Notice that the integral is the expected value of an exponential random variable with parameter $\lambda(1+w)$, so it equals $1/\lambda(1+w)$ (recall Example 3.5.6). Therefore,

$$f_W(w) = \frac{\lambda^2}{\lambda(1+w)} \frac{1}{\lambda(1+w)} = \frac{1}{(1+w)^2}, \quad w \geq 0$$

■

Theorem 3.8.5 *Let X and Y be independent continuous random variables with pdfs $f_X(x)$ and $f_Y(y)$, respectively. Let $W = XY$. Then*

$$f_W(w) = \int_{-\infty}^{\infty} \frac{1}{|x|} f_X(x) f_Y(w/x)\,dx = \int_{-\infty}^{\infty} \frac{1}{|x|} f_X(w/x) f_Y(x)\,dx$$

Proof A line-by-line, straightforward modification of the proof of Theorem 3.8.4 will provide a proof of Theorem 3.8.5. The details are left to the reader. □

Example 3.8.5 Suppose that X and Y are independent random variables with pdfs $f_X(x) = 1$, $0 \leq x \leq 1$, and $f_Y(y) = 2y$, $0 \leq y \leq 1$, respectively. Find $f_W(w)$, where $W = XY$.

According to Theorem 3.8.5,

$$f_W(w) = \int_{-\infty}^{\infty} \frac{1}{|x|} f_X(x) f_Y(w/x)\,dx$$

The region of integration, though, needs to be restricted to values of x for which the integrand is positive. But $f_Y(w/x)$ is positive only if $0 \leq w/x \leq 1$, which implies that $x \geq w$. Moreover, for $f_X(x)$ to be positive requires that $0 \leq x \leq 1$. Any x, then, from w to 1 will yield a positive integrand. Therefore,

$$f_W(w) = \int_w^1 \frac{1}{x}(1)(2w/x)\,dx = 2w \int_w^1 \frac{1}{x^2}\,dx = 2 - 2w, \quad 0 \leq w \leq 1$$

Comment Theorems 3.8.3, 3.8.4, and 3.8.5 can be adapted to situations where X and Y are not independent by replacing the product of the marginal pdfs with the joint pdf. ■

Questions

3.8.1. Let X and Y be two independent random variables. Given the marginal pdfs shown below, find the pdf of $X+Y$. In each case, check to see if $X+Y$ belongs to the same family of pdfs as do X and Y.

(a) $p_X(k) = e^{-\lambda}\dfrac{\lambda^k}{k!}$ and $p_Y(k) = e^{-\mu}\dfrac{\mu^k}{k!}$, $k = 0, 1, 2, \ldots$

(b) $p_X(k) = p_Y(k) = (1-p)^{k-1}p$, $k = 1, 2, \ldots$

3.8.2. Suppose $f_X(x) = xe^{-x}$, $x \geq 0$, and $f_Y(y) = e^{-y}$, $y \geq 0$, where X and Y are independent. Find the pdf of $X+Y$.

3.8.3. Let X and Y be two independent random variables, whose marginal pdfs are given below. Find the pdf of $X+Y$. (*Hint:* Consider two cases, $0 \leq w < 1$ and $1 \leq w \leq 2$.)

$$f_X(x) = 1,\ 0 \leq x \leq 1, \text{ and } f_Y(y) = 1,\ 0 \leq y \leq 1$$

3.8.4. If a random variable V is independent of two independent random variables X and Y, prove that V is independent of $X+Y$.

3.8.5. Let Y be a continuous nonnegative random variable. Show that $W = Y^2$ has pdf $f_W(w) = \frac{1}{2\sqrt{w}} f_Y(\sqrt{w})$. [*Hint:* First find $F_W(w)$.]

3.8.6. Let Y be a uniform random variable over the interval [0, 1]. Find the pdf of $W = Y^2$.

3.8.7. Let Y be a random variable with $f_Y(y) = 6y(1-y)$, $0 \leq y \leq 1$. Find the pdf of $W = Y^2$.

3.8.8. Suppose the velocity of a gas molecule of mass m is a random variable with pdf $f_Y(y) = ay^2e^{-by^2}$, $y \geq 0$, where a and b are positive constants depending on the gas. Find the pdf of the kinetic energy, $W = (m/2)Y^2$, of such a molecule.

3.8.9. Given that X and Y are independent random variables, find the pdf of XY for the following two sets of marginal pdfs:

(a) $f_X(x) = 1, 0 \le x \le 1$, and $f_Y(y) = 1, 0 \le y \le 1$
(b) $f_X(x) = 2x, 0 \le x \le 1$, and $f_Y(y) = 2y, 0 \le y \le 1$

3.8.10. Let X and Y be two independent random variables. Given the marginal pdfs indicated below, find the cdf of Y/X. (*Hint:* Consider two cases, $0 \le w \le 1$ and $1 < w$.)

