ACH2053 – Introdução à Estatística

Conteúdo Teórico:
04 – Esperança

Marcelo de Souza Lauretto
Sistemas de Informação – EACH
www.each.usp.br/lauretto

Referência:
Morris DeGroot, Mark Schervish.
Probability and Statistics. 4th Ed. – 4º capítulo

Abreviações e notações úteis

- Abreviações:
  - f.p.: função de probabilidade (variável discreta)
    - Inglês: p.f. (probability function)
  - f.d.p.: função de densidade de probabilidade (variável contínua)
    - Inglês: p.d.f. (probability density function)
  - f.d.a.: função de distribuição acumulada
    - Inglês: c.d.f. (cummulative distribution function)
- Notações:
  - $f(x)$: f.p. ou f.d.p. da variável aleatória *X*
  - $F(x)$: f.d.a. da variável aleatória *X*
  - $F^{-1}(p)$: função quantil de uma variável aleatória X; $p \in (0,1)$
  - $f(x, y)$: f.p. ou f.d.p. conjunta das variáveis aleatórias *X*,*Y*
  - $f_1(x)$: f.p. ou f.d.p. marginal de X; $f_2(y)$: f.p. ou f.d.p. marginal de *Y*;
  - $g_1(x|y)$: f.p. ou f.d.p. condicional de *X* dado $Y = y$; $g_2(y|x)$: f.p. ou f.d.p. condicional de *Y* dado $X = x$
- Slides com símbolo ☛ correspondem aos tópicos relevantes

☛ 4.1 Esperança de variáveis aleatórias

- A distribuição de uma variável aleatória *X* contém toda a informação probabilística a respeito de X;
- Contudo, a distribuição completa de *X* é difícil ou inconveniente de ser apresentada
  - Necessidade de medidas resumo;
- A medida resumo mais popular é a *média*, também denominada *valor esperado* ou *esperança*
  - Útil para dar às pessoas uma ideia de onde o valor esperado de X deve se situar, sem necessidade de tentar descrever a distribuição completa;
  - Intuitivamente, a esperança de uma variável aleatória é a média ponderada dos possíveis valores da variável aleatória, em que os pesos correspondem às probabilidades (ou densidades de probabilidade)
- O valor esperado também tem papel importante em métodos de aproximação (especialmente em simulação);

4.1 Esperança de variáveis aleatórias
Esperança de distribuições discretas

- Exemplo 4.1.1:

Fair Price for a Stock. An investor is considering whether or not to invest \$18 per share in a stock for one year. The value of the stock after one year, in dollars, will be $18 + X$, where X is the amount by which the price changes over the year. At present X is unknown, and the investor would like to compute an "average value" for X in order to compare the return she expects from the investment to what she would get by putting the \$18 in the bank at 5% interest. ◀

- Exemplo 4.1.2:

Stock Price Change. Suppose that the change in price of the stock in Example 4.1.1 is a random variable X that can assume only the four different values -2, 0, 1, and 4, and that $\Pr(X = -2) = 0.1$, $\Pr(X = 0) = 0.4$, $\Pr(X = 1) = 0.3$, and $\Pr(X = 4) = 0.2$. Then the weighted avarage of these values is

$$-2(0.1) + 0(0.4) + 1(0.3) + 4(0.2) = 0.9.$$

The investor now compares this with the interest that would be earned on \$18 at 5% for one year, which is $18 \times 0.05 = 0.9$ dollars. From this point of view, the price of \$18 seems fair. ◀

4.1 Esperança de variáveis aleatórias
Esperança de distribuições discretas

- O cálculo apresentado no Exemplo 4.1.2 é facilmente generalizável para variáveis aleatórias com números <u>finitos</u> de valores.

- **Definição 4.1.1: Média de variáveis aleatórias discretas limitadas**
  Seja *X* uma variável aleatória discreta limitada com f.p. f. A *esperança* de *X*, denotada por $E(X)$, é um número definido como segue:

$$E(X) = \sum_{\text{All } x} x f(x). \tag{4.1.1}$$

  A esperança de X é também referida como *média* de X ou *valor esperado* de X.

4.1 Esperança de variáveis aleatórias
Esperança de distribuições discretas

- No exemplo 4.1.2, $E(X) = 0.9$.
  - 0.9 não é um dos valores possíveis de *X* naquele exemplo.
  - Essa característica é típica com variáveis aleatórias discretas.

- Exemplo 4.1.3:

Bernoulli Random Variable. Let X have the Bernoulli distribution with parameter p, that is, assume that X takes only the two values 0 and 1 with $\Pr(X = 1) = p$. Then the mean of X is

$$E(X) = 0 \times (1 - p) + 1 \times p = p. \quad \blacktriangleleft$$

- Se *X* for ilimitada, ainda pode ser possível definir $E(X)$ como a média ponderada de seus possíveis valores, desde que satisfeita certa condição.

4.1 Esperança de variáveis aleatórias
Esperança de distribuições discretas

- **Definição 4.1.1: Média de variáveis aleatórias discretas gerais**
  Seja X uma variável aleatória discreta com f.p. f. Suponha que pelo menos uma das seguintes somas seja finita:

$$\sum_{\text{Positive } x} xf(x), \qquad \sum_{\text{Negative } x} xf(x). \tag{4.1.2}$$

  Então a média, esperança, ou valor esperado de X é dita existir e é definida como

$$E(X) = \sum_{\text{All } x} xf(x). \tag{4.1.3}$$

  Se as duas somas em (4.1.2) são infinitas, então $E(X)$ *não existe*.

4.1 Esperança de variáveis aleatórias
Esperança de distribuições discretas

- A razão pela qual a esperança falha em existir se ambas as somas em (4.1.2) são infinitas é que, em tais casos, a soma em (4.1.3) não é bem definida.
  - Ou a soma total em (4.1.3) não converge ou pode convergir para diferentes valores rearranjando-se os termos em ordens diferentes.
- Se apenas uma das duas somas em (4.1.2) é infinita, então o valor esperado também é infinito com o mesmo sinal da soma que for infinita.
- Se ambas as somas são finitas, então a soma em (4.1.3) converge e não depende da ordem na qual os termos são somados.

4.1 Esperança de variáveis aleatórias
Esperança de distribuições discretas

- Exemplo 4.1.5:

An Infinite Mean. Let X be a random variable whose p.f. is

$$f(x) = \begin{cases} \dfrac{1}{x(x+1)} & \text{if } x = 1, 2, 3, \ldots, \\ 0 & \text{otherwise.} \end{cases}$$

The sum over negative values in Eq. (4.1.2) is 0, so the mean of X exists and is

$$E(X) = \sum_{x=1}^{\infty} x \frac{1}{x(x+1)} = \infty.$$

We say that the mean of X is *infinite* in this case. ◀

☛ 4.1 Esperança de variáveis aleatórias
Esperança de distribuições discretas

- Observação:
  - Embora $E(X)$ seja denominada esperança de X, ela depende somente da distribuição de X.
  - Duas variáveis aleatórias com mesma distribuição terão a mesma esperança, mesmo que tenham significados diferentes.
  - Por esta razão, frequentemente nos referimos à esperança de uma distribuição mesmo que não tenhamos em mente uma variável aleatória com aquela distribuição.

☛ 4.1 Esperança de variáveis aleatórias
Esperança de distribuições contínuas

- A ideia de computar uma média ponderada dos possíveis valores pode ser generalizada para variáveis aleatórias contínuas usando-se integrais ao invés de somas.
- A distinção entre variáveis aleatórias limitadas e ilimitadas aparece nesse caso pelas mesmas razões.

- **Definição 4.1.1: Média de variáveis aleatórias contínuas limitadas**
  Seja X uma variável aleatória contínua limitada com f.d.p. f. A média, esperança ou valor esperado de X, denotada $E(X)$, é definida como segue:

$$E(X) = \int_{-\infty}^{\infty} x f(x)\, dx. \qquad (4.1.4)$$

4.1 Esperança de variáveis aleatórias
Esperança de distribuições contínuas

- Exemplo 4.1.6:

Expected Failure Time. An appliance has a maximum lifetime of one year. The time X until it fails is a random variable with a continuous distribution having p.d.f.

$$f(x) = \begin{cases} 2x & \text{for } 0 < x < 1, \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$E(X) = \int_0^1 x(2x)\,dx = \int_0^1 2x^2\,dx = \frac{2}{3}.$$

We can also say that the expectation of the distribution with p.d.f. f is 2/3. ◄

- Para variáveis aleatórias contínuas gerais, modifica-se a Definição 4.1.2.

4.1 Esperança de variáveis aleatórias
Esperança de distribuições contínuas

- **Definição 4.1.1: Média de variáveis aleatórias contínuas gerais**

  Seja *X* uma variável aleatória contínua com f.d.p. *f*. Suponha que pelo menos uma das seguintes integrais seja finita:

$$\int_0^\infty xf(x)dx, \quad \int_{-\infty}^0 xf(x)dx. \qquad (4.1.5)$$

  Então a média, esperança, ou valor esperado de X é dita existir e é definida como

$$E(X) = \int_{-\infty}^\infty xf(x)dx. \qquad (4.1.6)$$

  Se as duas integrais em (4.1.5) são infinitas, então $E(X)$ *não existe*.

☛ 4.1 Esperança de variáveis aleatórias
Esperança de distribuições contínuas

- Exemplo 4.1.7:

Failure after Warranty. A product has a warranty of one year. Let X be the time at which the product fails. Suppose that X has a continuous distribution with the p.d.f.

$$f(x) = \begin{cases} 0 & \text{for } x < 1, \\ \frac{2}{x^3} & \text{for } x \geq 1. \end{cases}$$

The expected time to failure is then

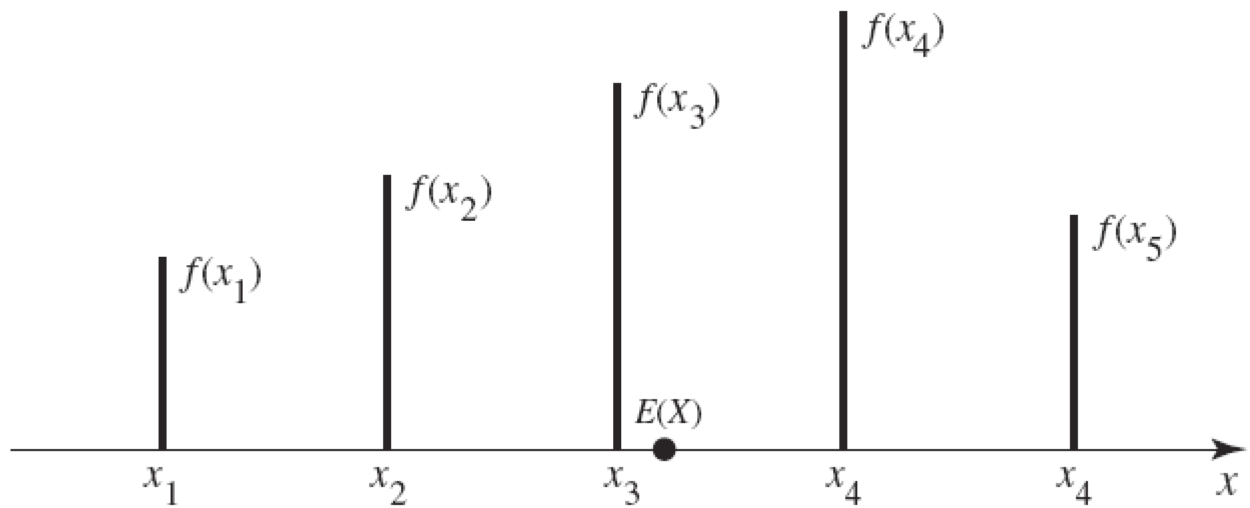
$$E(X) = \int_1^\infty x \frac{2}{x^3} dx = \int_1^\infty \frac{2}{x^2} dx = 2. \quad ◀$$

4.1 Esperança de variáveis aleatórias
Interpretação da esperança

- A esperança de uma variável aleatória ou, equivalentemente, a média de sua distribuição, pode ser vista como sendo o centro de gravidade daquela distribuição
- Considere a f.p. representada na Fig. 4.1.
  - O eixo x pode ser visto como um longo bastão sem massa ao longo do qual pesos são fixados.
  - Se cada peso igual a $f(x_j)$ é fixado neste bastão no ponto x_j, então o bastão estará equilibrado se for apoiado no ponto $E(X)$.
- Considere agora a f.d.p. representada na Fig. 4.2.
  - Nesse caso, o eixo x pode ser visto como um longo bastão sobre o qual a massa varia continuamente.
  - Se a densidade do bastão em cada ponto x é igual a $f(x)$, então o centro de gravidade do bastão será localizado no ponto $E(X)$.

