2}$$

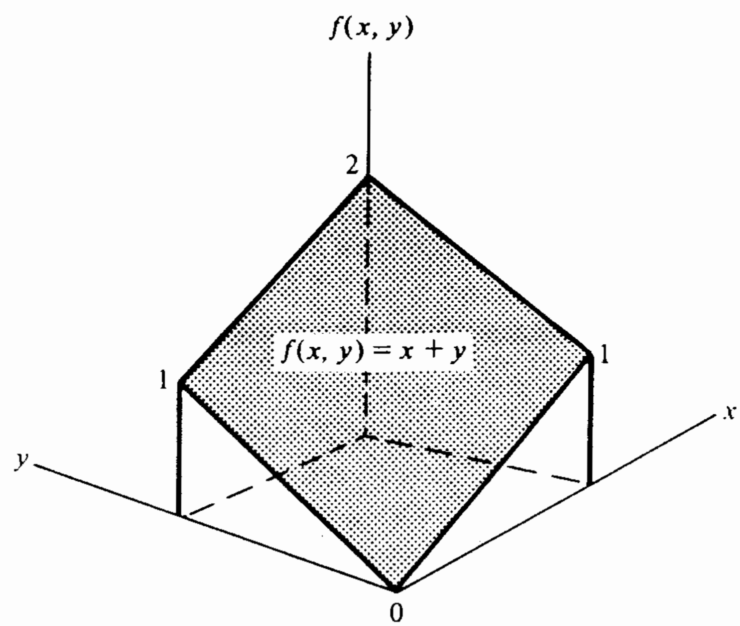

FIGURA 6.1 Gráfica de la función de densidad conjunta $f(x, y) = x + y$

y

$$\frac{\partial^2 F(x, y)}{\partial x\, \partial y} = x + y = f(x, y).$$

6.3 Distribuciones marginales de probabilidad

Es posible determinar varias distribuciones marginales para cualquier distribución de probabilidad que contenga más de dos variables aleatorias. Por ejemplo, si X y Y son variables aleatorias discretas, la suma de la función de probabilidad bivariada sobre todos los valores posibles de Y dará origen a la función de probabilidad univariada de X. Por otro lado, si X y Y son variables aleatorias continuas, la integración de la función de densidad de probabilidad bivariada sobre el intervalo completo de variación de Y generará la función de densidad de probabilidad univariada de X. De acuerdo con lo anterior, se formulan las siguientes definiciones:

Definición 6.3 Sean X y Y dos variables aleatorias discretas con una función de probabilidad conjunta $p(x, y)$. Las funciones marginales de probabilidad de X y de Y están dadas por

$$p_X(x) = \sum_y p(x, y)$$

y

$$p_Y(y) = \sum_x p(x, y),$$

respectivamente.

Definición 6.4 Sean X y Y dos variables aleatorias continuas con una función de densidad de probabilidad conjunta $f(x, y)$. Las funciones de densidad de probabilidad de X y de Y están dadas por

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)dy$$

y

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y)dx,$$

respectivamente.

Para variables aleatorias continuas conjuntas, si se conoce la función de distribución acumulativa $F(x, y)$, las distribuciones acumulativas marginales de X y Y se obtienen de la siguiente forma:

$$P(X \leq x) = F_X(x) = \int_{-\infty}^{x} \int_{-\infty}^{\infty} f(t, y)dydt,$$

y

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt = F(x, \infty). \tag{6.6}$$

De manera similar

$$P(Y \leq y) = F_Y(y) = \int_{-\infty}^{y} \int_{-\infty}^{\infty} f(x, t)dxdt = \int_{-\infty}^{y} f_Y(t)dt = F(\infty, y). \tag{6.7}$$

Así puede determinarse la distribución acumulativa marginal de X dejando que Y tome un valor igual al límite superior de la función de distribución conjunta de X y Y.

Ejemplo 6.3 Sean X y Y dos variables aleatorias continuas con una función de densidad de probabilidad conjunta:

$$f(x, y) = \begin{cases} 3x(1 - xy) & 0 \leq x, y \leq 1, \\ 0 & \text{para cualquier otro valor.} \end{cases}$$

Obtener las distribuciones de densidad marginal y acumulativa de X y Y.

La función de densidad marginal de X es

$$f_X(x) = 3 \int_0^1 x(1 - xy)dy = 3\left(xy - \frac{x^2y^2}{2}\right)\Bigg|_0^1 = 3x\left(1 - \frac{x}{2}\right).$$

De manera similar para Y

$$f_Y(y) = 3 \int_0^1 x(1 - xy)dx = 3\left(\frac{x^2}{2} - \frac{x^3y}{3}\right)\Bigg|_0^1 = (3 - 2y)/2.$$

La distribución acumulativa conjunta de X y Y es

$$F(x, y) = 3 \int_0^x \int_0^y u(1 - uv)dvdu = 3 \int_0^x \left(uy - \frac{u^2y^2}{2}\right)du$$

$$= x^2y(3 - xy)/2, \qquad 0 \leq x, y \leq 1.$$

Por lo tanto, las distribuciones acumulativas marginales de X y Y están dadas por

$$F_X(x) = F(x, 1) = x^2(3 - x)/2, \qquad 0 \leq x \leq 1,$$

y

$$F_Y(y) = F(1, y) = y(3 - y)/2, \qquad 0 \leq y \leq 1,$$

respectivamente.

y

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt = F(x, \infty). \tag{6.6}$$

De manera similar

$$P(Y \leq y) = F_Y(y) = \int_{-\infty}^{y}\int_{-\infty}^{\infty} f(x, t)dxdt = \int_{-\infty}^{y} f_Y(t)dt = F(\infty, y). \tag{6.7}$$

Así puede determinarse la distribución acumulativa marginal de X dejando que Y tome un valor igual al límite superior de la función de distribución conjunta de X y Y.

Ejemplo 6.3 Sean X y Y dos variables aleatorias continuas con una función de densidad de probabilidad conjunta:

$$f(x, y) = \begin{cases} 3x(1 - xy) & 0 \leq x, y \leq 1, \\ 0 & \text{para cualquier otro valor.} \end{cases}$$

Obtener las distribuciones de densidad marginal y acumulativa de X y Y.

La función de densidad marginal de X es

$$f_X(x) = 3\int_0^1 x(1 - xy)dy = 3\left(xy - \frac{x^2y^2}{2}\right)\Bigg|_0^1 = 3x\left(1 - \frac{x}{2}\right).$$

De manera similar para Y

$$f_Y(y) = 3\int_0^1 x(1 - xy)dx = 3\left(\frac{x^2}{2} - \frac{x^3y}{3}\right)\Bigg|_0^1 = (3 - 2y)/2.$$

La distribución acumulativa conjunta de X y Y es

$$F(x, y) = 3\int_0^x\int_0^y u(1 - uv)dvdu = 3\int_0^x \left(uy - \frac{u^2y^2}{2}\right)du$$

$$= x^2y(3 - xy)/2, \qquad 0 \leq x, y \leq 1.$$

Por lo tanto, las distribuciones acumulativas marginales de X y Y están dadas por

$$F_X(x) = F(x, 1) = x^2(3 - x)/2, \qquad 0 \leq x \leq 1,$$

y

$$F_Y(y) = F(1, y) = y(3 - y)/2, \qquad 0 \leq y \leq 1,$$

respectivamente.

6.4 Valores esperados y momentos para distribuciones bivariadas

En esta sección se tratarán los conceptos de valor esperado y momentos para distribuciones conjuntas de probabilidad.

Definición 6.5 Sean X y Y dos variables aleatorias que se distribuyen conjuntamente. El valor esperado de una función de X y de Y, $g(x, y)$, se define como

$$E[g(X, Y)] = \sum_x \sum_y g(x, y)p(x, y)$$

si X y Y son discretas, o

$$E[g(X, Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x, y)f(x, y)dydx$$

si X y Y son continuas, en donde $p(x, y)$ y $f(x, y)$ son las funciones de probabilidad y de densidad de probabilidad conjuntas, respectivamente.

