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

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

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

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

Com isso, se X e Y são variáveis aleatórias discretas, é natural definir a função discreta de probabilidade de X dado que $Y = y$ como

$$
\begin{aligned}
p_{X|Y}(x|y) &= P\{X = x|Y = y\} \\
&= \frac{P\{X = x, Y = y\}}{P\{Y = y\}} \\
&= \frac{p(x,y)}{p_Y(y)}
\end{aligned}
$$

para todos os valores de y tais que $p_Y(y) > 0$. Similarmente, a função distribuição de probabilidade condicional de X dado que $Y = y$ é definida, para todo y tal que $p_Y(y) > 0$, como

$$
\begin{aligned}
F_{X|Y}(x|y) &= P\{X \le x|Y = y\} \\
&= \sum_{a \le x} p_{X|Y}(a|y)
\end{aligned}
$$

Em outras palavras, as definições são exatamente iguais ao caso incondicional, exceto que tudo é agora condicionado no evento em que $Y = y$. Se X é independente de Y, então a função de probabilidade condicional e a função distribuição são iguais aos respectivos casos incondicionais. Isso resulta porque, se X é independente de Y, então

$$
\begin{aligned}
p_{X|Y}(x|y) &= P\{X = x|Y = y\} \\
&= \frac{P\{X = x, Y = y\}}{P\{Y = y\}} \\
&= \frac{P\{X = x\}P\{Y = y\}}{P\{Y = y\}} \\
&= P\{X = x\}
\end{aligned}
$$

Exemplo 4a

Suponha que $p(x,y)$, a função discreta de probabilidade conjunta de X e Y, seja dada por

$$
p(0,0) = 0{,}4 \qquad p(0,1) = 0{,}2 \qquad p(1,0) = 0{,}1 \qquad p(1,1) = 0{,}3
$$

Calcule a função de probabilidade condicional de X dado que $Y = 1$.

Solução Primeiros notamos que

$$
p_Y(1) = \sum_x p(x,1) = p(0,1) + p(1,1) = 0{,}5
$$

Assim,

$$
p_{X|Y}(0|1) = \frac{p(0,1)}{p_Y(1)} = \frac{2}{5}
$$

e

$$p_{X|Y}(1|1) = \frac{p(1,1)}{p_Y(1)} = \frac{3}{5}$$

■

Exemplo 4b

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

Solução Calculamos a função de probabilidade condicional de X dado que $X + Y = n$ da seguinte maneira:

$$\begin{aligned} P\{X = k|X + Y = n\} &= \frac{P\{X = k, X + Y = n\}}{P\{X + Y = n\}} \\ &= \frac{P\{X = k, Y = n - k\}}{P\{X + Y = n\}} \\ &= \frac{P\{X = k\}P\{Y = n - k\}}{P\{X + Y = n\}} \end{aligned}$$

onde a última igualdade resulta da hipótese de independência de X e Y. Lembrando (Exemplo 3e) que $X + Y$ tem uma distribuição de Poisson com parâmetro $\lambda_1 + \lambda_2$, vemos que a equação anterior é igual a

$$\begin{aligned} P\{X = k|X + Y = n\} &= \frac{e^{-\lambda_1}\lambda_1^k}{k!}\frac{e^{-\lambda_2}\lambda_2^{n-k}}{(n-k)!}\left[\frac{e^{-(\lambda_1+\lambda_2)}(\lambda_1 + \lambda_2)^n}{n!}\right]^{-1} \\ &= \frac{n!}{(n-k)!\,k!}\frac{\lambda_1^k\lambda_2^{n-k}}{(\lambda_1 + \lambda_2)^n} \\ &= \binom{n}{k}\left(\frac{\lambda_1}{\lambda_1 + \lambda_2}\right)^k\left(\frac{\lambda_2}{\lambda_1 + \lambda_2}\right)^{n-k} \end{aligned}$$

Em outras palavras, a distribuição condicional de X dado que $X + Y = n$ é binomial com parâmetros n e $\lambda_1/(\lambda_1 + \lambda_2)$. ■

Podemos também falar de distribuições condicionais conjuntas, conforme indicado nos próximos dois exemplos.