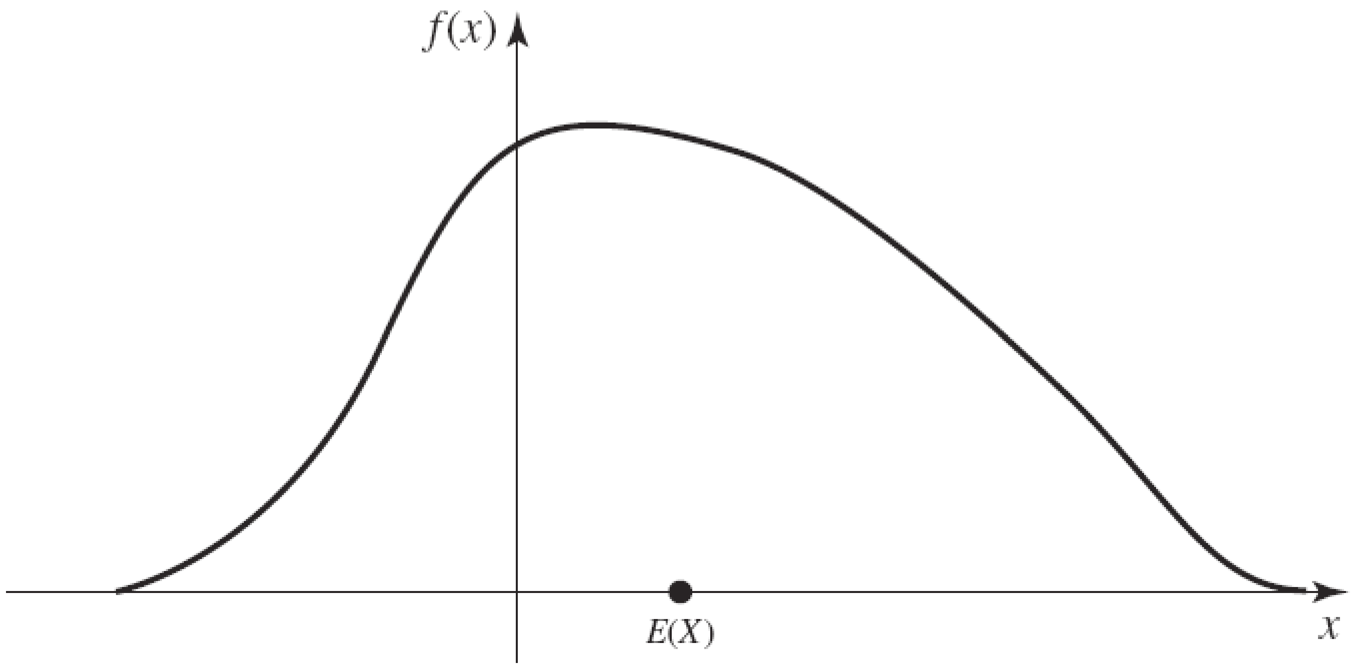
4.1 Esperança de variáveis aleatórias
Interpretação da esperança

Figure 4.1 The mean of a discrete distribution.



4.1 Esperança de variáveis aleatórias
Interpretação da esperança

Figure 4.2 The mean of a continuous distribution.



4.1 Esperança de variáveis aleatórias
Interpretação da esperança

- Note que a média de uma distribuição pode ser fortemente afetada mesmo por uma pequena mudança no montante de probabilidade atribuída por um valor grande de x.
  - Por exemplo, a média da distribuição representada pela f.p. na Fig. 4.1 pode ser movida para qualquer ponto no eixo x, não importando quão distante da origem aquele ponto pode estar
  - Basta remover um montante de probabilidade arbitrariamente pequeno mas positivo de um dos pontos x_j e adicionando este montante de probabilidade a um ponto suficientemente distante da origem.
- Outro fato é que, se uma distribuição for simétrica com respeito a um ponto dado x_0, ou seja, se $f(x_0 + \delta) = f(x_0 - \delta)$ para todo δ, <u>e se sua média</u> $E(X)$ <u>existe</u>, então $E(X) = x_0$.
  - A condição de existência da média é importante, pois há distribuições simétricas para as quais a média não existe.
    Ex: Distribuição de Cauchy

4.1 Esperança de variáveis aleatórias
A esperança de uma função

- Como calcular a esperança de uma variável $Y = r(X)$, dado que conhecemos a distribuição de X?

- Exemplo 4.1.10:

Failure Rate and Time to Failure. Suppose that appliances manufactured by a particular company fail at a rate of X per year, where X is currently unknown and hence is a random variable. If we are interested in predicting how long such an appliance will last before failure, we might use the mean of $1/X$. How can we calculate the mean of $Y = 1/X$? ◄

4.1 Esperança de variáveis aleatórias
A esperança de uma função

- **Funções de uma única variável aleatória:**
  Se X é uma variável aleatória cuja f.d.p. é f, então a esperança de cada função a valores reais $r(X)$ pode ser encontrada aplicando-se a definição da esperança de $r(X)$ como segue:
  - Se $Y = r(X)$, determine a distribuição de probabilidade de Y, e então determine $E(Y)$ aplicando ou a Eq. (4.1.1) ou (4.1.4).
  - Por exemplo, suponha que Y tenha uma distribuição contínua com f.d.p. g.Então

$$E[r(X)] = E(Y) = \int_{-\infty}^{\infty} yg(y)\, dy, \tag{4.1.8}$$

se a esperança existe.

4.1 Esperança de variáveis aleatórias
A esperança de uma função

- Exemplo 4.1.11:

Failure Rate and Time to Failure. In Example 4.1.10, suppose that the p.d.f. of X is

$$f(x) = \begin{cases} 3x^2 & \text{if } 0 < x < 1, \\ 0 & \text{otherwise.} \end{cases}$$

Let $r(x) = 1/x$. Using the methods of Sec. 3.8, we can find the p.d.f. of $Y = r(X)$ as

$$g(y) = \begin{cases} 3y^{-4} & \text{if } y > 1, \\ 0 & \text{otherwise.} \end{cases}$$

The mean of Y is then

$$E(Y) = \int_0^{\infty} y 3y^{-4} dy = \frac{3}{2}. \blacktriangleleft$$

4.1 Esperança de variáveis aleatórias
A esperança de uma função

- Embora o método apresentado no Exemplo 4.1.11 possa ser usado para encontrar a média de uma variável aleatória contínua, não é necessário determinar a f.d.p. de $r(X)$ para calcular a esperança $E[r(X)]$.
- De fato, pode-se mostrar que o valor de $E[r(X)]$ pode ser calculado diretamente usando-se o resultado a seguir.

4.1 Esperança de variáveis aleatórias
A esperança de uma função

- **Teorema 4.1.1: Lei do estatístico inconsciente**
  Seja *X* uma variável aleatória, e seja *r* uma função a valores reais com domínio nos reais.
  Se X tem uma distribuição contínua, então

$$E[r(X)] = \int_{-\infty}^{\infty} r(x) f(x)\, dx, \tag{4.1.9}$$

  se a média existe. Se X tem uma distribuição discreta,

$$E[r(X)] = \sum_{\text{All } x} r(x) f(x), \tag{4.1.10}$$

  se a média existe.

☛ 4.1 Esperança de variáveis aleatórias
A esperança de uma função

- **Teorema 4.1.1 (cont)**

Proof A general proof will not be given here. However, we shall provide a proof for two special cases. First, suppose that the distribution of X is discrete. Then the distribution of Y must also be discrete. Let g be the p.f. of Y. For this case,

$$\begin{aligned}\sum_y yg(y) &= \sum_y y \Pr[r(X) = y] \\ &= \sum_y y \sum_{x:r(x)=y} f(x) \\ &= \sum_y \sum_{x:r(x)=y} r(x) f(x) = \sum_x r(x) f(x).\end{aligned}$$

Hence, Eq. (4.1.10) yields the same value as one would obtain from Definition 4.1.1 applied to Y.

- Por que “lei do estatístico inconsciente”? Porque é possível calcular a esperança de Y sem necessidade de conhecer sua distribuição (basta conhecer a distribuição de X)

4.1 Esperança de variáveis aleatórias
A esperança de uma função

- Exemplo 4.1.12:

Failure Rate and Time to Failure. In Example 4.1.11, we can apply Theorem 4.1.1 to find

$$E(Y) = \int_0^1 \frac{1}{x} 3x^2 dx = \frac{3}{2},$$

the same result we got in Example 4.1.11. ◀

- Exemplo 4.1.13:

Determining the Expectation of $X^{1/2}$. Suppose that the p.d.f. of X is as given in Example 4.1.6 and that $Y = X^{1/2}$. Then, by Eq. (4.1.9),

$$E(Y) = \int_0^1 x^{1/2}(2x)\, dx = 2\int_0^1 x^{3/2}\, dx = \frac{4}{5}.$$ ◀

4.1 Esperança de variáveis aleatórias
A esperança de uma função

- **Funções de diversas variáveis aleatórias:**
- **Teorema 4.1.2: Lei do estatístico inconsciente**
  Suponha que $X_1, X_2, ..., X_n$ são variáveis aleatórias com f.d.p. conjunta $f(x_1, x_2, ..., x_n)$. Seja r uma função a valores reais de n variáveis reais. Suponha que $Y=r(X_1, X_2, ..., X_n)$. Então $E(Y)$ pode ser determinada diretamente da relação

$$E(Y) = \int \underset{R^n}{\cdots} \int r(x_1, \ldots, x_n) f(x_1, \ldots, x_n)\, dx_1 \cdots dx_n,$$

  se a média existe. Analogamente, se $X_1, X_2, ..., X_n$ tem uma distribuição conjunta discreta com f.p. conjunta $f(x_1, x_2, ..., x_n)$,

$$E(Y) = \sum_{\text{All } x_1,\ldots,x_n} r(x_1, \ldots, x_n) f(x_1, \ldots, x_n),$$

  se a média existe.

4.1 Esperança de variáveis aleatórias
A esperança de uma função

- **Funções de diversas variáveis aleatórias:**
- Exemplo 4.1.16:

Determining the Expectation of a Function of Two Variables. Suppose that a point (X, Y) is chosen at random from the square S containing all points (x, y) such that $0 \le x \le 1$ and $0 \le y \le 1$. We shall determine the expected value of $X^2 + Y^2$.

Since X and Y have the uniform distribution over the square S, and since the area of S is 1, the joint p.d.f. of X and Y is

$$f(x, y) = \begin{cases} 1 & \text{for } (x, y) \in S, \\ 0 & \text{otherwise.} \end{cases}$$

Therefore,

$$E(X^2 + Y^2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x^2 + y^2) f(x, y)\, dx\, dy$$
$$= \int_0^1 \int_0^1 (x^2 + y^2)\, dx\, dy = \frac{2}{3}. \quad \blacktriangleleft$$

☛ **4.2 Propriedades da esperança**
Teoremas básicos

- Apresentamos alguns resultados que simplificam o cálculo de esperanças para algumas funções comuns de variáveis aleatórias.

- Suponha que X seja uma variável aleatória para a qual a esperança E(X) existe. Serão apresentados alguns resultados das propriedades básicas da esperança.