(a) $f_X(x) = 1, 0 \le x \le 1$, and $f_Y(y) = 1, 0 \le y \le 1$
(b) $f_X(x) = 2x, 0 \le x \le 1$, and $f_Y(y) = 2y, 0 \le y \le 1$

3.8.11. Suppose that X and Y are two independent random variables, where $f_X(x) = xe^{-x}, x \ge 0$, and $f_Y(y) = e^{-y}$, $y \ge 0$. Find the pdf of Y/X.

3.9 Further Properties of the Mean and Variance

Sections 3.5 and 3.6 introduced the basic definitions related to the expected value and variance of *single* random variables. We learned how to calculate $E(W)$, $E[g(W)]$, $E(aW + b)$, $\text{Var}(W)$, and $\text{Var}(aW + b)$, where a and b are any constants and W could be either a discrete or a continuous random variable. The purpose of this section is to examine certain multivariable extensions of those results, based on the joint pdf material covered in Section 3.7.

We begin with a theorem that generalizes $E[g(W)]$. While it is stated here for the case of *two* random variables, it extends in a very straightforward way to include functions of n random variables.

Theorem 3.9.1

1. *Suppose X and Y are discrete random variables with joint pdf $p_{X,Y}(x, y)$, and let $g(X, Y)$ be a function of X and Y. Then the expected value of the random variable $g(X, Y)$ is given by*

$$E[g(X, Y)] = \sum_{\text{all } x} \sum_{\text{all } y} g(x, y) \cdot p_{X,Y}(x, y)$$

provided $\sum_{\text{all } x} \sum_{\text{all } y} |g(x, y)| \cdot p_{X,Y}(x, y) < \infty$.

2. *Suppose X and Y are continuous random variables with joint pdf $f_{X,Y}(x, y)$, and let $g(X, Y)$ be a continuous function. Then the expected value of the random variable $g(X, Y)$ is given by*

$$E[g(X, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) \cdot f_{X,Y}(x, y)\, dx\, dy$$

provided $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |g(x, y)| \cdot f_{X,Y}(x, y)\, dx\, dy < \infty$.

Proof The basic approach taken in deriving this result is similar to the method followed in the proof of Theorem 3.5.3. See (128) for details. □

Example 3.9.1

Consider the two random variables X and Y whose joint pdf is detailed in the 2×4 matrix shown in Table 3.9.1. Let

$$g(X, Y) = 3X - 2XY + Y$$

Find $E[g(X, Y)]$ two ways—first, by using the basic definition of an expected value, and second, by using Theorem 3.9.1.

Let $Z = 3X - 2XY + Y$. By inspection, Z takes on the values 0, 1, 2, and 3 according to the pdf $f_Z(z)$ shown in Table 3.9.2. Then from the basic definition

Theorem 3.9.3 *If X and Y are independent random variables,*

$$E(XY) = E(X) \cdot E(Y)$$

provided $E(X)$ and $E(Y)$ both exist.

Proof Suppose X and Y are both discrete random variables. Then their joint pdf, $p_{X,Y}(x, y)$, can be replaced by the product of their marginal pdfs, $p_X(x) \cdot p_Y(y)$, and the double summation required by Theorem 3.9.1 can be written as the product of two single summations:

$$\begin{aligned} E(XY) &= \sum_{\text{all } x} \sum_{\text{all } y} xy \cdot p_{X,Y}(x, y) \\ &= \sum_{\text{all } x} \sum_{\text{all } y} xy \cdot p_X(x) \cdot p_Y(y) \\ &= \sum_{\text{all } x} x \cdot p_X(x) \cdot \left[\sum_{\text{all } y} y \cdot p_Y(y) \right] \\ &= E(X) \cdot E(Y) \end{aligned}$$

The proof when X and Y are both continuous random variables is left as an exercise. □

Questions

3.9.1. Suppose that r chips are drawn with replacement from an urn containing n chips, numbered 1 through n. Let V denote the sum of the numbers drawn. Find $E(V)$.

3.9.2. Suppose that $f_{X,Y}(x, y) = \lambda^2 e^{-\lambda(x+y)}$, $0 \le x$, $0 \le y$. Find $E(X + Y)$.

3.9.3. Suppose that $f_{X,Y}(x, y) = \frac{2}{3}(x + 2y)$, $0 \le x \le 1$, $0 \le y \le 1$ [recall Question 3.7.19(c)]. Find $E(X + Y)$.

3.9.4. Marksmanship competition at a certain level requires each contestant to take ten shots with each of two different handguns. Final scores are computed by taking a weighted average of 4 times the number of bull's-eyes made with the first gun plus 6 times the number gotten with the second. If Cathie has a 30% chance of hitting the bull's-eye with each shot from the first gun and a 40% chance with each shot from the second gun, what is her expected score?

