B

Fundamentos da Probabilidade

Este apêndice trata dos conceitos-chave de probabilidade básica. Os Apêndices B e C são essencialmente de recapitulação; eles não pretendem substituir um curso sobre probabilidade ou estatística. Porém, todos os conceitos sobre probabilidade e estatística que usamos neste livro são discutidos nesses apêndices.

A probabilidade por si só é de interesse dos estudiosos de negócios, economia e outras ciências sociais. Por exemplo, considere o problema de uma empresa aérea que esteja tentando decidir quantas reservas aceitar para um vôo com 100 lugares disponíveis. Se menos de 100 pessoas quiserem fazer reservas, então todas deverão ser aceitas. Mas e se mais de 100 pessoas solicitarem reserva? Uma solução segura seria aceitar no máximo 100 reservas. Porém, como algumas pessoas fazem reservas e não comparecem para o embarque, existe alguma probabilidade de que o avião não lote mesmo que sejam feitas 100 reservas. Isso resultará em perda de receita para a empresa aérea. Uma estratégia diferente seria aceitar mais de 100 reservas e esperar que algumas pessoas não compareçam para embarque, e assim o número final de passageiros seria o mais próximo possível de 100. Essa decisão corre o risco de a companhia aérea ter de compensar as pessoas que não puderam embarcar devido à venda de um número de assentos maior que o da capacidade do avião.

Uma questão natural nesse contexto é: podemos decidir sobre o número ótimo (ou o melhor) de reservas que a companhia aérea deveria fazer? Esse não é um problema trivial. Contudo, levando-se em consideração certas informações (sobre os custos da empresa aérea e a freqüência das pessoas deixarem de comparecer para o embarque), podemos usar probabilidade básica para chegar a uma solução.

B.1 VARIÁVEIS ALEATÓRIAS E SUA DISTRIBUIÇÕES DE PROBABILIDADE

Suponha que joguemos para o alto uma moeda dez vezes e contemos o número de vezes em que dê cara. Esse é um exemplo de um **experimento**. De forma geral, um experimento é qualquer procedimento que possa, pelo menos em teoria, ser repetido indefinidamente, e tem um conjunto de resultados bem definido. Poderíamos, em princípio, continuar tirando cara ou coroa repetidamente. Antes de atirarmos a moeda, sabemos que o número de caras que aparecerá será um inteiro entre 0 e 10, e, portanto, os resultados do experimento são bem definidos.

Uma **variável aleatória** é aquela que assume valores numéricos e tem um resultado que é determinado por um experimento. No exemplo da moeda, o número de caras que aparecerá em dez lances de uma moeda é um exemplo de uma variável aleatória. Antes de atirarmos a moeda dez vezes, não sabemos quantas vezes vai dar cara. Ao lançarmos a moeda dez vezes e contarmos o número de vezes que deu cara, obteremos o resultado da variável aleatória para essa particular verificação do experimento. Outra verificação poderá produzir um resultado diferente.

No exemplo das reservas da empresa aérea mencionado anteriormente, o número de pessoas que comparece para o embarque é uma variável aleatória: antes de qualquer vôo, não sabemos quantas pessoas comparecerão para embarque.

Para analisar os dados coletados em economia e nas ciências sociais, é importante ter-se um conhecimento básico das variáveis aleatórias e de suas propriedades. Seguindo as convenções tradicionais de probabilidade e estatística, ao longo dos Apêndices B e C, representaremos as variáveis aleatórias com letras maiúsculas, em geral W, X, Y e Z; os resultados particulares das variáveis aleatórias são representados pelas minúsculas correspondentes, w, x, y e z. Por exemplo, no experimento da moeda, seja X o número de vezes que apareceu cara em dez lances da moeda. Nesse caso, X não está associado com qualquer valor em particular, mas sabemos que X assumirá um valor do conjunto $\{0, 1, 2, ..., 10\}$. Um resultado particular seria, digamos, $x = 6$.

Indicamos coleções grandes de variáveis aleatórias pelo uso de subscritos. Por exemplo, se registrarmos a renda do ano passado de 20 famílias escolhidas aleatoriamente nos Estados Unidos, poderemos representar essas variáveis aleatórias por $X_1, X_2, ..., X_{20}$; os resultados particulares seriam representados por $x_1, x_2, ..., x_{20}$.

Como afirmado na definição, as variáveis aleatórias sempre são estabelecidas para assumir valores numéricos, mesmo quando descrevem eventos qualitativos. Por exemplo, considere jogar uma única moeda, na qual os dois resultados são cara e coroa. Podemos definir uma variável aleatória da seguinte forma: $X = 1$ se der cara, e $X = 0$ se der coroa.

Uma variável aleatória que somente pode assumir os valores zero e um é chamada **variável aleatória de Bernoulli** (ou **binária**). Em probabilidade básica, é tradição chamar o evento $X = 1$ de "sucesso" e o evento $X = 0$ de "fracasso". A nomenclatura sucesso-fracasso pode não corresponder à nossa noção de sucesso e fracasso em determinadas aplicações, mas é uma terminologia útil que adotaremos.

Variáveis Aleatórias Discretas

Uma **variável aleatória discreta** é a que somente assume um número finito ou infinito enumerável de valores. A noção de "infinito enumerável" significa que, embora um número infinito de valores possa ser assumido por uma variável aleatória, esses valores podem ser postos em uma correspondência um-a-um com os números inteiros positivos. Como a distinção entre "infinito enumerável" e "infinito não-enumerável" é um pouco sutil, nos concentraremos nas variáveis aleatórias discretas que assumem somente um número finito de valores. Larsen e Marx (1986, Capítulo 3) apresentam uma abordagem detalhada sobre o assunto.

Uma variável aleatória de Bernoulli é o exemplo mais simples de uma variável aleatória discreta. A única coisa que precisamos para descrever completamente o comportamento de uma variável aleatória de Bernoulli é a probabilidade que ela assume no valor um. No exemplo da moeda, se ela for "justa", então, $P(X = 1) = 1/2$ (lê-se como "a probabilidade de que X seja igual a um é de 0,5). Como a soma das probabilidades deve ser igual à unidade, $P(X = 0) = 1/2$.

Os cientistas sociais estão interessados em mais do que cara ou coroa, e, portanto, devemos considerar situações mais gerais. Novamente, considere o exemplo em que a empresa aérea tem de decidir quantas reservas aceitar para um vôo com 100 lugares disponíveis. Esse problema pode ser analisado no contexto de diversas variáveis aleatórias de Bernoulli da seguinte maneira: para um passageiro selecionado aleatoriamente, defina uma variável aleatória de Bernoulli como $X = 1$ se a pessoa aparecer para embarque, e $X = 0$ se não aparecer.

Não há nenhuma razão para pensar que a probabilidade de qualquer passageiro em particular comparecer para embarque ser 1/2; em princípio, a probabilidade pode ser qualquer número entre zero e um. Chame esse número θ, de forma que

$$P(X = 1) = \theta \tag{B.1}$$

$$P(X = 0) = 1 - \theta. \tag{B.2}$$

Por exemplo, se $\theta = 0{,}75$, existirá 75% de probabilidade de que um passageiro apareça para o embarque após ter feito a reserva e 25% de probabilidade de que o passageiro não apareça. Intuitivamente, o valor de θ é fundamental para determinar a estratégia da companhia aérea quanto à aceitação de reservas. Os métodos para estimar θ, considerando os dados históricos de reservas das companhias aéreas, são tópicos de estatística matemática, que veremos no Apêndice C.

De forma mais geral, qualquer variável aleatória discreta é completamente descrita listando seus possíveis valores e a probabilidade associada que ela assume para cada valor. Se X assumir os k possíveis valores $\{x_1, ..., x_k\}$, as probabilidades $p_1, p_2, ..., p_k$ serão definidas por

$$p_j = P(X = x_j), j = 1, 2, ..., k, \tag{B.3}$$

onde cada p_j estará entre zero e um e

$$p_1 + p_2 + ... + p_k = 1. \tag{B.4}$$

A equação (B.3) é lida como: "A probabilidade de X assumir o valor x_j é igual a p_j".

As equações (B.1) e (B.2) mostram que as probabilidades de sucesso e fracasso de uma variável aleatória de Bernoulli são determinadas inteiramente pelo valor de θ. Como as variáveis aleatórias de Bernoulli são tão freqüentes, temos uma notação especial para elas: $X \sim$ Bernoulli(θ) é lida como "X tem uma distribuição de Bernoulli com probabilidade de sucesso igual a θ."

A **função densidade de probabilidade** (**fdp**) de X resume as informações relativas aos possíveis resultados de X e as probabilidades correspondentes:

$$f(x_j) = p_j, j = 1, 2, ..., k, \tag{B.5}$$

com $f(x) = 0$ de qualquer x não igual a x_j para algum j. Em outras palavras, para qualquer número real x, $f(x)$ será a probabilidade que a variável aleatória X assumirá para o valor particular de x. Quando lidamos com mais de uma variável aleatória, algumas vezes é útil subscrever a fdp em questão: f_X é a fdp de X, f_Y é a fdp de Y, e assim por diante.

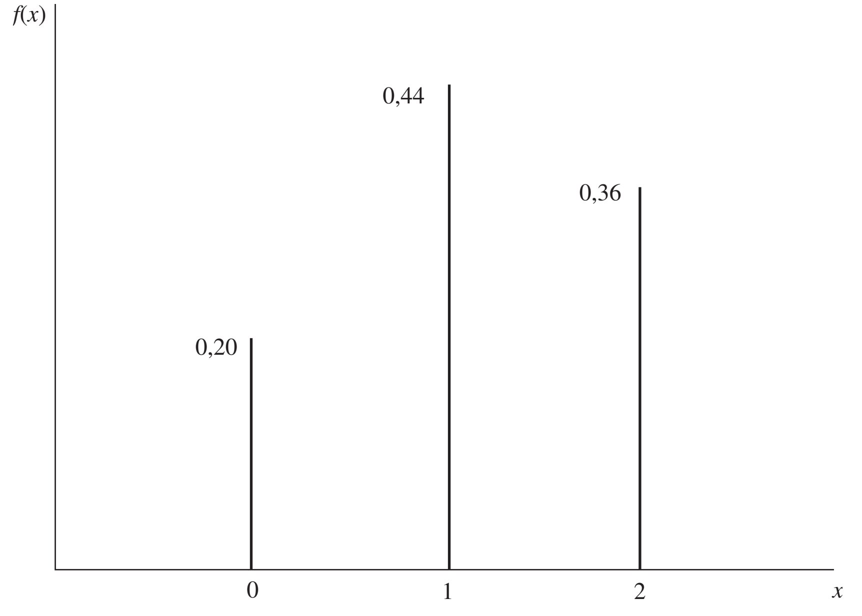
Dada a fdp de qualquer variável aleatória discreta, é simples calcular a probabilidade de qualquer evento envolvendo aquela variável aleatória. Por exemplo, suponha que X seja o número de pontos feitos por um jogador de basquetebol a cada dois lances livres, de forma que X pode assumir os três valores $\{0,1,2\}$. Assuma que a fdp de X seja dada por

$$f(0) = 0{,}20, f(1) = 0{,}44, \text{ e } f(2) = 0{,}36.$$

A soma das três probabilidades é igual a um, como deveria ser. Usando essa fdp, podemos calcular a probabilidade de que o jogador converta *pelo menos* um lance livre: $P(X \geq 1) = P(X = 1) + P(X = 2) = 0{,}44 + 0{,}36 = 0{,}80$. A fdp de X é mostrada na Figura B.1.

Figura B.1

A fdp do número de lances livres convertidos a cada duas tentativas.



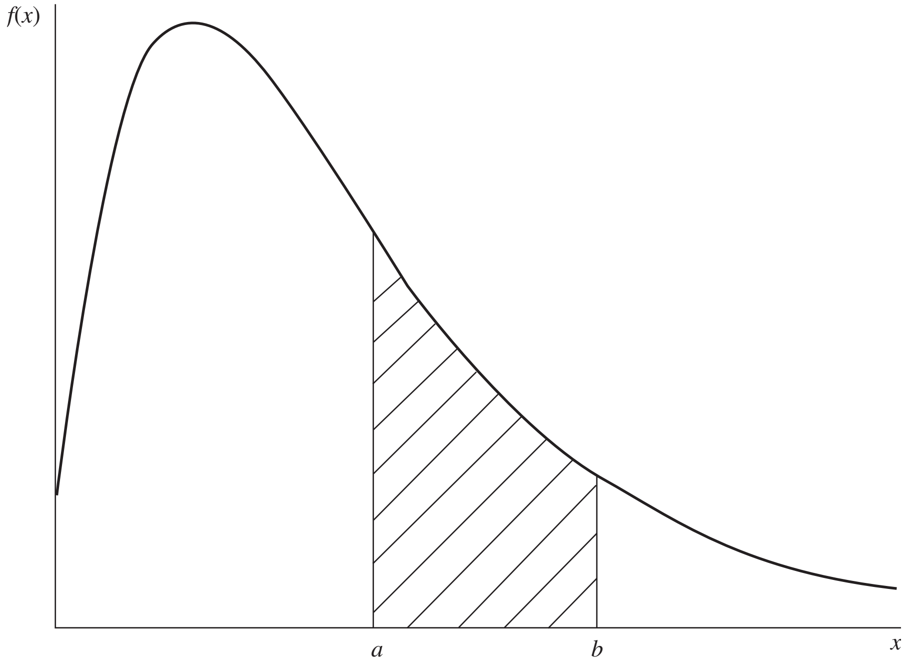
Variáveis Aleatórias Contínuas

Uma variável X será uma **variável aleatória contínua** se assumir qualquer valor real com probabilidade zero. Essa definição é um tanto quanto contra-intuitiva, já que em qualquer aplicação acabaremos observando algum resultado para uma variável aleatória. A idéia é que uma variável aleatória contínua X pode assumir tantos valores possíveis que não podemos enumerá-los ou compará-los com os inteiros positivos, de modo que a consistência lógica garante que X pode assumir cada valor com probabilidade zero. Embora as medidas sejam sempre discretas na prática, as variáveis aleatórias que assumem numerosos valores são melhor tratadas como contínuas. Por exemplo, a medida mais refinada do preço de um bem é em termos de centavos. Podemos nos imaginar relacionando todos os possíveis valores de preços ordenadamente (mesmo que a lista possa continuar indefinidamente), o que tecnicamente faz com que preço seja uma variável aleatória discreta. Porém, existem tantos valores possíveis de preços que o uso da mecânica das variáveis aleatórias discretas não é viável.