4.2 Propriedades da esperança
Teoremas básicos

- **Teorema 4.2.1: Função linear**
  Se Y=aX+b, onde a e b são constantes finitas, então

$$E(Y) = aE(X) + b.$$

Proof We first shall assume, for convenience, that X has a continuous distribution for which the p.d.f. is f. Then

$$\begin{aligned} E(Y) = E(aX + b) &= \int_{-\infty}^{\infty} (ax + b) f(x)\, dx \\ &= a \int_{-\infty}^{\infty} x f(x)\, dx + b \int_{-\infty}^{\infty} f(x)\, dx \\ &= aE(X) + b. \end{aligned}$$

A similar proof can be given for a discrete distribution. ■

- **Corolário 4.2.1:**
  Se X=c com probabilidade 1, então E(X) = c.

4.2 Propriedades da esperança
Teoremas básicos

- Exemplo 4.2.2:

Investment. An investor is trying to choose between two possible stocks to buy for a three-month investment. One stock costs \$50 per share and has a rate of return of R_1 dollars per share for the three-month period, where R_1 is a random variable. The second stock costs \$30 per share and has a rate of return of R_2 per share for the same three-month period. The investor has a total of \$6000 to invest. For this example, suppose that the investor will buy shares of only one stock. (In Example 4.2.3, we shall consider strategies in which the investor buys more than one stock.) Suppose that R_1 has the uniform distribution on the interval $[-10, 20]$ and that R_2 has the uniform distribution on the interval $[-4.5, 10]$. We shall first compute the expected dollar value of investing in each of the two stocks. For the first stock, the \$6000 will purchase 120 shares, so the return will be $120R_1$, whose mean is $120E(R_1) = 600$. (Solve Exercise 1 in Sec. 4.1 to see why $E(R_1) = 5$.) For the second stock, the \$6000 will purchase 200 shares, so the return will be $200R_2$, whose mean is $200E(R_2) = 550$. The first stock has a higher expected return.

4.2 Propriedades da esperança
Teoremas básicos

- **Teorema 4.2.2:**
Se existe uma constante tal que Pr(X≥a)=1, então E(X) ≥ a. Se existe uma constante tal que Pr(X ≤ b)=1, então E(X) ≤ b.

Proof We shall assume again, for convenience, that X has a continuous distribution for which the p.d.f. is f, and we shall suppose first that $\Pr(X \geq a) = 1$. Because X is bounded below, the second integral in (4.1.5) is finite. Then

$$E(X) = \int_{-\infty}^{\infty} x f(x)\, dx = \int_{a}^{\infty} x f(x)\, dx$$
$$\geq \int_{a}^{\infty} a f(x)\, dx = a \Pr(X \geq a) = a.$$

The proof of the other part of the theorem and the proof for a discrete distribution are similar. ■

- Segue do Teorema 4.2.2. que, se Pr(a ≤ X ≤ b)=1, então a ≤ E(X) ≤ b.

4.2 Propriedades da esperança
Teoremas básicos

- **Teorema 4.2.3:**
  Suponha que E(X)=a e que ou Pr(X≥a)=1 ou Pr(X ≤ a)=1.
  Então Pr(X = a) = 1.

Proof We shall provide a proof for the case in which X has a discrete distribution and $\Pr(X \geq a) = 1$. The other cases are similar. Let $x_1, x_2, \ldots$ include every value $x > a$ such that $\Pr(X = x) > 0$, if any. Let $p_0 = \Pr(X = a)$. Then,

$$E(X) = p_0 a + \sum_{j=1}^{\infty} x_j \Pr(X = x_j). \tag{4.2.1}$$

Each x_j in the sum on the right side of Eq. (4.2.1) is greater than a. If we replace all of the x_j's by a, the sum can't get larger, and hence

$$E(X) \geq p_0 a + \sum_{j=1}^{\infty} a \Pr(X = x_j) = a. \tag{4.2.2}$$

Furthermore, the inequality in Eq. (4.2.2) will be strict if there is even one $x > a$ with $\Pr(X = x) > 0$. This contradicts $E(X) = a$. Hence, there can be no $x > a$ such that $\Pr(X = x) > 0$. ■

4.2 Propriedades da esperança
Teoremas básicos

- **Teorema 4.2.4:**
  Suponha que $X_1, X_2, \ldots, X_n$ são n variáveis aleatórias tais que $E(X_1)$, $E(X_2)$ ... $E(X_n)$ existem. Então

$$E(X_1 + \cdots + X_n) = E(X_1) + \cdots + E(X_n).$$

Proof We shall first assume that $n = 2$ and also, for convenience, that X_1 and X_2 have a continuous joint distribution for which the joint p.d.f. is f. Then

$$\begin{aligned}
E(X_1 + X_2) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x_1 + x_2) f(x_1, x_2)\, dx_1\, dx_2 \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1 f(x_1, x_2)\, dx_1\, dx_2 + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_2 f(x_1, x_2)\, dx_1\, dx_2 \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1 f(x_1, x_2)\, dx_2\, dx_1 + \int_{-\infty}^{\infty} x_2 f_2(x_2)\, dx_2 \\
&= \int_{-\infty}^{\infty} x_1 f_1(x_1)\, dx_1 + \int_{-\infty}^{\infty} x_2 f_2(x_2)\, dx_2 \\
&= E(X_1) + E(X_2),
\end{aligned}$$

4.2 Propriedades da esperança
Teoremas básicos

- O Corolário a seguir decorre dos Teoremas 4.2.1 e 4.2.4, e afirma que a esperança de uma combinação linear entre diversas variáveis corresponde à combinação linear entre as respectivas esperanças:

- **Corolário 4.2.2:**
  Suponha que E(X_i) é finita para i=1,...,n. Para quaisquer constantes $a_1,\ldots,a_n$ e b,

$$E(a_1X_1 + \cdots + a_nX_n + b) = a_1E(X_1) + \cdots + a_nE(X_n) + b.$$

4.2 Propriedades da esperança
Teoremas básicos

- Exemplo 4.2.3:

Investment Portfolio. Suppose that the investor with \$6000 in Example 4.2.2 can buy shares of both of the two stocks. Suppose that the investor buys s_1 shares of the first stock at \$50 per share and s_2 shares of the second stock at \$30 per share. Such a combination of investments is called a *portfolio*. Ignoring possible problems with fractional shares, the values of s_1 and s_2 must satisfy

$$50s_1 + 30s_2 = 6000,$$

in order to invest the entire \$6000. The return on this portfolio will be $s_1R_1 + s_2R_2$. The mean return will be

$$s_1E(R_1) + s_2E(R_2) = 5s_1 + 2.75s_2.$$

For example, if $s_1 = 54$ and $s_2 = 110$, then the mean return is 572.5. ◄

4.2 Propriedades da esperança
Teoremas básicos

- Exemplo 4.2.5:
  - Esse exemplo mostra que, como uma variável com distribuição binomial é formada pela soma de variáveis de Bernoulli independentes, sua esperança pode ser calculada usando o teorema 4.2.4:

Sampling with Replacement. Suppose again that in a box containing red balls and blue balls, the proportion of red balls is p $(0 \le p \le 1)$. Suppose now, however, that a random sample of n balls is selected from the box *with replacement*. If X denotes the number of red balls in the sample, then X has the binomial distribution with parameters n and p, as described in Sec. 3.1. We shall now determine the value of $E(X)$.

As before, for $i = 1, \ldots, n$, let $X_i = 1$ if the ith ball that is selected is red, and let $X_i = 0$ otherwise. Then, as before, $X = X_1 + \cdots + X_n$. In this problem, the random variables $X_1, \ldots, X_n$ are independent, and the marginal distribution of each X_i is again given by Eq. (4.2.3). Therefore, $E(X_i) = p$ for $i = 1, \ldots, n$, and it follows from Theorem 4.2.4 that

$$E(X) = np. \tag{4.2.5}$$

4.2 Propriedades da esperança
Teoremas básicos

- Exemplo 4.2.5 (cont):
  - Note que, se calcularmos a esperança explicitamente pela definição 4.1.1, conforme Equação 4.2.6 abaixo, obteremos o mesmo valor para a esperança.

Thus, the mean of the binomial distribution with parameters n and p is np. The p.f. $f(x)$ of this binomial distribution is given by Eq. (3.1.4), and the mean can be computed directly from the p.f. as follows:

$$E(X) = \sum_{x=0}^{n} x \binom{n}{x} p^x q^{n-x}. \tag{4.2.6}$$

Hence, by Eq. (4.2.5), the value of the sum in Eq. (4.2.6) must be np. ◄

4.2 Propriedades da esperança
Teoremas básicos

- Exemplo 4.2.6:

Expected Number of Matches. Suppose that a person types n letters, types the addresses on n envelopes, and then places each letter in an envelope in a random manner. Let X be the number of letters that are placed in the correct envelopes. We shall find the mean of X. (In Sec. 1.10, we did a more difficult calculation with this same example.)

For $i = 1, \ldots, n$, let $X_i = 1$ if the ith letter is placed in the correct envelope, and let $X_i = 0$ otherwise. Then, for $i = 1, \ldots, n$,

$$\Pr(X_i = 1) = \frac{1}{n} \quad \text{and} \quad \Pr(X_i = 0) = 1 - \frac{1}{n}.$$

Therefore,

$$E(X_i) = \frac{1}{n} \quad \text{for } i = 1, \ldots, n.$$

4.2 Propriedades da esperança
Teoremas básicos

- Exemplo 4.2.6 (cont):

Since $X = X_1 + \cdots + X_n$, it follows that

$$E(X) = E(X_1) + \cdots + E(X_n)$$
$$= \frac{1}{n} + \cdots + \frac{1}{n} = 1.$$

Thus, the expected value of the number of correct matches of letters and envelopes is 1, regardless of the value of n. ◄

☛ 4.2 Propriedades da esperança

Esperança do produto de variáveis aleatórias independentes

- **Teorema 4.2.6:**
  Se $X_1, X_2, \ldots, X_n$ são n variáveis aleatórias independentes tais que a esperança $E(X_i)$ é finita (i=1,...,n), então

$$E\left(\prod_{i=1}^{n} X_i\right) = \prod_{i=1}^{n} E(X_i).$$

Proof We shall again assume, for convenience, that $X_1, \ldots, X_n$ have a continuous joint distribution for which the joint p.d.f. is f. Also, we shall let f_i denote the marginal p.d.f. of X_i ($i = 1, \ldots, n$). Then, since the variables $X_1, \ldots, X_n$ are independent, it follows that at every point $(x_1, \ldots, x_n) \in R^n$,

$$f(x_1, \ldots, x_n) = \prod_{i=1}^{n} f_i(x_i).$$

☛ 4.2 Propriedades da esperança
Esperança do produto de variáveis aleatórias independentes

- **Teorema 4.2.6 (cont):**

Therefore,

$$E\left(\prod_{i=1}^{n} X_i\right) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left(\prod_{i=1}^{n} x_i\right) f(x_1, \cdots, x_n)\, dx_1 \cdots dx_n$$

$$= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left[\prod_{i=1}^{n} x_i f_i(x_i)\right] dx_1 \ldots dx_n$$

$$= \prod_{i=1}^{n} \int_{-\infty}^{\infty} x_i f_i(x_i)\, dx_i = \prod_{i=1}^{n} E(X_i).$$

The proof for a discrete distribution is similar. ■

4.2 Propriedades da esperança
Esperança do produto de variáveis aleatórias independentes

- Diferença fundamental entre Teoremas 4.2.4 e 4.2.6:
  - Se cada esperança $E(X_i)$ é finita (i=1,...,n), então a soma de um grupo de variáveis aleatórias é sempre igual à soma de suas esperanças individuais (Teorema 4.2.4);
  - Contudo, a esperança do produto de um grupo de variáveis aleatórias não é sempre igual ao produto de suas esperanças individuais; Se as variáveis aleatórias são independentes, então essa igualdade também vale (Teorema 4.2.6).