3.9.5. Suppose that X_i is a random variable for which $E(X_i) = \mu$, $i = 1, 2, \ldots, n$. Under what conditions will the following be true?

$$E\left(\sum_{i=1}^{n} a_i X_i\right) = \mu$$

3.9.6. Suppose that the daily closing price of stock goes up an eighth of a point with probability p and down an eighth of a point with probability q, where $p > q$. After n days how much gain can we expect the stock to have achieved? Assume that the daily price fluctuations are independent events.

3.9.7. An urn contains r red balls and w white balls. A sample of n balls is drawn *in order* and *without* replacement. Let X_i be 1 if the ith draw is red and 0 otherwise, $i = 1, 2, \ldots, n$.

(a) Show that $E(X_i) = E(X_1)$, $i = 2, 3, \ldots, n$.
(b) Use the corollary to Theorem 3.9.2 to show that the expected number of red balls is $nr/(r + w)$.

3.9.8. Suppose two fair dice are tossed. Find the expected value of the product of the faces showing.

3.9.9. Find $E(R)$ for a two-resistor circuit similar to the one described in Example 3.9.2, where $f_{X,Y}(x, y) = k(x + y)$, $10 \le x \le 20$, $10 \le y \le 20$.

3.9.10. Suppose that X and Y are both uniformly distributed over the interval [0, 1]. Calculate the expected value of the square of the distance of the random point (X, Y) from the origin; that is, find $E(X^2 + Y^2)$. (*Hint:* See Question 3.8.6.)

3.9.11. Suppose X represents a point picked at random from the interval $[0, 1]$ on the x-axis, and Y is a point picked at random from the interval $[0, 1]$ on the y-axis. Assume that X and Y are independent. What is the expected value of the area of the triangle formed by the points $(X, 0)$, $(0, Y)$, and $(0, 0)$?

3.9.12. Suppose $Y_1, Y_2, \ldots, Y_n$ is a random sample from the uniform pdf over $[0, 1]$. The geometric mean of the numbers is the random variable $\sqrt[n]{Y_1 Y_2 \cdot \cdots \cdot Y_n}$. Compare the expected value of the geometric mean to that of the arithmetic mean $\bar{Y}$.

Calculating the Variance of a Sum of Random Variables

When random variables are not independent, a measure of the relationship between them, their *covariance*, enters into the picture.

Definition 3.9.1. Given random variables X and Y with finite variances, define the *covariance* of X and Y, written $\text{Cov}(X, Y)$, as

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y)$$

Theorem 3.9.4 *If X and Y are independent, then* $\text{Cov}(X, Y) = 0$.

Proof If X and Y are independent, by Theorem 3.9.3, $E(XY) = E(X)E(Y)$. Then

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y) = E(X)E(Y) - E(X)E(Y) = 0$$

The converse of Theorem 3.9.4 is *not* true. Just because $\text{Cov}(X, Y) = 0$, we cannot conclude that X and Y are independent. Example 3.9.7 is a case in point. □

Example 3.9.7 Consider the sample space $S = \{(-2, 4), (-1, 1), (0, 0), (1, 1), (2, 4)\}$, where each point is assumed to be equally likely. Define the random variable X to be the first component of a sample point and Y, the second. Then $X(-2, 4) = -2$, $Y(-2, 4) = 4$, and so on.

Notice that X and Y are dependent:

$$\frac{1}{5} = P(X = 1, Y = 1) \neq P(X = 1) \cdot P(Y = 1) = \frac{1}{5} \cdot \frac{2}{5} = \frac{2}{25}$$

However, the convariance of X and Y is zero:

$$E(XY) = [(-8) + (-1) + 0 + 1 + 8] \cdot \frac{1}{5} = 0$$

$$E(X) = [(-2) + (-1) + 0 + 1 + 2] \cdot \frac{1}{5} = 0$$

and

$$E(Y) = (4 + 1 + 0 + 1 + 4) \cdot \frac{1}{5} = 2$$

so

$$\text{Cov}(X, Y) = E(XY) - E(X) \cdot E(Y) = 0 - 0 \cdot 2 = 0$$ ■

Theorem 3.9.5 demonstrates the role of the covariance in finding the variance of a sum of random variables that are not necessarily independent.

Theorem 3.9.5 *Suppose X and Y are random variables with finite variances, and a and b are constants. Then*

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

From the first corollary to Theorem 3.9.5, then,

$$\begin{aligned}\operatorname{Var}(X) &= \sum_{i=1}^{n} \operatorname{Var}(X_i) + 2\sum_{j<k} \operatorname{Cov}(X_j, X_k) \\ &= np(1-p) - 2\binom{n}{2} p(1-p) \cdot \frac{1}{N-1} \\ &= p(1-p)\left[n - \frac{n(n-1)}{N-1}\right] \\ &= np(1-p) \cdot \frac{N-n}{N-1}\end{aligned}$$