Podemos definir uma função densidade de probabilidade para variáveis aleatórias contínuas, e, como acontece com as variáveis aleatórias discretas, a fdp fornecerá informações sobre os prováveis resultados da variável aleatória. Porém, como também não faz sentido discutir a probabilidade de que uma variável aleatória contínua assuma um valor em particular, usamos a fdp de uma variável aleatória contínua somente para computar eventos envolvendo uma diversidade de valores. Por exemplo, se a e b forem constantes onde $a < b$, a probabilidade de X estar entre os números a e b, $P(a \leq X \leq b)$, será a *área* sob a fdp entre os pontos a e b, como mostrado na Figura B.2. Se você estiver familiarizado com cálculo diferencial, você reconhecerá isso como a *integral* da função f entre os pontos a e b. A área total sob a fdp deve sempre ser igual a um.

Figura B.2

A probabilidade que X esteja entre os pontos *a* e *b*.



Ao computar probabilidades para variáveis aleatórias contínuas, é mais fácil trabalhar com a **função de distribuição cumulativa** (**fdc**). Se X for qualquer variável aleatória, então, sua fdc será definida por qualquer numero x real pela equação

$$F(x) \equiv \mathrm{P}(X \leq x). \tag{B.6}$$

Para variáveis aleatórias discretas, (B.6) será obtida somando a fdp para todos os valores x_j tais que $x_j \leq x$. Para uma variável aleatória contínua, $F(x)$ será a área sob a fdp, f, à esquerda do ponto x. Como $F(x)$ é simplesmente uma probabilidade, ela estará sempre entre 0 e 1. Além disso, se $x_1 < x_2$, então, $\mathrm{P}(X \leq x_1) \leq \mathrm{P}(X \leq x_2)$, isto é, $F(x_1) \leq F(x_2)$. Isso significa que uma fdc é uma função crescente (ou pelo menos não-decrescente) de x.

Duas propriedades importantes das fdcs que são úteis no cálculo de probabilidades são as seguintes:

$$\text{Para qualquer número } c,\ \mathrm{P}(X > c) = 1 - F(c). \tag{B.7}$$

$$\text{Para quaisquer números } a < b,\ \mathrm{P}(a < X \leq b) = F(b) - F(a). \tag{B.8}$$

Em nosso estudo da econometria, usaremos as fdcs para calcular probabilidades somente de variáveis aleatórias contínuas, caso em que não importa se as desigualdades nas especificações probabilísticas são estritas ou não. Ou seja, para uma variável aleatória contínua X,

$$P(X \geq c) = P(X > c) \tag{B.9}$$

e

$$P(a < X < b) = P(a \leq X \leq b) = P(a \leq X < b) = P(a < X \leq b). \tag{B.10}$$

Combinadas com (B.7) e (B.8), as equações (B.9) e (B.10) expandem bastante os cálculos de probabilidade que podem ser feitos com o uso de fdcs contínuas.

As funções de distribuições cumulativas foram tabuladas para todas as distribuições contínuas importantes em probabilidade e estatística. A mais conhecida delas é a distribuição normal, da qual trataremos com algumas distribuições relacionadas na Seção B.5

B.2 DISTRIBUIÇÕES CONJUNTAS, DISTRIBUIÇÕES CONDICIONAIS E INDEPENDÊNCIA

Em economia, geralmente estamos interessados na ocorrência de eventos que envolvem mais de uma variável aleatória. No exemplo das reservas da companhia aérea anteriormente referido, esta pode estar interessada na probabilidade de que uma pessoa que faz uma reserva compareça para embarque *e* que seja uma pessoa que viaje a negócios; esse é um exemplo de uma *probabilidade conjunta*. Ou a empresa aérea pode estar interessada na seguinte *probabilidade condicional*: condicional à pessoa ser uma pessoa que viaje a negócios, qual é a probabilidade de que ela compareça para embarque? Nas próximas duas subseções, formalizaremos a noção de distribuições conjunta e condicional e a importante noção de *independência* das variáveis aleatórias.

Distribuições Conjuntas e Independência

Sejam X e Y variáveis aleatórias discretas. Então, (X,Y) têm uma **distribuição conjunta**, que é totalmente definida pela *função densidade de probabilidade conjunta* de (X,Y):

$$f_{X,Y}(x,y) = P(X = x, Y = y), \tag{B.11}$$

onde o lado direito é a probabilidade de que $X = x$ e $Y = y$. Quando X e Y são contínuas, uma fdp conjunta também pode ser definida, mas não trataremos de tais detalhes, pois fdps conjuntas de variáveis aleatórias contínuas não são explicitamente usadas neste livro.

Em um caso, é fácil obter a fdp conjunta se forem dadas as fdps de X e Y. Em particular, as variáveis aleatórias X e Y são independentes se, e somente se,

$$f_{X,Y}(x,y) = f_X(x)f_y(y) \tag{B.12}$$

para todos os x e y, quando f_X for a fdp de X e f_Y for a fdp de Y. No contexto de mais de uma variável aleatória, as fdps f_X e f_Y são freqüentemente chamadas *funções de densidade de probabilidade marginal* para distingui-las da fdp conjunta $f_{X,Y}$. Essa definição de independência é válida para variáveis aleatórias discretas e contínuas.

Para entendermos o significado de (B.12) é mais fácil lidar com o caso discreto. Se X e Y forem discretas, então, (B.12) será a mesma coisa que

$$P(X = x, Y = y) = P(X = x)P(Y = y); \qquad \textbf{(B.13)}$$

em outras palavras, a probabilidade de que $X = x$ e $Y = y$ é o produto das duas probabilidades $P(X = x)$ e $P(Y = y)$. Uma implicação de (B.13) é que as probabilidades conjuntas são razoavelmente fáceis de serem calculadas, já que elas apenas exigem o conhecimento de $P(X = x)$ e $P(Y = y)$.

Se as variáveis aleatórias não forem independentes, então, elas são *dependentes*.

EXEMPLO B.1

(Arremessos de Lances Livres)

Considere um jogador de basquetebol fazendo dois lances livres. Seja X uma variável aleatória de Bernoulli igual a um se ele converter o primeiro arremesso, e zero, caso contrário. Seja Y uma variável aleatória de Bernoulli igual a um se ele converter o segundo arremesso. Suponha que ele seja um jogador que converte 80% dos arremessos, de forma que $P(X = 1) = P(Y = 1) = 0{,}8$. Qual é a probabilidade de o jogador converter os dois arremessos?

Se X e Y forem independentes, podemos facilmente responder a essa pergunta: $P(X = 1, Y = 1) = P(X = 1)(P(Y = 1) = (0{,}8)(0{,}8) = 0{,}64$. Portanto, existe 64% de probabilidade de converter ambos os lances livres. Se a probabilidade de converter o segundo arremesso depender de o primeiro arremesso ter sido convertido — isto é, X e Y não são independentes — então, esse cálculo simples não será válido.

A independência de variáveis aleatórias é um conceito muito importante. Na próxima subseção, mostraremos que, se X e Y forem independentes, conhecer resultado de X não altera as probabilidades dos possíveis resultados de Y, e vice-versa. Um fato útil sobre a questão da independência é que se X e Y forem independentes e definirmos novas variáveis aleatórias $g(X)$ e $h(Y)$ para quaisquer funções g e h, então, essas novas variáveis aleatórias também serão independentes.

Não há necessidade de parar em duas variáveis aleatórias. Se $X_1, X_2, ..., X_n$ forem variáveis aleatórias discretas, então, suas fdps conjuntas serão $f(x_1, x_2, ..., x_n) = P(X_1 = x_1, X_2 = x_2, ..., X_n = x_n)$. As variáveis aleatórias $X_1, X_2, ..., X_n$ serão **variáveis aleatórias independentes** se, e somente se, suas fdps conjuntas forem o produto das fdps individuais para quaisquer $(x_1, x_2, ..., x_n)$. Essa definição de independência também é válida para variáveis aleatórias contínuas.

A noção de independência desempenha um papel importante na obtenção de algumas distribuições clássicas em probabilidade e estatística. Anteriormente, definimos uma variável aleatória de Bernoulli como uma variável aleatória zero-um indicando se ocorre algum evento ou não. Freqüentemente, estamos interessados no número de sucessos em uma seqüência de ensaios de Bernoulli *independentes*.

Um exemplo padrão de ensaios de Bernoulli independentes é jogar repetidamente uma moeda. Como o resultado de qualquer lance particular não tem nada a ver com os resultados dos outros lances, a independência é uma hipótese apropriada.

A independência é muitas vezes uma aproximação razoável em situações mais complicadas. No exemplo das reservas da companhia aérea, suponha que a companhia aceite n reservas para determinado vôo. De cada $i = 1, 2, ..., n$, seja Y a variável aleatória de Bernoulli indicando se o passageiro i aparece para embarque: $Y_i = 1$ se o passageiro i aparecer para embarque, e $Y_i = 0$, caso contrário. Definindo θ novamente como a probabilidade de sucesso (usando as reservas), cada Y_i terá uma distribuição de Bernoulli (θ). Como uma aproximação, podemos assumir que os Y_i são independentes entre si, embora isso não seja exatamente verdadeiro na realidade: algumas pessoas viajam em grupo, o que significa que se uma pessoa comparecerá ou não para embarque não é verdadeiramente independente de se as outras pessoas comparecerão ou não. Porém, modelar esse tipo de dependência é complexo, de modo que podemos querer usar a independência como uma aproximação.

A variável de interesse principal é o número total de passageiros que comparecem para embarque das n reservas; chamemos essa variável de X. Como cada Y_i será igual à unidade quando uma pessoa comparece para embarque, podemos escrever $X = Y_1 + Y_2 + ... + Y_n$. Agora, assumindo que cada Y_i tem probabilidade de sucesso θ e que os Y_i são independentes, é possível mostrar que X tem uma **distribuição binomial**. Isto é, a função densidade de probabilidade de X é

$$f(x) = \binom{n}{x}\theta^x(1 - \theta)^{n-x}, \; x = 0, 1, 2, ..., n, \tag{B.14}$$

onde $\binom{n}{x} = \frac{n!}{x!(n-x)!}$, e para qualquer inteiro n, $n!$ (lê-se "fatorial de n") é definido como $n! = n \cdot (n-1)(n-2)...1$. Por convenção, $0! = 1$. Quando uma variável aleatória X tem a fdp dada em (B.14), escrevemos $X \sim \text{Binomial}(n,\theta)$. A equação (B.14) pode ser usada para calcular $P(X = x)$ para qualquer valor de x de 0 a n.

Se o vôo tiver 100 lugares disponíveis, a empresa aérea estará interessada em $P(X > 100)$. Suponha, inicialmente, que $n = 120$, de modo que a companhia aérea aceitará 120 reservas, e que a probabilidade de que cada pessoa compareça para embarque seja $\theta = 0{,}85$. Então, $P(X > 100) = P(X = 101) + P(X = 102) + ... + P(X = 120)$, e cada uma das probabilidades na soma poderá ser encontrada pela equação (B.14) com $n = 120$, $\theta = 0{,}85$ e o valor apropriado de x (101 a 120). Esse é um cálculo difícil de ser feito manualmente, mas muitos programas estatísticos possuem comandos para computar esse tipo de probabilidade. Nesse caso, a probabilidade de que mais de 100 pessoas comparecerão para embarque é cerca de 0,659, o que provavelmente é um risco de excesso de reservas maior do que a companhia aérea deseja tolerar. Se, em vez disso, o número de reservas for 110, a probabilidade de que mais de 100 pessoas comparecerão para embarque será de apenas 0,024.

Distribuições Condicionais

Em econometria, geralmente estamos interessados em como uma variável aleatória, vamos chamá-la Y, está relacionada com uma ou mais das outras variáveis. Por enquanto, suponha que haja somente uma variável em cujos efeitos estamos interessados, vamos chamá-la X. O máximo que podemos saber sobre como X afeta Y está contido na **distribuição condicional** da Y, dado X. Essa informação é resumida pela *função de densidade de probabilidade condicional*, definida por

$$f_{Y|X}(y|x) = f_{X,Y}(x,y)/f_X(x) \tag{B.15}$$

para todos os valores de x de tal forma que $f_X(x) > 0$. A interpretação de (B.15) é mais fácil de ser vista quando X e Y são discretas. Então,

$$f_{Y|X}(y|x) = \mathrm{P}(Y = y|X = x), \tag{B.16}$$

onde o lado direito é lido como "a probabilidade de $Y = y$ em decorrência de $X = x$." Quando Y é contínuo, $f_{Y|X}(y|x)$ não é interpretada diretamente como uma probabilidade, pelas razões explicadas anteriormente, mas as probabilidades condicionais são encontradas computando áreas sob a fdp condicional.