4.2 Propriedades da esperança
Esperança do produto de variáveis aleatórias independentes

- Exemplo 4.2.7:

Calculating the Expectation of a Combination of Random Variables. Suppose that X_1, X_2, and X_3 are independent random variables such that $E(X_i) = 0$ and $E(X_i^2) = 1$ for $i = 1, 2, 3$. We shall determine the value of $E[X_1^2(X_2 - 4X_3)^2]$.

Since X_1, X_2, and X_3 are independent, it follows that the two random variables X_1^2 and $(X_2 - 4X_3)^2$ are also independent. Therefore,

$$\begin{aligned} E[X_1^2(X_2 - 4X_3)^2] &= E(X_1^2)E[(X_2 - 4X_3)^2] \\ &= E(X_2^2 - 8X_2X_3 + 16X_3^2) \\ &= E(X_2^2) - 8E(X_2X_3) + 16E(X_3^2) \\ &= 1 - 8E(X_2)E(X_3) + 16 \\ &= 17. \end{aligned}$$

◄

4.2 Propriedades da esperança
Esperança do produto de variáveis aleatórias independentes

- Exemplo 4.2.8:

Repeated Filtering. A filtration process removes a random proportion of particulates in water to which it is applied. Suppose that a sample of water is subjected to this process twice. Let X_1 be the proportion of the particulates that are removed by the first pass. Let X_2 be the proportion of what remains after the first pass that is removed by the second pass. Assume that X_1 and X_2 are independent random variables with common p.d.f. $f(x) = 4x^3$ for $0 < x < 1$ and $f(x) = 0$ otherwise. Let Y be the proportion of the original particulates that remain in the sample after two passes. Then $Y = (1 - X_1)(1 - X_2)$. Because X_1 and X_2 are independent, so too are $1 - X_1$ and $1 - X_2$. Since $1 - X_1$ and $1 - X_2$ have the same distribution, they have the same mean, call it μ. It follows that Y has mean μ^2. We can find μ as

$$\mu = E(1 - X_1) = \int_0^1 (1 - x_1) 4x_1^3 dx_1 = 1 - \frac{4}{5} = 0.2.$$

It follows that $E(Y) = 0.2^2 = 0.04$. ◄

4.2 Propriedades da esperança
Esperança de distribuições não negativas

- **Teorema 4.2.7: Variáveis aleatórias com valores inteiros**
  Seja X uma variável aleatória que pode assumir apenas os valores 0,1,2,... Então

$$E(X) = \sum_{n=1}^{\infty} \Pr(X \geq n). \tag{4.2.7}$$

Proof First, we can write

$$E(X) = \sum_{n=0}^{\infty} n \Pr(X = n) = \sum_{n=1}^{\infty} n \Pr(X = n). \tag{4.2.8}$$

Next, consider the following triangular array of probabilities:

$$\begin{array}{cccc}
\Pr(X=1) & \Pr(X=2) & \Pr(X=3) & \cdots \\
 & \Pr(X=2) & \Pr(X=3) & \cdots \\
 & & \Pr(X=3) & \cdots \\
 & & & \ddots
\end{array}$$

4.2 Propriedades da esperança
Esperança de distribuições não negativas

- **Teorema 4.2.7: Variáveis aleatórias com valores inteiros (cont)**

We can compute the sum of all the elements in this array in two different ways because all of the summands are nonnegative. First, we can add the elements in each column of the array and then add these column totals. Thus, we obtain the value $\sum_{n=1}^{\infty} n \Pr(X = n)$. Second, we can add the elements in each row of the array and then add these row totals. In this way we obtain the value $\sum_{n=1}^{\infty} \Pr(X \geq n)$. Therefore,

$$\sum_{n=1}^{\infty} n \Pr(X = n) = \sum_{n=1}^{\infty} \Pr(X \geq n).$$

Eq. (4.2.7) now follows from Eq. (4.2.8). ■

- Outra forma de computar *E*(*X*) para variáveis aleatórias com a condição do Teorema 4.2.7 é $E(X) = \sum_{n=1}^{\infty}[1 - F(n)]$, onde F é a f.d.a. de *X*.

4.2 Propriedades da esperança
Esperança de distribuições não negativas

- Exemplo 4.2.9:

Expected Number of Trials. Suppose that a person repeatedly tries to perform a certain task until he is successful. Suppose also that the probability of success on each given trial is p $(0 < p < 1)$ and that all trials are independent. If X denotes the number of the trial on which the first success is obtained, then $E(X)$ can be determined as follows.

Since at least one trial is always required, $\Pr(X \geq 1) = 1$. Also, for $n = 2, 3, \ldots,$ at least n trials will be required if and only if none of the first $n - 1$ trials results in success. Therefore,

$$\Pr(X \geq n) = (1 - p)^{n-1}.$$

By Eq. (4.2.7), it follows that

$$E(X) = 1 + (1 - p) + (1 - p)^2 + \cdots = \frac{1}{1 - (1 - p)} = \frac{1}{p}. \quad \blacktriangleleft$$

- O Teorema 4.2.7 possui uma versão mais geral que se aplica a todas as variáveis aleatórias não negativas.

4.2 Propriedades da esperança
Esperança de distribuições não negativas

- O Teorema 4.2.7 possui uma versão mais geral que se aplica a todas as variáveis aleatórias não negativas.

- Teorema 4.2.8: Variáveis aleatórias gerais não negativas
  Seja X uma variável aleatória não negativa com f.d.a. *F*. Então

$$E(X) = \int_0^\infty [1 - F(x)]dx. \tag{4.2.9}$$

■

4.2 Propriedades da esperança
Esperança de distribuições não negativas

- Exemplo 4.2.10:

Expected Waiting Time. Let X be the time that a customer spends waiting for service in a queue. Suppose that the c.d.f. of X is

$$F(x) = \begin{cases} 0 & \text{if } x \le 0, \\ 1 - e^{-2x} & \text{if } x > 0. \end{cases}$$

Then the mean of X is

$$E(X) = \int_0^\infty e^{-2x} dx = \frac{1}{2}. \qquad \blacktriangleleft$$

4.3 Variância

- Embora a média de uma distribuição seja um resumo útil, ela não carrega muita informação a respeito da distribuição.
- Por exemplo, uma variável aleatória *X* com média 2 tem a mesma média de uma variável aleatória constante *Y* tal que $\Pr(Y = 2) = 1$, mesmo que X não seja constante.
- Para distinguir a distribuição de *X* da distribuição de *Y* neste caso, pode ser útil fornecer alguma medida de quão dispersa é a distribuição de *X*.
- A *variância* de *X* é uma medida desse tipo.
- O *desvio padrão* de *X* corresponde à raiz quadrada da variância.
- A variância também exerce um importante papel teórico para os métodos de aproximação a serem vistos posteriormente.

4.3 Variância

- Exemplo 4.3.1:

Stock Price Changes. Consider the prices A and B of two stocks at a time one month in the future. Assume that A has the uniform distribution on the interval $[25, 35]$ and B has the uniform distribution on the interval $[15, 45]$. It is easy to see (from Exercise 1 in Sec. 4.1) that both stocks have a mean price of 30. But the distributions are very different. For example, A will surely be worth at least 25 while $\Pr(B < 25) = 1/3$. But B has more upside potential also. The p.d.f.'s of these two random variables are plotted in Fig. 4.5. ◄

4.3 Variância

- Exemplo 4.3.1 (cont):

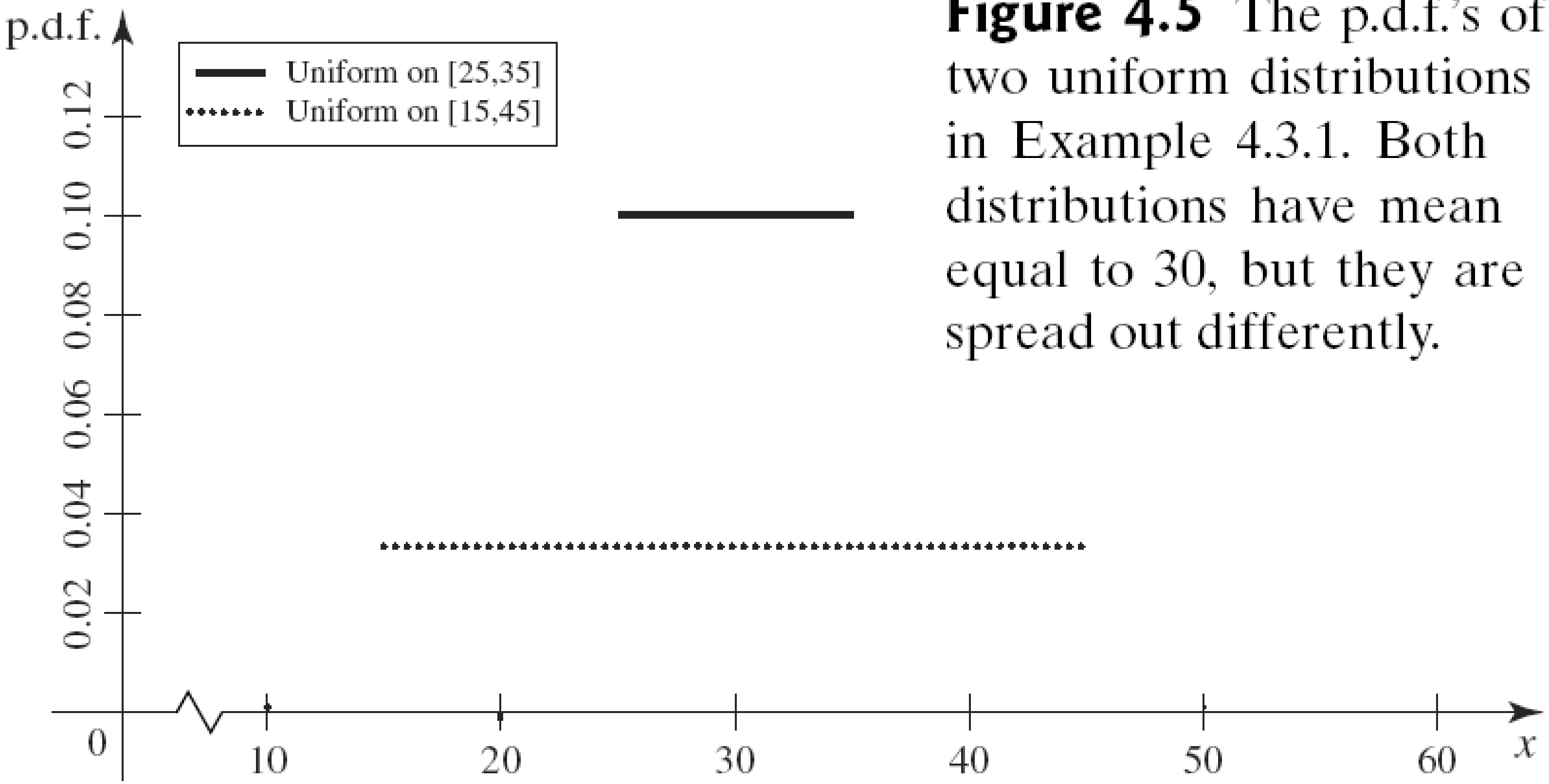


Figure 4.5 The p.d.f.'s of two uniform distributions in Example 4.3.1. Both distributions have mean equal to 30, but they are spread out differently.

  – Embora os preços dos dois papéis possuam a mesma média, os do papel B são mais dispersos do que os do papel A;
    É recomendável ter um resumo das distribuições que torne essa diferença mais fácil de se observar.

☛ 4.3 Variância
Definições da variância e do desvio padrão

- **Definição 4.3.1: Variância / desvio padrão**
  Seja X uma variável aleatória com média finita $\mu = E(X)$.
  A *variância de X*, denotada por $\mathrm{Var}(X)$, é definida como segue:

$$\mathrm{Var}(X) = E[(X - \mu)^2]. \tag{4.3.1}$$

  Se X possui média infinita ou se a média de X não existe, dizemos que $\mathrm{Var}(X)$ *não existe*.
  O *desvio padrão de X* é a raiz quadrada não negativa de $\mathrm{Var}(X)$ se a variância existir.
  Se a esperança na Eq. (4.3.1) é infinita, dizemos que $\mathrm{Var}(X)$ e o desvio padrão de X são infinitos.