Exemplo 4c

Considere a distribuição multinomial com função de probabilidade conjunta

$$P\{X_i = n_i, i = 1,\ldots,k\} = \frac{n!}{n_1!\cdots n_k!}p_1^{n_1}\cdots p_k^{n_k}, \quad n_i \geq 0, \quad \sum_{i=1}^{k} n_i = n$$

Obtém-se tal função de probabilidade quando n tentativas independentes são realizadas, com cada tentativa levando ao resultado i com probabilidade p_i, $\sum_{i=1}^k p_i = 1$. As variáveis aleatórias $X_i, i = 1,\ldots, k$, representam, respectivamente, o número de tentativas que levam ao resultado $i, i = 1,\ldots, k$. Suponha que saibamos que n_j das tentativas tenham levado ao resultado j, para $j = r + 1,\ldots, k$, onde

$\sum_{j=r+1}^{k} n_j = m \leq n$. Então, como cada uma das demais $n - m$ tentativas deve ter levado a um dos resultados 1,..., r, parece-nos que a distribuição condicional de $X_1,..., X_r$ é multinomial em $n - m$ tentativas com respectivas probabilidades

$$P\{\text{resultado } i|\text{resultado não é nenhum de } r + 1,..., k\} = \frac{p_i}{F_r}, i = 1,..., r$$

onde $F_r = \sum_{i=1}^{r} p_i$ é a probabilidade de que uma tentativa leve a um dos resultados 1,..., r.

Solução Para verificar essa intuição, faça com que $n_1,..., n_r$ sejam tais que $\sum_{i=1}^{r} n_i = n - m$. Então,

$$\begin{aligned}
P\{X_1 = n_1, \ldots, X_r = n_r | X_{r+1} = n_{r+1}, \ldots X_k = n_k\} \\
&= \frac{P\{X_1 = n_1, \ldots, X_k = n_k\}}{P\{X_{r+1} = n_{r+1}, \ldots X_k = n_k\}} \\
&= \frac{\frac{n!}{n_1! \cdots n_k!} p_1^{n_1} \cdots p_r^{n_r} p_{r+1}^{n_{r+1}} \cdots p_k^{n_k}}{\frac{n!}{(n-m)! n_{r+1}! \cdots n_k!} F_r^{n-m} p_{r+1}^{n_{r+1}} \cdots p_k^{n_k}}
\end{aligned}$$

onde a probabilidade no denominador foi obtida considerando-se os resultados 1,..., r como um único resultado com probabilidade F_r. Isso mostra que a probabilidade é multinomial em n tentativas com probabilidades de resultados $F_r, p_{r+1},..., p_k$. Como $\sum_{i=1}^{r} n_i = n - m$, pode-se escrever o resultado anterior como

$$\begin{aligned}
P\{X_1 = n_1, \ldots, X_r = n_r | X_{r+1} = n_{r+1}, \ldots X_k = n_k\} \\
&= \frac{(n - m)!}{n_1! \cdots n_r!} \left(\frac{p_1}{F_r}\right)^{n_1} \cdots \left(\frac{p_r}{F_r}\right)^{n_r}
\end{aligned}$$

e nossa intuição é confirmada. ■

Exemplo 4d

Considere n tentativas independentes, com cada tentativa sendo um sucesso com probabilidade p. Dado um total de k sucessos, mostre que todas as possíveis ordenações dos k sucessos e $n - k$ fracassos são igualmente prováveis.

