Variáveis Aleatórias Conjuntamente Distribuídas

Capítulo

6

6.1 FUNÇÕES CONJUNTAMENTE DISTRIBUÍDAS

Até agora, trabalhamos apenas com distribuições de probabilidade de uma única variável aleatória. Entretanto, com frequência estamos interessados em analisar probabilidades de duas ou mais variáveis aleatórias. Nesse caso, definimos, para quaisquer variáveis aleatórias X e Y, a *função distribuição de probabilidade cumulativa conjunta* de X e Y como

$$F(a, b) = P\{X \leq a, Y \leq b\} \qquad -\infty < a, b < \infty$$

A distribuição de X pode ser obtida a partir da distribuição conjunta de X e Y da seguinte maneira:

$$\begin{aligned}
F_X(a) &= P\{X \leq a\} \\
&= P\{X \leq a, Y < \infty\} \\
&= P\left(\lim_{b\to\infty} \{X \leq a, Y \leq b\}\right) \\
&= \lim_{b\to\infty} P\{X \leq a, Y \leq b\} \\
&= \lim_{b\to\infty} F(a, b) \\
&\equiv F(a, \infty)
\end{aligned}$$

Note que, nesse conjunto de igualdades, fizemos uso uma vez mais do fato de que a probabilidade é uma função contínua de um conjunto (isto é, evento). Similarmente, a função distribuição cumulativa de Y é dada por

$$\begin{aligned} F_Y(b) &= P\{Y \leq b\} \\ &= \lim_{a\to\infty} F(a,b) \\ &\equiv F(\infty, b) \end{aligned}$$

As funções distribuição F_X e F_Y são às vezes chamadas de distribuições *marginais* de X e Y.

Tudo o que se deseja saber sobre as probabilidades conjuntas de X e Y pode, em tese, ser respondido em termos de sua função distribuição conjunta. Por exemplo, suponha que queiramos calcular a probabilidade conjunta de que X seja maior que a, e Y, maior que b. Isso poderia ser feito como a seguir

$$\begin{aligned} P\{X > a, Y > b\} &= 1 - P(\{X > a, Y > b\}^c) \\ &= 1 - P(\{X > a\}^c \cup \{Y > b\}^c) \\ &= 1 - P(\{X \leq a\} \cup \{Y \leq b\}) \\ &= 1 - [P\{X \leq a\} + P\{Y \leq b\} - P\{X \leq a, Y \leq b\}] \\ &= 1 - F_X(a) - F_Y(b) + F(a,b) \end{aligned} \tag{1.1}$$

A Equação (1.1) é um caso especial da equação a seguir, cuja verificação é deixada como exercício:

$$\begin{aligned} &P\{a_1 < X \leq a_2, b_1 < Y \leq b_2\} \\ &\quad = F(a_2,b_2) + F(a_1,b_1) - F(a_1,b_2) - F(a_2,b_1) \end{aligned} \tag{1.2}$$

sempre que $a_1 < a_2, b_1 < b_2$.

No caso em que X e Y são variáveis aleatórias discretas, é conveniente definir a *função discreta de probabilidade conjunta* (ou simplesmente *função de probabilidade conjunta*) de X e Y como

$$p(x,y) = P\{X = x, Y = y\}$$

A função de probabilidade de X pode ser obtida de $p(x, y)$ por:

$$\begin{aligned} p_X(x) &= P\{X = x\} \\ &= \sum_{y:p(x,y)>0} p(x,y) \end{aligned}$$

Similarmente,

$$p_Y(y) = \sum_{x:p(x,y)>0} p(x,y)$$

Exemplo 1a

Suponha que 3 bolas sejam sorteadas de uma urna contendo 3 bolas vermelhas, 4 bolas brancas e 5 bolas azuis. Se X e Y representam, respectivamente, o número de bolas vermelhas e brancas escolhidas, então a função de probabilidade conjunta de X e $Y, p(i,j) = P\{X = i, Y = j\}$, é dada por

$$p(0,0) = \binom{5}{3} \bigg/ \binom{12}{3} = \frac{10}{220}$$
$$p(0,1) = \binom{4}{1}\binom{5}{2} \bigg/ \binom{12}{3} = \frac{40}{220}$$
$$p(0,2) = \binom{4}{2}\binom{5}{1} \bigg/ \binom{12}{3} = \frac{30}{220}$$
$$p(0,3) = \binom{4}{3} \bigg/ \binom{12}{3} = \frac{4}{220}$$
$$p(1,0) = \binom{3}{1}\binom{5}{2} \bigg/ \binom{12}{3} = \frac{30}{220}$$
$$p(1,1) = \binom{3}{1}\binom{4}{1}\binom{5}{1} \bigg/ \binom{12}{3} = \frac{60}{220}$$
$$p(1,2) = \binom{3}{1}\binom{4}{2} \bigg/ \binom{12}{3} = \frac{18}{220}$$
$$p(2,0) = \binom{3}{2}\binom{5}{1} \bigg/ \binom{12}{3} = \frac{15}{220}$$
$$p(2,1) = \binom{3}{2}\binom{4}{1} \bigg/ \binom{12}{3} = \frac{12}{220}$$
$$p(3,0) = \binom{3}{3} \bigg/ \binom{12}{3} = \frac{1}{220}$$

Essas probabilidades podem ser apresentadas mais claramente na forma de uma tabela, como na Tabela 6.1. O leitor deve observar que a função de probabilidade de X é obtida calculando-se a soma das linhas, enquanto que

Tabela 6.1 $P\{X = i, Y = j\}$

i \ j	0	1	2	3	Soma da linha = $P\{X = i\}$
0	$\frac{10}{220}$	$\frac{40}{220}$	$\frac{30}{220}$	$\frac{4}{220}$	$\frac{84}{220}$
1	$\frac{30}{220}$	$\frac{60}{220}$	$\frac{18}{220}$	0	$\frac{108}{220}$
2	$\frac{15}{220}$	$\frac{12}{220}$	0	0	$\frac{27}{220}$
3	$\frac{1}{220}$	0	0	0	$\frac{1}{220}$
Soma da coluna = $P\{Y = j\}$	$\frac{56}{220}$	$\frac{112}{220}$	$\frac{48}{220}$	$\frac{4}{220}$	

a função de probabilidade de Y é obtida calculando-se a soma das colunas. Como as funções de probabilidade individuais de X e Y aparecem na margem da tabela, elas são muitas vezes chamadas de *funções de probabilidade marginais* de X e Y, respectivamente. ■

Exemplo 1b

Suponha que 15% das famílias de certa comunidade não tenham filhos, 20% tenham 1 filho, 35% tenham 2 filhos e 30% tenham 3. Suponha também que, em cada família, cada filho tenha a mesma probabilidade (independente) de ser menino ou menina. Se uma família dessa comunidade é escolhida aleatoriamente, então B, o número de meninos, e G, o número de meninas nesta família, terão a função de probabilidade conjunta mostrada na Tabela 6.2.

As probabilidades mostradas na Tabela 6.2 são obtidas da maneira a seguir:

$$P\{B = 0, G = 0\} = P\{\text{família sem filhos}\} = 0{,}15$$

$$\begin{aligned} P\{B = 0, G = 1\} &= P\{\text{família com 1 menina e um total de 1 filho}\} \\ &= P\{\text{1 filho}\}P\{\text{1 menina}|\text{1 filho}\} = (0{,}20)\left(\frac{1}{2}\right) \end{aligned}$$

$$\begin{aligned} P\{B = 0, G = 2\} &= P\{\text{família com 2 meninas e um total de 2 filhos}\} \\ &= P\{\text{2 filhos}\}P\{\text{2 meninas}|\text{2 filhos}\} = (0{,}35)\left(\frac{1}{2}\right)^2 \end{aligned}$$

Deixamos a verificação das demais probabilidades na tabela para o leitor. ■

Dizemos que X e Y são *conjuntamente contínuas* se existir uma função $f(x, y)$, definida para todos os x e y reais, com a probabilidade de que, para todo conjunto C de pares de números reais (isto é, C é um conjunto no plano bidimensional),

$$P\{(X, Y) \in C\} = \iint_{(x,y)\in C} f(x, y)\, dx\, dy \tag{1.3}$$

A função $f(x, y)$ é chamada de *função densidade de probabilidade conjunta* de X e Y. Se A e B são quaisquer conjuntos de números reais, então, definindo $C = \{(x, y): x \in A, y \in B\}$, vemos da Equação (1.3) que

$$P\{X \in A, Y \in B\} = \int_B \int_A f(x, y)\, dx\, dy \tag{1.4}$$

Tabela 6.2 $P\{B = i, G = j\}$

i \ j	0	1	2	3	Soma da linha = $P\{B = i\}$
0	0,15	0,10	0,0875	0,0375	0,3750
1	0,10	0,175	0,1125	0	0,3875
2	0,0875	0,1125	0	0	0,2000
3	0,0375	0	0	0	0,0375
Soma da coluna = $P\{G = j\}$	0,3750	0,3875	0,2000	0,375	

Como

$$
\begin{aligned}
F(a,b) &= P\{X \in (-\infty,a], Y \in (-\infty,b]\} \\
&= \int_{-\infty}^{b} \int_{-\infty}^{a} f(x,y)\,dx\,dy
\end{aligned}
$$

vemos, calculando as derivadas, que

$$
f(a,b) = \frac{\partial^2}{\partial a\,\partial b} F(a,b)
$$

sempre que as derivadas parciais forem definidas. Outra interpretação da função densidade conjunta, obtida da Equação (1.4), é

$$
\begin{aligned}
P\{a < X < a + da, b < Y < b + db\} &= \int_{b}^{d+db} \int_{a}^{a+da} f(x,y)\,dx\,dy \\
&\approx f(a,b)\,da\,db
\end{aligned}
$$

onde da e db são pequenos e $f(x, y)$ é contínua em a, b. Com isso, $f(a, b)$ é uma medida de quão provável é a presença do vetor aleatório (X, Y) na vizinhança de (a, b).

Se X e Y são variáveis aleatórias conjuntamente contínuas, elas são individualmente contínuas, e suas funções densidade de probabilidade podem ser obtidas da seguinte maneira:

$$
\begin{aligned}
P\{X \in A\} &= P\{X \in A, Y \in (-\infty,\infty)\} \\
&= \int_A \int_{-\infty}^{\infty} f(x,y)\,dy\,dx \\
&= \int_A f_X(x)\,dx
\end{aligned}
$$

onde

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

é portanto a função densidade de probabilidade de X. Similarmente, a função densidade de probabilidade de Y é dada por

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

Exemplo 1c

A função densidade conjunta de X e Y é dada por

$$
f(x,y) = \begin{cases} 2e^{-x}e^{-2y} & 0 < x < \infty,\ 0 < y < \infty \\ 0 & \text{caso contrário} \end{cases}
$$

Calcule (a) $P\{X > 1, Y < 1\}$, (b) $P\{X < Y\}$ e (c) $P\{X < a\}$.