4.3 Variância
Definições da variância e do desvio padrão

- Quando apenas uma variável aleatória está em discussão, é comum denotar seu desvio padrão por σ, e a variância é denotada por σ^2.
- Quando mais de uma variável estão em discussão, o nome da variável aleatória é incluído como um subscrito do símbolo σ;
- P.ex. σ_X denota o desvio padrão de *X*, enquanto σ_Y^2 denota a variância de *Y*.

- Assim como ocorre com a esperança, a variância e o desvio padrão dependem apenas da distribuição de uma variável, independentemente da interpretação ou contexto daquela variável.

4.3 Variância
Definições da variância e do desvio padrão

- Exemplo 4.3.2:

Stock Price Changes. Return to the two random variables A and B in Example 4.3.1. Using Theorem 4.1.1, we can compute

$$\mathrm{Var}(A) = \int_{25}^{35} (a-30)^2 \frac{1}{10} da = \frac{1}{10} \int_{-5}^{5} x^2 dx = \frac{1}{10} \frac{x^3}{3} \bigg|_{x=-5}^{5} = \frac{25}{3},$$

$$\mathrm{Var}(B) = \int_{15}^{45} (b-30)^2 \frac{1}{30} db = \frac{1}{30} \int_{-15}^{15} y^2 dy = \frac{1}{30} \frac{y^3}{3} \bigg|_{y=-15}^{15} = 75.$$

So, $\mathrm{Var}(B)$ is nine times as large as $\mathrm{Var}(A)$. The standard deviations of A and B are $\sigma_A = 2.87$ and $\sigma_B = 8.66$. ◀

☛ 4.3 Variância
Definições da variância e do desvio padrão

- Exemplo 4.3.3:

Variance and Standard Deviation of a Discrete Distribution. Suppose that a random variable X can take each of the five values $-2, 0, 1, 3$, and 4 with equal probability. We shall determine the variance and standard deviation of X.

In this example,

$$E(X) = \frac{1}{5}(-2 + 0 + 1 + 3 + 4) = 1.2.$$

Let $\mu = E(X) = 1.2$, and define $W = (X - \mu)^2$. Then $\mathrm{Var}(X) = E(W)$. We can easily compute the p.f. f of W:

x	-2	0	1	3	4
w	10.24	1.44	0.04	3.24	7.84
$f(w)$	1/5	1/5	1/5	1/5	1/5

It follows that

$$\mathrm{Var}(X) = E(W) = \frac{1}{5}[10.24 + 1.44 + 0.04 + 3.24 + 7.84] = 4.56.$$

The standard deviation of X is the square root of the variance, namely, 2.135. ◀

4.3 Variância
Definições da variância e do desvio padrão

- Existe um método alternativo de calcular a variância de uma distribuição, que é usualmente mais fácil de usar.

- **Teorema 4.3.1: Método alternativo para calcular a variância**
  Para toda variável aleatória X,
$$\mathrm{Var}(X) = E(X^2) - [E(X)]^2 .$$

Proof Let $E(X) = \mu$. Then

$$\begin{aligned}
\mathrm{Var}(X) &= E[(X-\mu)^2] \\
&= E(X^2 - 2\mu X + \mu^2) \\
&= E(X^2) - 2\mu E(X) + \mu^2 \\
&= E(X^2) - \mu^2.
\end{aligned}$$

■

4.3 Variância
Definições da variância e do desvio padrão

- Exemplo 4.3.4:

Variance of a Discrete Distribution. Once again, consider the random variable X in Example 4.3.3, which takes each of the five values $-2, 0, 1, 3$, and 4 with equal probability. We shall use Theorem 4.3.1 to compute $\mathrm{Var}(X)$. In Example 4.3.3, we computed the mean of X as $\mu = 1.2$. To use Theorem 4.3.1, we need

$$E(X^2) = \frac{1}{5}[(-2)^2 + 0^2 + 1^2 + 3^2 + 4^2] = 6.$$

Because $E(X) = 1.2$, Theorem 4.3.1 says that

$$\mathrm{Var}(X) = 6 - (1.2)^2 = 4.56,$$

which agrees with the calculation in Example 4.3.3. ◄

4.3 Variância
Definições da variância e do desvio padrão

- A variância (assim como o desvio padrão) de uma distribuição fornece uma medida da dispersão da distribuição ao redor de sua média μ.
  - Um valor baixo da variância indica que a distribuição de probabilidade está concentrada proximamente ao redor de μ, enquanto um valor alto indica que a distribuição possui uma grande dispersão ao redor de μ.
- Contudo, a variância de uma distribuição, assim como sua média, pode ser arbitrariamente alta colocando-se mesmo um montante pequeno mas positivo de probabilidade longe o suficiente da origem na reta real.

4.3 Variância
Definições da variância e do desvio padrão

- Exemplo 4.3.5:

Slight Modification of a Bernoulli Distribution. Let X be a discrete random variable with the following p.d.f.:

$$f(x) = \begin{cases} 0.5 & \text{if } x = 0, \\ 0.499 & \text{if } x = 1, \\ 0.001 & \text{if } x = 10{,}000, \\ 0 & \text{otherwise.} \end{cases}$$

There is a sense in which the distribution of X differs very little from the Bernoulli distribution with parameter 0.5. However, the mean and variance of X are quite different from the mean and variance of the Bernoulli distribution with parameter 0.5. Let Y have the Bernoulli distribution with parameter 0.5. In Example 4.1.3, we computed the mean of Y as $E(Y) = 0.5$. Since $Y^2 = Y$, $E(Y^2) = E(Y) = 0.5$, so $\text{Var}(Y) = 0.5 - 0.5^2 = 0.25$. The means of X and X^2 are also straightforward calculations:

$$E(X) = 0.5 \times 0 + 0.499 \times 1 + 0.001 \times 10{,}000 = 10.499$$

$$E(X^2) = 0.5 \times 0^2 + 0.499 \times 1^2 + 0.001 \times 10{,}000^2 = 100{,}000.499.$$

So $\text{Var}(X) = 99{,}890.27$. The mean and variance of X are much larger than the mean and variance of Y. ◀

4.3 Variância
Propriedades da variância

- São apresentadas a seguir algumas propriedades básicas da variância.
- Nesses teoremas, assume-se que as variâncias das variáveis aleatórias existem.
- O primeiro teorema refere-se aos possíveis valores da variância.

- **Teorema 4.3.2:**
  Para cada X, $\mathrm{Var}(X) \geq 0$. Se X é uma variável aleatória limitada, então $\mathrm{Var}(X)$ existe e é finita.

Proof Because $\mathrm{Var}(X)$ is the mean of a nonnegative random variable $(X-\mu)^2$, it must be nonnegative according to Theorem 4.2.2. If X is bounded, then the mean exists, and hence the variance exists. Furthermore, if X is bounded the so too is $(X-\mu)^2$, so the variance must be finite. ■

4.3 Variância
Propriedades da variância

- O teorema abaixo mostra que a variância de uma variável aleatória não pode ser 0 a menos que a distribuição de probabilidade de X esteja concentrada em um único ponto.

- **Teorema 4.3.3:**
  $\mathrm{Var}(X) = 0$ se e somente se existe uma constante c tal que $\Pr(X = c) = 1$.

Proof Suppose first that there exists a constant c such that $\Pr(X = c) = 1$. Then $E(X) = c$, and $\Pr[(X - c)^2 = 0] = 1$. Therefore,

$$\mathrm{Var}(X) = E[(X - c)^2] = 0.$$

Conversely, suppose that $\mathrm{Var}(X) = 0$. Then $\Pr[(X - \mu)^2 \geq 0] = 1$ but $E[(X - \mu)^2] = 0$. It follows from Theorem 4.2.3 that

$$\Pr[(X - \mu)^2 = 0] = 1.$$

Hence, $\Pr(X = \mu) = 1$. ■

4.3 Variância
Propriedades da variância

- **Teorema 4.3.4:**
  Para constantes a e b, seja $Y = aX + b$. Então
  $$\mathrm{Var}(Y) = a^2\,\mathrm{Var}(X),$$
  e $\sigma_Y = |a|\sigma_X$.

Proof If $E(X) = \mu$, then $E(Y) = a\mu + b$ by Theorem 4.2.1. Therefore,

$$\begin{aligned}\mathrm{Var}(Y) &= E[(aX + b - a\mu - b)^2] = E[(aX - a\mu)^2]\\ &= a^2 E[(X-\mu)^2] = a^2\,\mathrm{Var}(X).\end{aligned}$$

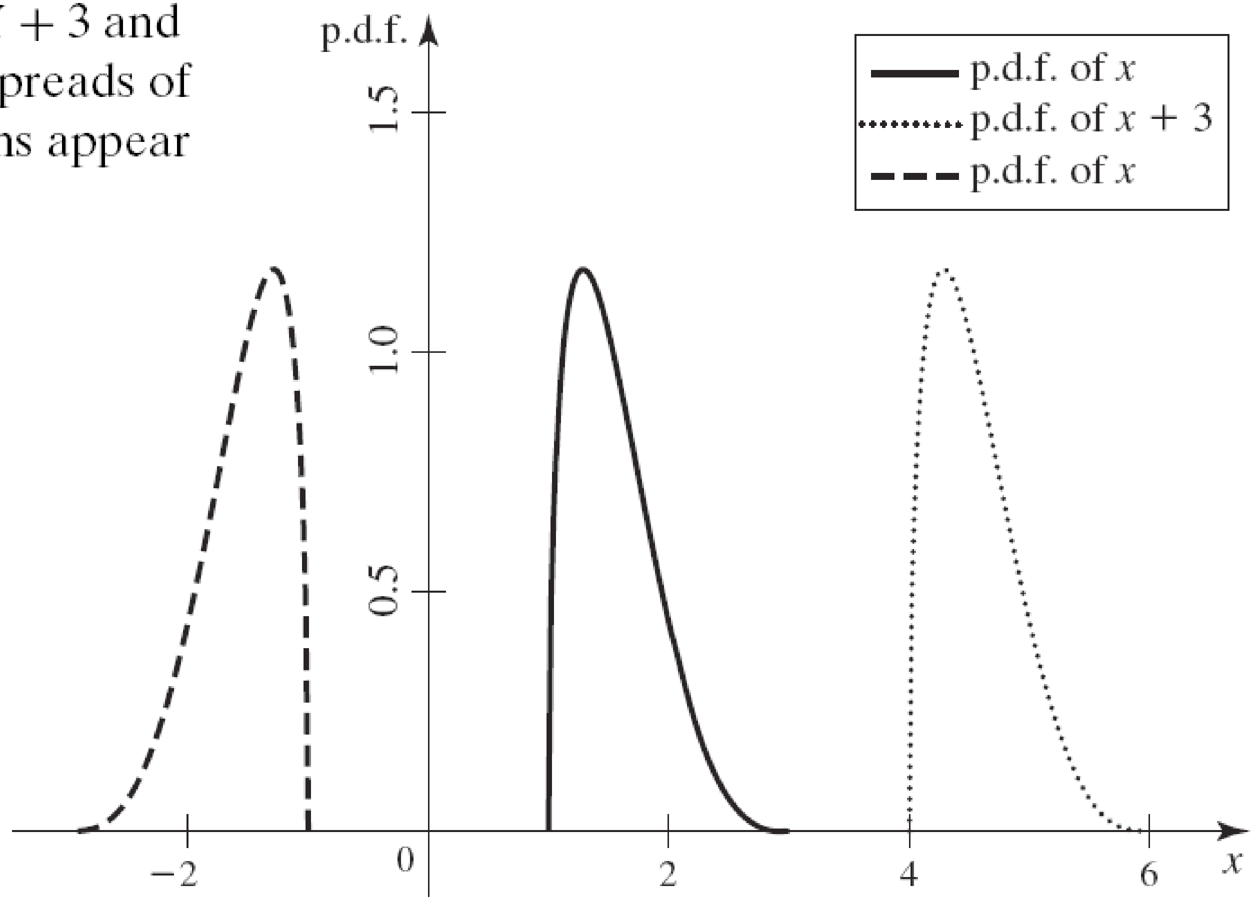
Taking the square root of $\mathrm{Var}(Y)$ yields $|a|\sigma_X$. ■

4.3 Variância
Propriedades da variância

- Segue do Teorema 4.3.4 que $\mathrm{Var}(X+b) = \mathrm{Var}(X)$ para toda constante *b*.
  - Deslocar a distribuição inteira de *X* a uma distância de *b* unidades na reta real deslocará a média de X por b unidades, mas tal deslocamento não alterará a dispersão da distribuição ao redor de sua média.
- Adicionalmente, segue do Teorema 4.3.4 que $\mathrm{Var}(-X) = \mathrm{Var}(X)$
  - Refletir a distribuição completa de X com respeito à origem na reta real resulta em uma nova distribuição, em que a média mudará de μ para $-\mu$, mas a dispersão da distribuição ao redor de sua média não é afetada.
- Figura 4.6 mostra a f.d.p. de uma variável aleatória *X* juntamente com as f.d.p's de *X*+3 e de –*X*, para ilustrar como o deslocamento ou inversão da distribuição não afeta sua dispersão.