Sin pérdida de generalidad, se restringirá la presentación al caso continuo. Como consecuencia de la definición 6.5, el r-ésimo momento de X alrededor del cero es

$$\begin{aligned} E(X^r) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^r f(x, y)dydx \\ &= \int_{-\infty}^{\infty} x^r f_X(x)dx. \end{aligned} \tag{6.8}$$

De manera similar

$$E(Y^r) = \int_{-\infty}^{\infty} y^r f_Y(y)dy. \tag{6.9}$$

El r y s-ésimo *momento producto* de X y Y alrededor del origen es:

$$E(X^r Y^s) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^r y^s f(x, y)dydx, \tag{6.10}$$

y alrededor de las medias es

$$E\{(X - \mu_X)^r (Y - \mu_Y)^s\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x - \mu_X)^r (y - \mu_Y)^s f(x, y)dydx, \tag{6.11}$$

en donde r y s son enteros, no negativos. Nótese que el r-ésimo momento de X alrededor del cero se obtiene de (6.10) con $s = 0$. De manera similar, el r-ésimo momento central de X puede determinarse a partir de (6.11) con $s = 0$.

De particular importancia es el momento producto alrededor de las medias cuando $r = s = 1$. Este momento producto recibe el nombre de *covarianza de* X y Y, y se encuentra definido por

$$Cov(X, Y) = E\{(X - \mu_X)(Y - \mu_Y)\}. \tag{6.12}$$

Al igual que la varianza, que es una medida de la dispersión de una variable aleatoria, la covarianza es una medida de la variabilidad conjunta de X y de Y. De esta forma, la covarianza es una medida de asociación entre los valores de X y de Y y sus respectivas dispersiones. Si, por ejemplo, se tiene una alta probabilidad de que valores grandes de X se encuentren asociados con valores grandes de Y, la covarianza será positiva. Por otro lado, si existe una alta probabilidad de que valores grandes de X se encuentren asociados con valores pequeños de Y o viceversa, la covarianza será negativa. Se demostrará posteriormente que la covarianza es cero si X y Y son estadísticamente independientes.

Desarrollando el miembro derecho de (6.12) se tiene

$$\begin{aligned} E\{(X - \mu_X)(Y - \mu_Y)\} &= E[XY - X\mu_Y - Y\mu_X + \mu_X\mu_Y] \\ &= E(XY) - \mu_X\mu_Y; \end{aligned}$$

de esta forma

$$Cov(X, Y) = E(XY) - E(X)E(Y). \tag{6.13}$$

Si la covarianza de X y de Y se divide por el producto de las desviaciones estándar de X y de Y, el resultado es una cantidad sin dimensiones que recibe el nombre de *coeficiente de correlación* y que se denota por $\rho(X, Y)$:*

$$\rho(X, Y) = Cov(X, Y)/\sigma_X\sigma_Y. \tag{6.14}$$

Se puede demostrar que el coeficiente de correlación se encuentra contenido en el intervalo $-1 \leq \rho \leq 1$. De hecho ρ es la covarianza de dos variables aleatorias estandarizadas X' y Y' en donde $X' = (X - \mu_X)/\sigma_X$ y $Y' = (Y - \mu_Y)/\sigma_Y$. Esto significa que el coeficiente de correlación es sólo una medida estandarizada de la asociación lineal que existe entre las variables aleatorias X y Y en relación con sus dispersiones. El valor $\rho = 0$ indica la ausencia de cualquier asociación lineal, mientras que los valores -1 y $+1$ indican relaciones lineales perfectas negativa y positiva, respectivamente. En este punto es necesario señalar que debe rechazarse cualquier otra interpretación de la palabra "correlación". Después se expondrá con detalle el coeficiente de correlación cuando se estudie el análisis de regresión.

Ejemplo 6.4 Sean X y Y dos variables aleatorias con una función de densidad conjunta de probabilidad.

$$f(x, y) = \begin{cases} \dfrac{2}{3}(x + y)\exp(-x) & x > 0, \quad 0 < y < 1, \\ 0 & \text{para cualquier otro valor.} \end{cases}$$

* Se omitirá la identificación de las variables aleatorias cuando sea necesario.

Obtener la covarianza y el coeficiente de correlación de X y de Y.

Si se toman los valores esperados apropiados, se tiene

$$
\begin{aligned}
E(X) &= \frac{2}{3}\int_0^\infty\int_0^1 (x^2 + xy)\exp(-x)dydx \\
&= \frac{2}{3}\int_0^\infty (x^2 + x/2)\exp(-x)dx \\
&= \frac{2}{3}\int_0^\infty x^2\exp(-x)dx + \frac{1}{3}\int_0^\infty x\exp(-x)dx \\
&= \frac{2\Gamma(3)}{3} + \frac{\Gamma(2)}{3} \\
&= 5/3;
\end{aligned}
$$

$$
\begin{aligned}
E(X^2) &= \frac{2}{3}\int_0^\infty\int_0^1 (x^3 + x^2y)\exp(-x)dydx \\
&= \frac{2}{3}\int_0^\infty x^3\exp(-x)dx + \frac{1}{3}\int_0^\infty x^2\exp(-x)dx \\
&= \frac{2\Gamma(4)}{3} + \frac{\Gamma(3)}{3} \\
&= 14/3;
\end{aligned}
$$

$$
\begin{aligned}
E(Y) &= \frac{2}{3}\int_0^\infty\int_0^1 (xy + y^2)\exp(-x)dydx \\
&= \frac{1}{3}\int_0^\infty x\exp(-x)dx + \frac{2}{9}\int_0^\infty \exp(-x)dx \\
&= \frac{\Gamma(2)}{3} + \frac{2}{9} \\
&= 5/9;
\end{aligned}
$$

$$
\begin{aligned}
E(Y^2) &= \frac{2}{3}\int_0^\infty\int_0^1 (xy^2 + y^3)\exp(-x)dydx \\
&= \frac{2}{9}\int_0^\infty x\exp(-x)dx + \frac{1}{6}\int_0^\infty \exp(-x)dx \\
&= \frac{2\Gamma(2)}{9} + \frac{1}{6} \\
&= 7/18;
\end{aligned}
$$

$$\begin{aligned} E(XY) &= \frac{2}{3}\int_0^\infty \int_0^1 (x^2y + xy^2)\exp(-x)dydx \\ &= \frac{1}{3}\int_0^\infty x^2 \exp(-x)dx + \frac{2}{9}\int_0^\infty x\exp(-x)dx \\ &= \frac{\Gamma(3)}{3} + \frac{2\Gamma(2)}{9} \\ &= 8/9. \end{aligned}$$

Por lo tanto

$$Cov(X, Y) = E(XY) - E(X)E(Y) = 8/9 - (5/3)(5/9) = -1/27.$$

Dado que

$$Var(X) = E(X^2) - E^2(X) = 17/9$$

y

$$Var(Y) = E(Y^2) - E^2(Y) = 13/162,$$

el coeficiente de correlación es

$$\rho(X, Y) = \frac{-1/27}{\sqrt{(17/9)(13/162)}} = -0.0951.$$

6.5 Variables aleatorias estadísticamente independientes

En el capítulo dos se mencionó que dos eventos son estadísticamente independientes si su probabilidad conjunta es igual al producto de sus probabilidades marginales. En esta sección se extenderá el concepto de independencia a variables aleatorias. A fin de asegurar la consistencia de la definición debe insistirse que para variables aleatorias estadísticamente independientes, la probabilidad conjunta $P(a < X < b, c < Y < d)$ es igual al producto de las probabilidades individuales $P(a < X < b)$ y $P(c < Y < d)$. En este punto se proporciona la siguiente definición:

Definición 6.6 Sean X y Y dos variables aleatorias con una distribución conjunta. Se dice que X y Y son estadísticas independientes si y sólo si,

$$p(x, y) = p_X(x)p_Y(y) \qquad \text{si } X \text{ y } Y \text{ son discretas}$$

o bien

$$f(x, y) = f_X(x)f_Y(y) \qquad \text{si } X \text{ y } Y \text{ son continuas,}$$

para toda x y y, en donde $p(x, y)$ y $f(x, y)$ son las funciones bivariadas de probabilidad y de densidad de probabilidad, respectivamente, y en donde $p_X(x)$, $p_Y(y)$, $f_X(x)$, y $f_Y(y)$ son las funciones de probabilidad marginal o de densidad de probabilidad marginal apropiadas.