Solução (a)

$$\begin{aligned}
P\{X > 1, Y < 1\} &= \int_0^1 \int_1^\infty 2e^{-x}e^{-2y}\,dx\,dy \\
&= \int_0^1 2e^{-2y}\left(-e^{-x}\Big|_1^\infty\right)dy \\
&= e^{-1}\int_0^1 2e^{-2y}\,dy \\
&= e^{-1}(1 - e^{-2})
\end{aligned}$$

(b)

$$\begin{aligned}
P\{X < Y\} &= \iint\limits_{(x,y):x<y} 2e^{-x}e^{-2y}\,dx\,dy \\
&= \int_0^\infty \int_0^y 2e^{-x}e^{-2y}\,dx\,dy \\
&= \int_0^\infty 2e^{-2y}(1 - e^{-y})dy \\
&= \int_0^\infty 2e^{-2y}dy - \int_0^\infty 2e^{-3y}dy \\
&= 1 - \frac{2}{3} \\
&= \frac{1}{3}
\end{aligned}$$

(c)

$$\begin{aligned}
P\{X < a\} &= \int_0^a \int_0^\infty 2e^{-2y}e^{-x}\,dy\,dx \\
&= \int_0^a e^{-x}\,dx \\
&= 1 - e^{-a}
\end{aligned}$$

■

Exemplo 1d

Considere um círculo de raio R e suponha que um ponto em seu interior seja escolhido aleatoriamente de maneira tal que todas as regiões do círculo tenham a mesma probabilidade de conter esse ponto (em outras palavras, o ponto está uniformemente distribuído no interior do círculo). Se o centro do círculo está na origem do sistema de coordenadas, e X e Y correspondem às coordenadas do ponto escolhido (Figura 6.1), então, como (X, Y) tem a mesma probabilidade de estar na vizinhança de qualquer ponto no círculo, a função densidade conjunta de X e Y é dada por

$$f(x,y) = \begin{cases} c \text{ se } x^2 + y^2 \le R^2 \\ 0 \text{ se } x^2 + y^2 > R^2 \end{cases}$$

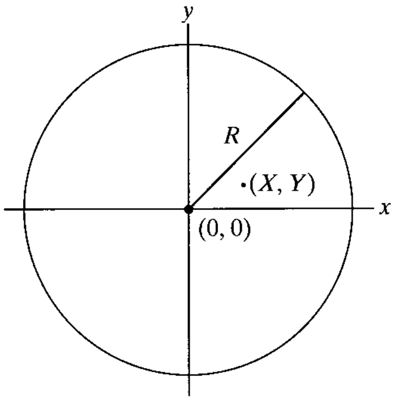

Figura 6.1 Distribuição de probabilidade conjunta.

para algum valor de c.

(a) Determine c.
(b) Determine as funções densidade marginais de X e Y.
(c) Calcule a probabilidade de que D, a distância da origem ao ponto selecionado, seja menor ou igual a a.
(d) Determine $E[D]$.

Solução (a) Como

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dy\,dx = 1$$

tem-se que

$$c \iint\limits_{x^2+y^2 \leq R^2} dy\,dx = 1$$

Podemos calcular $\iint_{x^2+y^2\leq R^2} dy\,dx$ usando coordenadas polares ou, mais simplesmente, notando que essa função representa a área do círculo, que é igual a πR^2. Com isso,

$$c = \frac{1}{\pi R^2}$$

(b)

$$\begin{aligned}
f_X(x) &= \int_{-\infty}^{\infty} f(x,y)\,dy \\
&= \frac{1}{\pi R^2}\int_{x^2+y^2\leq R^2} dy \\
&= \frac{1}{\pi R^2}\int_{-c}^{c} dy, \qquad \text{onde } c = \sqrt{R^2 - x^2} \\
&= \frac{2}{\pi R^2}\sqrt{R^2 - x^2} \quad x^2 \leq R^2
\end{aligned}$$

e essa expressão é igual a 0 quando $x^2 > R^2$. Por simetria, a densidade marginal de Y é dada por

$$\begin{aligned} f_Y(y) &= \frac{2}{\pi R^2}\sqrt{R^2 - y^2} \quad & y^2 \leq R^2 \\ &= 0 & y^2 > R^2 \end{aligned}$$

(c) A função distribuição de $D = \sqrt{X^2 + Y^2}$, a distância a partir da origem, é obtida como a seguir: para $0 \leq a \leq R$,

$$\begin{aligned} F_D(a) &= P\{\sqrt{X^2 + Y^2} \leq a\} \\ &= P\{X^2 + Y^2 \leq a^2\} \\ &= \iint\limits_{x^2+y^2 \leq a^2} f(x,y)\, dy\, dx \\ &= \frac{1}{\pi R^2} \iint\limits_{x^2+y^2 \leq a^2} dy\, dx \\ &= \frac{\pi a^2}{\pi R^2} \\ &= \frac{a^2}{R^2} \end{aligned}$$

onde usamos o fato de que $\iint_{x^2+y^2 \leq a^2} dy\, dx$ representa a área de um círculo de raio a, que é igual a πa^2.

(d) Da letra (c), a função densidade de D é

$$f_D(a) = \frac{2a}{R^2} \quad 0 \leq a \leq R$$

Portanto,

$$E[D] = \frac{2}{R^2}\int_0^R a^2 da = \frac{2R}{3}$$

■

Exemplo 1e

A função densidade de X e Y é dada por

$$f(x,y) = \begin{cases} e^{-(x+y)} & 0 < x < \infty,\ 0 < y < \infty \\ 0 & \text{caso contrário} \end{cases}$$

Determine a função densidade da variável aleatória X/Y.

Solução Começamos calculando a função distribuição de X/Y. Para $a > 0$.

$$\begin{aligned}
F_{X/Y}(a) &= P\left\{\frac{X}{Y} \leq a\right\} \\
&= \iint\limits_{x/y \leq a} e^{-(x+y)}\,dx\,dy \\
&= \int_0^{\infty}\int_0^{ay} e^{-(x+y)}\,dx\,dy \\
&= \int_0^{\infty} (1 - e^{-ay})e^{-y}dy \\
&= \left\{-e^{-y} + \frac{e^{-(a+1)y}}{a+1}\right\}\Bigg|_0^{\infty} \\
&= 1 - \frac{1}{a+1}
\end{aligned}$$

Se calcularmos a derivada dessa função, mostramos que a função densidade de X/Y é $f_{X/Y}(a) = 1/(a+1)^2, 0 < a < \infty$. ■

Também podemos definir funções distribuição de probabilidade conjunta para n variáveis da mesma maneira como fizemos para $n = 2$. Por exemplo, a função distribuição de probabilidade conjunta $F(a_1, a_2, \ldots, a_n)$ das n variáveis aleatórias $X_1, X_2, \ldots, X_n$ é definida como

$$F(a_1, a_2, \ldots, a_n) = P\{X_1 \leq a_1, X_2 \leq a_2, \ldots, X_n \leq a_n\}$$

Além disso, as n variáveis aleatórias são chamadas de conjuntamente contínuas se existir uma função $f(x_1, x_2, \ldots, x_n)$, chamada de *função densidade de probabilidade conjunta*, tal que, para qualquer conjunto C no n-espaço,

$$P\{(X_1, X_2, \ldots, X_n) \in C\} = \iint\limits_{(x_1,\ldots,x_n)\in C}\cdots\int f(x_1, \ldots, x_n)dx_1dx_2\cdots dx_n$$

Em particular, para quaisquer n conjuntos de números reais $A_1, A_2, \ldots, A_n$.

$$\begin{aligned}
&P\{X_1 \in A_1, X_2, \in A_2, \ldots, X_n \in A_n\} \\
&= \int_{A_n}\int_{A_{n-1}}\cdots\int_{A_1} f(x_1, \ldots, x_n)\,dx_1dx_2\cdots dx_n
\end{aligned}$$

Exemplo 1f A distribuição multinomial

Uma das mais importantes distribuições conjuntas é a distribuição multinomial, que surge quando uma sequência de n experimentos independentes e idênticos é realizada. Suponha que cada experimento possa levar a qualquer um de r

resultados possíveis, com respectivas probabilidades $p_1, p_2, \ldots, p_r$, $\sum_{i=1}^{r} p_i = 1$. Se X_i representa o número de experimentos que levam ao resultado i, então

$$P\{X_1 = n_1, X_2 = n_2, \ldots, X_r = n_r\} = \frac{n!}{n_1!n_2!\cdots n_r!} p_1^{n_1} p_2^{n_2} \cdots p_r^{n_r} \tag{1.5}$$

sempre que $\sum_{i=1}^{r} n_i = n$.

A Equação (1.5) é verificada notando-se que a sequência de resultados dos n experimentos que leva à ocorrência do resultado i um total de n_i vezes para $i = 1, 2, \ldots, r$ tem, pela hipótese de independência dos experimentos, probabilidade de ocorrência $p_1^{n_1} p_2^{n_2} \ldots p_r^{n_r}$. Como existem $n!/(n_1!n_2!\ldots n_r!)$ sequências de resultados assim (existem $n!/n_1!\ldots n_r!$ diferentes permutações de n coisas das quais n_1 são do mesmo tipo, n_2 são do mesmo tipo,..., n_r são do mesmo tipo), estabelece-se a Equação (1.5). A distribuição conjunta cuja função discreta de probabilidade conjunta é especificada pela Equação (1.5) é chamada de *distribuição multinomial*. Note que, quando $r = 2$, a distribuição multinomial reduz-se para a distribuição binomial.

Note também que qualquer soma de um conjunto fixo de X_i's tem distribuição binomial. Isto é, se $N \subset \{1, 2, \ldots, r\}$, então $\sum_{i \in N} X_i$ é uma variável aleatória binomial com parâmetros n e $p = \sum_{i \in N} p_i$. Isso é obtido porque $\sum_{i \in N} X_i$ representa o número de n experimentos cujo resultado está em N, e cada experimento leva independentemente a um resultado como esse com probabilidade $\sum_{i \in N} p_i$.

Como uma aplicação da distribuição multinomial, suponha que um dado honesto seja rolado 9 vezes. A probabilidade de que o 1 apareça três vezes, o 2 e o 3 apareçam duas vezes cada, e o 4 e o 5 apareçam 1 vez cada, e o 6 não apareça nenhuma vez é dada por

$$\frac{9!}{3!2!2!1!1!0!}\left(\frac{1}{6}\right)^3\left(\frac{1}{6}\right)^2\left(\frac{1}{6}\right)^2\left(\frac{1}{6}\right)^1\left(\frac{1}{6}\right)^1\left(\frac{1}{6}\right)^0 = \frac{9!}{3!2!2!}\left(\frac{1}{6}\right)^9 \qquad \square$$

6.2 VARIÁVEIS ALEATÓRIAS INDEPENDENTES

As variáveis aleatórias X e Y são independentes se, para quaisquer dois conjuntos de números reais A e B,

$$P\{X \in A, Y \in B\} = P\{X \in A\}P\{Y \in B\} \tag{2.1}$$

Em outras palavras, X e Y são independentes se, para todo A e B, os eventos $E_A = \{X \in A\}$ e $F_B = \{Y \in B\}$ forem independentes.