4.3 Variância
Propriedades da variância

Figure 4.6 The p.d.f. of a random variable X together with the p.d.f.'s of $X + 3$ and $-X$. Note that the spreads of all three distributions appear the same.



4.3 Variância
Propriedades da variância

- Exemplo 4.3.6:

Calculating the Variance and Standard Deviation of a Linear Function. Consider the same random variable X as in Example 4.3.3, which takes each of the five values -2, 0, 1, 3, and 4 with equal probability. We shall determine the variance and standard deviation of $Y = 4X - 7$.

In Example 4.3.3, we computed the mean of X as $\mu = 1.2$ and the variance as 4.56. By Theorem 4.3.4,

$$\mathrm{Var}(Y) = 16\,\mathrm{Var}(X) = 72.96.$$

Also, the standard deviation σ of Y is

$$\sigma_Y = 4\sigma_X = 4(4.56)^{1/2} = 8.54. \quad ◄$$

- O Teorema a seguir fornece um método alternativo para calcular a variância de uma soma de variáveis aleatórias <u>independentes</u>.

4.3 Variância
Propriedades da variância

- **Teorema 4.3.5:**
  Se X_1, X_2, ..., X_n são variáveis aleatórias independentes com médias finitas, então

$$\mathrm{Var}(X_1 + \cdots + X_n) = \mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_n).$$

Proof Suppose first that $n = 2$. If $E(X_1) = \mu_1$ and $E(X_2) = \mu_2$, then

$$E(X_1 + X_2) = \mu_1 + \mu_2.$$

Therefore,

$$\begin{aligned}\mathrm{Var}(X_1 + X_2) &= E[(X_1 + X_2 - \mu_1 - \mu_2)^2] \\ &= E[(X_1 - \mu_1)^2 + (X_2 - \mu_2)^2 + 2(X_1 - \mu_1)(X_2 - \mu_2)] \\ &= \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + 2E[(X_1 - \mu_1)(X_2 - \mu_2)].\end{aligned}$$

4.3 Variância
Propriedades da variância

- **Teorema 4.3.5 (cont):**

Since X_1 and X_2 are independent,

$$\begin{aligned} E[(X_1-\mu_1)(X_2-\mu_2)] &= E(X_1-\mu_1)E(X_2-\mu_2) \\ &= (\mu_1-\mu_1)(\mu_2-\mu_2) \\ &= 0. \end{aligned}$$

It follows, therefore, that

$$\mathrm{Var}(X_1+X_2) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2).$$

The theorem can now be established for each positive integer n by an induction argument. ■

- Importante: O Teorema 4.3.5 só vale para variáveis aleatórias <u>independentes</u>.

4.3 Variância
Propriedades da variância

- Combinando-se os Teoremas 4.3.4 e 4.3.5, obtém-se o seguinte corolário.

- **Corolário 4.3.1:**
  Se $X_1, X_2, \ldots, X_n$ são variáveis aleatórias independentes com médias finitas, e se $a_1, a_2, \ldots, a_n$ e b são constantes quaisquer, então

$$\mathrm{Var}(a_1 X_1 + \cdots + a_n X_n + b) = a_1^2 \,\mathrm{Var}(X_1) + \cdots + a_n^2 \,\mathrm{Var}(X_n).$$

4.3 Variância
Propriedades da variância

- Exemplo 4.3.7:

Investment Portfolio. An investor with \$100,000 to invest wishes to construct a portfolio consisting of shares of one or both of two available stocks and possibly some fixed-rate investments. Suppose that the two stocks have random rates of return R_1 and R_2 per share for a period of one year. Suppose that R_1 has a distribution with mean 6 and variance 55, while R_2 has mean 4 and variance 28. Suppose that the first stock costs \$60 per share and the second costs \$48 per share. Suppose that money can also be invested at a fixed rate of 3.6 percent per year. The portfolio will consist of s_1 shares of the first stock, s_2 shares of the second stock, and all remaining money ($\$s_3$) invested at the fixed rate. The return on this portfolio will be

$$s_1R_1 + s_2R_2 + 0.036s_3,$$

where the coefficients are constrained by

$$60s_1 + 48s_2 + s_3 = 100{,}000, \tag{4.3.2}$$

4.3 Variância
Propriedades da variância

- Exemplo 4.3.7 (cont):

as well as $s_1, s_2, s_3 \geq 0$. For now, we shall assume that R_1 and R_2 are independent. The mean and the variance of the return on the portfolio will be

$$E(s_1R_1 + s_2R_2 + 0.036s_3) = 6s_1 + 4s_2 + 0.036s_3,$$
$$\mathrm{Var}(s_1R_1 + s_2R_2 + 0.036s_3) = 55s_1^2 + 28s_2^2.$$

One method for comparing a class of portfolios is to say that portfolio A is at least as good as portfolio B if the mean return for A is at least as large as the mean return for B and if the variance for A is no larger than the variance of B. (See Markowitz, 1987, for a classic treatment of such methods.) The reason for preferring smaller variance is that large variance is associated with large deviations from the mean, and for portfolios with a common mean, some of the large deviations are going to have to be below the mean, leading to the risk of large losses. Figure 4.7 is a plot of the pairs (mean, variance) for all of the possible portfolios in this example. That

4.3 Variância
Propriedades da variância

- Exemplo 4.3.7 (cont):

is, for each (s_1, s_2, s_3) that satisfy (4.3.2), there is a point in the outlined region of Fig. 4.7. The points to the right and toward the bottom are those that have the largest mean return for a fixed variance, and the ones that have the smallest variance for a fixed mean return. These portfolios are called *efficient*. For example, suppose that the investor would like a mean return of 7000. The vertical line segment above 7000 on the horizontal axis in Fig. 4.7 indicates the possible variances of all portfolios with mean return of 7000. Among these, the portfolio with the smallest variance is efficient and is indicated in Fig. 4.7. This portfolio has $s_1 = 524.7$, $s_2 = 609.7$, $s_3 = 39{,}250$, and variance 2.55×10^7. So, every portfolio with mean return greater than 7000 must have variance larger than 2.55×10^7, and every portfolio with variance less than 2.55×10^7 must have mean return smaller than 7000. ◀

4.3 Variância
Propriedades da variância

- Exemplo 4.3.7 (cont):

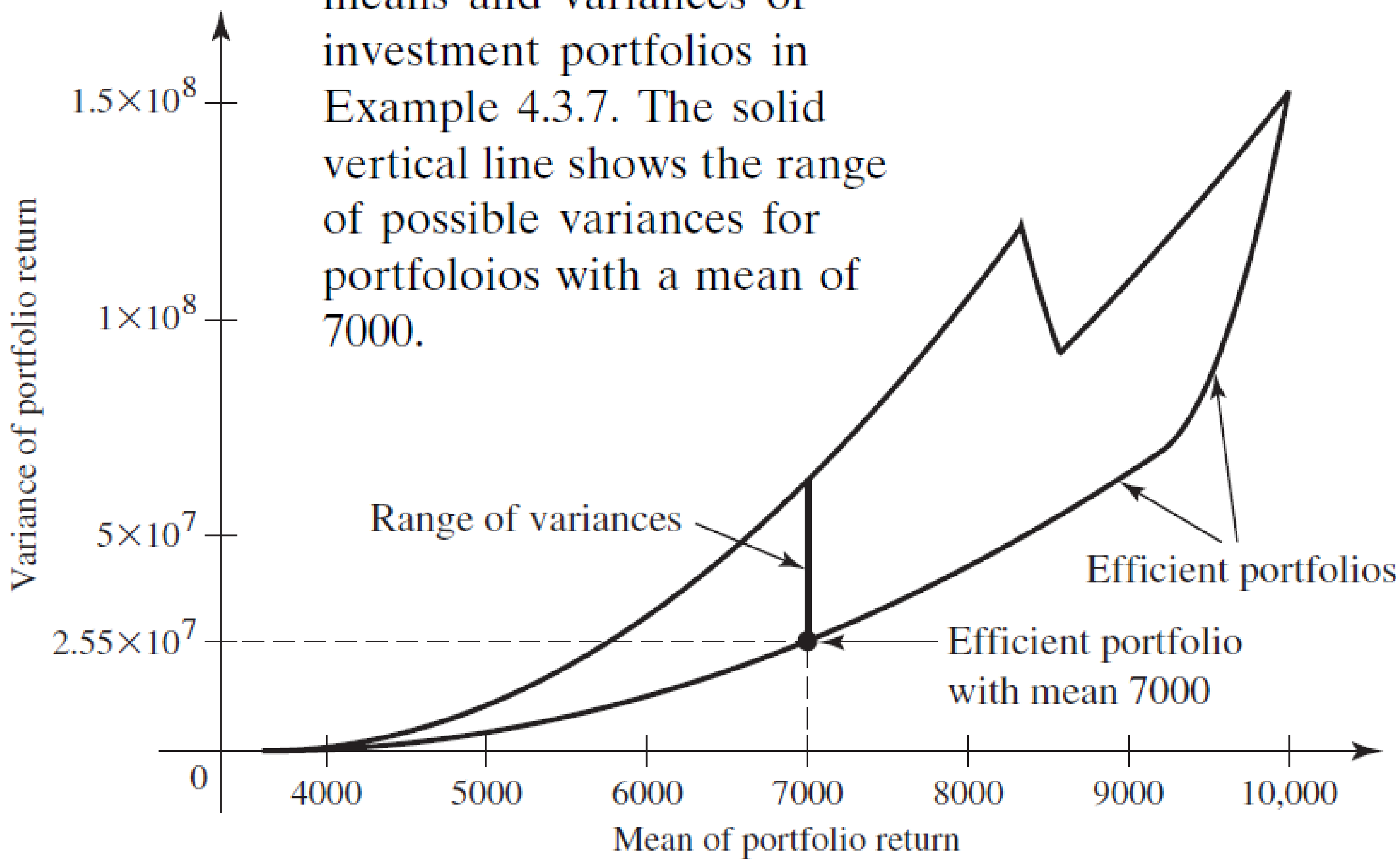


Figure 4.7 The set of all means and variances of investment portfolios in Example 4.3.7. The solid vertical line shows the range of possible variances for portfoloios with a mean of 7000.