■

Example 3.9.11 In statistics, it is often necessary to draw inferences based on $\overline{W}$, the average computed from a random sample of n observations. Two properties of $\overline{W}$ are especially important. First, if the W_i's come from a population where the mean is μ, the corollary to Theorem 3.9.2 implies that $E(\overline{W}) = \mu$. Second, if the W_i's come from a population whose variance is σ^2, then $\operatorname{Var}(\overline{W}) = \sigma^2/n$. To verify the latter, we can appeal again to Theorem 3.9.5. Write

$$\overline{W} = \frac{1}{n}\sum_{i=1}^{n} W_i = \frac{1}{n} \cdot W_1 + \frac{1}{n} \cdot W_2 + \cdots + \frac{1}{n} \cdot W_n$$

Then

$$\begin{aligned}\operatorname{Var}(\overline{W}) &= \left(\frac{1}{n}\right)^2 \cdot \operatorname{Var}(W_1) + \left(\frac{1}{n}\right)^2 \cdot \operatorname{Var}(W_2) + \cdots + \left(\frac{1}{n}\right)^2 \cdot \operatorname{Var}(W_n) \\ &= \left(\frac{1}{n}\right)^2 \sigma^2 + \left(\frac{1}{n}\right)^2 \sigma^2 + \cdots + \left(\frac{1}{n}\right)^2 \sigma^2 \\ &= \frac{\sigma^2}{n}\end{aligned}$$

■

Questions

3.9.13. Suppose that two dice are thrown. Let X be the number showing on the first die and let Y be the larger of the two numbers showing. Find $\operatorname{Cov}(X, Y)$.

3.9.14. Show that

$$\operatorname{Cov}(aX + b, cY + d) = ac\operatorname{Cov}(X, Y)$$

for any constants a, b, c, and d.

3.9.15. Let U be a random variable uniformly distributed over $[0, 2\pi]$. Define $X = \cos U$ and $Y = \sin U$. Show that X and Y are dependent but that $\operatorname{Cov}(X, Y) = 0$.

3.9.16. Let X and Y be random variables with

$$f_{X,Y}(x, y) = \begin{cases} 1, & -y < x < y, \quad 0 < y < 1 \\ 0, & \text{elsewhere} \end{cases}$$

Show that $\operatorname{Cov}(X, Y) = 0$ but that X and Y are dependent.

3.9.17. Suppose that $f_{X,Y}(x, y) = \lambda^2 e^{-\lambda(x+y)}$, $0 \le x$, $0 \le y$. Find $\operatorname{Var}(X + Y)$. (*Hint:* See Questions 3.6.11 and 3.9.2.)

3.9.18. Suppose that $f_{X,Y}(x, y) = \frac{2}{3}(x + 2y)$, $0 \le x \le 1$, $0 \le y \le 1$. Find $\operatorname{Var}(X + Y)$. (*Hint:* See Question 3.9.3.)

3.9.19. For the uniform pdf defined over [0, 1], find the variance of the geometric mean when $n = 2$ (see Question 3.9.12).

3.9.20. Let X be a binomial random variable based on n trials and a success probability of p_x; let Y be an independent binomial random variable based on m trials and a success probability of p_Y. Find $E(W)$ and $\operatorname{Var}(W)$, where $W = 4X + 6Y$.

3.9.21. Let the Poisson random variable U (see p. 227) be the number of calls for technical assistance received by a computer company during the firm's nine normal workday hours. Suppose the average number of calls per hour is 7.0 and that each call costs the company \$50. Let V be a Poisson random variable representing the number of calls for technical assistance received during a day's remaining fifteen hours. Suppose the average number of calls per hour is 4.0 for that time period and that each such call costs the company \$60. Find the expected cost and the variance of the cost associated with the calls received during a twenty-four-hour day.

3.9.22. A mason is contracted to build a patio retaining wall. Plans call for the base of the wall to be a row of fifty 10-inch bricks, each separated by $\frac{1}{2}$-inch-thick mortar. Suppose that the bricks used are randomly chosen from a population of bricks whose mean length is 10 inches and whose standard deviation is $\frac{1}{32}$ inch. Also, suppose that the mason, on the average, will make the mortar $\frac{1}{2}$ inch thick, but that the actual dimension will vary from brick to brick, the standard deviation of the thicknesses being $\frac{1}{16}$ inch. What is the standard deviation of L, the length of the first row of the wall? What assumption are you making?

3.9.23. An electric circuit has six resistors wired in series, each nominally being five ohms. What is the maximum standard deviation that can be allowed in the manufacture of these resistors if the combined circuit resistance is to have a standard deviation no greater than 0.4 ohm?

3.9.24. A gambler plays n hands of poker. If he wins the kth hand, he collects k dollars; if he loses the kth hand, he collects nothing. Let T denote his total winnings in n hands. Assuming that his chances of winning each hand are constant and independent of his success or failure at any other hand, find $E(T)$ and $\text{Var}(T)$.

3.10 Order Statistics

The single-variable transformation taken up in Section 3.4 involved a standard linear operation, $Y = aX + b$. The bivariate transformations in Section 3.8 were similarly arithmetic, typically being concerned with either sums or products. In this section we will consider a different sort of transformation, one involving the *ordering* of an entire *set* of random variables. This particular transformation has wide applicability in many areas of statistics, and we will see some of its consequences in later chapters.