Uma característica importante das distribuições condicionais é que, se X e Y forem variáveis aleatórias independentes, o conhecimento dos valores assumidos por X não nos diz nada sobre a probabilidade de que Y assuma diversos valores (e vice-versa). Isto é, $f_{Y|X}(y|x) = f_Y(y)$ e $f_{X|Y}(x|y) = f_X(x)$.

EXEMPLO B.2

(Arremessos de Lances Livres)

Considere novamente o exemplo dos lances livres no basquetebol, quando dois lances livres devem ser tentados. Assuma que a densidade condicional seja

$$f_{Y|X}(1|1) = 0{,}85, f_{Y|X}(0|1) = 0{,}15$$
$$f_{Y|X}(1|0) = 0{,}70, f_{Y|X}(0|0) = 0{,}30.$$

Isso significa que a probabilidade de o jogador converter o segundo lance depende de o primeiro lance ter sido convertido: se o primeiro lance foi convertido, a probabilidade de converter o segundo lance é de 0,85; se o primeiro lance foi perdido, a probabilidade de converter o segundo lance é de 0,70. Isso implica que X e Y não são independentes; eles são dependentes.

Ainda podemos computar $\mathrm{P}(X = 1, Y = 1)$ desde que conheçamos $\mathrm{P}(X = 1)$. Assuma que a probabilidade de converter o primeiro lance livre seja 0,8, isto é, $\mathrm{P}(X = 1) = 0{,}8$. Então, de (B.15), teremos

$$\mathrm{P}(X = 1, Y = 1) = \mathrm{P}(Y = 1|X = 1)\cdot\mathrm{P}(X = 1) = (0{,}85)(0{,}8) = 0{,}68.$$

B.3 CARACTERÍSTICAS DAS DISTRIBUIÇÕES DE PROBABILIDADE

Para muitos propósitos, estaremos interessados em somente alguns poucos aspectos das distribuições das variáveis aleatórias. As características de interesse podem ser classificadas em três categorias: medidas de tendência central, medidas de variabilidade ou intervalo e medidas de associação entre duas variáveis aleatórias. Trataremos desta última na Seção B.4.

Uma Medida de Tendência Central: O Valor Esperado

O valor esperado é um dos mais importantes conceitos da probabilidade que encontraremos em nosso estudo da econometria. Se X for uma variável aleatória, o **valor esperado** (ou esperança) de X, representado por $\mathrm{E}(X)$ e algumas vezes por μ_X, ou simplesmente μ, é uma média ponderada de todos os possíveis valores de X. Os pesos são determinados pela função de densidade de probabilidade. Algumas vezes, o valor esperado é chamado *média populacional*, especialmente quando queremos enfatizar que X representa alguma variável em uma população.

A definição precisa do valor esperado é mais simples no caso em que X é uma variável aleatória discreta assumindo um número finito de valores, digamos, $\{x_1, ..., x_k)$. Seja $f(x)$ a função de densidade de probabilidade de X. O valor esperado de X será a média ponderada

$$E(X) = x_1 f(x_1) + x_2 f(x_2) + ... + x_k f(x_k) \equiv \sum_{j=1}^{k} x_j f(x_j) \tag{B.17}$$

Essa expressão é facilmente calculada dados os valores da fdp de cada possível resultado de X.

EXEMPLO B.3

(Calculando um Valor Esperado)

Suponha que X assuma os valores -1, 0 e 2 com probabilidades 1/8, 1/2 e 3/8, respectivamente. Então,

$$E(X) = (-1)\cdot(1/8) + 0\cdot(1/2) + 2\cdot(3/8) = 5/8.$$

Esse exemplo ilustra uma coisa curiosa sobre os valores esperados: o valor esperado de X pode ser um número que não é sequer um possível resultado de X. Sabemos que X assume os valores -1, 0 e 2, ainda que seu valor esperado seja 5/8. Isso torna o valor esperado deficiente para resumir a tendência central de certas variáveis aleatórias discretas, mas cálculos como o que acabamos de mostrar podem ser úteis, como veremos mais tarde.

Se X for uma variável aleatória contínua, então, $E(X)$ será definido como uma integral:

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx, \tag{B.18}$$

que assumimos como bem definida. Isso ainda pode ser interpretado como uma média ponderada. Para as distribuições contínuas mais comuns, $E(X)$ é um número que é um possível resultado de X. Neste livro, não precisaremos calcular valores esperados usando integração, embora utilizemos de alguns resultados bem conhecidos de probabilidade de valores esperados de variáveis aleatórias especiais.

Dada uma variável aleatória X e uma função $g(\cdot)$, podemos criar uma nova variável aleatória $g(X)$. Por exemplo, se X for uma variável aleatória, então, X^2 e $\log(X)$ (se $X > 0$) também serão variáveis aleatórias. O valor esperado de $g(X)$ será, de novo, simplesmente uma média ponderada:

$$E[g(X)] = \sum_{j=1}^{k} g(x_j) f_X(x_j) \tag{B.19}$$

ou, para uma variável aleatória contínua,

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)\, f_X(x)dx. \quad \textbf{(B.20)}$$

EXEMPLO B.4

(Valor Esperado de X^2)

Para a variável aleatória no Exemplo B.3, seja $g(X) = X^2$. Então,

$$E(X^2) = (-1)^2(1/8) + (0)^2(1/2) + (2)^2(3/8) = 13/8.$$

No Exemplo B.3, calculamos $E(X) = 5/8$, de forma que $[E(X)]^2 = 25/64$. Isso mostra que $E(X^2)$ *não* é o mesmo que $[E(X)]^2$. De fato, para uma função não-linear $g(X)$, $E[g(X)] \neq g[E(X)]$ (exceto em casos muito especiais).

Se X e Y forem variáveis aleatórias, então, $g(X,Y)$ será uma variável aleatória para qualquer função g, e assim poderemos definir sua esperança. Quando X e Y são ambas discretas, assumindo os valores $\{x_1, x_2, ..., x_k\}$ e $\{y_1, y_2, ..., y_m\}$, respectivamente, o valor esperado será

$$E[g(X, Y)] = \sum_{h=1}^{k} \sum_{j=1}^{m} g(x_h,y_j)\, f_{X,Y}(x_h,y_j),$$

onde $f_{X,Y}$ será a fdp conjunta de (X,Y). A definição é mais complicada para variáveis aleatórias contínuas, pois envolve integração; não precisamos dela aqui. A extensão para mais de duas variáveis aleatórias é fácil de ser feita.

Propriedades dos Valores Esperados

Em econometria, não nos preocupamos muito em calcular os valores esperados de diversas distribuições; os cálculos principais já foram feitos muitas vezes, e em grande parte os aceitaremos sem questionar. Teremos que manipular alguns valores esperados usando umas poucas regras simples. Elas são tão importantes que as rotulamos:

PROPRIEDADE E.1
Para qualquer constante c, $E(c) = c$.

PROPRIEDADE E.2
Para quaisquer constantes a e b, $E(aX + b) = aE(X) + b$.

Uma implicação útil de E.2 é que, se $\mu = E(X)$, e definirmos uma nova variável aleatória como $Y = X - \mu$, então, $E(Y) = 0$; em E.2, considere $a = 1$ e $b = -\mu$.

Como um exemplo da propriedade E.2, seja X a temperatura medida em graus Celsius, ao meio dia de determinado dia, em determinada localidade; suponha que a temperatura esperada seja $E(X) = 25$. Se Y for a temperatura medida em graus Fahrenheit, então, $Y = 32 + (9/5)X$. Pela propriedade E.2, a temperatura esperada em Fahrenheit será $E(Y) = 32 + (9/5)\cdot E(X) = 32 + (9/5)\cdot 25 = 77$.

De forma geral, é fácil calcular o valor esperado de uma função linear de diversas variáveis aleatórias.

PROPRIEDADE E.3

Se $\{a_1, a_2, ..., a_n\}$ forem constantes e $\{X_1, X_2, ..., X_n\}$ forem variáveis aleatórias, então,

$$E(a_1X_1 + a_2X_2 + ... + a_nX_n) = a_1E(X_1) + a_2E(X_2) + ... + a_nE(X_n).$$

Ou, usando a notação de somatórios,

$$E(\sum_{i=1}^{n} a_iX_i) = \sum_{i=1}^{n} a_iE(X_i). \quad \textbf{(B.21)}$$

Como um caso especial dessa equação, temos (com cada $a_1 = 1$)

$$E(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} E(X_i), \quad \textbf{(B.22)}$$

de forma que o valor esperado da soma será a soma dos valores esperados. Essa propriedade é usada com freqüência para derivações em estatística matemática.

EXEMPLO B.5

(Encontrando a Receita Esperada)

Sejam X_1, X_2 e X_3 os números de pizzas pequenas, médias e grandes, respectivamente, vendidas durante o dia em uma pizzaria. Elas são variáveis aleatórias com valores esperados $E(X_1) = 25$, $E(X_2) = 57$ e $E(X_3) = 40$. Os preços das pizzas pequena, média e grande são 5,50, 7,60 e 9,15 (em dólares). Portanto, a receita esperada das vendas de pizzas em determinado dia será

$$\begin{aligned} E(5{,}50\,X_1 + 7{,}60\,X_2 + 9{,}15\,X_3) &= 5{,}50\,E(X_1) + 7{,}60\,E(X_2) + 9{,}15\,E(X_3) \\ &= 5{,}50(25) + 7{,}60(57) + 9{,}15(40) = 936{,}70, \end{aligned}$$

isto é, 936,70 dólares. A receita efetiva de qualquer dia particular geralmente será diferente desse valor, mas essa é a receita *esperada*.

Também podemos usar a Propriedade E.3 para mostrar que se X ~ Binomial(n,θ), então, $E(X) = n\theta$. Ou seja, o número esperado de sucessos em n ensaios de Bernoulli é simplesmente o número de ensaios vezes a probabilidade de sucesso de qualquer ensaio particular. Isso será facilmente observado escrevendo X como $X = Y_1 + Y_2 + ... Y_n$, onde cada $Y_i \sim$ Bernoulli(θ). Então,

$$E(X) = \sum_{i=1}^{n} E(Y_i) = \sum_{i=1}^{n} \theta = n\theta.$$

Podemos aplicar esse resultado no exemplo das reservas da companhia aérea, quando ela aceita $n = 120$ reservas, e a probabilidade de comparecimento para embarque é $\theta = 0{,}85$. O número *esperado* de pessoas comparecendo para embarque é $120(0{,}85) = 102$. Portanto, se há 100 lugares disponíveis, o número esperado de pessoas que comparecerão para embarque é grande demais; isso terá alguma influência na conclusão de ser uma boa idéia a companhia aceitar 120 reservas.

Na realidade, o que a companhia aérea poderia fazer seria definir uma função do lucro que considerasse a receita ganha por lugar vendido e o custo por passageiro que seja impedido de embarcar. Essa função do lucro será aleatória, pois o número efetivo de pessoas que comparecerão para embarque é aleatório. Seja r a receita líquida correspondente a cada passageiro. (Para simplificar, você pode pensar nisso como sendo o preço da passagem). Seja i a indenização devida a cada passageiro que não puder embarcar. Nem r nem i são aleatórios; eles são assumidos como conhecidos pela companhia aérea. Seja Y o lucro do vôo. Então, com 100 lugares disponíveis,

$$
\begin{aligned}
Y &= rX \text{ se } X \leq 100 \\
&= 100r - i(X - 100) \text{ se } X > 100.
\end{aligned}
$$

A primeira equação mostra o lucro se não mais que 100 pessoas comparecerem para embarque; a segunda equação é o lucro se mais de 100 pessoas comparecerem para embarque. (Nesse último caso, a receita líquida da venda de passagens é $100r$, pois foram vendidos todos os 100 lugares, e, então, $i(X - 100)$ é o custo de aceitar mais de 100 reservas). Usando o fato de que X tem uma distribuição Binomial (n,0,85), onde n é o número de reservas feitas, os lucros esperados, E(Y) poderão ser encontrados como uma função de n (e r e i). Calcular E(Y) diretamente seria muito difícil, mas poderá ser encontrado rapidamente usando-se um computador. Uma vez os valores de r e i tenham sido dados, o valor de n que maximiza os lucros esperados poderá ser encontrado pesquisando-se diferentes valores de n.

Outra Medida de Tendência Central: A Mediana

O valor esperado é somente uma possibilidade para definir a tendência central de uma variável aleatória. Outra medida de tendência central é a **mediana**. Uma definição geral de mediana é complicada demais para nosso propósito. Se X for uma variável contínua, então, a mediana de X, digamos m, será um valor tal que metade da área de uma fdp está à esquerda de m, e a outra metade está à direita de m.

Quando X for uma variável discreta e assumir um número ímpar finito de valores, a mediada será obtida ordenando-se os possíveis valores de X e então selecionando-se o valor que estiver no centro. Por exemplo, se X assumir os valores $\{-4,0,2,8,10,13,17\}$, então, o valor mediano de X será 8. Se X assumir um número par de valores, existirão, na realidade, dois valores medianos; algumas vezes calcula-se a média desses números para obter um único valor mediano. Assim, se X assumir os valores $\{-5,3,9,17\}$, os valores medianos serão 3 e 9; se calcularmos a média desses números obteremos uma mediana igual a 6.