4.3 Variância
Variância de uma distribuição binomial

- Consideraremos novamente o método de gerar uma distribuição apresentado na Seção 4.2 (Exemplo 4.2.5).
- Suponha que uma caixa contenha bolas vermelhas e azuis, e que a proporção de bolas vermelhas seja p ($0 \leq p \leq 1$).
- Suponha também que uma amostra aleatória de n bolas seja selecionada da caixa com reposição. Para $i = 1, \dots, n$, seja $X_i = 1$ se a i-ésima bola selecionada é vermelha, e seja $X_i = 0$ caso contrário.
- Se X denota o número de bolas vermelhas na amostra, então $X = X_1 + X_2 + \cdots + X_n$ e X terá distribuição binomial com parâmetros n e p.
- Como $X_1, X_2, \dots, X_n$ são independentes, segue do Teorema 4.3.5 que

$$\mathrm{Var}(X) = \sum_{i=1}^{n} \mathrm{Var}(X_i)$$

4.3 Variância
Variância de uma distribuição binomial

- De acordo com o Exemplo 4.1.3, $E(X_i) = p$ para $i = 1, \ldots, n$. Como $X_i{}^2 = X_i$ para todo *i*, $E(X_i{}^2) = E(X_i) = p$. Logo, pelo Teorema 4.3.1,

$$\begin{aligned} \mathrm{Var}(X_i) &= E(X_i^2) - [E(X_i)]^2 \\ &= p - p^2 = p(1-p). \end{aligned}$$

- Segue então que

$$\mathrm{Var}(X) = np(1-p). \qquad (4.3.3)$$

- Fig. 4.8 compara duas distribuições binomiais com mesma média (2.5) mas variâncias diferentes (1.25 e 1.875).
  - Note que a f.p. da distribuição com variância mais alta (n=10, p=0.25) é maior nos valores extremos e menor nos valores centrais, em comparação com a f.p. da distribuição com menor variância (n=5, p=0.5)

4.3 Variância
Variância de uma distribuição binomial

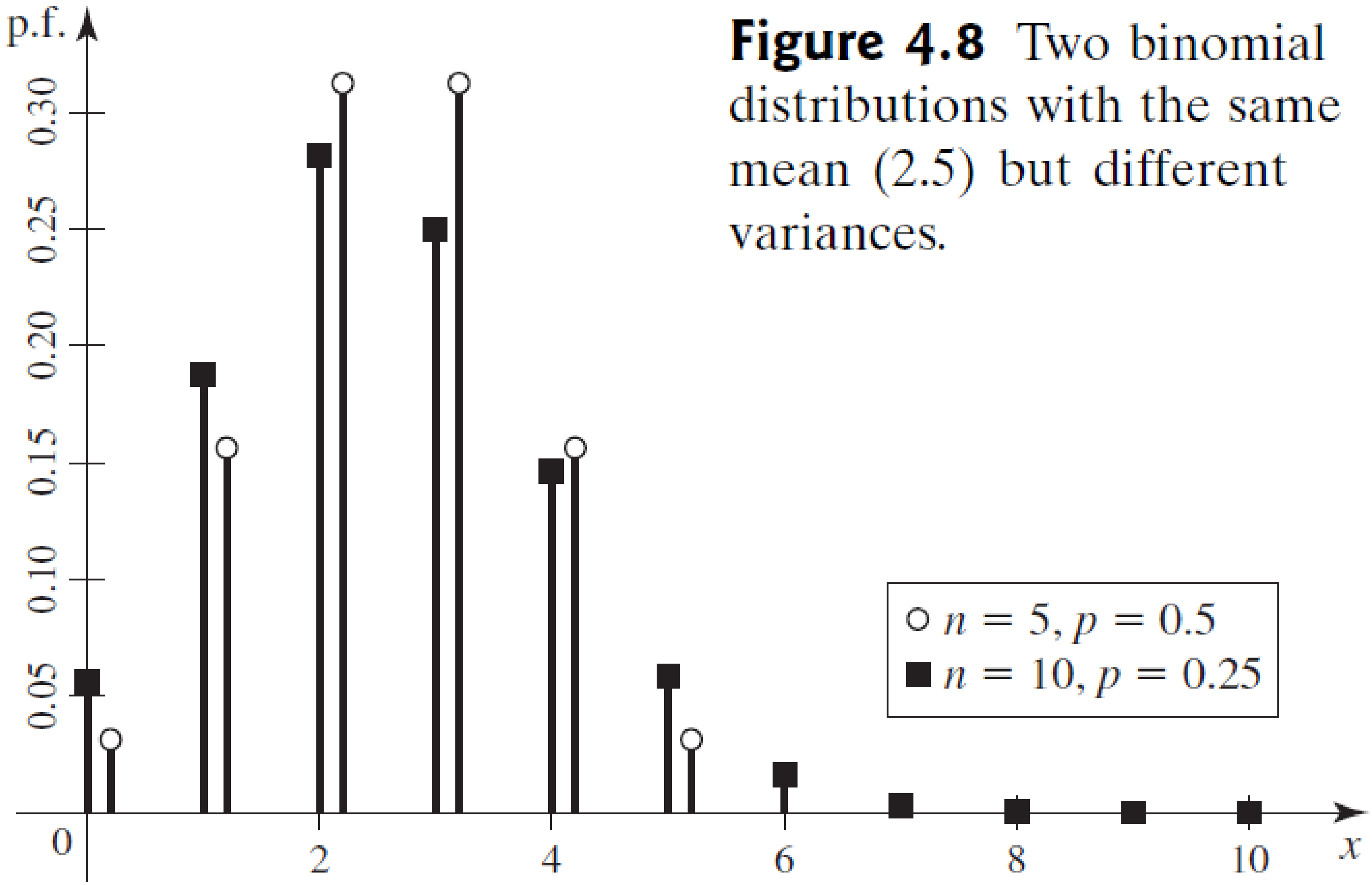


Figure 4.8 Two binomial distributions with the same mean (2.5) but different variances.

4.5 A média e a mediana
Minimização do erro quadrático médio

- Suponha que X seja uma variável aleatória com média μ e variância σ^2.
- Suponha também que o valor de X deverá ser observado em algum experimento, mas tal valor precisa ser previsto antes da observação ser feita.
- Uma base para realizar a previsão é selecionar algum número d que minimize o valor esperado do quadrado do erro $X - d$.

- **Definição 4.5.2: Erro quadrático médio / E.Q.M. / M.S.E.** O número $E[(X-d)^2]$ é denominado *erro quadrático médio* da predição *d*. Abreviação: E.Q.M. (port.); M.S.E. (inglês)

- O resultado a seguir mostra que o número d que minimiza o E.Q.M. é $d = \mu$.

4.5 A média e a mediana
Minimização do erro quadrático médio

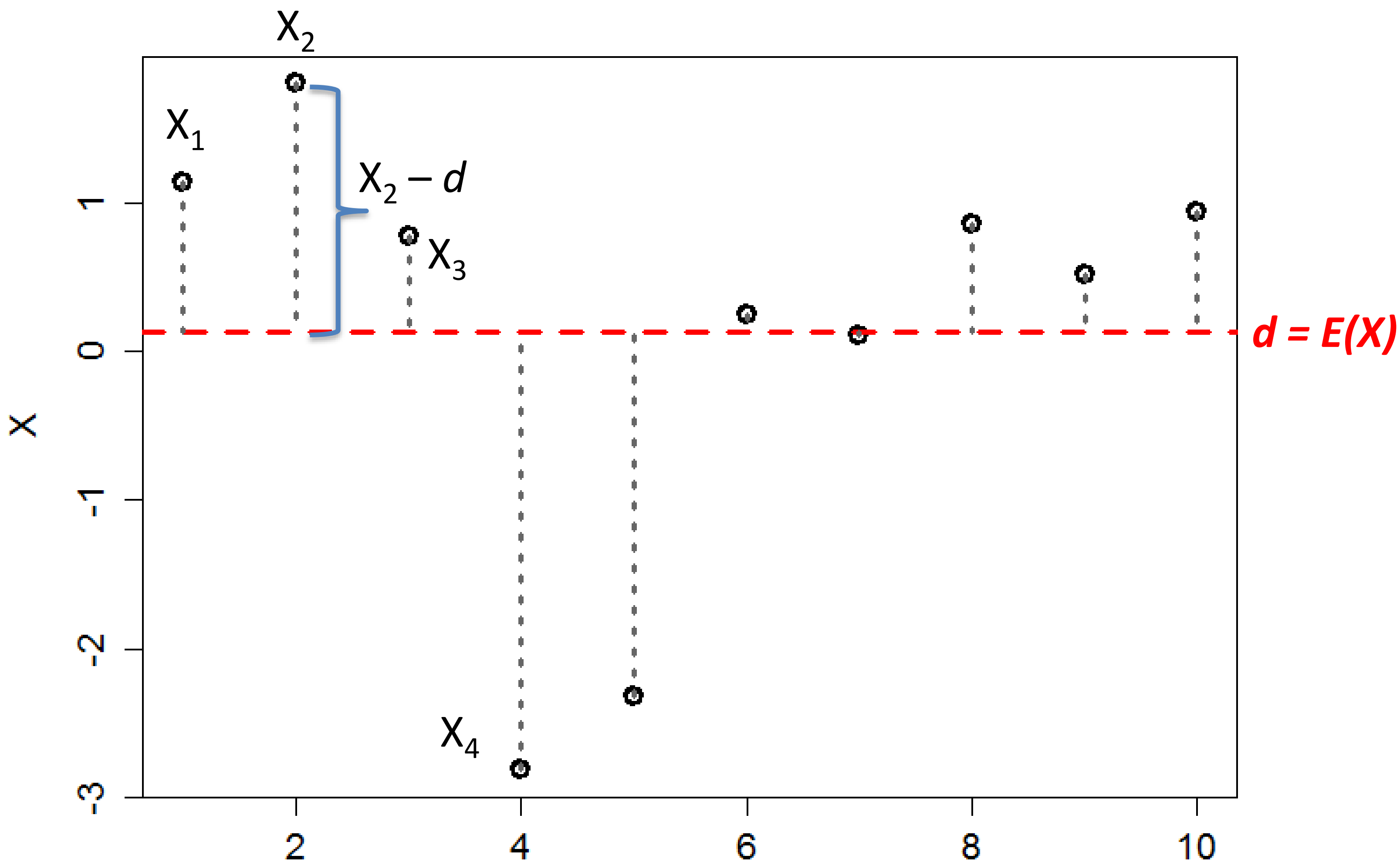
- Teorema 4.5.2:
  Seja X uma variável aleatória com variância finita σ^2, e seja $\mu = E(X)$. Para todo número d,
  $$E[(X-\mu)^2] \leq E[(X-d)^2]. \tag{4.5.1}$$
  Além disso, ocorrerá igualdade na relação (4.5.1) se e somente se $d = \mu$.

Proof For every value of d,

$$\begin{aligned} E[(X-d)^2] &= E(X^2 - 2\,dX + d^2) \\ &= E(X^2) - 2\,d\mu + d^2. \end{aligned} \tag{4.5.2}$$

The final expression in Eq. (4.5.2) is simply a quadratic function of d. By elementary differentiation it will be found that the minimum value of this function is attained when $d = \mu$. Hence, in order to minimize the M.S.E., the predicted value of X should be its mean μ. Furthermore, when this prediction is used, the M.S.E. is simply $E[(X-\mu)^2] = \sigma^2$. ■

4.5 A média e a mediana
Minimização do erro quadrático médio



4.7 Esperança condicional

- Dado que esperanças (incluindo variâncias) são propriedades das distribuições, existem versões condicionais de todos os resumos, bem como versões condicionais de todos os teoremas apresentados.

- Em particular, suponha que queremos prever uma variável aleatória Y usando uma função $d(X)$ de outra variável aleatória X de forma a minimizar $E([Y - d(X)]^2)$. Então $d(X)$ deveria ser a média condicional de Y dado X.

4.7 Esperança condicional
Definições e propriedades básicas

- Exemplo 4.7.1:

Household Survey. A collection of households were surveyed, and each household reported the number of members and the number of automobiles owned. The reported numbers are in Table 4.1.