Definition 3.10.1. Let Y be a continuous random variable for which $y_1, y_2, \ldots, y_n$ are the values of a random sample of size n. Reorder the y_i's from smallest to largest:

$$y_1' < y_2' < \cdots < y_n'$$

(No two of the y_i's are equal, except with probability zero, since Y is continuous.) Define the random variable Y_i' to have the value y_i', $1 \le i \le n$. Then Y_i' is called the ith *order statistic*. Sometimes Y_n' and Y_1' are denoted $Y_{\max}$ and $Y_{\min}$, respectively.

Example 3.10.1 Suppose that four measurements are made on the random variable Y: $y_1 = 3.4$, $y_2 = 4.6$, $y_3 = 2.6$, and $y_4 = 3.2$. The corresponding ordered sample would be

$$2.6 < 3.2 < 3.4 < 4.6$$

The random variable representing the smallest observation would be denoted Y_1', with its value for this particular sample being 2.6. Similarly, the value for the second order statistic, Y_2', is 3.2, and so on. ■

The Distribution of Extreme Order Statistics

By definition, every observation in a random sample has the same pdf. For example, if a set of four measurements is taken from a normal distribution with $\mu = 80$ and $\sigma = 15$, then $f_{Y_1}(y)$, $f_{Y_2}(y)$, $f_{Y_3}(y)$, and $f_{Y_4}(y)$ are all the same—each is a normal

Questions

3.10.1. Suppose the length of time, in minutes, that you have to wait at a bank teller's window is uniformly distributed over the interval (0, 10). If you go to the bank four times during the next month, what is the probability that your second longest wait will be less than five minutes?

3.10.2. A random sample of size $n = 6$ is taken from the pdf $f_Y(y) = 3y^2$, $0 \le y \le 1$. Find $P(Y_5' > 0.75)$.

3.10.3. What is the probability that the larger of two random observations drawn from any continuous pdf will exceed the sixtieth percentile?

3.10.4. A random sample of size 5 is drawn from the pdf $f_Y(y) = 2y$, $0 \le y \le 1$. Calculate $P(Y_1' < 0.6 < Y_5')$. (*Hint:* Consider the complement.)

3.10.5. Suppose that $Y_1, Y_2, \ldots, Y_n$ is a random sample of size n drawn from a continuous pdf, $f_Y(y)$, whose median is m. Is $P(Y_1' > m)$ less than, equal to, or greater than $P(Y_n' > m)$?

3.10.6. Let $Y_1, Y_2, \ldots, Y_n$ be a random sample from the exponential pdf $f_y(y) = e^{-y}$, $y \ge 0$. What is the smallest n for which $P(Y_{\min} < 0.2) > 0.9$?

3.10.7. Calculate $P(0.6 < Y_4' < 0.7)$ if a random sample of size 6 is drawn from the uniform pdf defined over the interval [0, 1].

3.10.8. A random sample of size $n = 5$ is drawn from the pdf $f_Y(y) = 2y$, $0 \le y \le 1$. On the same set of axes, graph the pdfs for Y_2, Y_1', and Y_5'.

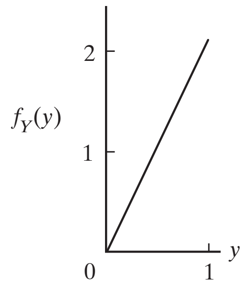
3.10.9. Suppose that n observations are taken at random from the pdf

$$f_Y(y) = \frac{1}{\sqrt{2\pi}(6)} e^{-\frac{1}{2}\left(\frac{y-20}{6}\right)^2}, \quad -\infty < y < \infty$$

What is the probability that the smallest observation is larger than twenty?

3.10.10. Suppose that n observations are chosen at random from a continuous pdf $f_Y(y)$. What is the probability that the last observation recorded will be the smallest number in the entire sample?

3.10.11. In a certain large metropolitan area, the proportion, Y, of students bused varies widely from school to school. The distribution of proportions is roughly described by the following pdf:



Suppose the enrollment figures for five schools selected at random are examined. What is the probability that the school with the fourth highest proportion of bused children will have a Y value in excess of 0.75? What is the probability that none of the schools will have fewer than 10% of their students bused?

3.10.12. Consider a system containing n components, where the lifetimes of the components are independent random variables and each has pdf $f_Y(y) = \lambda e^{-\lambda y}$, $y > 0$. Show that the average time elapsing before the first component failure occurs is $1/n\lambda$.

3.10.13. Let $Y_1, Y_2, \ldots, Y_n$ be a random sample from a uniform pdf over [0, 1]. Use Theorem 3.10.2 to show that $\int_0^1 y^{i-1}(1-y)^{n-i}dy = \dfrac{(i-1)!(n-i)!}{n!}$.