Solução Queremos mostrar que, dado um total de k sucessos, cada uma das $\binom{n}{k}$ ordenações possíveis de k sucessos e $n - k$ fracassos é igualmente provável. Suponha que X represente o número de sucessos, e considere qualquer ordenação de k sucessos e $n - k$ fracassos, digamos, $\mathbf{o} = (s, s, f, f,..., f)$. Então,

$$\begin{aligned}
P(\mathbf{o}|X = k) &= \frac{P(\mathbf{o}, X = k)}{P(X = k)} \\
&= \frac{P(\mathbf{o})}{P(X = k)} \\
&= \frac{p^k(1 - p)^{n-k}}{\binom{n}{k} p^k (1 - p)^{n-k}} \\
&= \frac{1}{\binom{n}{k}}
\end{aligned}$$

■

6.5 DISTRIBUIÇÕES CONDICIONAIS: CASO CONTÍNUO

Se X e Y têm função densidade de probabilidade conjunta $f(x, y)$, então a função densidade de probabilidade condicional de X dado que $Y = y$ é definida, para todos os valores de y tais que $f_Y(y) > 0$, como

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)}$$

Para motivar essa definição, multiplique o lado esquerdo por dx e o lado direito por $(dx\, dy)/dy$ para obter

$$\begin{aligned} f_{X|Y}(x|y)\, dx &= \frac{f(x,y)\, dx\, dy}{f_Y(y)\, dy} \\ &\approx \frac{P\{x \le X \le x + dx, y \le Y \le y + dy\}}{P\{y \le Y \le y + dy\}} \\ &= P\{x \le X \le x + dx | y \le Y \le y + dy\} \end{aligned}$$

Em outras palavras, para valores pequenos de dx e dy, $f_{x|y}(x|y)dx$ representa a probabilidade condicional de que X esteja entre x e $x + dx$ dado que Y esteja entre y e $y + dy$.

O uso de densidades condicionais nos permite definir probabilidades condicionais de eventos associados a uma variável aleatória quando conhecemos o valor de uma segunda variável aleatória. Isto é, se X e Y são conjuntamente contínuas, então, para qualquer conjunto A,

$$P\{X \in A | Y = y\} = \int_A f_{X|Y}(x|y)\, dx$$

Em particular, fazendo $A = (-\infty, a]$, podemos definir a função distribuição cumulativa condicional de X dado que $Y = y$ como

$$F_{X|Y}(a|y) \equiv P\{X \le a | Y = y\} = \int_{-\infty}^{a} f_{X|Y}(x|y)\, dx$$

O leitor deve notar que, ao usarmos as idéias apresentadas na discussão anterior, obtivemos expressões para probabilidades condicionais com as quais podemos trabalhar, muito embora o evento no qual estejamos colocando a condição (isto é, o evento $\{Y = y\}$) tenha probabilidade 0.

Exemplo 5a

A função densidade conjunta de X e Y é dada por

$$f(x,y) = \begin{cases} \frac{12}{5}x(2 - x - y) & 0 < x < 1, 0 < y < 1 \\ 0 & \text{caso contrário} \end{cases}$$

Calcule a densidade condicional de X dado que $Y = y$, onde $0 < y < 1$.

Solução: Para $0 < x < 1, 0 < y < 1$, temos

$$\begin{aligned}
f_{X|Y}(x|y) &= \frac{f(x,y)}{f_Y(y)} \\
&= \frac{f(x,y)}{\int_{-\infty}^{\infty} f(x,y)\,dx} \\
&= \frac{x(2-x-y)}{\int_0^1 x(2-x-y)\,dx} \\
&= \frac{x(2-x-y)}{\frac{2}{3} - y/2} \\
&= \frac{6x(2-x-y)}{4-3y}
\end{aligned}$$

■

Exemplo 5b

Suponha que a densidade conjunta de X e Y seja dada por

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

Determine $P\{X > 1|Y = y\}$

Solução Primeiro obtemos a densidade condicional de X dado que $Y = y$.

$$\begin{aligned}
f_{X|Y}(x|y) &= \frac{f(x,y)}{f_Y(y)} \\
&= \frac{e^{-x/y}e^{-y}/y}{e^{-y}\int_0^{\infty}(1/y)e^{-x/y}\,dx} \\
&= \frac{1}{y}e^{-x/y}
\end{aligned}$$

Portanto,

$$\begin{aligned}
P\{X > 1|Y = y\} &= \int_1^{\infty} \frac{1}{y}e^{-x/y}\,dx \\
&= -e^{-x/y}\Big|_1^{\infty} \\
&= e^{-1/y}
\end{aligned}$$