Pode-se mostrar, usando-se os três axiomas da probabilidade, que a Equação (2.1) é obtida se e somente se, para todo a, b,

$$P\{X \leq a, Y \leq b\} = P\{X \leq a\}P\{Y \leq b\}$$

Portanto, em termos da função distribuição conjunta F de X e Y, X e Y são independentes se

$$F(a, b) = F_X(a)F_Y(b) \quad \text{para todo } a, b$$

Quando X e Y são variáveis aleatórias discretas, a condição de independência (2.1) é equivalente a

$$p(x, y) = p_X(x)p_Y(y) \quad \text{para todo } x, y \tag{2.2}$$

Tem-se essa equivalência porque, se a Equação (2.1) é satisfeita, então obtemos a Equação (2.2) fazendo com que A e B sejam, respectivamente, os conjuntos unitários $A = \{x\}$ e $B = \{y\}$. Além disso, se a Equação (2.2) é válida, então, para quaisquer conjuntos A, B,

$$\begin{aligned} P\{X \in A, Y \in B\} &= \sum_{y \in B} \sum_{x \in A} p(x, y) \\ &= \sum_{y \in B} \sum_{x \in A} p_X(x)p_Y(y) \\ &= \sum_{y \in B} p_Y(y) \sum_{x \in A} p_X(x) \\ &= P\{Y \in B\}P\{X \in A\} \end{aligned}$$

e a Equação (2.1) é estabelecida.

No caso conjuntamente contínuo, a condição de independência é equivalente a

$$f(x, y) = f_X(x)f_Y(y) \quad \text{para todo } x, y$$

Assim, de maneira informal, X e Y são independentes se o conhecimento do valor de um não mudar a distribuição do outro. Variáveis aleatórias que não são independentes são chamadas de *dependentes*.

Exemplo 2a

Suponha que $n + m$ tentativas independentes com probabilidade comum de sucesso p sejam realizadas. Se X é o número de sucessos nas primeiras n tentativas e Y é o número de sucessos nas m tentativas finais, então X e Y são independentes, já que o conhecimento do número de sucessos nas primeiras tentativas não afeta a distribuição do número de sucessos nas m tentativas finais (pela hipótese de tentativas independentes). De fato, para x e y inteiros,

$$\begin{aligned} P\{X = x, Y = y\} &= \binom{n}{x} p^x (1 - p)^{n-x} \binom{m}{y} p^y (1 - p)^{m-y} \qquad \begin{matrix} 0 \le x \le n, \\ 0 \le y \le m \end{matrix} \\ &= P\{X = x\}P\{Y = y\} \end{aligned}$$

Por outro lado, X e Z são dependentes, onde Z é o número total de sucessos em $n + m$ tentativas. (Por quê?) ■

Exemplo 2b

Suponha que o número de pessoas que entram em uma agência de correio em certo dia seja uma variável aleatória de Poisson com parâmetro λ. Mostre que, se cada pessoa que entra na agência de correio for homem com probabilidade p e mulher com probabilidade $1 - p$, então pode-se representar o número de homens e mulheres entrando na agência por variáveis aleatórias de Poisson com respectivos parâmetros λp e $\lambda(1 - p)$.

Solução Suponha que X e Y representem, respectivamente, o número de homens e mulheres que entram na agência de correios. Vamos demonstrar a independência de X e Y estabelecendo a Equação (2.2). Para obtermos uma expressão para $P\{X = i, Y = j\}$, condicionamos em $X + Y$ da maneira a seguir:

$$P\{X = i, Y = j\} = P\{X = i, Y = j | X + Y = i + j\}P\{X + Y = i + j\} + P\{X = i, Y = j | X + Y \neq i + j\}P\{X + Y \neq i + j\}$$

[Note que essa equação é meramente um caso especial da fórmula $P(E) = P(E|F) P(F) + P(E|F^c)P(F^c)$].

Como $P\{X = i, Y = j | X + Y \neq i + j\}$ é claramente 0, obtemos

$$P\{X = i, Y = j\} = P\{X = i, Y = j | X + Y = i + j\}P\{X + Y = i + j\} \tag{2.3}$$

Agora, como $X + Y$ é o número total de pessoas que entram na agência de correios, tem-se, por hipótese, que

$$P\{X + Y = i + j\} = e^{-\lambda}\frac{\lambda^{i+j}}{(i + j)!} \tag{2.4}$$

Além disso, dado que $i + j$ pessoas entrem na agência de correios, como cada pessoa que entra tem probabilidade p de ser homem, tem-se que a probabilidade de que exatamente i dessas pessoas sejam homens (e portanto j sejam mulheres) é somente a probabilidade binomial $\binom{i + j}{i} p^i(1 - p)^j$ Isto é,

$$P\{X = i, Y = j | X + Y = i + j\} = \binom{i + j}{i} p^i(1 - p)^j \tag{2.5}$$

A substituição das Equações (2.4) e (2.5) na Equação (2.3) resulta em

$$\begin{aligned} P\{X = i, Y = j\} &= \binom{i + j}{i} p^i(1 - p)^j e^{-\lambda}\frac{\lambda^{i+j}}{(i + j)!} \\ &= e^{-\lambda}\frac{(\lambda p)^i}{i!j!}[\lambda(1 - p)]^j \\ &= \frac{e^{-\lambda p}(\lambda p)^i}{i!} e^{-\lambda(1-p)}\frac{[\lambda(1 - p)]^j}{j!} \end{aligned} \tag{2.6}$$

Assim,

$$P\{X = i\} = e^{-\lambda p}\frac{(\lambda p)^i}{i!}\sum_j e^{-\lambda(1-p)}\frac{[\lambda(1 - p)]^j}{j!} = e^{-\lambda p}\frac{(\lambda p)^i}{i!} \tag{2.7}$$

e similarmente

$$P\{Y = j\} = e^{-\lambda(1-p)} \frac{[\lambda(1 - p)]^j}{j!} \tag{2.8}$$

As Equações (2.6), (2.7) e (2.8) estabelecem o resultado desejado. ■

Exemplo 2c

Um homem e uma mulher decidem se encontrar em certo lugar. Se cada um deles chega independentemente em um tempo uniformemente distribuído entre 12:00 e 13:00, determine a probabilidade de que o primeiro a chegar tenha que esperar mais de 10 minutos.

Solução Se X e Y representam, respectivamente, o tempo após o meio-dia em que chegam o homem e a mulher, então X e Y são variáveis aleatórias independentes, cada uma uniformemente distribuída no intervalo $(0, 60)$. A probabilidade desejada, $P\{X + 10 < Y\} + P\{Y + 10 < X\}$, que, por simetria, é igual a $2P\{X + 10 < Y\}$, é obtida da seguinte maneira:

$$\begin{aligned}
2P\{X + 10 < Y\} &= 2 \iint\limits_{x+10<y} f(x,y)\,dx\,dy \\
&= 2 \iint\limits_{x+10<y} f_X(x) f_Y(y)\,dx\,dy \\
&= 2 \int_{10}^{60} \int_{0}^{y-10} \left(\frac{1}{60}\right)^2 dx\,dy \\
&= \frac{2}{(60)^2} \int_{10}^{60} (y - 10)\,dy \\
&= \frac{25}{36}
\end{aligned}$$

■

Nosso próximo exemplo apresenta o mais antigo problema relacionado a probabilidades geométricas. Ele foi analisado e resolvido pela primeira vez por Buffon, um naturalista francês do século dezoito. Por esse motivo, é usualmente chamado de *problema da agulha de Buffon*.

Exemplo 2d Problema da agulha de Buffon

Uma tábua é graduada com linhas paralelas com espaçamento D entre si. Uma agulha de comprimento L, onde $L \leq D$, é jogada aleatoriamente sobre a mesa. Qual é a probabilidade de que a agulha intercepte uma das linhas (sendo a outra possibilidade a de que a agulha fique completamente contida no espaço existente entre as linhas)?

Solução Vamos determinar a posição da agulha especificando (1) a distância X do ponto central da agulha até a linha paralela mais próxima e (2) o ângulo θ entre a agulha e a linha de comprimento X indicada na Figura 6.2. A agulha interceptará uma linha se a hipotenusa do triângulo reto na Figura 6.2 for menor que $L/2$ – isto é, se

$$\frac{X}{\cos\theta} < \frac{L}{2} \quad \text{ou} \quad X < \frac{L}{2}\cos\theta$$

Como X varia entre 0 e $D/2$ e θ varia entre 0 e $\pi/2$, é razoável supor que essas variáveis aleatórias sejam independentes e uniformemente distribuídas ao longo de seus respectivos intervalos. Com isso,

$$\begin{aligned}
P\left\{X < \frac{L}{2}\cos\theta\right\} &= \iint\limits_{x<L/2\cos y} f_X(x)f_\theta(y)\,dx\,dy \\
&= \frac{4}{\pi D}\int_0^{\pi/2}\int_0^{L/2\cos y} dx\,dy \\
&= \frac{4}{\pi D}\int_0^{\pi/2}\frac{L}{2}\cos y\,dy \\
&= \frac{2L}{\pi D}
\end{aligned}$$

■

*Exemplo 2e Caracterização da distribuição normal

Suponha que X e Y representem as distâncias de erro horizontal e vertical quando uma bala é disparada contra um alvo, e suponha que

1. X e Y sejam variáveis aleatórias independentes contínuas com funções densidade deriváveis.
2. A densidade conjunta $f(x, y) = f_X(x)f_Y(y)$ de X e Y depende de (x, y) somente através de $x^2 + y^2$.

De maneira informal, a hipótese 2 diz que a probabilidade da bala acertar qualquer ponto no plano x-y depende somente da distância do ponto até o alvo e não de seu ângulo de orientação. Uma maneira equivalente de enunciar essa hipótese é dizer que a função densidade conjunta é invariante com a rotação.