3.10.14. Use Question 3.10.13 to find the expected value of Y_i', where $Y_1, Y_2, \ldots, Y_n$ is a random sample from a uniform pdf defined over the interval [0, 1].

3.10.15. Suppose three points are picked randomly from the unit interval. What is the probability that the three are within a half unit of one another?

3.10.16. Suppose a device has three independent components, all of whose lifetimes (in months) are modeled by the exponential pdf, $f_Y(y) = e^{-y}$, $y > 0$. What is the probability that all three components will fail within two months of one another?

3.11 Conditional Densities

We have already seen that many of the concepts defined in Chapter 2 relating to the probabilities of *events*—for example, independence—have random variable counterparts. Another of these carryovers is the notion of a conditional probability, or, in what will be our present terminology, a *conditional probability density function.* Applications of conditional pdfs are not uncommon. The height and girth of a tree, Suppose $z = 2$. Then

$$\begin{aligned}
p_Z(2) = P(Z=2) &= P[(2,1,2) \cup (1,2,2) \cup (2,2,2)] \\
&= 2 \cdot \frac{1}{18} + 1 \cdot \frac{2}{18} + 2 \cdot \frac{2}{18} \\
&= \frac{8}{18}
\end{aligned}$$

so

$$\begin{aligned}
p_{X,Y|2}(x,y) &= \frac{p_{X,Y,Z}(x,y,2)}{p_Z(2)} \\
&= \frac{x \cdot y/18}{\frac{8}{18}} \\
&= \frac{xy}{8} \quad \text{for} \quad (x,y) = (2,1), (1,2), \text{ and } (2,2)
\end{aligned}$$

■

Questions

3.11.1. Suppose X and Y have the joint pdf $p_{X,Y}(x,y) = \frac{x+y+xy}{21}$ for the points (1, 1), (1, 2), (2, 1), (2, 2), where X denotes a "message" sent (either $x = 1$ or $x = 2$) and Y denotes a "message" received. Find the probability that the message sent was the message received—that is, find $p_{Y|x}(y)$.

3.11.2. Suppose a die is rolled six times. Let X be the total number of 4's that occur and let Y be the number of 4's in the first two tosses. Find $p_{Y|x}(y)$.

3.11.3. An urn contains eight red chips, six white chips, and four blue chips. A sample of size 3 is drawn without replacement. Let X denote the number of red chips in the sample and Y, the number of white chips. Find an expression for $p_{Y|x}(y)$.

3.11.4. Five cards are dealt from a standard poker deck. Let X be the number of aces received, and Y the number of kings. Compute $P(X = 2|Y = 2)$.

3.11.5. Given that two discrete random variables X and Y follow the joint pdf $p_{X,Y}(x,y) = k(x+y)$, for $x = 1, 2, 3$ and $y = 1, 2, 3$,

(a) Find k.
(b) Evaluate $p_{Y|x}(1)$ for all values of x for which $p_x(x) > 0$.

3.11.6. Let X denote the number on a chip drawn at random from an urn containing three chips, numbered 1, 2, and 3. Let Y be the number of heads that occur when a fair coin is tossed X times.

(a) Find $p_{X,Y}(x,y)$.
(b) Find the marginal pdf of Y by summing out the x values.

3.11.7. Suppose X, Y, and Z have a trivariate distribution described by the joint pdf

$$p_{X,Y,Z}(x,y,z) = \frac{xy + xz + yz}{54}$$

where x, y, and z can be 1 or 2. Tabulate the joint conditional pdf of X and Y given each of the two values of z.

3.11.8. In Question 3.11.7 define the random variable W to be the "majority" of x, y, and z. For example, $W(2,2,1) = 2$ and $W(1,1,1) = 1$. Find the pdf of $W|x$.

3.11.9. Let X and Y be independent random variables where $p_x(k) = e^{-\lambda}\frac{\lambda^k}{k!}$ and $p_Y(k) = e^{-\mu}\frac{\mu^k}{k!}$ for $k = 0, 1, \ldots$. Show that the conditional pdf of X given that $X + Y = n$ is binomial with parameters n and $\frac{\lambda}{\lambda+\mu}$. (*Hint:* See Question 3.8.1.)

3.11.10. Suppose Compositor A is preparing a manuscript to be published. Assume that she makes X errors on a given page, where X has the Poisson pdf, $p_X(k) = e^{-2}2^k/k!$, $k = 0, 1, 2, \ldots$. A second compositor, B, is also working on the book. He makes Y errors on a page, where $p_Y(k) = e^{-3}3^k/k!$, $k = 0, 1, 2, \ldots$. Assume that Compositor A prepares the first one hundred pages of the text and Compositor B, the last one hundred pages. After the book is completed, reviewers (with too much time on their hands!) find that the text contains a total of 520 errors. Write a formula for the exact probability that fewer than half of the errors are due to Compositor A.