Suppose that we were to sample a household at random from those households in the survey and learn the number of members. What would then be the expected number of automobiles that they own? ◄

Table 4.1 Reported numbers of household members and automobiles in Example 4.7.1

Number of automobiles	Number of members							
	1	2	3	4	5	6	7	8
0	10	7	3	2	2	1	0	0
1	12	21	25	30	25	15	5	1
2	1	5	10	15	20	11	5	3
3	0	2	3	5	5	3	2	1

4.7 Esperança condicional
Definições e propriedades básicas

- A questão levantada no Exemplo 4.7.1 está intimamente relacionada com a distribuição condicional de uma variável aleatória dada a outra, como definido na Seção 3.6.

- **Definição 4.7.1: Esperança condicional**
  Sejam *X* e *Y* variáveis aleatórias tais que a média de *Y* existe e é finita. A *esperança condicional* (ou *média condicional*) *de Y dado* $X = x$ é denotada por $E(Y|x)$ e é definida como a esperança da distribuição condicional de *Y* dado $X = x$.

4.7 Esperança condicional
Definições e propriedades básicas

- Por exemplo, se Y possui uma distribuição condicional contínua dado $X = x$ com f.d.p. condicional $g_2(y|x)$, então

$$E(Y|x) = \int_{-\infty}^{\infty} y g_2(y|x)\, dy. \tag{4.7.1}$$

Analogamente, se Y possui uma distribuição condicional discreta dado $X = x$ com f.p. condicional $g_2(y|x)$, então

$$E(Y|x) = \sum_{\text{All } y} y g_2(y|x). \tag{4.7.2}$$

- As expressões nas Eqs. (4.7.1) e (4.7.2) são funções de *x*, e podem ser computadas antes de *X* ser observada. Essa ideia leva ao seguinte conceito.

4.7 Esperança condicional
Definições e propriedades básicas

- **Definição 4.7.2: Médias condicionais como variáveis aleatórias**
  Seja $h(x)$ a função de *x* que é denotada por $E(Y|x)$ nas Eqs. (4.7.1) ou (4.7.2). Defina o símbolo $E(Y|X)$ como $h(X)$ e chame-o de *média condicional* de Y dado X.

- Em outras palavras, $E(Y|X)$ é uma variável aleatória (uma função de *X*) cujo valor quando $X = x$ é $E(Y|x)$.

- Pode-se definir $E(X|Y)$ de forma análoga.

4.7 Esperança condicional
Definições e propriedades básicas

- Exemplo 4.7.2:

Household Survey. Consider the household survey in Example 4.7.1. Let X be the number of members in a randomly selected household from the survey, and let Y be the number of cars owned by that household. The 250 surveyed households are all equally likely to be selected, so $\Pr(X = x, Y = y)$ is the number of households with x members and y cars, divided by 250. Those probabilities are reported in Table 4.2. Suppose that the sampled household has $X = 4$ members. The conditional p.f. of Y given $X = 4$ is $g_2(y|4) = f(4, y)/f_1(4)$, which is the $x = 4$ column of Table 4.2 divided by $f_1(4) = 0.208$, namely,

$$g_2(0|4) = 0.0385, \quad g_2(1|4) = 0.5769, \quad g_2(2|4) = 0.2885, \quad g_2(3|4) = 0.0962.$$

The conditional mean of Y given $X = 4$ is then

$$E(Y|4) = 0 \times 0.0385 + 1 \times 0.5769 + 2 \times 0.2885 + 3 \times 0.0962 = 1.442.$$

Similarly, we can compute $E(Y|x)$ for all eight values of x. They are

x	1	2	3	4	5	6	7	8
$E(Y\|x)$	0.609	1.057	1.317	1.442	1.538	1.533	1.75	2

4.7 Esperança condicional
Definições e propriedades básicas

- Exemplo 4.7.2 (cont):

The random variable that takes the value 0.609 when the sampled household has one member, takes the value 1.057 when the sampled household has two members, and so on, is the random variable $E(Y|X)$. ◀

Table 4.2 Joint p.f. $f(x, y)$ of X and Y in Example 4.7.2 together with marginal p.f.'s $f_1(x)$ and $f_2(y)$

	x								
y	1	2	3	4	5	6	7	8	$f_2(y)$
0	0.040	0.028	0.012	0.008	0.008	0.004	0	0	0.100
1	0.048	0.084	0.100	0.120	0.100	0.060	0.020	0.004	0.536
2	0.004	0.020	0.040	0.060	0.080	0.044	0.020	0.012	0.280
3	0	0.008	0.012	0.020	0.020	0.012	0.008	0.004	0.084
$f_1(x)$	0.092	0.140	0.164	0.208	0.208	0.120	0.048	0.020	

4.7 Esperança condicional
Definições e propriedades básicas

- Exemplo 4.7.3:

A Clinical Trial. Consider a clinical trial in which a number of patients will be treated and each patient will have one of two possible outcomes: success or failure. Let P be the proportion of successes in a very large collection of patients, and let $X_i = 1$ if the ith patient is a success and $X_i = 0$ if not. Assume that the random variables $X_1, X_2, \ldots$ are conditionally independent given $P = p$ with $\Pr(X_i = 1|P = p) = p$. Let $X = X_1 + \cdots + X_n$, which is the number of patients out of the first n who are successes. We now compute the conditional mean of X given P. The patients are independent and identically distributed conditional on $P = p$. Hence, the conditional distribution of X given $P = p$ is the binomial distribution with parameters n and p. As we saw in Sec. 4.2, the mean of this binomial distribution is np, so $E(X|p) = np$ and $E(X|P) = nP$. Later, we will show how to compute the conditional mean of P given X. This can be used to predict P after observing X. ◀

4.7 Esperança condicional
Definições e propriedades básicas

- **Observação: A média condicional de Y dado X é uma variável aleatória.**
  - Como $E(Y|X)$ é uma função da variável aleatória X, ela é por si mesma uma variável aleatória com sua própria distribuição de probabilidade, que pode ser derivada da distribuição de X.
  - Por outro lado, $h(x) = E(Y|x)$ é uma função de *x* que pode ser manipulada como qualquer outra.
  - A conexão entre ambas é que quando se substitui x em $h(x)$ por X, o resultado é $h(X) = E(Y|X)$.

- Mostraremos que a média da variável aleatória $E(Y|X)$ deve ser $E(Y)$. Cálculo similar mostra que a média de $E(X|Y)$ deve ser $E(X)$.

4.7 Esperança condicional
Definições e propriedades básicas

- **Teorema 4.7.1: Lei da probabilidade total para esperanças**
  Sejam X e Y variáveis aleatórias tais que Y possui média finita. Então

$$E[E(Y|X)] = E(Y). \tag{4.7.3}$$

Proof We shall assume, for convenience, that X and Y have a continuous joint distribution. Then

$$\begin{aligned} E[E(Y|X)] &= \int_{-\infty}^{\infty} E(Y|x) f_1(x)\, dx \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y g_2(y|x) f_1(x)\, dy\, dx. \end{aligned}$$

Since $g_2(y|x) = f(x, y)/f_1(x)$, it follows that

$$E[E(Y|X)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y f(x, y)\, dy\, dx = E(Y).$$

The proof for a discrete distribution or a more general type of distribution is similar. ∎

4.7 Esperança condicional
Definições e propriedades básicas

- Exemplo 4.7.4:

Household Survey. At the end of Example 4.7.2, we described the random variable $E(Y|X)$. Its distribution can be constructed from that description. It has a discrete distribution that takes the eight values of $E(Y|x)$ listed near the end of that example with corresponding probabilities $f_1(x)$ for $x = 1, \ldots, 8$. To be specific, let $Z = E(Y|X)$, then $\Pr[Z = E(Y|x)] = f_1(x)$ for $x = 1, \ldots, 8$. The specific values are

z	0.609	1.057	1.317	1.442	1.538	1.533	1.75	2
$\Pr(Z = z)$	0.092	0.140	0.164	0.208	0.208	0.120	0.048	0.020

We can compute $E(Z) = 0.609 \times 0.092 + \cdots + 2 \times 0.020 = 1.348$. The reader can verify that $E(Y) = 1.348$ by using the values of $f_2(y)$ in Table 4.2. ◀

4.7 Esperança condicional
Definições e propriedades básicas

- Exemplo 4.7.5:

A Clinical Trial. In Example 4.7.3, we let X be the number of patients out of the first n who are successes. The conditional mean of X given $P = p$ was computed as $E(X|p) = np$, where P is the proportion of successes in a large population of patients. If the distribution of P is uniform on the interval [0, 1], then the marginal expected value of X is $E[E(X|P)] = E(nP) = n/2$. We will see how to calculate $E(P|X)$ in Example 4.7.8. ◀

- Exemplo 4.7.6:

Choosing Points from Uniform Distributions. Suppose that a point X is chosen in accordance with the uniform distribution on the interval [0, 1]. Also, suppose that after the value $X = x$ has been observed $(0 < x < 1)$, a point Y is chosen in accordance with a uniform distribution on the interval $[x, 1]$. We shall determine the value of $E(Y)$.

For each given value of x $(0 < x < 1)$, $E(Y|x)$ will be equal to the midpoint $(1/2)(x + 1)$ of the interval $[x, 1]$. Therefore, $E(Y|X) = (1/2)(X + 1)$ and

$$E(Y) = E[E(Y|X)] = \frac{1}{2}[E(X) + 1] = \frac{1}{2}\left(\frac{1}{2} + 1\right) = \frac{3}{4}.$$ ◀

4.7 Esperança condicional
Definições e propriedades básicas

- Quando falamos de distribuições condicionais dado X=x, usualmente tratamos X como uma constante x.
- Este fato pode simplificar o cálculo de certas médias condicionais.

- **Teorema 4.7.2:**
  Sejam *X* e *Y* variáveis aleatórias, e seja $Z = r(X, Y)$ para alguma função *r*. A distribuição condicional de Z dado $X = x$ é a mesma da distribuição condicional $r(x, Y)$ dado $X = x$.

4.7 Esperança condicional
Definições e propriedades básicas

- Uma consequência do Teorema 4.7.2 quando X e Y possuem uma distribuição conjunta contínua é que

$$E(Z|x) = E(r(x, Y)|x) = \int_{-\infty}^{\infty} r(x, y) g_2(y|x)\, dy.$$

- O Teorema 4.7.1 também implica que para duas variáveis aleatórias X e Y,

$$E\{E[r(X, Y)|X]\} = E[r(X, Y)], \tag{4.7.4}$$

  denotando $Z = r(X, Y)$ e observando-se que $E\{E(Z|X)\} = E(Z)$.

- De maneira similar, pode-se definir a esperança condicional de $r(X, Y)$ dado Y e a esperança condicional de uma função $r(X_1, X_2, \ldots, X_n)$ de várias variáveis aleatórias dada uma ou mais das variáveis $X_1, X_2, \ldots, X_n$.

4.7 Esperança condicional
Definições e propriedades básicas

- Exemplo 4.7.7:

Linear Conditional Expectation. Suppose that $E(Y|X) = aX + b$ for some constants a and b. We shall determine the value of $E(XY)$ in terms of $E(X)$ and $E(X^2)$.

By Eq. (4.7.4), $E(XY) = E[E(XY|X)]$. Furthermore, since X is considered to be given and fixed in the conditional expectation,

$$E(XY|X) = XE(Y|X) = X(aX + b) = aX^2 + bX.$$

Therefore,

$$E(XY) = E(aX^2 + bX) = aE(X^2) + bE(X). \quad \blacktriangleleft$$

- Além da média condicional, pode-se também definir a variância condicional.

4.7 Esperança condicional
Definições e propriedades básicas

- **Definição 4.7.3: Variância condicional**
  Para todo valor dado *x*, $Var(Y|x)$ denota a variância da distribuição condicional de *Y* dado que $X = x$. Ou seja,

$$\text{Var}(Y|x) = E\{[Y - E(Y|x)]^2|x\}. \qquad (4.7.5)$$

  Denominamos $Var(Y|x)$ a variância condicional de *Y* dado $X = x$.