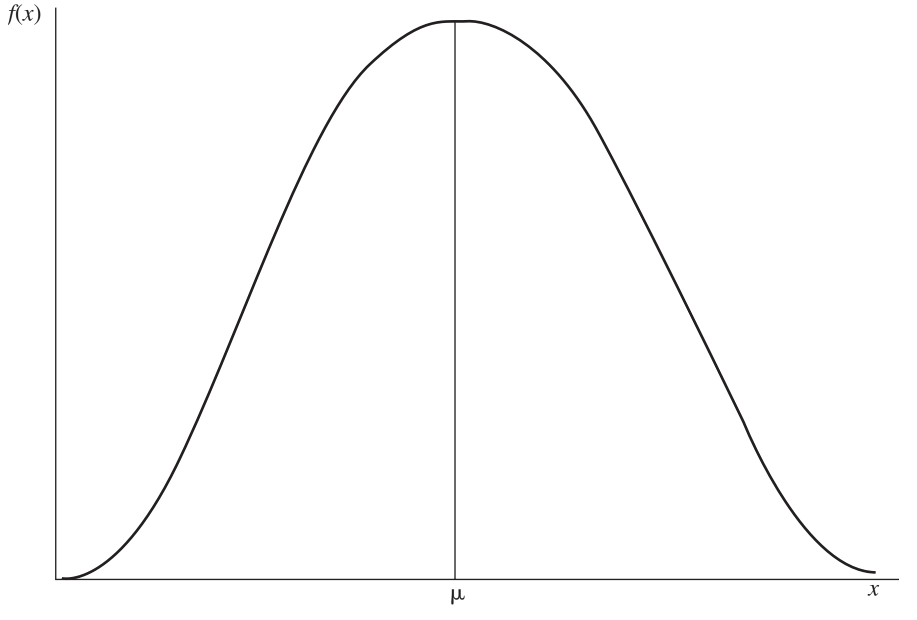
Em geral, a mediana, algumas vezes indicada por Med(X), e o valor esperado E(X), são diferentes. Nenhum é "melhor" que o outro como uma medida de tendência central; ambos são maneiras válidas de indicar o centro da distribuição de X. Em um caso especial, a mediana e o valor esperado (ou média) são os mesmos. Se X tiver uma **distribuição simétrica** em torno do valor μ, então, μ será tanto o valor esperado como a mediana. Matematicamente, a condição será $f(\mu + x) = f(\mu - x)$ para todo x. Esse caso está ilustrado na Figura B.3.

Medidas de Variabilidade: Variância e Desvio-Padrão

Embora a tendência central de uma variável aleatória seja valiosa, ela não nos diz tudo que queremos saber sobre a distribuição de uma variável aleatória. A Figura B.4 mostra as fdp de duas variáveis aleatórias com a mesma média. Claramente, a distribuição de X é mais concentrada em relação à sua média que a distribuição de Y. Gostaríamos de ter uma maneira simples de resumir isso.

Figura B.3

Uma distribuição de probabilidade simétrica.



Variância

Para uma variável aleatória X, seja $\mu = E(X)$. Há várias maneiras de medir o quanto X está distante de seu valor esperado, mas a mais simples de trabalhar algebricamente é a diferença elevada ao quadrado, $(X - \mu)^2$. (A elevação ao quadrado serve para eliminar o sinal da medida da distância; o valor positivo resultante corresponde à nossa noção intuitiva de distância.) Essa distância em si é uma variável aleatória, já que ela pode mudar a cada resultado de X. Da mesma forma que precisamos de um número para resumir a tendência central de X, precisamos de um número que nos informe o quanto X está distante de μ, *em média*. Um desses números é a **variância**, que nos informa a distância esperada de X até sua média:

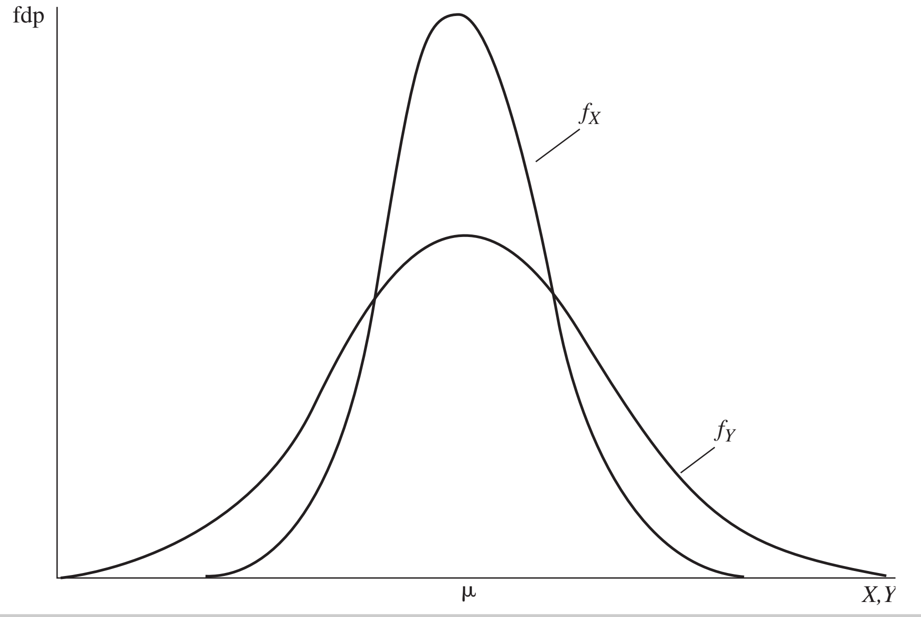
$$\mathrm{Var}(X) \equiv E(X - \mu)^2]. \qquad \textbf{(B.23)}$$

A variância é algumas vezes representada por σ_x^2, ou simplesmente σ^2, quando o contexto é claro. De (B.3), deduz-se que a variância é sempre não-negativa.

Como um instrumento computacional, é interessante observar que

Figura B.4

Variáveis aleatórias com a mesma média, mas com distribuições diferentes.



$$\sigma^2 = \mathrm{E}(X^2 - 2X\mu + \mu^2) = \mathrm{E}(X^2) - 2\mu^2 + \mu^2 = \mathrm{E}(X^2) - \mu^2. \quad \textbf{(B.24)}$$

Usando (B.23) ou (B.24), não precisamos fazer a distinção entre variáveis aleatórias discretas e contínuas: a definição de variância é a mesma em qualquer dos casos. Na maioria da vezes, primeiro calculamos $\mathrm{E}(X)$, depois $\mathrm{E}(X^2)$, e, então, usamos a fórmula de (B.4). Por exemplo, se $X \sim \text{Bernoulli}(\theta)$, então, $\mathrm{E}(X) = \theta$, e como $X^2 = X$, $\mathrm{E}(X^2) = \theta$. Deduz-se da equação (B.24) que $\mathrm{Var}(X) = \mathrm{E}(X^2) - \mu^2 = \theta - \theta^2 = \theta(1 - \theta)$.

São apresentadas a seguir duas importantes propriedades da variância.

PROPRIEDADE VAR.1

$\mathrm{Var}(X) = 0$ se, e somente se, houver uma constante c, de tal forma que $\mathrm{P}(X = c) = 1$, em cujo caso, $\mathrm{E}(X) = c$.

Essa primeira propriedade diz que a variância de qualquer constante é zero, e se uma variável aleatória tiver variância zero, então, ela será essencialmente constante.

PROPRIEDADE VAR.2

Para quaisquer constantes a e b, $\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X)$.

Isso significa que a adição de uma constante a uma variável aleatória não altera a variância, mas a multiplicação de uma variável aleatória por uma constante aumenta a variância por um fator igual ao *quadrado* daquela constante. Por exemplo, se X representar a temperatura em graus Celsius e $Y = 32 + (9/5)X$ for a temperatura em graus Fahrenheit, então, $\text{Var}(Y) = (9/5)^2\text{Var}(X) = (81/25)\text{Var}(X)$.

Desvio-Padrão

O **desvio-padrão** de uma variável aleatória, representado por $\text{dp}(X)$, é simplesmente a raiz quadrada positiva da variância: $\text{dp}(X) \equiv +\sqrt{\text{Var}(X)}$. O desvio-padrão algumas vezes é representado por σ_x, ou simplesmente σ, quando a variável aleatória é entendida. Duas propriedades do desvio-padrão resultam das propriedades VAR.1 e VAR.2.

PROPRIEDADE DP.1
Para qualquer constante c, $\text{dp}(c) = 0$.

PROPRIEDADE DP.2
Para quaisquer constantes a e b,

$$\text{dp}(aX + b) = |a|\text{dp}(X).$$

Em particular, se $a > 0$, então, $\text{dp}(aX) = a \cdot \text{dp}(X)$.

Essa última propriedade faz com que seja mais natural trabalhar com o desvio-padrão do que com a variância. Por exemplo, suponha que X seja uma variável aleatória medida em milhares de dólares, digamos renda. Se definirmos $Y = 1.000X$, então, Y será a renda medida em dólares. Suponha que $\text{E}(X) = 20$ e $\text{dp}(X) = 6$. Então, $\text{E}(Y) = 1{,}000\text{E}(X) = 20.000$ e $\text{dp}(Y) = 1.000 \cdot \text{dp}(X) = 6.000$, de forma que o valor esperado e o desvio-padrão crescem pelo mesmo fator, 1.000. Se tivéssemos trabalhado com a variância, teríamos $\text{Var}(Y) = (1.000)^2\text{Var}(X)$, de forma que a variância de Y é um milhão de vezes maior que a variância de X.

Padronizando uma Variável Aleatória

Como uma aplicação das propriedades da variância e do desvio-padrão — e um tópico de interesse prático por si mesmo — suponha que, dada uma variável aleatória X, definamos uma nova variável aleatória subtraindo sua média μ e dividindo o resultado por seu desvio-padrão σ:

$$Z \equiv \frac{X - \mu}{\sigma}, \tag{B.25}$$

que podemos escrever como $Z = aX + b$, onde $a \equiv (1/\sigma)$ e $b \equiv -(\mu/\sigma)$. Então, de acordo com a propriedade E.2,

$$\text{E}(Z) = a\text{E}(X) + b = (\mu/\sigma) - (\mu/\sigma) = 0.$$

Da propriedade Var.2,

$$\text{Var}(Z) = a^2\text{Var}(X) = (\sigma^2/\sigma^2) = 1.$$

Portanto, a variável aleatória Z tem uma média zero e uma variância (e portanto um desvio-padrão) igual a um. Esse procedimento algumas vezes é referido como *padronização* da variável aleatória X, e Z é chamado uma **variável aleatória padronizada**. (Em cursos introdutórios de estatística, ele algumas vezes é chamado de *transformação-z* de X). É importante lembrar que o desvio-padrão, não a variância, aparece no denominador de (B.25). Como veremos, essa transformação é freqüentemente utilizada na inferência estatística.

Como um exemplo específico, suponha que $\text{E}(X) = 2$ e $\text{Var}(X) = 9$. Então, $Z = (X - 2)/3$ terá um valor esperado igual a zero e variância igual a um.

B.4 CARACTERÍSTICAS DAS DISTRIBUIÇÕES CONJUNTAS E CONDICIONAIS

Medidas de Associação: Covariância e Correlação

Embora a fdp conjunta de duas variáveis aleatórias descreva completamente a relação entre elas, é útil ter medidas resumidas de como, em média, duas variáveis aleatórias variam entre si. Como acontece com o valor esperado e a variância, isso é semelhante ao usar um único número para resumir alguma coisa de uma distribuição inteira, que, nesse caso, é uma distribuição conjunta de duas variáveis aleatórias.

Covariância

Seja $\mu_X = \text{E}(X)$, e $\mu_Y = \text{E}(Y)$, e considere a variável aleatória $(X - \mu_X)(Y - \mu_Y)$. Agora, se X e Y estiveram acima de suas respectivas médias, então, $(X - \mu_X)(Y - \mu_Y) > 0$. Isso também será verdadeiro se $X < \mu_X$ e $Y < \mu_Y$. Por outro lado, se $X > \mu_X$ e $Y < \mu_Y$, ou vice-versa, então, $(X - \mu_X)(Y - \mu_Y) < 0$. Como, então, esse produto poderá nos dar qualquer informação sobre a relação entre X e Y?

A **covariância** entre duas variáveis aleatórias X e Y, algumas vezes chamada a covariância populacional para enfatizar que ela se refere à relação entre duas variáveis descrevendo uma população, é definida como o valor esperado do produto $(X - \mu_X)(Y - \mu_Y)$:

$$\text{Cov}(X,Y) \equiv \text{E}[X - \mu_X)(Y - \mu_Y)], \tag{B.26}$$

que algumas vezes é representado por σ_{XY}. Se $\sigma_{XY} > 0$, então, em média, quando X estiver acima de sua média, Y também estará acima de sua média. Se $\sigma_{XY} < 0$, então, em média, quando X estiver acima de sua média, Y estará abaixo de sua média.

Algumas expressões úteis para calcular $\text{Cov}(X,Y)$ são as seguintes:

$$\begin{aligned}\text{Cov}(X,Y) &= \text{E}[X - \mu_X)(Y - \mu_Y)] = \text{E}[(X - \mu_X)Y] \\ &= \text{E}[X(Y - \mu_Y)] = \text{E}(XY) - \mu_X\mu_Y.\end{aligned} \tag{B.27}$$

Decorre de (B.27) que, se $\text{E}(X) = 0$ ou $\text{E}(Y) = 0$, então, $\text{Cov}(X,Y) = \text{E}(X,Y)$.