Se desprende de esta definición que si X y Y son estadísticamente independientes, la probabilidad conjunta

$$\begin{aligned}
P(a < X < b, c < Y < d) &= \int_a^b \int_c^d f(x, y)dydx \\
&= \int_a^b \int_c^d f_X(x)f_Y(y)dydx \\
&= \int_a^b f_X(x)dx \int_c^d f_Y(y)dy \\
&= P(a < X < b)P(c < Y < d).
\end{aligned}$$

Para la misma condición,

$$\begin{aligned}
E(XY) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xyf(x, y)dydx \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xyf_X(x)f_Y(y)dydx \\
&= \int_{-\infty}^{\infty} xf_X(x)dx \int_{-\infty}^{\infty} yf_Y(y)dy \\
&= E(X)E(Y).
\end{aligned}$$

Si X y Y son estadísticamente independientes, entonces $Cov(X, Y) = \rho(X, Y) = 0$. Sin embargo debe hacerse hincapié en que la proposición inversa no es cierta. Es decir, una covarianza igual a cero no es una condición suficiente para asegurar la independencia entre variables aleatorias. Debe notarse que si X y Y no son estadísticamente independientes, son estadísticamente dependientes.

Se establecerán algunos resultados útiles con base en las definiciones 6.5 y 6.6. Sean X y Y dos variables aleatorias continuas con una función de densidad conjunta de probabilidad $f(x, y)$.

El valor esperado de una función lineal de X y Y es

$$\begin{aligned}
E(aX + bY) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (ax + by)f(x, y)dydx \\
&= a \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xf(x, y)dydx + b \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} yf(x, y)dydx \\
&= a\,E(X) + b\,E(Y) \qquad (6.15)
\end{aligned}$$

para cualquier valor de las constantes a y b.

La varianza de una función lineal de X y Y es

$$\begin{aligned}
Var(aX + bY) &= E(aX + bY)^2 - E^2(aX + bY) \\
&= E(a^2X^2 + 2abXY + b^2Y^2) - [aE(X) + bE(Y)]^2
\end{aligned}$$

$$= a^2E(X^2) + 2abE(XY) + b^2E(Y^2)$$
$$- a^2E^2(X) - 2abE(X)E(Y) - b^2E^2(Y)$$
$$= a^2Var(X) + b^2Var(Y) + 2abCov(X, Y). \quad (6.16)$$

Como consecuencia de los resultados anteriores, se tiene que el valor esperado de la suma de X y Y es la suma de los correspondientes valores esperados de X y Y, y la varianza de la suma de X y Y es igual a la suma de las respectivas varianzas más la covarianza de X y Y. Además, si X y Y son estadísticamente independientes.

$$Var(aX + bY) = a^2Var(X) + b^2Var(Y). \quad (6.17)$$

La generalización de estos resultados a n variables aleatorias se hace por inducción y se establece en el siguiente teorema:

Teorema 6.1 Sean $X_1, X_2, \ldots, X_n$ n variables aleatorias con una función de densidad conjunta de probabilidad $f(x_1, x_2, \ldots, x_n)$. Entonces

$$E\left[\sum_{i=1}^{n} a_iX_i\right] = \sum_{i=1}^{n}\left[a_iE(X_i)\right]$$

$$Var\left[\sum_{i=1}^{n} a_iX_i\right] = \sum_{i=1}^{n} a_i^2Var(X_i) + \sum_{i=1}^{n}\sum_{\substack{j=1 \\ i\neq j}}^{n} a_ia_jCov(X_i, X_j)$$

para cualquier constante a_i, $i = 1, 2, \ldots, n$.

Ejemplo 6.5 Un vendedor obtiene sus ingresos mediante la venta de dos productos distintos. Por experiencia sabe que el volumen de ventas de A no tiene ninguna influencia sobre el de B. Su ingreso mensual es el 10% del volumen, en dólares, del producto A y el 15% del volumen de B. Si en promedio las ventas del producto A ascienden a $10 000 con una desviación estándar de $2 000 y las de B a $8 000 con una desviación estándar de $1 000, obténgase el valor esperado y la desviación estándar del ingreso mensual del vendedor.

Sean X y Y dos variables aleatorias que representan el volumen de ventas en dólares de los productos A y B, respectivamente. Por hipótesis:

$$E(X) = 10\,000, \quad d.e.(X) = 2\,000; \quad E(Y) = 8\,000, \quad d.e.(Y) = 1\,000.$$

De esta forma se tiene

$$E(0.1X + 0.15Y) = 0.1\,E(X) + 0.15\,E(Y) = \$2\,200,$$

y

$$Var(0.1X + 0.15Y) = 0.01\,Var(X) + 0.0225\,Var(Y) = 62\,500.$$

La desviación estándar es de $250.

6.6 Distribuciones de probabilidad condicional

Supóngase que un tanque de agua contiene dos contaminantes. Sean X y Y dos variables aleatorias que representan el nivel de estos contaminantes en una porción del tanque que a su vez se encuentra representada por una superficie rectangular. Supóngase que el nivel observado de concentración de Y es y, pero no se observa X. Si se conoce la función de densidad conjunta de probabilidad $f(x, y)$, se necesita obtener una función que proporcione la probabilidad de que el nivel de concentración de X esté contenido en un intervalo (a, b) dado el valor observado de Y. Considere la función

$$f(x, y)/f_Y(y),$$

en donde $f_Y(y)$ es la densidad marginal de Y. Si se mantiene constante a la variable aleatoria Y en el valor observado y de manera tal que $f_Y(y) > 0$, entonces $f(x, y)/f_Y(y)$ define una función no negativa de X cuya integral es 1, dado que por definición

$$\int_{-\infty}^{\infty} \frac{f(x, y)}{f_Y(y)}\, dx = \frac{1}{f_Y(y)} \int_{-\infty}^{\infty} f(x, y) dx = f_Y(y)/f_Y(y) = 1.$$

De esta forma, $f(x, y)/f_Y(y)$ es una función de densidad de probabilidad y la probabilidad de que $a < X < b$, dado que el nivel de concentración de Y es y, está dada por:

$$P(a < X < b \mid y) = \int_a^b \frac{f(x, y)}{f_Y(y)}\, dx. \tag{6.18}$$

Definición 6.7 Sean X y Y dos variables aleatorias con una función de densidad conjunta de probabilidad $f(x, y)$. La función de densidad de probabilidad condicional de la variable aleatoria X, denotada por $f(x \mid y)$, para un valor fijo y de Y, está definida por

$$f(x \mid y) = f(x, y)/f_Y(y),$$

en donde $f_Y(y)$ es la función de densidad de probabilidad de Y de manera tal que $f_Y(y) > 0$.

De manera análoga, la función de densidad de probabilidad condicional de Y para un valor fijo x de X se define como

$$f(y \mid x) = f(x, y)/f_X(x) \qquad f_X(x) > 0, \tag{6.19}$$

en donde $f_X(x)$ es la densidad marginal de X. Puede pensarse a $f(x \mid y)$ como una función que da la densidad de probabilidad a lo largo de una línea horizontal en el plano (x, y) correspondiente a un valor fijo y de Y. De manera similar, $f(y \mid x)$ es una función que da la densidad de probabilidad a lo largo de una línea vertical en el plano (x, y) correspondiente a un valor x de X.

Nótese que si la densidad condicional $f(x \mid y)$ por ejemplo, no contiene a y, entonces X es estadísticamente independiente de Y. Esto es, si X y Y son estadísticamente independientes, entonces

$$f(x, y) = f_X(x)f_Y(y)$$

y

$$\begin{aligned} f(x \mid y) &= f(x, y)/f_Y(y) \\ &= f_X(x)f_Y(y)/f_Y(y) \\ &= f_X(x). \end{aligned}$$

De manera similar, si

$$f(x, y) = f_X(x)f_Y(y),$$

entonces

$$\begin{aligned} f(y \mid x) &= f_X(x)f_Y(y)/f_X(x) \\ &= f_Y(y). \end{aligned}$$

Los valores esperados condicionales se definen de manera análoga a la señalada en la definición 6.5. Por ejemplo, los valores esperados condicionales de X puesto que $Y = y$, y de Y, ya que $X = x$, se definen como

$$E(X \mid y) = \int_{-\infty}^{\infty} xf(x \mid y)dx$$

y (6.20)

$$E(Y \mid x) = \int_{-\infty}^{\infty} yf(y \mid x)dy,$$

respectivamente. El valor esperado de X dado y es una función del punto fijo y y representa la media de X a lo largo de la línea correspondiente a y. Por simetría, el valor esperado condicional de Y dado x es una función de x y representa la media de Y a lo largo de la línea correspondiente a x. De manera similar,

$$Var(X \mid y) = E(X^2 \mid y) - E^2(X \mid y)$$

y (6.21)

$$Var(Y \mid x) = E(Y^2 \mid x) - E^2(Y \mid x),$$

en donde

$$E(X^2 \mid y) = \int_{-\infty}^{\infty} x^2f(x \mid y)dx$$

y (6.22)

$$E(Y^2 \mid x) = \int_{-\infty}^{\infty} y^2f(y \mid x)dy.$$

Ejemplo 6.6 Sean X y Y los niveles de concentración en ppm de dos contaminantes en una determinada porción de un tanque de agua. Si la función de densidad conjunta de probabilidad está dada por

$$f(x, y) = \begin{cases} (x + y)/8000 & 0 < x, y < 20, \\ 0 & \text{para cualquier otro valor,} \end{cases}$$

y si el nivel de concentración observado de Y es de 10 ppm, obtener la probabilidad de que el nivel de concentración de X sea, a lo más, 14 ppm. Obtener la media y la varianza condicional de X para $Y = 10$ ppm.