É bastante interessante o fato de as hipóteses 1 e 2 implicarem X e Y como sendo variáveis aleatórias normalmente distribuídas. Para demonstrar isso, note primeiro que as hipóteses levam à relação

$$f(x, y) = f_X(x)f_Y(y) = g(x^2 + y^2) \tag{2.9}$$

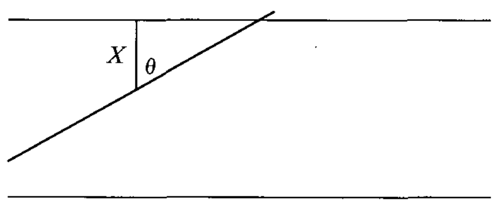

Figura 6.2

para alguma função g. Derivando-se a Equação (2.9) em relação a x, obtém-se

$$f'_X(x)f_Y(y) = 2xg'(x^2 + y^2) \tag{2.10}$$

Dividir a Equação (2.10) pela Equação (2.9) resulta em

$$\frac{f'_X(x)}{f_X(x)} = \frac{2xg'(x^2 + y^2)}{g(x^2 + y^2)}$$

ou

$$\frac{f'_X(x)}{2xf_X(x)} = \frac{g'(x^2 + y^2)}{g(x^2 + y^2)} \tag{2.11}$$

Como o valor do lado esquerdo da Equação (2.11) depende somente de x, enquanto o valor do lado direito depende de $x^2 + y^2$, tem-se como resultado que o lado esquerdo deve ser o mesmo para todo x. Para ver isso, considere quaisquer x_1, x_2 e escolha y_1, y_2 de forma que $x_1^2 + y_1^2 = x_2^2 + y_2^2$. Então, da Equação (2.11), obtemos

$$\frac{f'_X(x_1)}{2x_1f_X(x_1)} = \frac{g'(x_1^2 + y_1^2)}{g(x_1^2 + y_1^2)} = \frac{g'(x_2^2 + y_2^2)}{g(x_2^2 + y_2^2)} = \frac{f'_X(x_2)}{2x_2f_X(x_2)}$$

Portanto,

$$\frac{f'_X(x)}{xf_X(x)} = c \quad \text{ou} \quad \frac{d}{dx}(\log f_X(x)) = cx$$

o que implica, com a integração de ambos os lados, que

$$\log f_X(x) = a + \frac{cx^2}{2} \quad \text{ou} \quad f_X(x) = ke^{cx^2/2}$$

Como $\int_{-\infty}^{\infty} f_X(x)\,dx = 1$, então c é necessariamente negativo. Com isso, podemos escrever $c = -1/\sigma^2$. Assim,

$$f_X(x) = ke^{-x^2/2\sigma^2}$$

Isto é, X é uma variável aleatória normal com parâmetros $\mu = 0$ e σ^2. Um argumento parecido pode ser aplicado a $f_Y(y)$ para mostrar que

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\,\overline{\sigma}} e^{-y^2/2\overline{\sigma}^2}$$

Além disso, resulta da hipótese 2 que $\sigma^2 = \overline{\sigma}^2$ e que X e Y são portanto variáveis aleatórias normais identicamente distribuídas com parâmetros $\mu = 0$ e σ^2. ■

Uma condição necessária e suficiente para que as variáveis aleatórias X e Y sejam independentes é a de que sua função densidade de probabilidade conjunta (ou função de probabilidade conjunta no caso discreto) $f(x, y)$ possa ser fatorada em dois termos, um dependendo somente de x e o outro dependendo somente de y.

Proposição 2.1 As variáveis contínuas (discretas) X e Y são independentes se e somente se sua função densidade de probabilidade conjunta (ou discreta de probabilidade conjunta) puder ser escrita como

$$f_{X,Y}(x,y) = h(x)g(x) \qquad -\infty < x < \infty, -\infty < y < \infty$$

Demonstração Vamos demonstrar o caso contínuo. Primeiramente, note que a independência implica o fato da densidade conjunta ser o produto das densidades marginais de X e Y. Com isso, a fatoração anterior é verdadeira quando as variáveis aleatórias são independentes. Agora, suponha que

$$f_{X,Y}(x,y) = h(x)g(y)$$

Então,

$$\begin{aligned} 1 &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx\,dy \\ &= \int_{-\infty}^{\infty} h(x)\,dx \int_{-\infty}^{\infty} g(y)\,dy \\ &= C_1C_2 \end{aligned}$$

onde $C_1 = \int_{-\infty}^{\infty} h(x)\,dx$ e $C_2 = \int_{-\infty}^{\infty} g(y)\,dy$. Também,

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy = C_2h(x) \\ f_Y(y) &= \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx = C_1g(y) \end{aligned}$$

Como $C_1C_2 = 1$, obtemos

$$f_{X,Y}(x,y) = f_X(x)f_Y(y)$$

e a demonstração está completa. □

Exemplo 2f

Se a função densidade conjunta de X e Y é

$$f(x,y) = 6e^{-2x}e^{-3y} \text{ na região } 0 < x < \infty, 0 < y < \infty$$

e é igual a 0 fora dessa região, são as variáveis aleatórias independentes? E se a função densidade conjunta é

$$f(x,y) = 24xy \text{ na região } 0 < x < 1, 0 < y < 1, 0 < x + y < 1$$

e é igual a 0 caso contrário?

Solução No primeiro caso, a função densidade conjunta pode ser fatorada, e portanto as variáveis aleatórias são independentes (uma é exponencial com taxa 2 e a outra é exponencial com taxa 3). No segundo caso, como a região em que a densidade conjunta é diferente de zero não pode ser escrita na forma $x \in A, y \in B$, a densidade conjunta não pode ser fatorada. Com isso, as variáveis aleatórias não são independentes. Isso pode ser visto claramente se fizermos

$$I(x,y) = \begin{cases} 1 & \text{se } 0 < x < 1,\ 0 < y < 1,\ 0 < x + y < 1 \\ 0 & \text{caso contrário} \end{cases}$$

e escrevermos

$$f(x,y) = 24xy\, I(x,y)$$

o que claramente não pode ser fatorado em um termo que depende somente de x e em outro que depende somente de y. ■

O conceito de independência pode, naturalmente, ser definido para mais de duas variáveis aleatórias. Em geral, as n variáveis aleatórias $X_1, X_2,\ldots, X_n$ são independentes se, para todos os conjuntos de números reais $A_1, A_2,\ldots, A_n$,

$$P\{X_1 \in A_1, X_2 \in A_2, \ldots, X_n \in A_n\} = \prod_{i=1}^{n} P\{X_i \in A_i\}$$

Como antes, pode-se mostrar que essa condição é equivalente a

$$P\{X_1 \le a_1, X_2 \le a_2,\ldots, X_n \le a_n\} = \prod_{i=1}^{n} P\{X_i \le a_i\} \text{ para todo } a_1, a_2,\ldots, a_n$$

Finalmente, dizemos que uma coleção infinita de variáveis aleatórias é independente se cada uma de suas subcoleções finitas for independente.

Exemplo 2g

A maioria dos computadores é capaz de gerar, ou *simular*, o valor de uma variável aleatória uniforme no intervalo (0, 1) por meio de uma sub-rotina residente que (com certa aproximação) produz tais "números aleatórios". Como resultado, é muito fácil para um computador simular uma variável aleatória indicadora (isto é, de Bernoulli). Suponha que I seja uma variável indicadora tal que

$$P\{I = 1\} = p = 1 - P\{I = 0\}$$

O computador pode simular I sorteando um número aleatório U no intervalo (0, 1) e então fazendo

$$I = \begin{matrix} 1 \text{ se } U < p \\ 0 \text{ se } U \ge p \end{matrix}$$

Suponha que estejamos interessados em que o computador selecione $k, k \le n$, dos números $1, 2,\ldots, n$ de forma tal que cada um dos subconjuntos $\binom{n}{k}$ de tamanho k tenha a mesma probabilidade de ser escolhido. Agora, apresentamos um método que permitirá ao computador resolver essa tarefa. Para gerar um subconjunto como esse, vamos primeiro simular, em sequência, n variáveis indicadoras $I_1, I_2,..., I_n$, das quais exatamente k são iguais a 1. Os valores de i para os quais $I_i = 1$ vão então constituir o subconjunto desejado.

Para gerar as variáveis aleatórias $I_1,..., I_n$, comece simulando n variáveis aleatórias uniformes independentes $U_1, U_2,..., U_n$. Definimos agora

$$I_1 = \begin{cases} 1 & \text{se } U_1 < \dfrac{k}{n} \\ 0 & \text{caso contrário} \end{cases}$$

e então, assim que $I_1,..., I_i$ forem determinados, ajustamos recursivamente

$$I_{i+1} = \begin{cases} 1 & \text{se } U_{i+1} < \dfrac{k - (I_1 + \cdots + I_i)}{n - i} \\ 0 & \text{caso contrário} \end{cases}$$

Colocando em palavras, na $(i + 1)$-ésima etapa fazemos I_{i+1} igual a 1 (e então colocamos $i + 1$ no subconjunto desejado) com probabilidade igual ao número restante de vagas no subconjunto $\left(\text{isto é}, k - \sum_{j=1}^{i} I_j\right)$, dividido pelo número de possibilidades restantes (isto é, $n - i$). Com isso, a distribuição de probabilidade conjunta de $I_1, I_2,..., I_n$ é determinada a partir de

$$P\{I_1 = 1\} = \frac{k}{n}$$

$$P\{I_{i+1} = 1 | I_1, \ldots, I_i\} = \frac{k - \sum_{j=1}^{i} I_j}{n - i} \qquad 1 < i < n$$

A demonstração de que a última fórmula resulta em que todos subconjuntos de tamanho k tenham a mesma probabilidade de serem escolhidos é feita por indução em $k + n$. Ela é imediata quando $k + n = 2$ (isto é, quando $k = 1, n = 1$), então suponha que ela seja verdade sempre que $k + n \leq l$. Agora, suponha que $k + n = l + 1$ e considere qualquer subconjunto de tamanho k – digamos, $i_1 \leq i_2 \leq \cdots \leq i_k$ – e considere os dois casos a seguir.

Caso 1: $i_1 = 1$

$$\begin{aligned} &P\{I_1 = I_{i_2} = \cdots = I_{i_k} = 1, I_j = 0 \text{ caso contrário}\} \\ &\quad = P\{I_1 = 1\}P\{I_{i_2} = \cdots = I_{i_k} = 1, I_j = 0 \text{ caso contrário} | I_1 = 1\} \end{aligned}$$

Agora, dado que $I_1 = 1$, os elementos restantes do subconjunto são escolhidos como se um subconjunto de tamanho $k - 1$ tivesse que ser escolhido a partir dos $n - 1$ elementos $2, 3,..., n$. Portanto, pela hipótese de indução, a probabilidade condicional de que isso resulte na seleção de um dado subconjunto de tamanho $k - 1$ é igual a $1/\binom{n-1}{k-1}$. Com isso,

$$P\{I_1 = I_{i_2} = \cdots = I_{i_k} = 1, I_j = 0 \text{ caso contrário}\}$$
$$= \frac{k}{n} \frac{1}{\binom{n-1}{k-1}} = \frac{1}{\binom{n}{k}}$$

Caso 2: $i_1 \neq 1$

$$P\{I_{i_1} = I_{i_2} = \cdots = I_{i_k} = 1, I_j = 0 \text{ caso contrário}\}$$
$$= P\{I_{i_1} = \cdots = I_{i_k} = 1, I_j = 0 \text{ caso contrário} | I_1 = 0\} P\{I_1 = 0\}$$
$$= \frac{1}{\binom{n-1}{k}} \left(1 - \frac{k}{n}\right) = \frac{1}{\binom{n}{k}}$$

onde a hipótese de indução foi usada para avaliar a probabilidade condicional anterior.