A covariância mede o grau de dependência *linear* entre duas variáveis aleatórias. Uma covariância positiva indica que duas variáveis aleatórias se movem na mesma direção, enquanto uma covariância negativa indica que elas se movem em direções opostas. Interpretar a *magnitude* de uma covariância pode ser um pouco difícil, como veremos brevemente.

Como a covariância é uma medida de como duas variáveis aleatórias estão relacionadas, é natural perguntar como a covariância está relacionada à noção de independência. Isso é dado pela seguinte propriedade.

PROPRIEDADE COV.1

Se X e Y forem independentes, então, $\mathrm{Cov}(X,Y) = 0$.

Essa propriedade decorre da equação (B.27) e do fato de que $\mathrm{E}(XY) = \mathrm{E}(X)\mathrm{E}(Y)$ quando X e Y são independentes. É importante lembrar que a inversa de COV.1 *não* é verdadeira: covariância zero entre X e Y não implica que X e Y sejam independentes. De fato, existem variáveis aleatórias X de tal forma que, se $Y = X^2$, $\mathrm{Cov}(X,Y) = 0$. [Qualquer variável aleatória com $\mathrm{E}(X) = 0$ e $\mathrm{E}(X^3) = 0$ tem essa propriedade]. Se $Y = X^2$, então, X e Y serão claramente não-independentes: conhecendo X conheceremos Y. Parece bastante estranho que X e X^2 possam ter covariância zero, e isso revela um ponto fraco da covariância como uma medida geral de associação entre duas variáveis aleatórias. A covariância é útil em contextos em que as relações são pelo menos aproximadamente lineares.

A segunda mais importante propriedade da covariância envolve covariâncias entre funções lineares.

PROPRIEDADE COV.2

Para quaisquer constantes a_1, b_1 , a_2 e b_2,

$$\mathrm{Cov}(a_1X + b_1, a_2Y + b_2) = a_1a_2\mathrm{Cov}(X,Y). \tag{B.28}$$

Uma implicação importante de COV.2 é que a covariância entre duas variáveis aleatórias pode ser alterada simplesmente pela multiplicação de uma ou de ambas as variáveis aleatórias por uma constante. Isso é importante em economia, pois variáveis monetárias, taxas de inflação etc. podem ser definidas com diferentes unidades de medida sem que seja alterado seu significado.

Finalmente, é útil saber que o valor absoluto da covariância entre quaisquer duas variáveis aleatórias está limitado pelo produto de seus desvios-padrão; isso é conhecido como a *desigualdade de Cauchy-Schwartz*.

PROPRIEDADE COV.3

$|\mathrm{Cov}(X,Y)| \leq \mathrm{dp}(X)\mathrm{dp}(Y)$.

Coeficiente de Correlação

Suponha que queremos conhecer a relação entre o grau de educação e os rendimentos anuais da população que trabalha. Poderíamos chamar de X a educação e de Y os rendimentos, computando a seguir sua covariância. Entretanto, a resposta que obteremos dependerá de como escolheremos medir a educação e os rendimentos. A propriedade COV.2 implica que a covariância entre educação e rendimentos dependerá de se os rendimentos são medidos em dólares ou milhares de dólares, ou se a educação é medida em meses ou anos. É bastante claro que a maneira como mediremos essas variáveis não terá influência no quanto elas estão fortemente relacionadas. Mas a covariância entre elas efetivamente depende das unidades de medida.

O fato de a covariância depender das unidades de medida é uma deficiência que é compensada pelo **coeficiente de correlação** entre X e Y:

$$\text{Corr}(X,Y) \equiv \frac{\text{Cov}(X,Y)}{\text{dp}(X)\cdot\text{dp}(Y)} = \frac{\sigma_{XY}}{\sigma_X\sigma_Y}; \tag{B.29}$$

o coeficiente de correlação entre X e Y algumas vezes é representado por ρ_{XY} (e é algumas vezes chamado de correlação populacional).

Como σ_X e σ_Y são positivos, Cov(X,Y) e Corr(X,Y) sempre têm o mesmo sinal, e Corr(X,Y) $= 0$ se, e somente se, Cov(X,Y) $= 0$. Algumas das propriedades da covariância são transferidas para a correlação. Se X e Y forem independentes, então, Corr(X,Y) $= 0$, mas correlação zero não implica independência. (Como a covariância, o coeficiente de correlação também é uma medida de dependência linear). Porém, a magnitude do coeficiente de correlação é mais fácil de interpretar do que o tamanho da covariância, devido à seguinte propriedade.

PROPRIEDADE CORR.1

$-1 \leq \text{Corr}(X,Y) \leq 1$.

Se Corr(X,Y) $= 0$ ou, equivalentemente, Cov(X,Y) $= 0$, não haverá relação linear entre X e Y, e X e Y são chamadas de variáveis *não-correlacionadas*; caso contrário, X e Y serão correlacionadas. Corr(X,Y) $= 1$ indica uma relação linear positiva perfeita, o que significa que podemos escrever $Y = a + bX$ para alguma constante a e alguma constante $b > 0$. Corr(X,Y) $= -1$ indica uma relação linear negativa perfeita, de forma que $Y = a + bX$ para alguma constante $b < 0$. Os casos extremos de correlação positiva ou negativa igual à unidade raramente ocorre. Valores de RHO_{XY} próximos de 1 ou -1 indicam fortes relações lineares.

Como mencionado antes, a correlação entre X e Y não varia em relação às unidades de medida de X ou Y. Isso é especificado de forma mais geral como segue.

PROPRIEDADE CORR.2

Para as constantes a_1, b_1, a_2 e b_2, com $a_1a_2 > 0$,

$$\text{Corr}(a_1X + b_1, a_2Y + b_2) = \text{Corr}(X,Y).$$

Se $a_1a_2 < 0$, então,

$$\text{Corr}(a_1X + b_1, a_2Y + b_2) = -\text{Corr}(X,Y).$$

Como um exemplo, suponha que a correlação entre rendimentos e educação da população que trabalha seja 0,15. Essa medida não depende de se os rendimentos estão medidos em dólares, milhares de dólares, ou qualquer outra unidade; ela também não depende de se a educação está medida em anos, trimestres, meses etc.

Variância da Soma de Variáveis Aleatórias

Agora que já definimos covariância e correlação, podemos completar nossa relação das principais propriedades da variância.

PROPRIEDADE VAR.3

Para as constante a e b,

$$\text{Var}(aX + bY) = a^2\text{Var}(X) + b^2\text{Var}(Y) + 2ab\text{Cov}(X,Y).$$

Segue imediatamente que, se X e Y forem não-correlacionadas — de forma que $\text{Cov}(X,Y) = 0$ —, então,

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) \tag{B.30}$$

e

$$\text{Var}(X - Y) = \text{Var}(X) + \text{Var}(Y). \tag{B.31}$$

Neste último caso, observe como a variância da diferença é a *soma*, não a diferença, das variâncias.

Como um exemplo de (B.30), seja X os lucros ganhos por um restaurante durante uma noite de sexta-feira e Y os lucros ganhos na noite do sábado seguinte. Então, $Z = X + Y$ será os lucros das duas noites. Suponha que X e Y tenham, cada uma, um valor esperado de 300 dólares e um desvio-padrão de 15 dólares (de forma que a variância será 255). O lucro esperado das duas noites será $\text{E}(Z) + \text{E}(Y) = 2{\cdot}(300) = 600$ dólares. Se X e Y forem independentes e, portanto, não-correlacionadas, a variância do lucro total será a soma das variâncias: $\text{Var}(Z) = \text{Var}(X) + \text{Var}(Y) = 2\cdot(225) = 450$. Portanto, o desvio-padrão do lucro total será $\sqrt{450}$ ou aproximadamente 21,21 dólares.

As expressões (B.30) e (B.31) estendem-se para mais de duas variáveis aleatórias. Para especificar essa extensão, precisamos de uma definição. As variáveis aleatórias X e Y serão **variáveis aleatórias não-correlacionadas duas a duas** se cada variável no conjunto for não-correlacionada com cada outra variável do conjunto. Isto é, $\text{Cov}(X_i,X_j) = 0$ para todo $i \neq j$.

PROPRIEDADE VAR.4

Se $\{X_1, ..., X_n)$ forem variáveis aleatórias não-correlacionadas duas a duas e $\{a_i\text{: } i = 1, ..., n\}$ forem constantes, então,

$$\text{Var}(a_1X_1 + ... + a_nX_n) = a_1^2\text{Var}(X_1) + ... + a_n^2\text{Var}(X_n).$$

Em notação de somatórios, podemos escrever

$$\text{Var}(\sum_{i=1}^{n} a_iX_i) = \sum_{i=1}^{n} a_i^2\text{Var}(X_i). \tag{B.32}$$

Um caso especial da Propriedade VAR.4 ocorre quando consideramos $a_i = 1$ para todos os i. Então, para variáveis aleatórias não-correlacionadas duas a duas, a variância da soma será a soma das variâncias:

$$\text{Var}(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} \text{Var}(X_i). \tag{B.33}$$

Como variáveis aleatórias independentes são não-correlacionadas (veja a propriedade COV.1), a variância de uma soma de variáveis aleatórias independentes é a soma das variâncias.

Se as variáveis X_i não forem não-correlacionadas duas a duas, então, a expressão $\mathrm{Var}(\sum_{i=1}^{n} a_i X_i)$ será muito mais complicada; precisaremos adicionar no lado direito de (B.32) os termos $2a_i a_j \mathrm{Cov}(x_i, x_j)$ para todo $i > j$.

Podemos usar (B.33) para derivarmos a variância de uma variável aleatória binomial. Definimos $X \sim \mathrm{Binomial}(n,\theta)$ e escrevemos $X = Y_1, + ... + Y_n$, onde Y_i são variáveis aleatórias independentes de Bernoulli(θ). Então, de (B.33), $\mathrm{Var}(X) = \mathrm{Var}(Y_1) + ... + \mathrm{Var}(Y_n) = \mathrm{n}\theta(1 - \theta)$.

No exemplo das reservas da companhia aérea com $n = 120$ e $\theta = 0{,}85$, a variância do número de passageiros que comparecem para embarque seria $120(0{,}85)(0{,}15) = 15{,}3$, e, assim, o desvio-padrão seria aproximadamente 3,9.

Esperança Condicional

A covariância e a correlação medem a relação linear entre duas variáveis aleatórias e as tratam simetricamente. Muitas vezes, em ciências sociais, gostaríamos de explicar uma variável, chamada Y, em termos de outra variável, digamos X. Além disso, se Y for relacionada com X de uma maneira não linear, gostaríamos de ser informados sobre isso. Chamemos Y de variável explicada e X de variável explicativa. Por exemplo, Y poderia ser o salário por hora e X poderia ser o número de anos de educação formal.

Já introduzimos a noção de função de densidade de probabilidade condicional de Y, dado X. Assim, poderíamos querer ver como a distribuição dos salários é alterada pelo nível de educação. Porém, em geral, queremos ter uma maneira simples de resumir essa distribuição. Um único número não será suficiente, visto que a distribuição de Y, dado $X = x$, geralmente depende do valor de x. No entanto, podemos resumir a relação entre Y e X verificando a **esperança condicional** de Y, dado X, algumas vezes chamada *média condicional*. A idéia é a seguinte: suponha que saibamos que X assumiu um valor particular, digamos x. Então, poderemos calcular o valor esperado de Y em decorrência de conhecermos esse resultado de X. Representamos esse valor esperado por $\mathrm{E}(Y|X = x)$, ou algumas vezes $\mathrm{E}(Y|x)$ como forma abreviada. De forma geral, quando x muda, $\mathrm{E}(Y|x)$ também muda.

Quando Y for uma variável aleatória discreta assumindo valores $\{y_1, ..., y_n\}$, então,

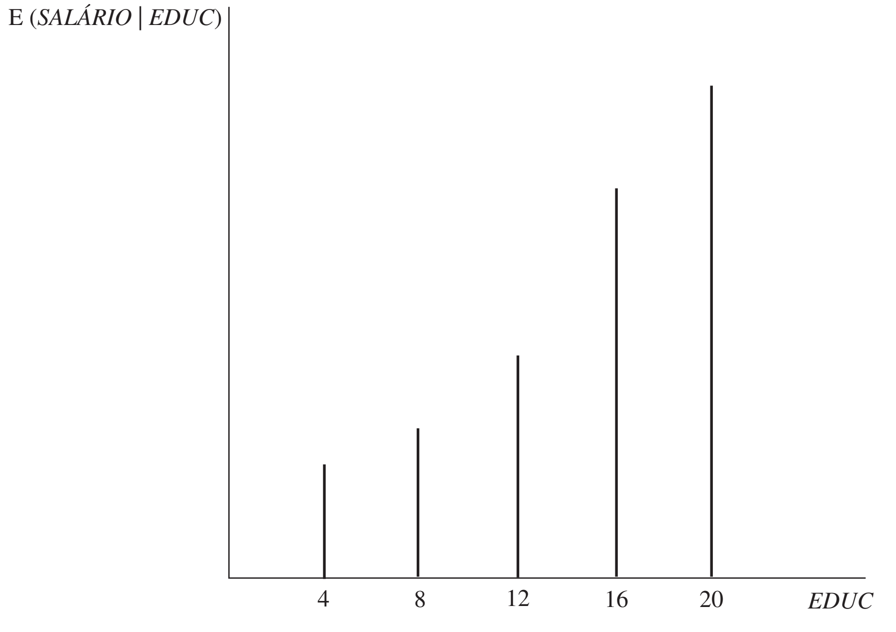
$$\mathrm{E}(Y|x) = \sum_{j=1}^{m} y_j f_{Y|X}(y_j|x).$$

Quando Y for contínua, $\mathrm{E}(Y|x)$ será definida pela integração de $y f_{Y|X}(y|x)$ sobre todos os valores possíveis de y. Assim como no caso da esperança incondicional, a esperança condicional é uma média ponderada de possíveis valores de Y, mas agora os pesos refletem o fato de que X assumiu um valor específico. Assim, $\mathrm{E}(Y|x)$ é apenas alguma função de x, que nos diz como o valor esperado de Y varia com x.