Dado que

$$f(x, y) = (x + y)/8000 \qquad 0 < x, y < 20,$$

se tiene

$$f_Y(y) = \frac{1}{8000}\int_0^{20} (x + y)dx = (y + 10)/400,$$

y la densidad de probabilidad condicional de X es

$$f(x \mid y) = (x + y)/20(y + 10),$$

la que se reduce a

$$f(x \mid Y = 10) = (x + 10)/400$$

para $Y = 10$. Por lo tanto,

$$\begin{aligned} P(X \leq 14 \mid Y = 10) &= \int_0^{14} f(x \mid Y = 10)dx \\ &= \frac{1}{400}\int_0^{14} (x + 10)dx \\ &= 0.595. \end{aligned}$$

Para la media y varianza condicional de X en $Y = 10$ se tiene

$$\begin{aligned} E(X \mid Y = 10) &= \int_0^{20} xf(x \mid Y = 10)dx \\ &= \frac{1}{400}\int_0^{20} (x^2 + 10x)dx \\ &= 11.67; \end{aligned}$$

$$E(X^2 \mid Y = 10) = \int_0^{20} x^2 f(x \mid Y = 10)dx$$

$$= \frac{1}{400}\int_0^{20} (x^3 + 10x^2)dx$$

$$= 166.67;$$

$$Var(X \mid Y = 10) = 30.56.$$

6.7 Análisis bayesiano: las distribuciones *a priori* y *a posteriori*

Se estableció en la sección 2.8 el teorema de Bayes para probabilidades condicionales de eventos discretos. En este contexto se examinará de manera breve cómo emplearlo para modificar el grado de creencia con respecto a los resultados de un fenómeno al tenerse nueva información de éste. Sin embargo, es más importante la representación que proporciona el teorema de Bayes para la distribución condicional de una variable aleatoria ya sea ésta cotinua o discreta. Tal representación es importante debido a que, como se verá en el capítulo 8, proporciona el mecanismo necesario sobre el cual se basa la inferencia bayesiana. En esta sección se examinarán los conceptos de distribución *a priori* y distribución *a posteriori* y se volverá a plantear el teorema de Bayes con estos conceptos.

Sea Y una variable aleatoria (discreta o continua) definida de manera tal que sus valores representan las posibles opciones en que puede ocurrir un fenómeno aleatorio antes de llevar a cabo un experimento. El grado de creencia del investigador con respecto a estas posibilidades se encuentra expresado por una función de probabilidad $p_Y(y)$, que recibe el nombre de *función de probabilidad a priori* de Y, si Y es discreta, o una función de densidad $f_Y(y)$, denominada *función de densidad de probabilidad a priori* de Y, si Y es continua. La especificación de la forma de $p_Y(y)$ o $f_Y(y)$ depende de la convicción del investigador con respecto a los valores de Y antes de que la información muestral se encuentre disponible. Esta convicción se puede basar en cualquier tipo de información que se encuentre disponible, incluyendo el juicio subjetivo. Sea $f(x \mid y)$ la función de densidad de probabilidad condicional de cualquier variable aleatoria X*, la cual representa evidencia muestral en función de una alternativa fija y de Y. La función $f(x \mid y)$ recibe el nombre de *función de verosimilitud* debido a que representa el grado de concordancia del resultado muestral x, dado el valor y de Y.

Cuando la información *a priori* con respecto a los valores de Y se combina con la información que proporcionó la muestra, el resultado es un conjunto de información modificada con respecto a la variable aleatoria Y. En otras palabras, la combinación de la distribución *a priori* y de la función de verosimilitud origina una distribución condicional para Y, dado el resultado muestral, que se conoce como la *distribución a posteriori* de Y. Esta combinación se hace de acuerdo con el teorema de Bayes, mismo que se replantea de la siguiente forma:

Teorema 6.2 Sea $p_Y(y)$ o $f_Y(y)$ la función de probabilidad o de densidad de probabilidad *a priori* de Y, respectivamente, y sea $f(x \mid y)$ la función de verosimilitud.

* Se supone que la variable aleatoria X es continua aunque también puede ser discreta.

Entonces la probabilidad ***a posteriori*** o función de densidad de probabilidad ***a posteriori*** de Y dada la evidencia muestral x, es

$$p(y \mid x) = \frac{f(x \mid y)p_Y(y)}{\sum_Y f(x \mid y)p_Y(y)} \qquad \text{si } Y \text{ es discreta,} \tag{6.23}$$

$$f(y \mid x) = \frac{f(x \mid y)f_Y(y)}{\int_Y f(x \mid y)f_Y(y)dy} \qquad \text{si } Y \text{ es continua.} \tag{6.24}$$

La función de probabilidad *a posteriori* $p(y \mid x)$ o la función de densidad de probabilidad *a posteriori* $f(y \mid x)$ reflejan el grado de creencia modificado del investigador con respecto a la variable aleatoria Y después de obtener información muestral. Dado que esta información se puede verificar de manera periódica, puede adoptarse fácilmente un punto de vista secuencial. En este contexto, la distribución *a posteriori* actual puede convertirse, en un futuro, en una distribución *a priori* cuando sea necesario llevar a cabo otra revisión con respecto a la variable aleatoria. La revisión periódica de las probabilidades se hace posible mediante el empleo sucesivo del teorema 6.2.

Es interesante notar que el denominador de (6.23) o (6.24) es la función de densidad de probabilidad marginal o no condicional de X; esto es,

$$f_X(x) = \sum_Y f(x \mid y)p_Y(y) \tag{6.25}$$

o

$$f_X(x) = \int_Y f(x \mid y)f_Y(y)dy, \tag{6.26}$$

dependiendo de cuando Y es discreta o continua, respectivamente. Además, el numerador de (6.23) o (6.24) es el producto de la función de verosimilitud y la función de probabilidad *a priori* y, de esta manera, es la probabilidad conjunta de X y Y expresada como

$$f(x, y) = f(x \mid y)p_Y(y) \qquad \text{si } Y \text{ es discreta,} \tag{6.27}$$

o

$$f(x, y) = f(x \mid y)f_Y(y) \qquad \text{si } Y \text{ es continua.} \tag{6.28}$$

Nótese que para (6.27) la función $f(x, y)$ es una mezcla bivariada de una variable aleatoria continua y otra discreta.

Ejemplo 6.7 Un vendedor de artículos domésticos nota que el número de personas que compran determinada marca de televisores varía aleatoriamente en el tiempo. El vendedor concluye que esta proporción es una variable aleatoria discreta que puede tomar los valores de 0.3, 0.35, 0.4 y 0.45, dependiendo de diversas consideraciones de tipo económico. Con base en información previa, les asigna las probabilidades *a priori* 0.4, 0.3, 0.2 y 0.1, respectivamente. Una muestra de tamaño $n = 15$ revela que ocho de los televisores que se venden son de la marca de interés. Si se supone que para una proporción en particular p, el número de televisores de la marca que se vende para una muestra fija n es una variable aleatoria binomial, obtener las probabilidades *a posteriori*.