Assim, em todos os casos, a probabilidade de que um dado subconjunto de tamanho k seja o subconjunto escolhido é igual a $1/\binom{n}{k}$. ■

Observação: O método anterior para gerar um subconjunto aleatório requer muito pouca memória. Um algoritmo mais rápido que requer um pouco mais de memória é apresentado na Seção 10.1 (este usa os últimos k elementos de uma permutação aleatória de $1, 2, ..., n$). ■

Exemplo 2h

Sejam X, Y, Z variáveis aleatórias independentes e uniformemente distribuídas ao longo de $(0, 1)$. Calcule $P\{X \geq YZ\}$.

Solução Como

$$f_{X,Y,Z}(x, y, z) = f_X(x) f_Y(y) f_Z(z) = 1 \qquad 0 \leq x \leq 1, 0 \leq y \leq 1, 0 \leq z \leq 1$$

temos

$$P\{X \geq YZ\} = \iiint\limits_{x \geq yz} f_{X,Y,Z}(x, y, z)\, dx\, dy\, dz$$
$$= \int_0^1 \int_0^1 \int_{yz}^1 dx\, dy\, dz$$
$$= \int_0^1 \int_0^1 (1 - yz)\, dy\, dz$$
$$= \int_0^1 \left(1 - \frac{z}{2}\right) dz$$
$$= \frac{3}{4}$$

■

Exemplo 2i Interpretação probabilística do tempo de meia-vida

Suponha que $N(t)$ represente o número de núcleos contidos em uma massa de material radioativo no instante de tempo t. O conceito de tempo de meia-vida é muitas vezes definido de uma maneira determinística com o argumento empírico de que, para algum valor h, chamado de *tempo de meia-vida*,

$$N(t) = 2^{-t/h}N(0) \quad t > 0$$

[Observe que $N(h) = N(0)/2$.] Como o resultado anterior implica que, para qualquer s e t não negativos,

$$N(t + s) = 2^{-(s+t)/h}N(0) = 2^{-t/h}N(s)$$

tem-se que, independentemente do tempo s já decorrido, em um tempo adicional t o número de núcleos existentes decairá de acordo com o fator $2^{-t/h}$.

Como a relação determinística acima resulta de observações de massas radioativas contendo números imensos de núcleos, ela parece ser consistente com uma interpretação probabilística. A dica para deduzir-se o modelo probabilístico apropriado para o tempo de meia-vida reside na observação empírica de que a proporção de decaimento em qualquer intervalo de tempo não depende nem do número de núcleos no início desse intervalo nem do instante de início desse intervalo (já que $N(t + s)/N(s)$ não depende nem de $N(s)$, nem de s). Assim, parece que cada núcleo age independentemente e de acordo com uma distribuição de tempo de vida sem memória. Consequentemente, como a distribuição exponencial é a única que possui a propriedade de ausência de memória, e como exatamente metade de uma dada quantidade de massa decai a cada h unidades de tempo, propomos o seguinte modelo probabilístico para o decaimento radioativo.

Interpretação probabilística do tempo de meia-vida h: Os tempos de vida dos núcleos individuais são variáveis aleatórias independentes com distribuição exponencial de mediana h. Isto é, se L representa o tempo de vida de um dado núcleo, então

$$P\{L < t\} = 1 - 2^{-t/h}$$

(Como $P\{L < h\} = \frac{1}{2}$ e a equação anterior pode ser escrita como

$$P\{L < t\} = 1 - \exp\left\{-t\frac{\log 2}{h}\right\}$$

pode-se ver que L tem de fato distribuição exponencial com mediana h.)

Note que, de acordo com a interpretação probabilística fornecida, se começamos com $N(0)$ núcleos no tempo 0, então $N(t)$, o número de núcleos que permanecem no tempo t, terão distribuição binomial com parâmetros $n = N(0)$ e $p = 2^{-t/h}$. Resultados do Capítulo 8 mostrarão que essa interpretação do tempo de meia-vida é consistente com o modelo determinístico quando se considera a proporção de um grande número de núcleos que decaem ao longo de um dado intervalo de tempo. Entretanto, a diferença entre as interpretações determinística e probabilística fica clara quando se considera o número real de núcleos que sofreram decaimento. Vamos agora indicar isso considerando a questão do decaimento ou não de prótons.

Há alguma controvérsia quanto ao fato de prótons decaírem ou não. De fato, uma teoria prediz que prótons decaem com um tempo de meia-vida da ordem de $h = 10^{30}$ anos. Para verificar essa predição empiricamente, sugeriu-se que um grande número de prótons fosse monitorado ao longo de, digamos, um ou dois anos, de forma a observar-se a ocorrência de qualquer decaimento ao longo deste período (claramente, não seria possível monitorar uma massa de prótons por 10^{30} anos para ver se metade dela terá decaído). Vamos supor que possamos a monitorar $N(0) = 10^{30}$ prótons por c anos. O número de decaimentos predito pelo modelo determinístico seria então dado por

$$\begin{aligned}
N(0) - N(c) &= h(1 - 2^{-c/h}) \\
&= \frac{1 - 2^{-c/h}}{1/h} \\
&\approx \lim_{x\to 0} \frac{1 - 2^{-cx}}{x} \qquad \text{como } \frac{1}{h} = 10^{-30} \approx 0 \\
&= \lim_{x\to 0} (c2^{-cx} \log 2) \qquad \text{pela regra de L'Hôpital} \\
&= c \log 2 \approx 0{,}6931c
\end{aligned}$$

Por exemplo, o modelo determinístico prediz que em 2 anos deverão ocorrer 1,3863 decaimentos. De fato, seria uma grande falha para o modelo de decaimento de prótons proposto se nenhum decaimento fosse observado ao longo desses 2 anos.

Vamos agora contrastar as conclusões que acabamos de obter com aquelas obtidas com o modelo probabilístico. Novamente, vamos considerar a hipótese de que o tempo de meia-vida dos prótons seja de $h = 10^{30}$ anos, e supor que monitoremos h prótons por c anos. Como há um número enorme de prótons independentes, cada um deles com probabilidade muito pequena de decair ao longo deste período de tempo, tem-se que o número de prótons que apresentam decaimento tem (com uma aproximação muito boa) uma distribuição de Poisson com parâmetro $h(1 - 2^{-c/h}) \approx c \log 2$. Assim,

$$\begin{aligned}
P\{0 \text{ decaimentos}\} &= e^{-c \log 2} \\
&= e^{-\log(2^c)} = \frac{1}{2^c}
\end{aligned}$$

e, em geral,

$$P\{n \text{ decaimentos}\} = \frac{2^{-c}[c \log 2]^n}{n!} \qquad n \geq 0$$

Assim vemos que, muito embora o número médio de decaimentos ao longo de 2 anos seja (conforme predito pelo modelo determinístico) de 1,3863, há 1 chance em 4 de que não ocorram quaisquer decaimentos. Este resultado de forma alguma invalida a hipótese original do decaimento de prótons. ■

Observação: *Independência é uma relação simétrica.* As variáveis aleatórias X e Y são independentes se sua função densidade conjunta (ou função de probabilidade conjunta, no caso discreto) é o produto de suas funções densidade (ou de probabilidade) individuais. Portanto, dizer que X é independente de Y é equivalente a dizer que Y é independente de X – ou somente que X e Y são independentes. Como resultado, ao considerar se X é independente ou não de Y em situações em que não é intuitivo saber que o valor de Y não muda as probabilidades relacionadas a X, pode ser útil inverter os papéis de X e Y e perguntar se Y é independente de X. O próximo exemplo ilustra este ponto.

Exemplo 2j

Se o primeiro resultado obtido em um jogo de dados resultar na soma dos dados ser igual a 4, então o jogador continua a jogar os dados até que a soma dê 4 ou 7. Se a soma der 4, então o jogador vence; se der 7, ele perde. Suponha que N represente o número de jogadas necessárias até que a soma seja 7 ou 4, e que X represente o valor (4 ou 7) da jogada final. N é independente de X? Isto é, saber se o resultado foi 4 ou 7 afeta a distribuição do número de jogadas necessárias para que um desses dois números apareça? Muitas pessoas não consideram a resposta para essa questão intuitivamente óbvia. Entretanto, suponha que invertamos a questão e perguntemos se X é independente de N. Isto é, saber quantas jogadas são necessárias para se obter uma soma igual a 4 ou 7 afeta a probabilidade de que a soma seja igual a 4? Por exemplo, suponha que saibamos que são necessárias n jogadas para que se obtenha uma soma igual a 4 ou 7. Isso afeta a distribuição de probabilidade da soma final? Claramente não, já que a única coisa que interessa é que o valor da soma seja 4 ou 7, e o fato de que nenhuma das primeiras $n - 1$ jogadas tenha dado 4 ou 7 não muda as probabilidades da n-ésima jogada. Assim podemos concluir que X é independente de N, ou equivalentemente, que N é independente de X.

Como outro exemplo, seja $X_1, X_2, \ldots$ uma série de variáveis aleatórias independentes e identicamente distribuídas, e suponha que observemos essas variáveis aleatórias em sequência. Se $X_n > X_i$ para cada $i = 1, \ldots, n - 1$, dizemos que X_n é um *valor recorde*. Isto é, cada variável aleatória que é maior que todas aquelas que a precedem é chamada de valor recorde. Suponha que A_n represente o evento em que X_n é um valor recorde. É A_{n+1} independente de A_n? Isto é, saber que a n-ésima variável aleatória é a maior das n primeiras variáveis muda a probabilidade de que a $(n + 1)$-ésima variável seja a maior das primeiras $(n + 1)$ variáveis? Embora seja verdade que A_{n+1} é independente de A_n, isso pode não ser intuitivamente óbvio. Entretanto, se invertermos a questão e perguntarmos se A_n é independente de A_{n+1}, então o resultado é mais facilmente entendido. Pois saber que o $(n + 1)$-ésimo valor é maior que $X_1, \ldots, X_n$ claramente não nos fornece qualquer informação sobre o tamanho relativo de X_n entre as primeiras n variáveis aleatórias. De fato, por simetria, fica claro que cada uma dessas n variáveis aleatórias tem a mesma probabilidade de ser a maior do conjunto, então $P\{A_n|A_{n+1}\} = P(A_n) = 1/n$. Com isso, podemos concluir que A_n e A_{n+1} são eventos independentes. □

Observação: Resulta da identidade

$$P\{X_1 \le a_1, \ldots, X_n \le a_n\}$$
$$= P\{X_1 \le a_1\}P\{X_2 \le a_2 | X_1 \le a_1\} \ldots P\{X_n \le a_n | X_1 \le a_1, \ldots, X_{n-1} \le a_{n-1}\}$$

que a independência de $X_1, \ldots, X_n$ pode ser estabelecida sequencialmente. Isto é, podemos mostrar que tais variáveis aleatórias são independentes mostrando que

X_2 é independente de X_1
X_3 é independente de X_1, X_2
X_4 é independente de X_1, X_2, X_3
.
.
.
X_n é independente de $X_1, \ldots, X_{n-1}$

6.3 SOMAS DE VARIÁVEIS ALEATÓRIAS INDEPENDENTES

É muitas vezes importante poder calcular a distribuição de $X + Y$ a partir das distribuições de X e Y quando X e Y são independentes. Suponha que X e Y sejam variáveis aleatórias independentes contínuas com funções densidade de probabilidade f_X e f_Y. A função distribuição cumulativa de $X + Y$ é obtida da seguinte maneira:

$$\begin{aligned} F_{X+Y}(a) &= P\{X + Y \le a\} \\ &= \iint_{x+y \le a} f_X(x) f_Y(y)\, dx\, dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{a-y} f_X(x) f_Y(y)\, dx\, dy \qquad (3.1) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{a-y} f_X(x)\, dx f_Y(y)\, dy \\ &= \int_{-\infty}^{\infty} F_X(a - y) f_Y(y)\, dy \end{aligned}$$

A função distribuição cumulativa F_{X+Y} é chamada de *convolução* das distribuições F_X e F_Y (que são as funções distribuição cumulativa de X e Y, respectivamente).