Como um exemplo, seja (X,Y) a população de todas as pessoas que trabalham, na qual X é anos de educação, e Y é o salário por hora. Então, $\mathrm{E}(Y|X = 12)$ será o salário médio por hora de todas as pessoas da população com 12 anos de educação (em termos gerais, correspondente à educação de ensino médio). $\mathrm{E}(Y|X = 16)$ será o salário médio por hora de todas as pessoas com 16 anos de educação. O gráfico de valores esperados com vários níveis de educação fornece informações importantes sobre como os salários e a educação estão relacionados. Veja a Figura B.5, para uma ilustração.

Figura B.5

O valor esperado do salário por hora considerando vários níveis de educação.



Em princípio, o valor esperado do salário por hora pode ser encontrado a cada nível de educação, e essas esperanças podem ser resumidas em uma tabela. Como a educação pode variar amplamente — e pode até mesmo ser medida em frações de um ano —, essa é uma maneira excessivamente trabalhosa de se mostrar a relação entre o salário médio e o grau de educação. Em econometria, geralmente especificamos funções simples que capturam essa relação. Como um exemplo, suponha que o valor esperado de *SALÁRIO*, dado *EDUC*, seja a função linear

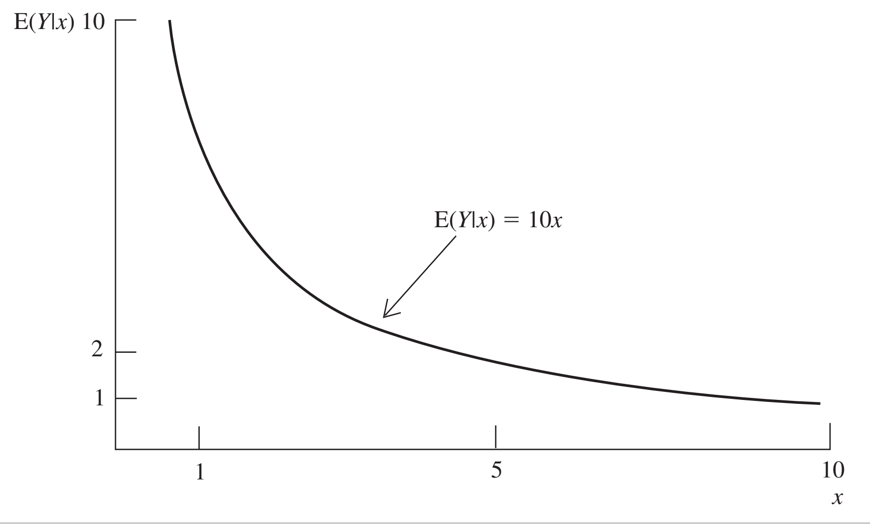
$$E(SALÁRIO|EDUC) = 1{,}05 + 0{,}45\ EDUC.$$

Se essa relação for válida na população das pessoas que trabalham, o salário médio das pessoas com 8 anos de educação será $1{,}05 + 0{,}45(8) = 4{,}65$, ou 4,65 dólares. O salário médio das pessoas com 16 anos de educação será 8,25 dólares. O coeficiente de *EDUC* implica que cada ano de educação aumenta o salário por hora esperado em 0,45, ou 45 centavos de dólar.

As esperanças condicionais também podem ser funções não-lineares. Por exemplo, suponha que $E(Y|x) = 10/x$, onde X é uma variável aleatória que sempre será maior que zero. Essa função está traçada na Figura B.6. Isso poderia representar uma função de demanda, na qual Y seria a quantidade demandada e X seria o preço. Se Y e X forem relacionadas nesta forma, uma análise de associação linear, tal como uma análise de correlação, seria incompleta.

Figura B.6

Gráfico de $E(Y|x) = 10/x$.



Propriedades da Esperança Condicional

Várias propriedades básicas das esperanças condicionais são úteis para derivações em análise econométrica.

PROPRIEDADE EC.1

$E[c(X)|X] = c(X)$, para qualquer função $c(X)$.

Essa primeira propriedade significa que funções de X comportam-se como constantes quando calculamos a esperança condicional de X. Por exemplo, $E(X^2|X) = X^2$. Intuitivamente, isso simplesmente significa que, se conhecermos X, também conheceremos X^2.

PROPRIEDADE EC.2

Para as funções $a(X)$ e $b(X)$,

$$E[a(X)Y + b(X)|X] = a(X)E(Y|X) + b(X).$$

Por exemplo, podemos calcular com facilidade a esperança condicional de uma função tal como $XY + 2X^2$: $E(XY + 2X^2|X)\ XE(Y|X) + 2X^2$.

A próxima propriedade interliga as noções de independência e esperanças condicionais.

PROPRIEDADE EC.3

Se X e Y forem independentes, então, $E(Y|X) = E(Y)$.

Essa propriedade significa que, se X e Y forem independentes, então, o valor esperado de Y, dado X, não dependerá de X, caso em que $E(Y|X)$ sempre será igual ao valor esperado (incondicional) de Y. No exemplo do salário e educação, se salário fosse independente de educação, então, os salários médios das pessoas com educação de ensino médio e com cursos superiores seriam os mesmos. Como quase certamente esse resultado seria falso, não podemos assumir que salário e educação sejam independentes.

Um caso especial da propriedade EC.3 é o seguinte: se U e X forem independentes e $E(U) = 0$, então, $E(U|X) = 0$.

Também existem propriedades da esperança condicional que têm a ver com o fato de $E(Y|X)$ ser uma função de X, digamos $E(Y|X) = \mu(X)$. Como X é uma variável aleatória, $\mu(X)$ também será uma variável aleatória. Além disso, $\mu(X)$ tem uma distribuição de probabilidade e, portanto, um valor esperado. De forma geral, o valor esperado de $\mu(X)$ pode ser muito difícil de ser calculado de forma direta. A **lei das expectativas iteradas** diz que o valor esperado de $\mu(X)$ é simplesmente igual ao valor esperado de Y. Escrevemos isso da seguinte maneira.

PROPRIEDADE EC.4

$E[E(Y|X)] = E(Y)$.

Essa propriedade é de difícil compreensão à primeira vista. Ela significa que, se primeiro obtivermos $E(Y|X)$ como uma função de X e considerarmos seu valor esperado (em relação à distribuição de X, é claro), então, acabaremos obtendo $E(Y)$. Isso não é tão óbvio, mas pode ser derivado utilizando a definição dos valores esperados.

Suponha que $Y = SALÁRIO$ e $X = EDUC$, onde *SALÁRIO* está medido em horas e *EDUC* em anos. Suponha que o valor esperado de *SALÁRIO*, dado *EDUC*, seja $E(SALÁRIO|EDUC) = 4 + 0{,}60\ EDUC$. Além disso, $E(EDUC) = 11{,}5$. Então, a lei das expectativas iteradas sugere que $E(SALÁRIO) = E(4 + 0{,}60\ EDUC) = 4 + 0{,}60\ E(EDUC) = 4 + 0{,}60(11{,}5) = 10{,}90$, ou 10,90 dólares por hora.

A próxima propriedade especifica uma versão mais geral da lei das expectativas iteradas.

PROPRIEDADE EC.4′

$E(Y|X) = E[E(Y|X,Z)|X]$.

Em outras palavras, podemos encontrar $E(Y|X)$ em duas etapas. Primeiro, encontramos $E(Y|X,Z)$ para qualquer outra variável aleatória Z. Em seguida, encontramos o valor esperado de $E(Y|X,Z)$, condicional em X.

PROPRIEDADE EC.5

Se $E(Y|X) = E(Y)$, então, $\mathrm{Cov}(X,Y) = 0$ [como também $\mathrm{Corr}(X,Y) = 0$]. De fato, *qualquer* função de X é não-correlacionada com Y.

Essa propriedade significa que, se o conhecimento de X não altera o valor esperado de Y, então, X e Y *devem* ser não-correlacionadas, o que implica que, se X e Y forem correlacionadas, então, $E(Y|X)$ deve depender de X. A inversa da propriedade EC.5 não é verdadeira: se X e Y forem não-correlacionadas, $E(Y|X)$ *poderá* ainda depender de X. Por exemplo, suponha que $Y = X^2$. Então, $E(Y|X) = X^2$, que claramente é uma função de X. Porém, como mencionado em nossa discussão sobre covariância e correlação, é possível que X e X^2 sejam não-correlacionadas. A esperança condicional captura a relação não linear entre X e Y que uma análise de correlação deixaria passar despercebida.

As propriedades EC.4 e EC.5 têm duas implicações importantes: se U e X forem variáveis aleatórias, de forma que $E(U|X) = 0$, então, $E(U) = 0$, e U e X serão não-correlacionadas.

PROPRIEDADE EC.6

Se $E(Y^2) < \infty$ e $E[g(X)^2] < \infty$ para alguma função g, então, $E\{[Y - \mu(X)]^2|X\} \leq E\{[Y - g(X)]^2|X\}$ e $E\{[Y - \mu(X)]^2\} \leq E\{[Y - g(X)]^2\}$.

A propriedade EC.6 é muito útil em contextos de previsão ou de projeções. A primeira desigualdade diz que, se medirmos a inexatidão da previsão como o erro quadrático de previsão esperado, condicional em X, então, a média condicional será melhor que qualquer outra função de X para prever Y. A média condicional também minimiza o erro quadrático de previsão esperado incondicional.

Variância Condicional

Dadas as variáveis aleatórias X e Y, a variância de Y, condicional em $X = x$, será simplesmente a variância associada à distribuição condicional de Y, dado $X = x$: $E\{[Y - E(Y|x)]^2|x\}$. A fórmula

$$\mathrm{Var}(Y|X = x) = E(Y^2|x) - [E(Y|x)]^2$$

é freqüentemente útil para os cálculos. Somente ocasionalmente teremos que calcular uma variância condicional. Entretanto, teremos que fazer hipóteses a respeito e manipular as variâncias condicionais para certos tópicos na análise de regressão.

Como um exemplo, defina $Y = POUPANÇA$ e $X = RENDA$ (ambas medidas em termos anuais, para a população de todas as famílias). Suponha que $\mathrm{Var}(POUPANÇA|RENDA) = 400 + 0{,}25\ RENDA$. Isso diz que, conforme aumente a renda, a variância dos níveis de poupança também aumenta. É importante verificar que a relação entre as variâncias de *POUPANÇA* e *RENDA* é totalmente separada da relação entre os *valores esperados* de *POUPANÇA* e *RENDA*.

Estabelecemos, portanto, uma propriedade importante da variância condicional.

PROPRIEDADE VC.1

Se X e Y forem independentes, então, $\mathrm{Var}(Y|X) = \mathrm{Var}(Y)$.

Essa propriedade é bastante clara, pois a distribuição de Y, dado X, não depende de X, e $\mathrm{Var}(Y|X)$ é apenas uma característica dessa distribuição.

B.5 A DISTRIBUIÇÃO NORMAL E OUTRAS DISTRIBUIÇÕES A ELA RELACIONADAS

A Distribuição Normal

A distribuição normal e as derivadas dela são as distribuições mais amplamente usadas em estatística e econometria. Assumir que variáveis aleatórias definidas para populações são normalmente distribuídas simplifica os cálculos de probabilidade. Além disso, nos valeremos pesadamente da distribuição normal e de outras a ela relacionadas para conduzir inferência em estatística e econometria — mesmo quando a população básica não for necessariamente normal. Precisamos adiar os detalhes, mas tenha certeza de que essas distribuições irão surgir muitas vezes ao longo deste livro.

Uma variável aleatória normal é uma variável aleatória contínua que pode assumir qualquer valor. Sua função de densidade de probabilidade tem a forma familiar de um sino traçada na Figura B.7.

Matematicamente, a fdp de X pode ser escrita como

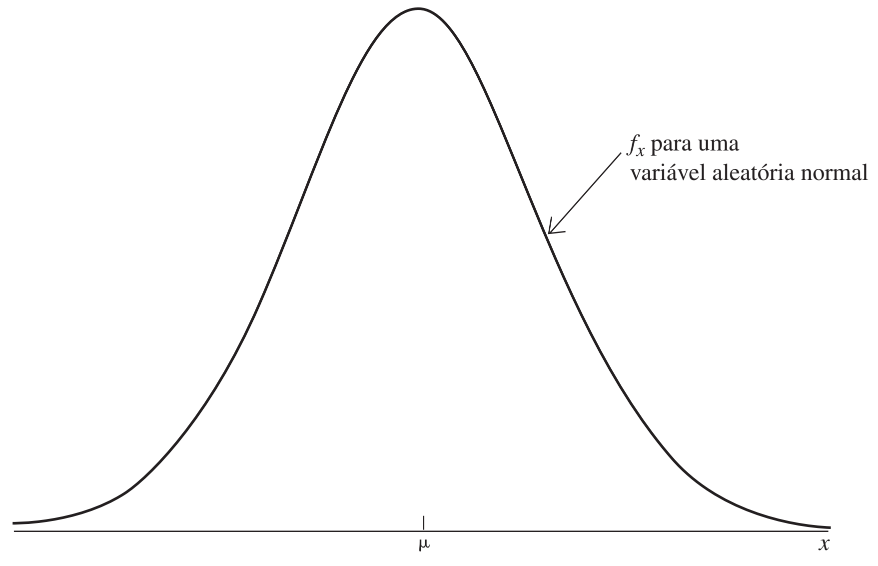
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp[-(x - \mu)^2/2\sigma^2], \ -\infty < x < \infty, \quad \textbf{(B.34)}$$

onde $\mu = \mathrm{E}(X)$ e $\sigma^2 = \mathrm{Var}(X)$. Dizemos que X tem uma **distribuição normal** com valor esperado μ e variância σ^2, escrita como $X \sim \mathrm{Normal}(\mu,\sigma^2)$. Como a distribuição normal é simétrica em relação a μ, μ também é a mediana de X. A distribuição normal é algumas vezes chamada de *distribuição de Gauss* em homenagem ao famoso estatístico C. F. Gauss.