Sea X la variable aleatoria que representa el número de aparatos de la marca de interés que se venden de una muestra de tamaño n. El valor $X = 8$ para $n = 15$, representa la evidencia muestral condicionada sobre una proporción en particular p de preferencia del consumidor para esta marca. Por hipótesis X es binomial y su función de verosimilitud es

$$p(x; 15 \mid p) = \frac{15!}{(15 - x)!x!} p^x(1 - p)^{15-x}, \qquad x = 0, 1, 2, \ldots, 15.$$

Si $p = 0.3$, el valor de verosimilitud de la muestra es

$$P(X = 8 \mid p = 0.3) = p(8; 15 \mid 0.3) = \frac{15!}{(15 - 8)!8!} (0.3)^8(0.7)^{15-8} = 0.0348.$$

Para los demás valores de p se tiene

$$P(X = 8 \mid p = 0.35) = 0.071,$$
$$P(X = 8 \mid p = 0.4) = 0.1181,$$
$$P(X = 8 \mid p = 0.45) = 0.1647.$$

Nótese que las dos variables aleatorias son discretas. A pesar de lo anterior, puede emplearse el teorema de Bayes (6.23) para obtener las probabilidades *a posteriori*. La tabla 6.1 proporciona los detalles computacionales. La suma de las probabilidades tanto *a priori* como *a posteriori* debe ser igual a uno, dado que cada una de éstas es una distribución de probabilidad. En la figura 6.2 se ilustran las gráficas

TABLA 6.1 Determinación de las probabilidades *a posteriori* para el ejemplo 6.7

Valores de la proporción	*Probabilidad* a priori	*Verosimilitud de la muestra*	*Probabilidad a priori* × *verosimilitud*	*Probabilidad* a posteriori
0.3	0.4	0.0348	0.01392	0.01392/0.07531 = 0.1848
0.35	0.3	0.071	0.02130	0.02130/0.07531 = 0.2828
0.4	0.2	0.1181	0.02362	0.02362/0.07531 = 0.3137
0.45	0.1	0.1647	0.01647	0.01647/0.07531 = 0.2187
Totales	1.0		0.07531	1.0000

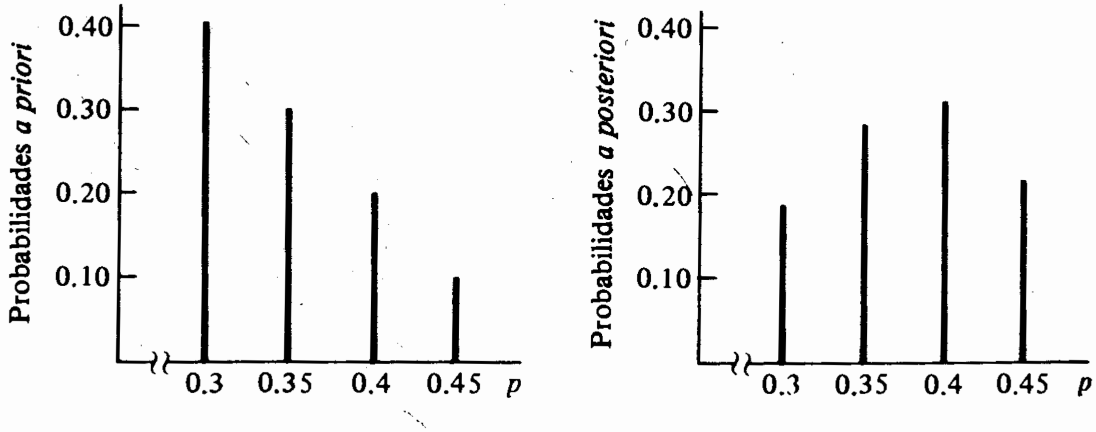

FIGURA 6.2 Probabilidades *a priori* y *a posteriori* para el ejemplo 6.7

de ambas distribuciones de probabilidad, las que muestran un desarrollo notable en las probabilidades para los cuatro valores posibles de p. También existe un desarrollo en los valores esperados de la preferencia del consumidor para esta marca. El valor esperado *a priori* es 0.35 y el valor esperado *a posteriori* es 0.3783.

Se mencionó en la sección 4.5 que la distribución binomial negativa es una alternativa adecuada del modelo de Poisson cuando la frecuencia de ocurrencia no es constante sobre el tiempo o el espacio. Por ejemplo, en las estadísticas de accidentes es poco probable que la frecuencia con que éstos se presentan entre grupos distintos sea constante e independiente sobre un lapso fijo. Lo anterior tiene como consecuencia que el punto de vista bayesiano sea una forma de análisis de estos datos mucho más apropiada.

Supóngase que todas las posibles frecuencias de ocurrencia pueden considerarse como valores de una variable aleatoria continua Λ, cuya distribución *a priori* es una distribución gama con una función de densidad dada por

$$f(\lambda; k, \theta) = \frac{1}{\Gamma(k)\theta^k} \lambda^{k-1}\exp(-\lambda/\theta), \qquad \lambda > 0. \tag{6.29}$$

Sea X una variable aleatoria que representa el número de accidentes que se observan en un grupo específico. Entonces puede argumentarse que X es una variable aleatoria de Poisson que depende de una λ específica de Λ, con una función de verosimilitud dada por

$$p(x \mid \lambda) = \lambda^x \exp(-\lambda)/x! \qquad x = 0, 1, 2, \ldots . \tag{6.30}$$

Antes de obtener la distribución *a posteriori* de Λ, se demostrará que la función de probabilidad marginal de X es la binomial negativa. Esto es, si para cada valor λ de Λ, X tiene una distribución de Poisson, entonces la distribución no condicional de X sobre todos los posibles valores de λ es la binomial negativa.

De (6.26) se desprende que la función de probabilidad marginal de X es:

$$p_X(x) = \int_0^{\infty} p(x \mid \lambda) f_\Lambda(\lambda) d\lambda. \tag{6.31}$$

Nótese que el integrando de (6.31) es la función de densidad conjunta de probabilidad de X y Λ, lo que da como resultado una mezcla bivariada de una variable aleatoria discreta con una continua.

La sustitución de (6.29) y (6.30) en (6.31) conduce a:

$$p_X(x) = \frac{1}{\Gamma(k)\theta^k x!} \int_0^{\infty} \lambda^{x+k-1} \exp\left[-\lambda \left(\frac{\theta + 1}{\theta} \right) \right] d\lambda. \tag{6.32}$$

En el integrando de (6.32) sea $u = \lambda\ [(\theta + 1)/\theta]$; de esta forma $\lambda = [\theta/(\theta + 1)]u$ y $d\lambda = [\theta/(\theta + 1)]du$. Entonces

$$\begin{aligned} p_X(x) &= \frac{1}{x!\Gamma(k)\theta^k} \int_0^{\infty} \theta/(\theta + 1)^{x+k} u^{x+k-1} \exp(-u) du \\ &= \frac{\theta/(\theta + 1)^{x+k}\Gamma(x + k)}{x!\Gamma(k)\theta^k} \\ &= \frac{\Gamma(x + k)}{x!\Gamma(k)} \left(\frac{1}{\theta + 1} \right)^k \left(\frac{\theta}{\theta + 1} \right)^x, \qquad \begin{array}{l} x = 0, 1, 2, \ldots, \\ k, \theta > 0. \end{array} \end{aligned} \tag{6.33}$$

La expresión (6.33) es idéntica a la dada por (4.35), que es la función de probabilidad de la distribución binomial negativa para $k > 0$. Nótese que en (6.33), $p = 1/(\theta + 1)$ y $1 - p = \theta/(\theta + 1)$, de forma tal que $0 < p < 1$ dado que $\theta > 0$. Además, de (4.39) la media de X es

$$E(X) = \frac{k\theta/(\theta + 1)}{1/(\theta + 1)} = k\theta = E(\Lambda).$$

De esta manera, la distribución binomial negativa es una combinación de distribuciones de Poisson donde la frecuencia aleatoria de ocurrencia tiene una distribución gama cuya media es igual a la media de Poisson. Por esta razón la distribución binomial negativa también se conoce como una distribución compuesta de Poisson.

Mediante el empleo del teorema 6.2 y, en particular, de la expresión (6.24), se puede obtener la densidad de probabilidad *a posteriori* de Λ condicionada al resultado muestral x de la siguiente forma:

$$\begin{aligned} f(\lambda \mid x) &= \frac{\lambda^{x+k-1}\exp\{-[(\theta + 1)/\theta]\lambda\}}{\Gamma(k)\theta^k x!} \Big/ \frac{\Gamma(x + k)}{x!\Gamma(k)} \left(\frac{1}{\theta + 1} \right)^k \left(\frac{\theta}{\theta + 1} \right)^x \\ &= \frac{\lambda^{x+k-1}\exp\{-[(\theta + 1)/\theta]\lambda\}}{\Gamma(k)x!\theta^k} \cdot \frac{\Gamma(k)x!(\theta + 1)^{x+k}}{\Gamma(x + k)\theta^x} \\ &= \frac{[(\theta + 1)/\theta]^{x+k}\lambda^{x+k-1}\exp\{-[(\theta + 1)/\theta]\lambda\}}{\Gamma(x + k)}, \qquad \lambda > 0. \end{aligned} \tag{6.34}$$

La comparación de (6.34) con la función de densidad de probabilidad de la distribución gama, dada por (5.45), muestra que la distribución *a posteriori* de Λ es una distribución gama con parámetros de forma $x + k$ y de escala $\theta/(\theta + 1)$. Debe notarse que si las distribuciones *a priori* y *a posteriori* pertenecen a la misma familia de distribuciones, como en el presente caso, ésta recibe el nombre de familia conjugada con respecto a la distribución de la muestra de datos. En este caso, la familia gama se conjuga con respecto a la distribución de Poisson.