a. From Theorem 3.7.2,

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dy = \int_2^4 \left(\frac{1}{8}\right)(6 - x - y)\, dy$$
$$= \left(\frac{1}{8}\right)(6 - 2x), \quad 0 \le x \le 2$$

b. Substituting into the "continuous" statement of Definition 3.11.1, we can write

$$f_{Y|x}(y) = \frac{f_{X,Y}(x, y)}{f_X(x)} = \frac{\left(\frac{1}{8}\right)(6 - x - y)}{\left(\frac{1}{8}\right)(6 - 2x)}$$

$$= \frac{6 - x - y}{6 - 2x}, \quad 0 \le x \le 2, \quad 2 \le y \le 4$$

c. To find $P(2 < Y < 3|x = 1)$, we simply integrate $f_{Y|1}(y)$ over the interval $2 < Y < 3$:

$$P(2 < Y < 3|x = 1) = \int_2^3 f_{Y|1}(y)\, dy$$
$$= \int_2^3 \frac{5 - y}{4}\, dy$$
$$= \frac{5}{8}$$

[A partial check that the derivation of a conditional pdf is correct can be performed by integrating $f_{Y|x}(y)$ over the entire range of Y. That integral should be 1. Here, for example, when $x = 1$, $\int_{-\infty}^{\infty} f_{Y|1}(y)\, dy = \int_2^4 [(5 - y)/4]\, dy$ *does* equal 1.] ■

Questions

3.11.11. Let X be a nonnegative random variable. We say that X is *memoryless* if

$$P(X > s + t|X > t) = P(X > s) \quad \text{for all } s, t \ge 0$$

Show that a random variable with pdf $f_X(x) = (1/\lambda)e^{-x/\lambda}$, $x > 0$, is memoryless.

3.11.12. Given the joint pdf

$$f_{X,Y}(x, y) = 2e^{-(x+y)}, \quad 0 \le x \le y, \quad y \ge 0$$

find

(a) $P(Y < 1|X < 1)$.
(b) $P(Y < 1|X = 1)$.
(c) $f_{Y|x}(y)$.
(d) $E(Y|x)$.

3.11.13. Find the conditional pdf of Y given x if

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

for $0 \le x \le 1$ and $0 \le y \le 1$.

3.11.14. If

$$f_{X,Y}(x, y) = 2, \quad x \ge 0, \quad y \ge 0, \quad x + y \le 1$$

show that the conditional pdf of Y given x is uniform.

3.11.15. Suppose that

$$f_{Y|x}(y) = \frac{2y + 4x}{1 + 4x} \quad \text{and} \quad f_X(x) = \frac{1}{3} \cdot (1 + 4x)$$

for $0 < x < 1$ and $0 < y < 1$. Find the marginal pdf for Y.

3.11.16. Suppose that X and Y are distributed according to the joint pdf

$$f_{X,Y}(x, y) = \frac{2}{5} \cdot (2x + 3y), \quad 0 \le x \le 1, \quad 0 \le y \le 1$$

Find

(a) $f_X(x)$.
(b) $f_{Y|x}(y)$.
(c) $P\left(\frac{1}{4} \le Y \le \frac{3}{4}|X = \frac{1}{2}\right)$.
(d) $E(Y|x)$.

3.11.17. If X and Y have the joint pdf

$$f_{X,Y}(x, y) = 2, \quad 0 \le x < y \le 1$$

find $P\left(0 < X < \frac{1}{2} | Y = \frac{3}{4}\right)$.

3.11.18. Find $P\left(X < 1 | Y = 1\frac{1}{2}\right)$ if X and Y have the joint Pdf

$$f_{X,Y}(x, y) = xy/2, \quad 0 \le x < y \le 2$$

3.11.19. Suppose that X_1, X_2, X_3, X_4, and X_5 have the joint pdf

$$f_{X_1,X_2,X_3,X_4,X_5}(x_1, x_2, x_3, x_4, x_5) = 32x_1x_2x_3x_4x_5$$

for $0 < x_i < 1$, $i = 1, 2, \dots, 5$. Find the joint conditional pdf of X_1, X_2, and X_3 given that $X_4 = x_4$ and $X_5 = x_5$.

3.11.20. Suppose the random variables X and Y are jointly distributed according to the Pdf

$$f_{X,Y}(x, y) = \frac{6}{7}\left(x^2 + \frac{xy}{2}\right), \quad 0 \le x \le 1, \quad 0 \le y \le 2$$

Find

(a) $f_X(x)$.
(b) $P(X > 2Y)$.
(c) $P\left(Y > 1 | X > \frac{1}{2}\right)$.

3.12 Moment-Generating Functions

Finding moments of random variables directly, particularly the higher moments defined in Section 3.6, is conceptually straightforward but can be quite problematic: Depending on the nature of the pdf, integrals and sums of the form $\int_{-\infty}^{\infty} y^r f_Y(y)\,dy$ and $\sum\limits_{\text{all } k} k^r p_X(k)$ can be very difficult to evaluate. Fortunately, an alternative method is available. For many pdfs, we can find a *moment-generating function* (or *mgf*), $M_W(t)$, one of whose properties is that the rth derivative of $M_W(t)$ evaluated at zero is equal to $E(W^r)$.