Derivando a Equação (3.1), vemos que a função densidade de probabilidade f_{X+Y} de $X + Y$ é dada por

$$\begin{aligned} f_{X+Y}(a) &= \frac{d}{da} \int_{-\infty}^{\infty} F_X(a - y) f_Y(y)\, dy \\ &= \int_{-\infty}^{\infty} \frac{d}{da} F_X(a - y) f_Y(y)\, dy \qquad (3.2) \\ &= \int_{-\infty}^{\infty} f_X(a - y) f_Y(y)\, dy \end{aligned}$$

6.3.1 Variáveis aleatórias uniformes identicamente distribuídas

Não é difícil determinar a função densidade da soma de duas variáveis uniformes independentes no intervalo $(0, 1)$.

Exemplo 3a Soma de duas variáveis aleatórias uniformes independentes

Se X e Y são variáveis aleatórias independentes, ambas uniformemente distribuídas em $(0, 1)$, calcule a função densidade de probabilidade de $X + Y$.

Solução Da Equação (3.2), como

$$f_X(a) = f_Y(a) = \begin{cases} 1 & 0 < a < 1 \\ 0 & \text{caso contrário} \end{cases}$$

obtemos

$$f_{X+Y}(a) = \int_0^1 f_X(a - y)\, dy$$

Para $0 \leq a \leq 1$, isso resulta em

$$f_{X+Y}(a) = \int_0^a dy = a$$

Para $1 < a < 2$, obtemos

$$f_{X+Y}(a) = \int_{a-1}^1 dy = 2 - a$$

Com isso,

$$f_{X+Y}(a) = \begin{cases} a & 0 \leq a \leq 1 \\ 2 - a & 1 < a < 2 \\ 0 & \text{caso contrário} \end{cases}$$

Por causa da forma de sua função densidade (veja a Figura 6.3), diz-se que a variável aleatória $X + Y$ tem distribuição *triangular*. ■

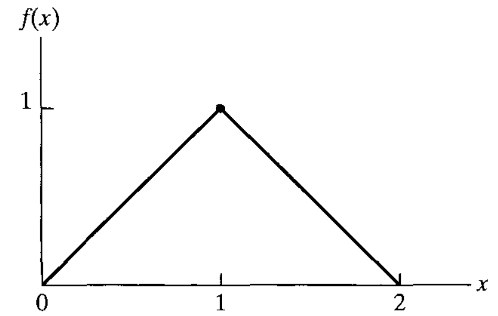

Figura 6.3 Função densidade triangular.

Agora, suponha que $X_1, X_2,..., X_n$ sejam variáveis aleatórias independentes no intervalo $(0, 1)$ e considere

$$F_n(x) = P\{X_1 +... + X_n \le x\}$$

Embora a fórmula geral para $F_n(x)$ seja complicada, ela tem uma forma particularmente simples quando $x \le 1$. De fato, usamos agora indução matemática para demonstrar que

$$F_n(x) = x^n/n!, 0 \le x \le 1$$

Como a equação anterior é verdadeira para $n = 1$, suponha que

$$F_{n-1}(x) = x^{n-1}/(n-1)!, 0 \le x \le 1$$

Agora, escrevendo

$$\sum_{i=1}^{n} X_i = \sum_{i=1}^{n-1} X_i + X_n$$

e usando o fato de que os X_i's são todos não negativos, vemos da Equação 3.1 que, para $0 \le x \le 1$,

$$\begin{aligned} F_n(x) &= \int_0^1 F_{n-1}(x - y) f_{X_n}(y) dy \\ &= \frac{1}{(n-1)!} \int_0^x (x - y)^{n-1} dy \quad \text{pela hipótese de indução} \\ &= x^n/n! \end{aligned}$$

o que completa a demonstração.

Para uma aplicação interessante da fórmula anterior, vamos usá-la para determinar o número esperado de variáveis aleatórias uniformes independentes no intervalo $(0, 1)$ que precisam ser somadas para que o resultado seja maior que 1. Isto é, com $X_1, X_2,...$ sendo variáveis aleatórias uniformes independentes no intervalo $(0, 1)$, queremos determinar $E[N]$, onde

$$N = \min[n: X_1 +... + X_n > 1]$$

Notando que N é maior que $n > 0$ se e somente se $X_1 +... + X_n \le 1$, vemos que

$$P\{N > n\} = F_n(1) = 1/n!, n > 0$$

Como

$$P\{N > 0\} = 1 = 1/0!$$

vemos que, para $n > 0$,

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

Portanto,

$$\begin{aligned}E[N] &= \sum_{n=1}^{\infty} \frac{n(n-1)}{n!} \\ &= \sum_{n=2}^{\infty} \frac{1}{(n-2)!} \\ &= e\end{aligned}$$

Isto é, o número de variáveis aleatórias uniformes independentes no intervalo (0, 1) que precisam ser somadas para que o resultado seja maior que 1 é igual a e.

6.3.2 Variáveis aleatórias gama

Lembre-se da variável aleatória gama, que tem função densidade da forma

$$f(y) = \frac{\lambda e^{-\lambda y}(\lambda y)^{t-1}}{\Gamma(t)} \quad 0 < y < \infty$$

Uma importante propriedade desta família de distribuições é a de que, para um valor fixo de λ, ela é fechada em convoluções.

Proposição 3.1 Se X e Y são variáveis aleatórias gama independentes com respectivos parâmetros (s, λ) e (t, λ), então $X + Y$ é uma variável aleatória gama com parâmetros $(s + t, \lambda)$.

Demonstração Usando a Equação (3.2), obtemos

$$\begin{aligned}f_{X+Y}(a) &= \frac{1}{\Gamma(s)\Gamma(t)} \int_0^a \lambda e^{-\lambda(a-y)}[\lambda(a-y)]^{s-1}\lambda e^{-\lambda y}(\lambda y)^{t-1}\,dy \\ &= Ke^{-\lambda a}\int_0^a (a-y)^{s-1}y^{t-1}dy \\ &= Ke^{-\lambda a}a^{s+t-1}\int_0^1 (1-x)^{s-1}x^{t-1}\,dx \quad \text{fazendo } x = \frac{y}{a} \\ &= Ce^{-\lambda a}a^{s+t-1}\end{aligned}$$

onde C é uma constante que não depende de a. Mas, como a fórmula anterior é uma função densidade, e portanto sua integral deve ser igual a 1, pode-se determinar o valor de C. Com isso, temos

$$f_{X+Y}(a) = \frac{\lambda e^{-\lambda a}(\lambda a)^{s+t-1}}{\Gamma(s+t)}$$

e o resultado está demonstrado. □

É fácil agora estabelecer, por indução e usando a Proposição 3.1, que, se X_i, $i = 1,\ldots, n$ são variáveis aleatórias gama com respectivos parâmetros (t_i, λ), $i = 1,\ldots, n$, então $\sum_{i=1}^{n} X_i$ é gama com parâmetros $\left(\sum_{i=1}^{n} t_i, \lambda\right)$. Deixamos a demonstração disso como exercício.

Exemplo 3b

Suponha que $X_1, X_2,..., X_n$ sejam n variáveis aleatórias exponenciais independentes, cada uma com parâmetro λ. Então, como uma variável aleatória exponencial com parâmetro λ é igual a uma variável aleatória gama com parâmetros $(1, \lambda)$, resulta da Proposição 3.1 que $X_1 + X_2 + \cdots + X_n$ é uma variável aleatória gama com parâmetros (n, λ). ■

Se $Z_1, Z_2,..., Z_n$ são variáveis aleatórias normais padrão independentes, então $\sum_{i=1}^{n} Z_i^2$ é chamada de distribuição *qui-quadrado* (às vezes representada como χ^2) com n graus de liberdade. Vamos calcular a função densidade de Y. Quando $n = 1, Y = Z_1^2$, e do Exemplo 7b do Capítulo 5 vemos que sua função densidade de probabilidade é dada por

$$\begin{aligned} f_{Z^2}(y) &= \frac{1}{2\sqrt{y}}[fz(\sqrt{y}) + fz(-\sqrt{y})] \\ &= \frac{1}{2\sqrt{y}}\frac{2}{\sqrt{2\pi}}e^{-y/2} \\ &= \frac{\frac{1}{2}e^{-y/2}(y/2)^{1/2-1}}{\sqrt{\pi}} \end{aligned}$$

Mas reconhecemos a equação anterior como sendo a distribuição gama com parâmetros $\left(\frac{1}{2}, \frac{1}{2}\right)$ [um subproduto dessa análise é que $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$]. Mas como cada Z_i^2 é gama $\left(\frac{1}{2}, \frac{1}{2}\right)$, resulta da Proposição 3.1 que a distribuição χ^2 com n graus de liberdade é tão somente a distribuição gama com parâmetros $\left(n/2, \frac{1}{2}\right)$. Ela portanto tem a função densidade de probabilidade dada por

$$\begin{aligned} f_{\chi^2}(y) &= \frac{\frac{1}{2}e^{-y/2}\left(\frac{y}{2}\right)^{n/2-1}}{\Gamma\left(\frac{n}{2}\right)} \qquad y > 0 \\ &= \frac{e^{-y/2}y^{n/2-1}}{2^{n/2}\Gamma\left(\frac{n}{2}\right)} \qquad y > 0 \end{aligned}$$

Quando n é um inteiro par, $\Gamma(n/2) = [(n/2) - 1]!$. Por outro lado, quando n é ímpar, $\Gamma(n/2)$ pode ser obtido fazendo-se iterações com $\Gamma(t) = (t-1)\Gamma(t-1)$ e então empregando-se o resultado previamente obtido de que $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$ [por exemplo, $\Gamma\left(\frac{5}{2}\right) = \frac{3}{2}\Gamma\left(\frac{3}{2}\right) = \frac{3}{2}\frac{1}{2}\Gamma\left(\frac{1}{2}\right) = \frac{3}{4}\sqrt{\pi}$].

Na prática, a distribuição qui-quadrado aparece com frequência como a distribuição do quadrado do erro envolvido quando se tenta acertar um alvo em um espaço n-dimensional se os erros das coordenadas são representados como variáveis aleatórias normais padrão. Ela também é importante em análises estatísticas.