Ejemplo 6.8 Supóngase que para las estadísticas de accidentes se decide asignar a la frecuencia de ocurrencia una distribución *a priori* gama con parámetro de forma dos y de escala tres. Supóngase que posteriormente se observan dos accidentes para una frecuencia en particular. Obtener la función de densidad *a posteriori* de la frecuencia, dado el resultado muestral, y compararla con la densidad *a priori*.

Sea Λ la frecuencia de ocurrencia. De (5.45) la densidad *a priori* de Λ es

$$f_{\Lambda}(\lambda; 2, 3) = \frac{1}{9}\lambda \exp(-\lambda/3), \qquad \lambda > 0.$$

Dado un resultado muestral $X = 2$, la densidad *a posteriori* de Λ que se obtiene de (6.34) es

$$f(\lambda; 4, 3/4 \mid x) = \frac{1}{6}(4/3)^4\lambda^3 \exp\left(-\frac{4}{3}\lambda\right), \qquad \lambda > 0.$$

En la figura 6.3 se proporciona una comparación entre las funciones de densidad *a priori* y *a posteriori*. De ésta es evidente que la densidad *a posteriori* se encuentra menos asimétrica que la densidad *a priori*. Nótese que la frecuencia media *a priori* es seis mientras que ésta misma *a posteriori* es tres.

En la sección 5.4 se mencionó que la distribución beta tiene un papel muy importante en la estadística bayesiana. Para ilustrar lo anterior considérese de nuevo el análisis bayesiano del parárametro de proporción de la distribución binomial.

Ejemplo 6.9 En un proceso de manufactura, el interés se centra alrededor de la proporción de artículos defectuosos. Dado que es poco probable que el proceso tenga cambios menores en un lapso determinado como distintos desarrollos, variaciones en la materia prima y otros que pueden influir en la proporción de artículos defectuosos, es razonable pensar la proporción de éstos como una variable aleatoria cuyos posibles valores se encuentran en el intervalo $(0, 1)$. Para una proporción dada de artículos defectuosos p, se sabe que el número x de éstos que se observa en una muestra aleatoria fija de n artículos es binomial. Esto es, la función de probabilidad condicional de X para n fijo, dado p, es

$$p(x; n \mid p) = \frac{n!}{(n - x)!x!} p^x(1 - p)^{n-x}, \qquad x = 0, 1, 2, \ldots n.$$

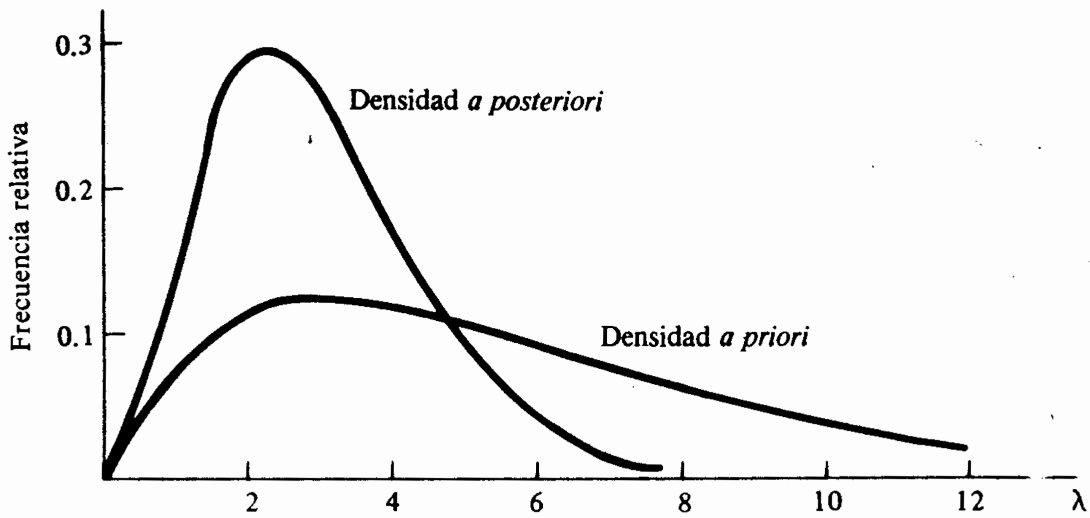

FIGURA 6.3 Densidades *a priori* y *a posteriori* para el ejemplo 6.8

Si la distribución *a priori* de la proporción de artículos defectuosos es una distribución beta con una función de densidad de probabilidad

$$f_P(p;\alpha,\beta)=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)}\,p^{\alpha-1}(1-p)^{\beta-1}\qquad 0\leq p\leq 1, \tag{6.35}$$

demostrar que la distribución *a posteriori* de la proporción de artículos defectuosos, dado el número x de éstos, también es una distribución beta.

De (6.24) la densidad de probabilidad *a posteriori* de la proporción de artículos defectuosos es:

$$f(p\mid x)=\frac{p(x;n\mid p)f_P(p;\alpha,\beta)}{\int_0^1 p(x;n\mid p)f_P(p;\alpha,\beta)dp}$$

$$=\frac{\dfrac{n!}{(n-x)!x!}\,p^x(1-p)^{n-x}\,\dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)}\,p^{\alpha-1}(1-p)^{\beta-1}}{\dfrac{n!}{(n-x)!x!}\cdot\dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)}\displaystyle\int_0^1 p^{x+\alpha-1}(1-p)^{n+\beta-x-1}dp}$$

$$=\frac{p^{x+\alpha-1}(1-p)^{n+\beta-x-1}}{\int_0^1 p^{x+\alpha-1}(1-p)^{n+\beta-x-1}dp};$$

pero de (5.33), la integral $\int_0^1 p^{x+\alpha-1}(1-p)^{n+\beta-x-1}dp = B(x+\alpha, n+\beta-x)$. Por lo tanto, la densidad *a posteriori* es:

$$f(p \mid x) = \frac{p^{x+\alpha-1}(1-p)^{n+\beta-x-1}}{B(x+\alpha, n+\beta-x)}$$

$$= \frac{\Gamma(n+\alpha+\beta)}{\Gamma(x+\alpha)\,\Gamma(n+\beta-x)}\, p^{x+\alpha-1}(1-p)^{n+\beta-x-1} \qquad 0 \leq p \leq 1, \qquad (6.36)$$

que es una densidad beta con parámetros $(x+\alpha)$ y $(n+\beta-x)$. Por lo tanto, la familia conjugada para la distribución binomial es la familia de distribuciones beta.

6.8 La distribución normal bivariada

En el capítulo cinco se estudió la distribución normal de una variable aleatoria. El concepto de distribución normal puede extenderse para incluir variables aleatorias. En particular, la distribución normal bivariada se emplea de manera extensa para describir el comportamiento probabilístico de dos variables aleatorias.

Definición 6.8 Se dice que las variables aleatorias X y Y tienen una *distribución normal bivariada* si su función de densidad conjunta de probabilidad está dada por

$$f(x, y) = \frac{1}{2\pi\,\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]\right\} \quad -\infty < x, y < \infty, \qquad (6.37)$$

en donde

$$\mu_X = E(X), \quad \mu_Y = E(Y), \quad \sigma_X^2 = Var(X), \quad \sigma_Y^2 = Var(Y),$$

y ρ es el coeficiente de correlación de X y Y, definido en la sección 6.4.

La figura 6.4 ilustra la función de densidad normal bivariada que es una superficie tridimensional con forma de campana. Cualquier corte a través de la superficie da origen a una curva de forma normal univariada, mientras que planos paralelos al plano xy interceptan la superficie en elipses que reciben el nombre de *contornos de probabilidad constante*.