Algumas variáveis aleatórias parecem seguir em linhas gerais uma distribuição normal. As alturas e pesos do ser humano, resultados de provas e índices de desemprego municipais possuem fdps com aproximadamente a forma na Figura B.7. Outras distribuições, como as da renda, não parecem seguir a função de probabilidade normal. Na maioria dos países, a renda não é simetricamente distribuída em torno de qualquer valor; a distribuição é inclinada em direção à extremidade superior. Em alguns casos, uma variável pode ser transformada para atingir a normalidade. Uma transformação popular é o log natural, que faz sentido para variáveis aleatórias positivas. Se X for uma variável aleatória positiva, tal como a renda, e $Y = \log(X)$ tiver uma distribuição normal, então, dizemos que X tem uma *distribuição lognormal*. É possível constatar que a distribuição lognormal se ajusta bastante bem à distribuição de renda de muitos países. Outras variáveis, como preços de mercadorias, parecem ser bem descritas quando utilizada a distribuição lognormal.

Figura B.7

A forma geral de uma função de densidade de probabilidade normal.



A Distribuição Normal Padrão

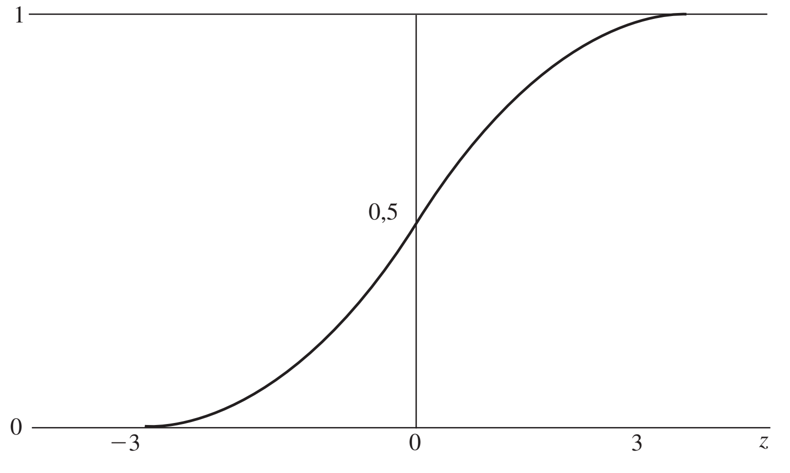
Um caso especial da distribuição normal ocorre quando a média for zero e a variância (e, portanto, o desvio-padrão) for a unidade. Se uma variável aleatória Z tiver uma distribuição Normal(0,1), dizemos que ela tem um **distribuição normal padrão**. A fdp de uma variável aleatória normal padrão é representada por PHI(z); de (B.34), com $\mu = 0$ e $\sigma^2 = 1$, ela é dada por

$$\phi(z) = \frac{1}{\sqrt{2\pi}} \exp(-z^2/2), \; -\infty < z < \infty. \qquad \textbf{(B.35)}$$

A função de distribuição cumulativa normal padrão é representada por $\Phi(z)$ e é obtida como a área sob ϕ, à esquerda de z; veja a Figura B.8. Lembre-se de que $\Phi(z) = P(Z \leq z)$; como Z é contínua, $\Phi(z) = P(Z < z)$ também é contínua.

Figura B.8

A função de distribuição cumulativa normal padrão.



Não existe fórmula simples que possa ser usada para obter os valores de $\Phi(z)$ [pois $\Phi(z)$ é a integral da função em (B.35), e essa integral não tem forma única]. No entanto, os valores de $\Phi(z)$ são facilmente tabulados; eles estão dados para z entre $-3{,}1$ e $3{,}1$ na Tabela G.1 no Apêndice G. Para $z \leq -3{,}1$, $\Phi(z)$ será menor que 0,001, e para $z \geq 3{,}1$, $\Phi(z)$ será maior que 0,999. A maioria dos programas estatísticos e econométricos inclui comandos simples para computar valores da fdc normal padrão, e assim podemos freqüentemente evitar totalmente as tabelas impressas para obter as probabilidades para qualquer valor de z.

Usando fatos básicos da probabilidade — e, em particular, as propriedades (B.7) e (B.8) em relação às fdcs —, podemos usar a fdc normal padrão para calcular a probabilidade de qualquer evento envolvendo uma variável aleatória normal padrão. As fórmulas mais importantes são

$$P(Z > z) = 1 - \Phi(z), \qquad \textbf{(B.36)}$$

$$P(Z < -z) = P(Z > z) \qquad \textbf{(B.37)}$$

e

$$P(a \leq Z \leq b) = \Phi(b) - \Phi(a). \qquad \textbf{(B.38)}$$

Como Z é uma variável aleatória contínua, todas as três fórmulas são válidas, sejam ou não restritas as desigualdades. Citamos alguns exemplos: $P(Z > 0{,}44) = 1 - 0{,}67 = 0{,}33$, $P(Z < -0{,}92) = P(Z > 0{,}92) = 1 - 0{,}821 = 0{,}179$, e $P(-1 < Z \leq 0{,}5) = 0{,}692 - 0{,}159 = 0{,}533$.

Outra expressão útil é que, para qualquer $c > 0$,

$$\begin{aligned} P(|Z| > c) &= P(Z > c) + P(Z < -c) \\ &= 2 \cdot P(Z > c) = 2[1 - \Phi(c)]. \end{aligned} \qquad \textbf{(B.39)}$$

Assim, a probabilidade de que o valor absoluto de Z seja maior que alguma constante c positiva será simplesmente duas vezes a probabilidade $P(Z > c)$; isso reflete a simetria da distribuição normal padrão.

Na maioria das aplicações, iniciamos com uma variável aleatória normalmente distribuída, $X \sim \text{Normal}(\mu,\sigma^2)$, onde μ é diferente de zero e $\sigma^2 \neq 1$. Qualquer variável aleatória normal pode ser transformada em uma normal padrão usando a seguinte propriedade.

PROPRIEDADE NORMAL.1

Se $X \sim \text{Normal}(\mu,\sigma^2)$, então, $(X - \mu)/\sigma \sim \text{Normal}(0,1)$.

A propriedade Normal.1 mostra como transformar qualquer variável aleatória normal em uma normal padrão. Assim, suponha que $X \sim \text{Normal}(3,4)$, e que gostaríamos de calcular $P(X \leq 1)$. As etapas sempre envolvem a normalização de X para uma normal padrão:

$$\begin{aligned} P(X \leq 1) &= P(X - 3 \leq 1 - 3) = P\left(\frac{X-3}{2} \leq -1\right) \\ &= P(Z \leq -1) = \Phi(-1) = 0{,}159. \end{aligned}$$

EXEMPLO B.6

(Probabilidades para uma Variável Aleatória Normal)

Primeiro, vamos calcular $P(2 < X \leq 6)$ quando $X \sim \text{Normal}(4,9)$ (usar $<$ ou $\leq$ é irrelevante, pois X é uma variável aleatória contínua). Agora

$$\begin{aligned} P(2 < X \leq 6) &= P\left(\frac{2-4}{3} < \frac{X-4}{3} \leq \frac{6-4}{3}\right) = P(-2/3 < Z \leq 2/3) \\ &= \Phi(0{,}67) - \Phi(-0{,}67) = 0{,}749 - 0{,}251 = 0{,}498. \end{aligned}$$

Agora, vamos calcular $P(|X| > 2)$:

$$\begin{aligned} P(|X| > 2) &= P(X > 2) + P(X < -2) \\ &= P[(X - 4)/3 > (2 - 4)/3] + P[(X - 4)/3 < (-2 - 4)/3] \\ &= 1 - \Phi(-2/3) + \Phi(-2) \\ &= 1 - 0{,}251 + 0{,}023 = 0{,}772. \end{aligned}$$

Propriedades Adicionais da Distribuição Normal

Terminamos esta subseção reunindo vários outros fatos sobre as distribuições normais que usaremos mais tarde.

PROPRIEDADE NORMAL.2
Se $X \sim \text{Normal}(\mu,\sigma^2)$, então, $aX + b \sim \text{Normal}(a\mu + b, a^2\mu^2)$.

Assim, se $X \sim \text{Normal}(1,9)$, então, $Y = 2X + 3$ será distribuída como normal com média $2\text{E}(X) + 3 = 5$ e variância $2^2 \cdot 9 = 36$; $\text{dp}(Y) = 2\text{dp}(X) = 2 \cdot 3 = 6$.

Anteriormente, explicamos como, em geral, correlação zero e independência não são a mesma coisa. No caso de variáveis aleatórias normalmente distribuídas, é possível constatar que a correlação zero é suficiente para a independência.

PROPRIEDADE NORMAL.3
Se X e Y forem conjunta e normalmente distribuídas, então, elas serão independentes se, e somente se, $\text{Cov}(X,Y) = 0$.

PROPRIEDADE NORMAL.4
Qualquer combinação linear de variáveis aleatórias independentes e identicamente normalmente distribuídas tem uma distribuição normal.

Por exemplo, sejam X_i, para $i = 1$, 2 e 3, variáveis aleatórias independentes distribuídas como $\text{Normal}(\mu,\sigma^2)$. Defina $W = X_1 + 2X_2 - 3X_3$. Então, W será normalmente distribuída; precisamos simplesmente encontrar sua média e sua variância. Agora,

$$\text{E}(W) = \text{E}(X_1) + 2\text{E}(X_2) - 3\text{E}(X_3) = \mu + 2\mu - 3\mu = 0.$$

Além disso,

$$\text{Var}(W) = \text{Var}(X_1) + 4\text{Var}(X_2) + 9\text{Var}(X_3) = 14\sigma^2.$$

A propriedade Normal.4 também conclui que a média das variáveis aleatórias independentes normalmente distribuídas tem uma distribuição normal. Se $Y_1, Y_2, ..., Y_n$ forem variáveis aleatórias independentes e cada uma for distribuída como $\text{Normal}(\mu,\sigma^2)$, então,

$$\bar{Y} \sim \text{Normal}(\mu,\sigma^2/n). \quad \textbf{(B.40)}$$

Esse resultado é crítico para a inferência com respeito à média em uma população normal.

A Distribuição Qui-Quadrado

A distribuição qui-quadrado é obtida diretamente das variáveis aleatórias independentes normais padrões. Sejam Z_i, $i = 1,2,...,n$ variáveis aleatórias independentes, cada uma distribuída como normal padrão. Defina uma nova variável aleatória como a soma dos quadrados de Z_i:

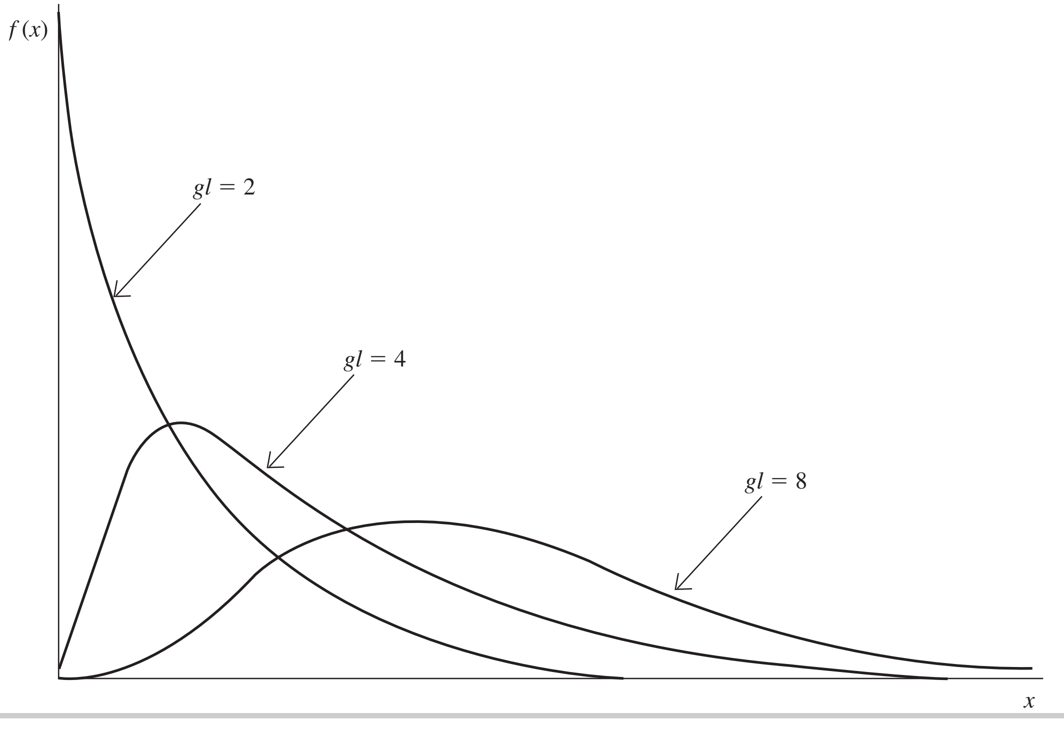
$$X = \sum_{i=1}^{n} Z_i^2. \quad \textbf{(B.41)}$$

Então, X terá o que é conhecido como **distribuição qui-quadrado** com n **graus de liberdade** (ou *gl* abreviadamente). Escrevemos isso como $X \sim \chi^2_n$. Os *gl* em uma distribuição qui-quadrado correspondem ao número de termos na soma em (B.41). O conceito de graus de liberdade desempenhará um papel importante em nossas análises estatísticas e econométricas.