6.3.3 Variáveis aleatórias normais

Também podemos usar a Equação (3.2) para demonstrar o seguinte resultado importante sobre as variáveis aleatórias normais.

Proposição 3.2 Se $X_i, i = 1,\ldots, n$, são variáveis aleatórias normalmente distribuídas com respectivos parâmetros $\mu_i, \sigma_i^2, i = 1,\ldots, n$, então $\sum_{i=1}^{n} X_i$ é normalmente distribuída com parâmetros $\sum_{i=1}^{n} \mu_i$ e $\sum_{i=1}^{n} \sigma_i^2$.

Demonstração da Proposição 3.2 Para começar, suponha que X e Y sejam variáveis aleatórias normais independentes com X tendo média 0 e variância σ^2, e Y tendo média 0 e variância 1. Vamos determinar a função densidade de $X + Y$ utilizando a Equação (3.2). Agora, com

$$c = \frac{1}{2\sigma^2} + \frac{1}{2} = \frac{1 + \sigma^2}{2\sigma^2}$$

temos

$$\begin{aligned} f_X(a - y)f_Y(y) &= \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(a - y)^2}{2\sigma^2}\right\} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{y^2}{2}\right\} \\ &= \frac{1}{2\pi\sigma} \exp\left\{-\frac{a^2}{2\sigma^2}\right\} \exp\left\{-c\left(y^2 - 2y\frac{a}{1 + \sigma^2}\right)\right\} \end{aligned}$$

Portanto, da Equação (3.2),

$$\begin{aligned} f_{X+Y}(a) &= \frac{1}{2\pi\sigma} \exp\left\{-\frac{a^2}{2\sigma^2}\right\} \exp\left\{\frac{a^2}{2\sigma^2(1 + \sigma^2)}\right\} \\ &\quad \times \int_{-\infty}^{\infty} \exp\left\{-c\left(y - \frac{a}{1 + \sigma^2}\right)^2\right\} dy \\ &= \frac{1}{2\pi\sigma} \exp\left\{-\frac{a^2}{2(1 + \sigma^2)}\right\} \int_{-\infty}^{\infty} \exp\{-cx^2\}\, dx \\ &= C \exp\left\{-\frac{a^2}{2(1 + \sigma^2)}\right\} \end{aligned}$$

onde C não depende de a. Mas isso implica que $X + Y$ é normal com média 0 e variância $1 + \sigma^2$.

Agora, suponha que X_1 e X_2 sejam variáveis aleatórias normais independentes X_i com média μ_i e variância $\sigma_i^2, i = 1, 2$. Então,

$$X_1 + X_2 = \sigma_2\left(\frac{X_1 - \mu_1}{\sigma_2} + \frac{X_2 - \mu_2}{\sigma_2}\right) + \mu_1 + \mu_2$$

Mas como $(X_1 - \mu_1)/\sigma_2$ é normal com média 0 e variância σ_1^2/σ_2^2, e $(X_2 - \mu_2)/\sigma_2$ é normal com média 0 e variância 1, segue de nosso resultado prévio que $(X_1 - \mu_1)/\sigma_2 + (X_2 - \mu_2)/\sigma_2$ é normal com média 0 e variância $1 + \sigma_1^2/\sigma_2^2$. Isso implica $X_1 + X_2$ ser normal com média $\mu_1 + \mu_2$ e variância $\sigma_2^2(1 + \sigma_1^2/\sigma_2^2) = \sigma_1^2 + \sigma_2^2$.

Assim, estabelece-se a Proposição 3.2 quando $n = 2$. O caso geral é agora obtido por indução. Suponha que a Proposição 3.2 seja verdadeira quando há $n - 1$ variáveis aleatórias. Agora considere o caso n e escreva

$$\sum_{i=1}^{n} X_i = \sum_{i=1}^{n-1} X_i + X_n$$

Pela hipótese de indução, $\sum_{i=1}^{n-1} X_i$ é normal com média $\sum_{i=1}^{n-1} \mu_i$ e variância $\sum_{i=1}^{n-1} \sigma_i^2$. Portanto, pelo resultado para $n = 2$, $\sum_{i=1}^{n} X_i$ é normal com média $\sum_{i=1}^{n} \mu_i$ e variância $\sum_{i=1}^{n} \sigma_i^2$.

Exemplo 3c

Um time de basquete jogará uma temporada com 44 partidas. Vinte e seis dessas partidas serão contra times da divisão A, e 18 contra times da divisão B. Suponha que o time ganhe cada partida disputada contra um adversário da divisão A com probabilidade 0,4, e ganhe cada partida disputada contra um adversário da divisão B com probabilidade 0,7. Suponha também que os resultados das diferentes partidas sejam independentes. Obtenha um valor aproximado para a probabilidade de que

(a) o time vença 25 partidas ou mais;

(b) o time vença mais partidas contra times da divisão A do que contra times da divisão B.

Solução (a) Suponha que X_A e X_B representem, respectivamente, o número de partidas que o time vence contra equipes da divisão A e B. Note que X_A e X_B são variáveis aleatórias binomiais independentes e

$$E[X_A] = 26(0{,}4) = 10{,}4 \quad \text{Var}(X_A) = 26(0{,}4)(0{,}6) = 6{,}24$$
$$E[X_B] = 18(0{,}7) = 12{,}6 \quad \text{Var}(X_B) = 18(0{,}7)(0{,}3) = 3{,}78$$

Pela aproximação normal para a distribuição binomial, X_A e X_B têm aproximadamente a mesma distribuição que teriam variáveis aleatórias normais independentes com os valores esperados e as variâncias acima. Portanto, pela Proposição 3.2, $X_A + X_B$ terá aproximadamente uma distribuição normal com média 23 e variância 10,02. Assim, fazendo com que Z represente uma variável aleatória normal padrão, temos

$$\begin{aligned}
P\{X_A + X_B \geq 25\} &= P\{X_A + X_B \geq 24{,}5\} \\
&= P\left\{\frac{X_A + X_B - 23}{\sqrt{10{,}02}} \geq \frac{24{,}5 - 23}{\sqrt{10{,}02}}\right\} \\
&\approx P\left\{Z \geq \frac{1{,}5}{\sqrt{10{,}02}}\right\} \\
&\approx 1 - P\{Z < 0{,}4739\} \\
&\approx 0{,}3178
\end{aligned}$$

(b) Notamos que $X_A - X_B$ tem aproximadamente uma distribuição normal com média −2,2 e variância 10,02. Com isso,

$$\begin{aligned}
P\{X_A - X_B \geq 1\} &= P\{X_A - X_B \geq 0{,}5\} \\
&= P\left\{\frac{X_A - X_B + 2{,}2}{\sqrt{10{,}02}} \geq \frac{0{,}5 + 2{,}2}{\sqrt{10{,}02}}\right\} \\
&\approx P\left\{Z \geq \frac{2{,}7}{\sqrt{10{,}02}}\right\} \\
&\approx 1 - P\{Z < 0{,}8530\} \\
&\approx 0{,}1968
\end{aligned}$$

Portanto, há aproximadamente 31,78% de chances de que o time ganhe pelo menos 25 partidas, e aproximadamente 19,68% de chances de que o time ganhe mais partidas contra times da divisão A do que contra times da divisão B. ■

A variável aleatória Y é chamada de *log-normal* com parâmetros μ e σ se $\log(Y)$ é uma variável aleatória normal com média μ e variância σ^2. Isto é, Y é log-normal se puder ser escrita como

$$Y = e^X$$

onde X é uma variável aleatória normal.

Exemplo 3d

Começando de certo instante fixo, suponha que $S(n)$ represente o preço de um depósito ao final de n semanas, $n \geq 1$. Um modelo popular para a evolução desses preços supõe que as relações entre os preços $S(n)/S(n-1)$, $n \geq 1$, sejam variáveis aleatórias log-normais independentes e identicamente distribuídas. Considerando esse modelo com parâmetros $\mu = 0{,}0165$, $\sigma = 0{,}0730$, qual é a probabilidade de que

(a) o preço do depósito suba ao longo de cada uma das próximas duas semanas?
(b) o preço no final de duas semanas seja maior do que é hoje?

Solução Seja Z uma variável aleatória normal padrão. Para resolver a letra (a), usamos o fato de que $\log(x)$ aumenta em x para concluir que $x > 1$ se e somente se $\log(x) > \log(1) = 0$. Como resultado, temos

$$P\left\{\frac{S(1)}{S(0)} > 1\right\} = P\left\{\log\left(\frac{S(1)}{S(0)}\right) > 0\right\}$$
$$= P\left\{Z > \frac{-0{,}0165}{0{,}0730}\right\}$$
$$= P\{Z < 0{,}2260\}$$
$$= 0{,}5894$$

Em outras palavras, a probabilidade de que o preço suba após 1 semana é de 0,5894. Como as relações de preço sucessivas são independentes, a probabilidade de que o preço suba ao longo de cada uma das próximas duas semanas é de $(0{,}5894)^2 = 0{,}3474$.

Para resolver a letra (b), raciocinamos da seguinte maneira:

$$P\left\{\frac{S(2)}{S(0)} > 1\right\} = P\left\{\frac{S(2)}{S(1)}\frac{S(1)}{S(0)} > 1\right\}$$
$$= P\left\{\log\left(\frac{S(2)}{S(1)}\right) + \log\left(\frac{S(1)}{S(0)}\right) > 0\right\}$$

Entretanto, $\log\left(\frac{S(2)}{S(1)}\right) + \log\left(\frac{S(1)}{S(0)}\right)$, sendo a soma de duas variáveis aleatórias normais independentes, ambas com média 0,0165 e desvio padrão 0,0730, é uma variável aleatória normal com média 0,0330 e variância $2(0{,}0730)^2$. Consequentemente,

$$P\left\{\frac{S(2)}{S(0)} > 1\right\} = P\left\{Z > \frac{-0{,}0330}{0{,}0730\sqrt{2}}\right\}$$
$$= P\{Z < 0{,}31965\}$$
$$= 0{,}6254$$

■

6.3.4 Variáveis aleatórias binomiais e de Poisson

Em vez de tentarmos deduzir uma expressão geral para a distribuição de $X + Y$ no caso discreto, vamos considerar alguns exemplos.

Exemplo 3e ***Somas de variáveis aleatórias de Poisson independentes***

Se X e Y são variáveis aleatórias de Poisson independentes com respectivos parâmetros λ_1 e λ_2, calcule a distribuição de $X + Y$.