■

Se X e Y são variáveis aleatórias independentes contínuas, a densidade condicional de X dado que $Y = y$ é somente a densidade incondicional de X. Isso ocorre porque, no caso independente,

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)} = \frac{f_X(x)f_Y(y)}{f_Y(y)} = f_X(x)$$

Também podemos falar de distribuições condicionais quando as variáveis aleatórias não são nem conjuntamente contínuas, nem conjuntamente discretas. Por exemplo, suponha que X seja uma variável aleatória contínua com função densidade de probabilidade f, e que N uma variável aleatória discreta, e considere a distribuição condicional de X dado que $N = n$. Então,

$$\frac{P\{x < X < x + dx | N = n\}}{dx} = \frac{P\{N = n | x < X < x + dx\}}{P\{N = n\}} \frac{P\{x < X < x + dx\}}{dx}$$

e fazendo dx tender a 0, obtemos

$$\lim_{dx \to 0} \frac{P\{x < X < x + dx | N = n\}}{dx} = \frac{P\{N = n | X = x\}}{P\{N = n\}} f(x)$$

o que mostra que a densidade condicional de X dado que $N = n$ é dada por

$$f_{X|N}(x|n) = \frac{P\{N = n | X = x\}}{P\{N = n\}} f(x)$$

Exemplo 5c A distribuição normal bivariada

Uma das mais importantes distribuições conjuntas é a distribuição normal bivariada. Dizemos que as variáveis aleatórias X e Y têm distribuição normal bivariada se, para as constantes $\mu_x, \mu_y, \sigma_x > 0, \sigma_y > 0, -1 < \rho < 1$, sua função densidade conjunta é dada, para todo $-\infty < x, y < \infty$, por

$$f(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1 - \rho^2}} \exp\left\{ -\frac{1}{2(1 - \rho^2)} \left[\left(\frac{x - \mu_x}{\sigma_x} \right)^2 + \left(\frac{y - \mu_y}{\sigma_y} \right)^2 - 2\rho \frac{(x - \mu_x)(y - \mu_y)}{\sigma_x \sigma_y} \right] \right\}$$

Determinamos agora a densidade condicional de X dado que $Y = y$. Ao fazer isso, vamos coletar continuamente todos os fatores que não dependem de x e representá-los pelas constantes C_i. A constante final é então determinada usando-se $\int_{-\infty}^{\infty} f_{X|Y}(x|y)\, dx = 1$. Temos

$$\begin{aligned} f_{X|Y}(x|y) &= \frac{f(x, y)}{f_Y(y)} \\ &= C_1 f(x, y) \\ &= C_2 \exp\left\{ -\frac{1}{2(1 - \rho^2)} \left[\left(\frac{x - \mu_x}{\sigma_x} \right)^2 - 2\rho \frac{x(y - \mu_y)}{\sigma_x \sigma_y} \right] \right\} \end{aligned}$$

$$= C_3 \exp\left\{-\frac{1}{2\sigma_x^2(1-\rho^2)}\left[x^2 - 2x\left(\mu_x + \rho\frac{\sigma_x}{\sigma_y}(y-\mu_y)\right)\right]\right\}$$

$$= C_4 \exp\left\{-\frac{1}{2\sigma_x^2(1-\rho^2)}\left[x - \left(\mu_x + \rho\frac{\sigma_x}{\sigma_y}(y-\mu_y)\right)\right]^2\right\}$$

Reconhecendo a equação anterior como uma função de densidade normal, podemos concluir que, dado que $Y = y$, a variável aleatória X é normalmente distribuída com média $\mu_x + \rho\frac{\sigma_x}{\sigma_y}(y - \mu_y)$ e variância $\sigma_x^2(1 - \rho^2)$. Além disso, como a densidade conjunta de Y, X é exatamente igual à de X, Y, exceto que μ_x, σ_x são trocadas por μ_y, σ_y, tem-se similarmente que a distribuição condicional de Y dado que $X = x$ é uma distribuição normal com média $\mu_y + \rho\frac{\sigma_y}{\sigma_x}(x - \mu_x)$ e variância $\sigma_y^2(1 - \rho^2)$. Como consequência desses resultados, tem-se que a condição necessária e suficiente para que as variáveis aleatórias normais bivariadas X e Y sejam independentes é que $\rho = 0$ (um resultado que também é consequência direta de sua densidade conjunta, porque somente quando $\rho = 0$ a função densidade pode ser fatorada em dois termos, um dependendo apenas de x e o outro dependendo apenas de y).