Es interesante notar que, a pesar de que $\rho = 0$ es una condición necesaria de independencia, para la distribución normal bivariada también es una condición suficiente. Eso es, si $\rho = 0$, entonces

$$f(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y} \exp\left[-\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - \frac{1}{2}\left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]$$

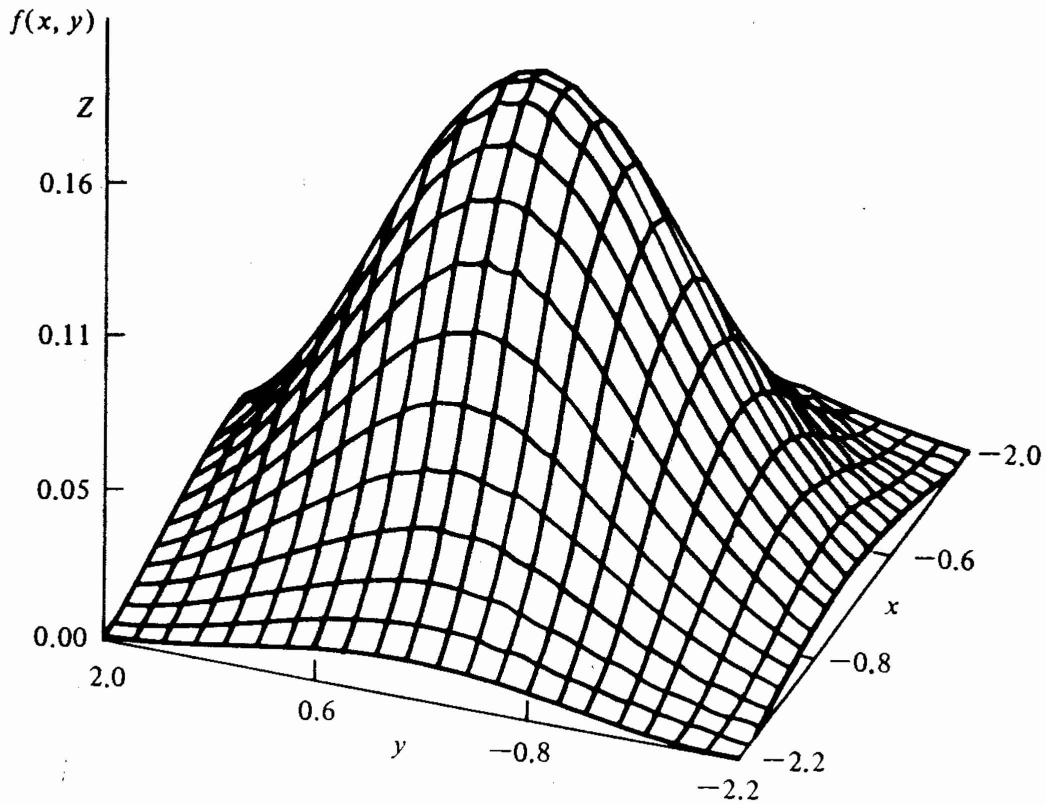

FIGURA 6.4 Densidad normal bivariada con $E(X) = E(Y) = 0$, $Var(X) = Var(Y) = 1$, y $\rho = 0$

$$= \frac{1}{\sqrt{2\pi}\,\sigma_X} \exp\left[-(x - \mu_X)^2/2\sigma_X^2 \right] \cdot \frac{1}{\sqrt{2\pi}\,\sigma_Y} \exp\left[-(y - \mu_Y)^2/2\sigma_Y^2 \right]$$

$$= f_X(x) f_Y(y),$$

en donde $f_X(x)$ y $f_Y(y)$ son las densidades normales univariadas de X y Y, respectivamente.

Se puede demostrar que, mediante el empleo de (6.37) e integrando con respecto a y, la densidad marginal de X es normal con media μ_X varianza σ_X^2. De manera similar, la densidad marginal de Y es normal con media μ_Y y varianza σ_Y^2. Por la definición 6.7, la densidad de probabilidad condicional de X dado el valor y de Y es

$$f(x \mid y) = \frac{1}{\sqrt{2\pi\,\sigma_X^2(1 - \rho^2)}}$$

$$\times \exp\left\{ -\frac{1}{2\sigma_X^2(1 - \rho^2)} \left[x - \mu_X - \frac{\rho\,\sigma_X}{\sigma_Y}(y - \mu_Y) \right]^2 \right\}. \qquad (6.38)$$

La expresión (6.38) es una función de densidad de probabilidad normal con

$$E(X \mid y) = \mu_X + \frac{\rho\sigma_X}{\sigma_Y}(y - \mu_Y) \quad \text{y} \quad Var(X \mid y) = \sigma_X^2(1 - \rho^2).$$

Se puede obtener una expresión similar para la densidad condicional de Y dado el valor x de X.

Ejemplo 6.10 Sean X y Y las desviaciones horizontal y vertical (sobre un plano), respectivamente, de un vehículo espacial tripulado con respecto al punto de aterrizaje de éste en el mar de la Tranquilidad. Supóngase que X y Y son dos variables aleatorias, independientes cada una, con una distribución normal bivariada y medias $\mu_X = \mu_Y = 0$ y varianzas iguales. ¿Cuál es la máxima desviación estándar permisible de X y Y, que cumplirá con el requisito de la NASA de tener una probabilidad de 0.99, de que el vehículo aterrice a no más de 500 ft del punto elegido, tanto en dirección vertical como horizontal?

Debido a la independencia y a la hipótesis de que $\sigma_X = \sigma_Y = \sigma$, la probabilidad conjunta es

$$
\begin{aligned}
P(-500 < X < 500,\ -500 < Y < 500) &= P(-500 < X < 500) \\
&\quad \cdot P(-500 < Y < 500) \\
&= P\left(-\frac{500}{\sigma} < Z < \frac{500}{\sigma}\right) \\
&\quad \cdot P\left(-\frac{500}{\sigma} < Z < \frac{500}{\sigma}\right) \\
&= P^2\left(-\frac{500}{\sigma} < Z < \frac{500}{\sigma}\right).
\end{aligned}
$$

Puesto que por hipótesis es

$$P^2\left(-\frac{500}{\sigma} < Z < \frac{500}{\sigma}\right) = 0.99,$$

$$P\left(-\frac{500}{\sigma} < Z < \frac{500}{\sigma}\right) = 0.99499$$

o

$$P\left(Z > \frac{500}{\sigma}\right) = 0.0025,$$

pero

$$P(Z > 2.81) = 0.0025;$$

por lo tanto $500/\sigma = 2.81$, y $\sigma_X = \sigma_Y \leqslant 177.94$ pies

Referencias

1. P. G. Hoel, *Introduction to mathematical statistics*, 4th ed., Wiley, New York, 1971.
2. R. V. Hogg and A. T. Craig, *Introduction to mathematical statistics*, 4th ed., Macmillan, New York, 1978.
3. B. W. Lindgren, *Statistical theory*, 3rd ed., Macmillan, New York, 1976.

Ejercicios

6.1. Se seleccionaron, aleatoriamente, 60 personas y se les preguntó su preferencia con respecto a tres marcas A, B y C. Éstas fueron de 27, 18 y 15 respectivamente. ¿Qué tan probable es este resultado si no existen otras marcas en el mercado y la preferencia se comparte por igual entre las tres?

6.2. Supóngase que de un proceso de producción se seleccionan, de manera aleatoria, 25 artículos. Este proceso de producción por lo general produce un 90% de artículos listos para venderse y un 7% reprocesables. ¿Cuál es la probabilidad de que 22 de los 25 artículos estén listos para venderse y que dos sean reprocesables?

6.3. Sean X y Y dos variables aleatorias continuas con una función de densidad conjunta de probabilidad dada por:

$$f(x, y) = \begin{cases} (3x - y)/5 & 1 < x < 2, \quad 1 < y < 3, \\ 0 & \text{para cualquier otro valor.} \end{cases}$$

a) Obtener la función de distribución conjunta acumulativa.
b) ¿Cuál es la probabilidad conjunta de que $X < 3/2$ y $Y < 2$?
c) Mediante el empleo de sus respuestas a la parte *a*, obtener las distribuciones acumulativas marginales de X y Y.
d) Obtener las funciones de densidad marginal de X y de Y.