A fdp de uma distribuição qui-quadrado com diversos graus de liberdade é mostrada na Figura B.9; não precisaremos da fórmula dessa fdp e, portanto, não a reproduzimos aqui. Pela equação (B.41), fica claro que uma variável aleatória qui-quadrada é sempre não-negativa, e que, diferentemente da distribuição normal, a distribuição qui-quadrado não é simétrica em torno de qualquer ponto. É possível mostrar que, se $X \sim \chi^2_n$, o valor esperado de X será n [o número de termos em (B.41)], e a variância de X será $2n$.

Figura B.9

A distribuição qui-quadrado com vários graus de liberdade.



A Distribuição *t*

A distribuição t é o burro de carga na estatística clássica e nas análises de regressão múltipla. Obtemos uma distribuição t a partir de uma variável aleatória normal padrão e de uma variável aleatória qui-quadrada.

Suponhamos que Z tenha uma distribuição normal padrão e que X tenha uma distribuição qui-quadrado com n graus de liberdade. Adicionalmente, suponhamos que Z e X sejam independentes. Então, a variável aleatória

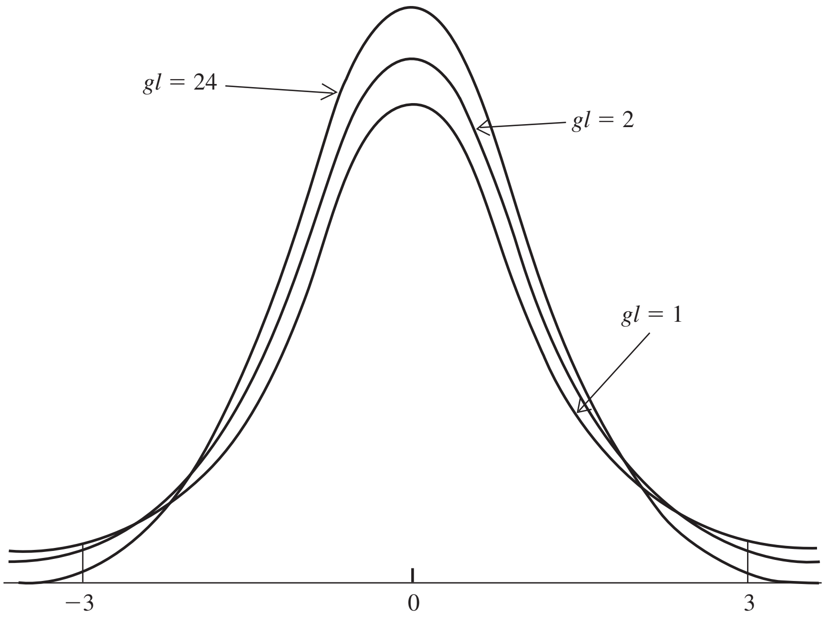
$$T = \frac{Z}{\sqrt{X/n}} \tag{B.42}$$

terá uma **distribuição *t*** com n graus de liberdade. Vamos representar isso como $T \sim t_n$. A distribuição t obtém seus graus de liberdade da variável aleatória qui-quadrada no denominador de (B.42).

A fdp da distribuição t tem uma forma semelhante à da distribuição normal padrão, exceto pelo fato de que ela é mais espalhada e, portanto, tem mais área nos extremos. O valor esperado de uma variável aleatória com distribuição t é zero (no sentido exato, o valor esperado somente existirá para $n > 1$), e a variância será $n/(n - 2)$ para $n > 2$. (Não existe variância de $n \leq 2$ devido à distribuição ser tão espalhada.) A fdp da distribuição t está traçada na Figura B.10 para vários graus de liberdade. Conforme os graus de liberdade vão ficando maiores, a distribuição t se aproxima da distribuição normal padrão.

Figura B.10

A distribuição *t* com vários graus de liberdade.



A Distribuição *F*

Outra distribuição importante na estatística e na econometria é a distribuição F. Em particular, a distribuição F será usada para testar hipóteses no contexto de análise de regressão múltipla.

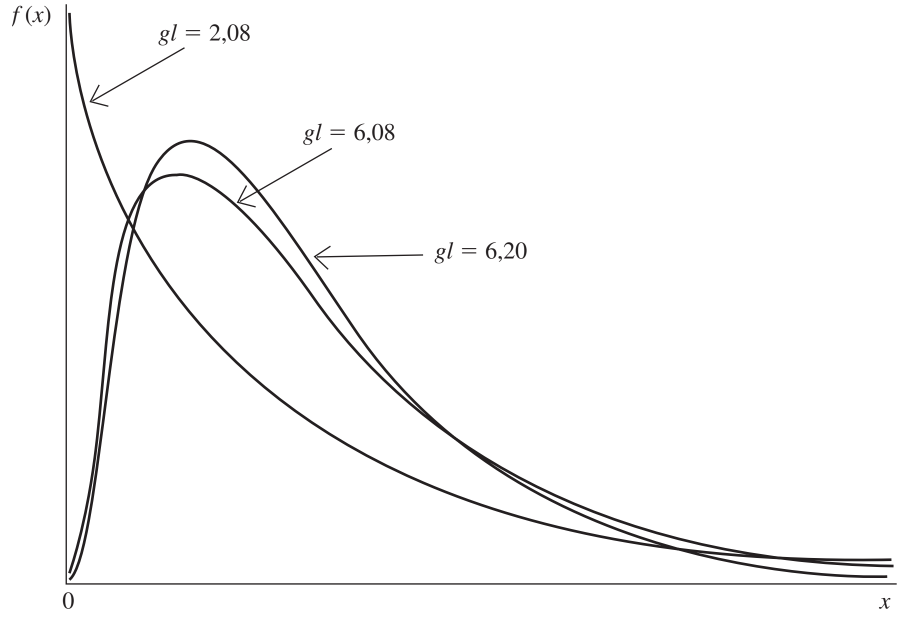
Para definir uma variável aleatória F, sejam $X_1 \sim \chi^2_{k_1}$ e $X_2 \sim \chi^2_{k_2}$ e X_1 e X_2 sejam independentes. Então, a variável aleatória

$$F = \frac{(X_1/k_1)}{(X_2/k_2)} \tag{B.43}$$

terá uma **distribuição *F*** com (k_1, k_2) graus de liberdade. Representamos isso como $F \sim F_{k_1,k_2}$. A fdp da distribuição F com diferentes graus de liberdade é mostrada na Figura B.11.

Figura B.11

A distribuição F_{k_1,k_2} para vários graus de liberdade, k_1 e k_2.



A ordem dos graus de liberdade em $F_{k1,k2}$ é crítica. O inteiro k_1 é chamado de *numerador dos graus de liberdade*, por ele estar associado à variável qui-quadrada no numerador. Da mesma forma, o inteiro k_2 é chamado de *denominador dos graus de liberdade*, por ele estar associado à variável qui-quadrada no denominador. Isso pode ser um pouco complicado, pois (B.43) também pode ser escrita como $(X_1k_2)/(X_2k_1)$, de forma que k_1 aparece no denominador. Apenas lembre-se de que o numerador *gl* é o inteiro associado à variável qui-quadrada no numerador de (B.43); de forma semelhante, uma associação é válida para o denominador *gl*.

RESUMO

Neste apêndice, revisamos os conceitos de probabilidade que são necessários em econometria. A maioria desses conceitos deve ser familiar a você de seus cursos introdutórios de probabilidade e estatística. Alguns dos tópicos mais avançados, como as características das esperanças condicionais, não precisam ser dominados agora — haverá tempo para tal quando esses conceitos surgirem, no contexto da análise de regressão na Parte 1.

Em um curso introdutório de estatística, o foco está no cálculo de médias, variâncias, covariâncias, e assim por diante, para distribuições particulares. Na Parte 1, não precisaremos de tais cálculos:

na maioria das vezes, recorremos às *propriedades* das esperanças, das variâncias etc., que explicamos neste apêndice.

PROBLEMAS

B.1 Suponha que um aluno do ensino médio está se preparando para prestar o exame vestibular. Explique por que a nota do vestibular dele é adequadamente vista como uma variável aleatória.

B.2 Defina X como uma variável aleatória distribuída como Normal(5,4). Encontre as probabilidades dos seguintes eventos

(i) $P(X \leq 6)$

(ii) $P(X > 4)$

(iii) $P(|X - 5| > 1)$

B.3 Muito se fala sobre o fato de que certos fundos mútuos têm desempenho superior ao do mercado ano após ano (isto é, o retorno por manter quotas nos fundos mútuos é maior que o retorno de possuir um portfólio como o da S&P 500). Em termos concretos, considere um período de dez anos e que a população de fundos mútuos seja os 4.170 reportados no *The Wall Street Journal* de 1º de janeiro de 1995. Ao dizermos que o desempenho relativo ao mercado é aleatório, queremos dizer que cada fundo tem uma possibilidade 50-50 de ter desempenho superior ao do mercado em qualquer ano e que o desempenho é independente de ano para ano.

(i) Se o desempenho relativo ao mercado for realmente aleatório, qual será a probabilidade de que qualquer fundo particular tenha um desempenho superior ao do mercado em todos os dez anos?

(ii) Encontre a probabilidade de que pelo menos um fundo dos 4.170 tenha um desempenho superior ao do mercado em todos os dez anos. Qual sua conclusão sobre sua resposta?

(iii) Se você possuir um programa estatístico que calcule probabilidades binomiais, encontre a probabilidade de que pelo menos cinco fundos tenham desempenho superior ao do mercado em todos os dez anos.

B.4 Para um município selecionado aleatoriamente nos Estados Unidos, seja X a proporção de adultos com mais de 65 anos que estejam empregados, ou a taxa de emprego das pessoas mais velhas. Então, X estará restrita a um valor entre zero e um. Suponha que a função de distribuição cumulativa de X seja dada por $F(x) = 3x^2 - 2x^3$ para $0 \leq x \leq 1$. Encontre a probabilidade de que a taxa de emprego das pessoas mais velhas seja de pelo menos 0,6 (60%).

B.5 Um pouco antes da seleção dos jurados para o julgamento por assassinato de O. J. Simpson em 1995, uma pesquisa de opinião constatou que 20% da população adulta acreditava que Simpson era inocente (após a maioria das provas físicas do caso ter se tornada pública). Ignore o fato de que esses 20% sejam uma estimativa baseada em uma subamostra da população; a título ilustrativo, considere esse número como a porcentagem verdadeira das pessoas que achavam que Simpson era inocente, antes da seleção dos jurados. Assuma que os 12 jurados tenham sido selecionados aleatória e independentemente da população (embora isso não tenha sido verdade).

(i) Encontre a probabilidade de que no júri havia pelo menos um membro que acreditava na inocência de Simpson antes da seleção dos jurados. [*Sugestão*: Defina a variável aleatória

X Binomial(12;0,20) como o número de jurados que acreditavam na inocência de Simpson.]

(ii) Encontre a probabilidade de que no júri havia pelo menos dois membros que acreditavam na inocência de Simpson. [*Sugestão*: $P(X \geq 2) = 1 - P(X \leq 1)$, e $P(X \leq 1) = P(X = 0) + P(X = 1)$.]

B.6 (Exige cálculo integral.) Seja X a sentença prisional, em anos, das pessoas condenadas por roubo de veículos em determinado estado dos Estados Unidos. Suponha que a fdp de X seja dada por

$$f(x) = (1/9)x^2,\ 0 < x < 3.$$

Use integração para encontrar a sentença prisional esperada.

B.7 Se um jogador de basquetebol converte 74% dos lances livres que faz, então, em média, quantos ele converterá em um jogo com oito lances livres?

B.8 Suponha que um aluno de uma universidade esteja fazendo três cursos, em determinado semestre: um curso de dois créditos, um curso de três créditos e um curso de quatro créditos. A nota esperada no curso de dois créditos é 3,5, enquanto a nota esperada nos cursos de três e quatro créditos é 3,0. Qual será a nota média global esperada no semestre? (Lembre-se de que cada curso é ponderado pela sua participação no número total de unidades de créditos.)

B.9 Seja X o salário anual dos professores universitários nos Estados Unidos, medido em milhares de dólares. Suponha que a média salarial seja 52,3 com um desvio-padrão de 14,6. Encontre a média e o desvio-padrão quando o salário é medido em dólares.

B.10 Suponha que, em uma grande universidade, a nota média de graduação, *nmgrad*, e a nota do exame vestibular, *sat*, estejam relacionadas pela esperança condicional $E(nmgrad|sat) = 0{,}70 + 0{,}002\ sat$.

(i) Encontre a *nmgrad* esperada quando $sat = 800$. Encontre $E(nmgrad|sat = 1.400)$. Comente sobre a diferença.

(ii) Se a *sat* média da universidade for 1.100, qual será a *nmgrad* média? (*Sugestão*: use a propriedade EC.4.)