Com $C = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}$, a densidade marginal de X pode ser obtida de

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

$$= C\int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 + \left(\frac{y-\mu_y}{\sigma_y}\right)^2 - 2\rho\frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}\right]\right\} dy$$

Fazendo a mudança de variáveis $w = \frac{y-\mu_y}{\sigma_y}$, obtemos

$$f_X(x) = C\sigma_y \exp\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{x-\mu_x}{\sigma_x}\right)^2\right\}$$
$$\times \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[w^2 - 2\rho\frac{x-\mu_x}{\sigma_x}w\right]\right\} dw$$

$$= C\sigma_y \exp\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{x-\mu_x}{\sigma_x}\right)^2 (1-\rho^2)\right\}$$
$$\times \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[w - \rho\frac{x-\mu_x}{\sigma_x}\right]^2\right\} dw$$

Como

$$\frac{1}{\sqrt{2\pi(1-\rho^2)}}\int_{-\infty}^{\infty}\exp\left\{-\frac{1}{2(1-\rho^2)}\left[w-\frac{\rho}{\sigma_x}(x-\mu_x)\right]^2\right\}dw=1$$

vemos que

$$\begin{aligned}f_X(x) &= C\sigma_y\sqrt{2\pi(1-\rho^2)}\,e^{-(x-\mu_x)^2/2\sigma_x^2}\\ &= \frac{1}{\sqrt{2\pi}\sigma_x}\,e^{-(x-\mu_x)^2/2\sigma_x^2}\end{aligned}$$

Isto é, X é normal com média μ_x e variância σ_x^2. Similarmente, Y é normal com média μ_y e variância σ_y^2. ■

Exemplo 5d

Considere $n + m$ tentativas com mesma probabilidade de sucesso. Suponha, no entanto, que essa probabilidade de sucesso não seja fixada antecipadamente, mas escolhida de uma população uniforme no intervalo $(0, 1)$. Qual é a distribuição condicional das probabilidades de sucesso dado que as $n + m$ tentativas resultam em n sucessos?

Solução Se X representa a probabilidade de que uma dada tentativa seja um sucesso, então X é uma variável aleatória uniforme no intervalo $(0,1)$. Também, dado que $X = x$, as $n + m$ tentativas são independentes com probabilidade de sucesso x. Com isso, N, o número de sucessos, é uma variável binomial com parâmetros $(n + m, x)$, e a densidade condicional de X dado que $N = n$ é

$$\begin{aligned}f_{X|N}(x|n) &= \frac{P\{N=n|X=x\}f_X(x)}{P\{N=n\}}\\ &= \frac{\binom{n+m}{n}x^n(1-x)^m}{P\{N=n\}} \qquad 0<x<1\\ &= cx^n(1-x)^m\end{aligned}$$

onde c não depende de x. Assim, a densidade condicional é igual à de uma variável aleatória beta com parâmetros $n + 1, m + 1$.

O resultado anterior é bastante interessante, pois ele diz que, se a distribuição original ou *a priori* (para aquele conjunto de dados) de uma probabilidade de sucesso em uma tentativa é uniformemente distribuída no intervalo $(0, 1)$ [ou, equivalentemente, é beta com parâmetros $(1, 1)$], então a distribuição posterior (ou condicional) dado que um total de n sucessos tenha ocorrido em $n + m$ tentativas é beta com parâmetros $(1 + n, 1 + m)$. Isto é valioso porque aumenta a nossa intuição sobre o que significa supor que uma variável aleatória tenha uma distribuição beta. ■