6.4. Sean X y Y dos variables aleatorias continuas con una función de densidad conjunta de probabilidad dada por

$$f(x, y) = \begin{cases} x \exp[-x(y + 1)] & x, y > 0, \\ 0 & \text{para cualquier otro valor.} \end{cases}$$

a) Demostrar que $f(x, y)$ es una función de densidad conjunta de probabilidad.
b) ¿Cuál es la probabilidad conjunta de que $X < 2$ y $Y < 1$?
c) Obtener las funciones de densidad marginal de X y de Y.
d) ¿Son X y Y estadísticamente independientes?

6.5. Sean X y Y dos variables aleatorias discretas en donde los posibles valores que éstas pueden tomar son -1, 0, y 1. En la siguiente tabla se dan las probabilidades conjuntas para todos los posibles valores de X y Y.

		X		
		−1	0	1
Y	−1	1/16	3/16	1/16
	0	3/16	0	3/16
	1	1/16	3/16	1/16

a) Obtener las funciones de probabilidad marginal $p_X(x)$ y $p_Y(y)$.
b) ¿Las variables aleatorias X y Y son estadísticamente independientes?
c) Obtener $Cov(X, Y)$.

6.6. Para la función de densidad conjunta de probabilidad del ejercicio 6.3, obtener $Cov(X, Y)$ y $\rho(X, Y)$.

6.7. En función de su prioridad, un programa para computadora espera en la fila de entrada cierto tiempo, depués del cual lo ejecuta el procesador central en un lapso dado. La función de densidad conjunta para los tiempos de espera y ejecución se determina por

$$f(t_1, t_2) = \begin{cases} 2\exp\left[-\left(\dfrac{t_1}{5} + 10t_2\right)\right] & t_1, t_2 > 0, \\ 0 & \text{para cualquier otro valor.} \end{cases}$$

Dada la distribución conjunta acumulativa:

$$F(t_1, t_2) = \begin{cases} [1 - \exp(-t_1/5)][1 - \exp(-10t_2)] & t_1, t_2 > 0, \\ 0 & \text{para cualquier otro valor.} \end{cases}$$

a) Obtener la probabilidad conjunta de que el tiempo de espera no sea mayor de ocho minutos y el de ejecución no sea mayor de 12 segundos.
b) Obtener las funciones de densidad marginal y deducir que estos lapsos son variables aleatorias independientes.

6.8. Las variables aleatorias X y Y representan las proporciones de los mercados correspondientes a dos productos distintos fabricados por la misma compañía y cuya función de densidad conjunta de probabilidad está dada por

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

a) Obtener las funciones de densidad marginal de X y Y.
b) ¿Las variables aleatorias X y Y son estadísticamente independientes?
c) Si $X = 0.2$, obtener la función de densidad de probabilidad condicional de Y.

6.9. Las variables aleatorias X y Y representan el largo y ancho (en cm) de una hoja de acero. Si X y Y son independientes con funciones de densidad de probabilidad dadas por

$$f_X(x) = \begin{cases} 1, & 99 < x < 100, \\ 0 & \text{para cualquier otro valor.} \end{cases} \qquad f_Y(y) = \begin{cases} 1, & 49 < y < 50, \\ 0 & \text{para cualquier otro valor.} \end{cases}$$

úsese la definición de la varianza para obtener la varianza del área de la hoja de acero XY.

6.10. Sea X una variable aleatoria continua y Y discreta.
a) Si $f(x, y) = x^y \exp(-2x)/y!$, $x > 0$, $y = 0, 1, 2, \ldots$, obtener la función de probabilidad marginal de Y.
b) Obtener la función de probabilidad condicional de X para $Y = 2$.
c) Obtener $E(X \mid 2)$ y $Var(X \mid 2)$.

6.11. Sean X y Y dos variables aleatorias. Demostrar que $Var(aX - bY) = a^2\,Var(X) + b^2\,Var(Y) - 2ab\,Cov(X, Y)$, en donde a y b son constantes.

6.12. Sean X y Y dos variables aleatorias. Demostrar que $Cov(aX, bY) = ab\, Cov(X, Y)$, en donde a y b son constantes.

6.13. Si X y Y son dos variables aleatorias independientes $Var(X + Y) = Var(X - Y) = Var(X) + Var(Y)$. Comparar este resultado con $Var(X + Y)$ cuando $Var(X - Y)$ $Cov(X, Y) > 0$ o $Cov(X, Y) < 0$. ¿Qué puede concluirse?

6.14. Supóngase que la frecuencia Λ a la que ocurren accidentes automovilísticos en un lapso fijo es una variable aleatoria con una ditribución gama y parámetros de forma y escala igual a dos. Si para cada valor λ de Λ la distribución condicional del número de accidentes es una distribución de Poisson, obtener la función de probabilidad marginal de X y calcular las probabilidades para $X = 0, 1, 2 \ldots 10$. ¿Cómo son estas probabilidades al compararlas con las que se obtienen bajo la suposición de una frecuencia constante $\lambda = 4$?

6.15. Supóngase que la incidencia de cáncer pulmonar para un determinado número de personas adultas, sin importar sus hábitos de fumador, su edad, etc., es una variable aleatoria con distribución gama con parámetros de forma y escala iguales a dos. Para un grupo específico de personas, el número que presentarán cáncer pulmonar es una variable aleatoria de Poisson en donde el valor del parámetro de ésta depende de la incidencia de cáncer en este grupo. Obtener la probabilidad no condicional de que no más de dos personas desarrollen cáncer en este grupo.

6.16 En el ejercicio 6.15 supóngase que $x = 5$ adultos, de cierto número, desarrollarán cáncer. Obtener la densidad *a posteriori* de Λ dado x, calcular las medias y varianzas tanto *a priori* como *a posteriori* y comparar los resultados.

6.17 Supóngase que el gerente de una planta descubre que la proporción de artículos defectuosos en su proceso de producción no es constante sino que se comporta como una variable aleatoria. Sin ninguna evidencia, decide asignar una distribución beta con $\alpha = 1$ y $\beta = 24$ para la producción de artículos defectuosos.

a) Graficar la función de densidad *a priori* y obtener su media y su varianza.
b) Supóngase que el gerente toma una muestra $n = 12$ artículos y encuentra uno defectuoso. Bajo las hipótesis necesarias, obtener y graficar la función de densidad de probabilidad *a posteriori*.
c) Encontrar la media y la varianza *a posteriori* y compararlas con la media y la varianza *a priori*.
d) Hágase uso del ejercicio 5.24 para obtener la probabilidad *a posteriori* de que la proporción de artículos defectuosos sea a lo más 0.05.

6.18. Supóngase que la proporción de lanzamientos exitosos de satélites de comunicaciones es una variable aleatoria con distribución beta y parámetros $\alpha = 21$ y $\beta = 1$. Si de los últimos 12 lanzamientos uno ha fracasado, obtener la función de probabilidad *a posteriori* de la proporción de lanzamientos exitosos y calcular la probabilidad *a posteriori* para que la proporción de éstos sea mayor de 0.95. Empleé la expresión 5.44.

6.19. La función de densidad conjunta de probabilidad para la demanda mensual de dos productos es una distribución normal bivariada dada por

$$f(x, y) = \frac{1}{100\pi\sqrt{3}} \exp\left\{-\frac{2}{3}\left[\left(\frac{x - 50}{10}\right)^2 - \left(\frac{x - 50}{10}\right)\left(\frac{y - 25}{10}\right) + \left(\frac{y - 25}{10}\right)^2\right]\right\}.$$

a) ¿Cuál es el coeficiente de correlación entre X y Y?
b) ¿Cuál es la covarianza entre X y Y?
c) Obtener la función de densidad de probabilidad condicional $f(x \mid y)$.
d) Supóngase que la demanda de Y es 30. ¿Cuál es la probabilidad condicional de que X sea menor que 65?

6.20. Supóngase que el CI(X) y la calificación promedio de estudiantes no graduados de licenciatura Y son variables aleatorias que se encuentran distribuidas de manera conjunta como una distribución normal bivariada $\mu_X = 100$, $\sigma_X = 10$, $\mu_Y = 3$, $\sigma_Y = 0.3$, y $Cov(X, Y) = 2.25$.

a) Si algún estudiante posee un CI de 120, ¿cuáles son los valores de la media y la desviación estándar condicionales para Y?
b) Dado que el CI es 120, obtener la probabilidad de que Y sea mayor de 3.5.
c) Supóngase que la calificación promedio de un estudiante es 2.8. ¿Cuál es la probabilidad de que esta persona tenga un CI mayor de 115?