Solução Como o evento $\{X + Y = n\}$ pode ser escrito como a união dos eventos disjuntos $\{X = k, Y = n - k\}, 0 \leq k \leq n$, temos

$$\begin{aligned}
P\{X + Y = n\} &= \sum_{k=0}^{n} P\{X = k, Y = n - k\} \\
&= \sum_{k=0}^{n} P\{X = k\}P\{Y = n - k\} \\
&= \sum_{k=0}^{n} e^{-\lambda_1}\frac{\lambda_1^k}{k!}e^{-\lambda_2}\frac{\lambda_2^{n-k}}{(n - k)!} \\
&= e^{-(\lambda_1+\lambda_2)}\sum_{k=0}^{n}\frac{\lambda_1^k\lambda_2^{n-k}}{k!(n - k)!} \\
&= \frac{e^{-(\lambda_1+\lambda_2)}}{n!}\sum_{k=0}^{n}\frac{n!}{k!(n - k)!}\lambda_1^k\lambda_2^{n-k} \\
&= \frac{e^{-(\lambda_1+\lambda_2)}}{n!}(\lambda_1 + \lambda_2)^n
\end{aligned}$$

Assim, $X_1 + X_2$ tem uma distribuição de Poisson com parâmetros $\lambda_1 + \lambda_2$. ■

Exemplo 3f Somas de variáveis aleatórias binomiais independentes

Sejam X e Y variáveis aleatórias binomiais independentes com respectivos parâmetros (n,p) e (m,p). Calcule a distribuição de $X + Y$.

Solução Relembrando a interpretação de uma variável aleatória binomial, e sem qualquer cálculo, podemos imediatamente concluir que $X + Y$ é binomial com parâmetros $(n + m,p)$. Isso ocorre porque X representa o número de sucessos em n tentativas independentes, cada uma das quais com probabilidade de sucesso p; similarmente, Y representa o número de sucessos em m tentativas independentes, cada uma das quais com probabilidade de sucesso p. Com isso, dada a independência de X e Y, tem-se que $X + Y$ representa o número de sucessos em $n + m$ tentativas independentes quando cada tentativa tem probabilidade de sucesso p. Entretanto, $X + Y$ é uma variável aleatória binomial com parâmetros $(n + m,p)$. Para verificar essa conclusão analiticamente, note que

$$\begin{aligned}
P\{X + Y = k\} &= \sum_{i=0}^{n} P\{X = i, Y = k - i\} \\
&= \sum_{i=0}^{n} P\{X = i\}P\{Y = k - i\} \\
&= \sum_{i=0}^{n} \binom{n}{i} p^i q^{n-i} \binom{m}{k - i} p^{k-i} q^{m-k+i}
\end{aligned}$$

onde $q = 1 - p$ e onde $\binom{r}{j} = 0$ quando $j < 0$. Assim,

$$P\{X + Y = k\} = p^k q^{n+m-k} \sum_{i=0}^{n} \binom{n}{i} \binom{m}{k-i}$$

e a conclusão resulta da aplicação da identidade combinatória

$$\binom{n+m}{k} = \sum_{i=0}^{n} \binom{n}{i} \binom{m}{k-i}$$

■

6.3.5 Variáveis aleatórias geométricas

Sejam $X_1,..., X_n$ variáveis aleatórias geométricas independentes, com X_i tendo parâmetro p_i, para $i = 1,..., n$. Estamos interessados em calcular a função discreta de probabilidade da soma $S_n = \sum_{i=1}^{n} X_i$. Para uma aplicação, considere n moedas, cada uma com probabilidade p_i de dar cara, $i = 1,..., n$. Suponha que a moeda 1 seja jogada até que dê cara, instante a partir do qual a moeda 2 começa a ser jogada até que dê cara, e então a moeda 3 é jogada até que dê cara, e assim por diante. Se X_i representa o número de jogadas feitas com a moeda i, então $X_1, X_2,..., X_n$ são variáveis aleatórias geométricas com respectivos parâmetros $p_1, p_2,..., p_n$, e $S_n = \sum_{i=1}^{n} X_i$ representa o número total de jogadas. Se todos os p_i são iguais – digamos, $p_i = p$ – então S_n tem a mesma distribuição que o número de jogadas necessárias para que se obtenha um total de n caras com uma moeda com probabilidade p de dar cara, e então S_n é uma variável aleatória binomial negativa com função discreta de probabilidade

$$P\{S_n = k\} = \binom{k-1}{n-1} p^n (1-p)^{k-n}, \quad k \geq n$$

Como um prelúdio para a determinação da função discreta de probabilidade de S_n quando todos os p_i são distintos, vamos primeiro considerar o caso $n = 2$. Fazendo $q_j = 1 - p_j, j = 1, 2$, obtemos

$$\begin{aligned}
P(S_2 = k) &= \sum_{j=1}^{k-1} P\{X_1 = j, X_2 = k - j\} \\
&= \sum_{j=1}^{k-1} P\{X_1 = j\}\, P\{X_2 = k - j\} \quad \text{(pela independência)} \\
&= \sum_{j=1}^{k-1} p_1 q_1^{j-1} p_2 q_2^{k-j-1} \\
&= p_1 p_2 q_2^{k-2} \sum_{j=1}^{k-1} (q_1/q_2)^{j-1}
\end{aligned}$$

$$
\begin{aligned}
&= p_1 p_2 q_2^{k-2} \frac{1 - (q_1/q_2)^{k-1}}{1 - q_1/q_2} \\
&= \frac{p_1 p_2 q_2^{k-1}}{q_2 - q_1} - \frac{p_1 p_2 q_1^{k-1}}{q_2 - q_1} \\
&= p_2 q_2^{k-1} \frac{p_1}{p_1 - p_2} + p_1 q_1^{k-1} \frac{p_2}{p_2 - p_1}
\end{aligned}
$$

Se agora fizermos $n = 3$ e calcularmos $P\{S_3 = k\}$ começando com a identidade

$$
P\{S_3 = k\} = \sum_{j=1}^{k-1} P\{S_2 = j, X_3 = k - j\} = \sum_{j=1}^{k-1} P\{S_2 = j\} P\{X_3 = k - j\}
$$

e então substituindo a fórmula deduzida para a função de probabilidade de S_2, obtemos, após algumas manipulações,

$$
\begin{aligned}
P\{S_3 = k\} &= p_1 q_1^{k-1} \frac{p_2}{p_2 - p_1} \frac{p_3}{p_3 - p_1} + p_2 q_2^{k-1} \frac{p_1}{p_1 - p_2} \frac{p_3}{p_3 - p_2} \\
&\quad + p_3 q_3^{k-1} \frac{p_1}{p_1 - p_3} \frac{p_2}{p_2 - p_3}
\end{aligned}
$$

As funções de probabilidade de S_2 e S_3 levam à seguinte conjectura para a função de probabilidade de S_n.

Proposição 3.3 Suponha que $X_1, \ldots, X_n$ sejam variáveis aleatórias geométricas independentes, com X_i tendo parâmetro p_i para $i = 1, \ldots, n$. Se todos os p_i's são distintos, então, para $k \geq n$,

$$
P\{S_n = k\} = \sum_{i=1}^{n} p_i q_i^{k-1} \prod_{j \neq i} \frac{p_j}{p_j - p_i}
$$

Demonstração da Proposição 3.3 Vamos demonstrar essa proposição por indução no valor de $n + k$. Como a proposição é verdadeira quando $n = 2, k = 2$, suponha, como hipótese de indução, que isso seja verdade para qualquer $k \geq n$ no qual $n + k \leq r$. Agora, suponha que $k \geq n$ seja tal que $n + k = r + 1$. Para calcular $P\{S_n = k\}$, condicionamos na ocorrência de $X_n = 1$. Isso resulta em

$$
\begin{aligned}
P\{S_n = k\} &= P\{S_n = k | X_n = 1\} P\{X_n = 1\} + P\{S_n = k | X_n > 1\} P\{X_n > 1\} \\
&= P\{S_n = k | X_n = 1\} p_n + P\{S_n = k | X_n > 1\} q_n
\end{aligned}
$$

Agora,

$$
\begin{aligned}
P\{S_n = k | X_n = 1\} &= P\{S_{n-1} = k - 1 | X_n = 1\} \\
&= P\{S_{n-1} = k - 1\} \quad \text{(pela independência)} \\
&= \sum_{i=1}^{n-1} p_i q_i^{k-2} \prod_{i \neq j \leq n-1} \frac{p_j}{p_j - p_i} \quad \text{(pela hipótese de indução)}
\end{aligned}
$$

Agora, se X é uma variável aleatória geométrica com parâmetro p, então a distribuição condicional de X dado que ele é maior que 1 é igual à distribuição de

1 (a primeira tentativa malsucedida) mais uma variável geométrica com parâmetro p (o número de tentativas adicionais após a primeira até que um sucesso ocorra). Consequentemente,

$$\begin{aligned} P\{S_n = k|X_n > 1\} &= P\{X_1 + \ldots + X_{n-1} + X_n + 1 = k\} \\ &= P\{S_n = k - 1\} \\ &= \sum_{i=1}^{n} p_i q_i^{k-2} \prod_{i \neq j \leq n} \frac{p_j}{p_j - p_i} \end{aligned}$$

onde a última igualdade resulta da hipótese de indução. Assim, do desenvolvimento anterior, obtemos

$$\begin{aligned} P\{S_n = k\} &= p_n \sum_{i=1}^{n-1} p_i q_i^{k-2} \prod_{i \neq j \leq n-1} \frac{p_j}{p_j - p_i} + q_n \sum_{i=1}^{n} p_i q_i^{k-2} \prod_{i \neq j \leq n} \frac{p_j}{p_j - p_i} \\ &= p_n \sum_{i=1}^{n-1} p_i q_i^{k-2} \prod_{i \neq j \leq n-1} \frac{p_j}{p_j - p_i} + q_n \sum_{i=1}^{n-1} p_i q_i^{k-2} \prod_{i \neq j \leq n} \frac{p_j}{p_j - p_i} \\ &\quad + q_n p_n q_n^{k-2} \prod_{j<n} \frac{p_j}{p_j - p_n} \\ &= \sum_{i=1}^{n-1} p_i q_i^{k-2} p_n \left(1 + \frac{q_n}{p_n - p_i}\right) \prod_{i \neq j \leq n-1} \frac{p_j}{p_j - p_i} + p_n q_n^{k-1} \prod_{j<n} \frac{p_j}{p_j - p_n} \end{aligned}$$

Agora, usando

$$1 + \frac{q_n}{p_n - p_i} = \frac{p_n - p_i + q_n}{p_n - p_i} = \frac{q_i}{p_n - p_i}$$

obtemos

$$\begin{aligned} P\{S_n = k\} &= \sum_{i=1}^{n-1} p_i q_i^{k-1} \prod_{i \neq j \leq n} \frac{p_j}{p_j - p_i} + p_n q_n^{k-1} \prod_{j<n} \frac{p_j}{p_j - p_n} \\ &= \sum_{i=1}^{n} p_i q_i^{k-1} \prod_{j \neq i} \frac{p_j}{p_j - p_i} \end{aligned}$$

e a demonstração por indução está completa. ∎

6.4 DISTRIBUIÇÕES CONDICIONAIS: CASO DISCRETO

Lembre que, para dois eventos E e F, a probabilidade condicional de E dado F é definida, com $P(F) > 0$, por

$$P(E|F) = \frac{P(EF)}{P(F)}$$