Calculating a Random Variable's Moment-Generating Function

In principle, what we call a moment-generating function is a direct application of Theorem 3.5.3.

Definition 3.12.1. Let W be a random variable. The *moment-generating function (mgf) for W* is denoted $M_W(t)$ and given by

$$M_W(t) = E(e^{tW}) = \begin{cases} \sum\limits_{\text{all } k} e^{tk} p_W(k) & \text{if } W \text{ is discrete} \\ \int_{-\infty}^{\infty} e^{tw} f_W(w)\,dw & \text{if } W \text{ is continuous} \end{cases}$$

at all values of t for which the expected value exists.

Example 3.12.1

Suppose the random variable X has a *geometric pdf*,

$$p_X(k) = (1 - p)^{k-1} p, \quad k = 1, 2, \dots$$

[In practice, this is the pdf that models the occurrence of the first success in a series of independent trials, where each trial has a probability p of ending in success (recall Example 3.3.2)]. Find $M_X(t)$, the moment-generating function for X.

Since X is discrete, the first part of Definition 3.12.1 applies, so

$$M_X(t) = E(e^{tX}) = \sum_{k=1}^{\infty} e^{tk}(1 - p)^{k-1} p$$

$$= \frac{p}{1 - p} \sum_{k=1}^{\infty} e^{tk}(1 - p)^k = \frac{p}{1 - p} \sum_{k=1}^{\infty} [(1 - p)e^t]^k \tag{3.12.1}$$

Questions

3.12.1. Let X be a random variable with pdf $p_X(k) = 1/n$, for $k = 0, 1, 2, \ldots, n-1$ and 0 otherwise. Show that $M_X(t) = \frac{1-e^{nt}}{n(1-e^t)}$.

3.12.2. Two chips are drawn at random and without replacement from an urn that contains five chips, numbered 1 through 5. If the sum of the chips drawn is even, the random variable X equals 5; if the sum of the chips drawn is odd, $X = -3$. Find the moment-generating function for X.

3.12.3. Find the expected value of e^{3X} if X is a binominal random variable with $n = 10$ and $p = \frac{1}{3}$.

3.12.4. Find the moment-generating function for the discrete random variable X whose probability function is given by

$$p_X(k) = \left(\frac{3}{4}\right)^k \left(\frac{1}{4}\right), \quad k = 0, 1, 2, \ldots$$

3.12.5. Which pdfs would have the following moment-generating functions?

(a) $M_Y(t) = e^{6t^2}$
(b) $M_Y(t) = 2/(2-t)$
(c) $M_X(t) = \left(\frac{1}{2} + \frac{1}{2}e^t\right)^4$
(d) $M_X(t) = 0.3e^t/(1-0.7e^t)$

3.12.6. Let Y have pdf

$$f_Y(y) = \begin{cases} y, & 0 \le y \le 1 \\ 2-y, & 1 \le y \le 2 \\ 0, & \text{elsewhere} \end{cases}$$

Find $M_Y(t)$.

3.12.7. A random variable X is said to have a *Poisson distribution* if $p_X(k) = P(X = k) = e^{-\lambda}\lambda^k/k!$, $k = 0, 1, 2, \ldots$. Find the moment-generating function for a Poisson random variable. Recall that

$$e^r = \sum_{k=0}^{\infty} \frac{r^k}{k!}$$

3.12.8. Let Y be a continuous random variable with $f_Y(y) = ye^{-y}$, $0 \le y$. Show that $M_Y(t) = \frac{1}{(1-t)^2}$.

Using Moment-Generating Functions to Find Moments

Having practiced *finding* the functions $M_X(t)$ and $M_Y(t)$, we now turn to the theorem that spells out their relationship to X^r and Y^r.

Theorem 3.12.1 *Let W be a random variable with probability density function $f_W(w)$. [If W is continuous, $f_W(w)$ must be sufficiently smooth to allow the order of differentiation and integration to be interchanged.] Let $M_W(t)$ be the moment-generating function for W. Then, provided the rth moment exists,*

$$M_W^{(r)}(0) = E(W^r)$$

Proof We will verify the theorem for the continuous case where r is either 1 or 2. The extensions to discrete random variables and to an arbitrary positive integer r are straightforward.

For $r = 1$,

$$\begin{aligned} M_Y^{(1)}(0) &= \frac{d}{dt}\int_{-\infty}^{\infty} e^{ty} f_Y(y)\,dy\bigg|_{t=0} = \int_{-\infty}^{\infty} \frac{d}{dt} e^{ty} f_Y(y)\,dy\bigg|_{t=0} \\ &= \int_{-\infty}^{\infty} ye^{ty} f_Y(y)\,dy\bigg|_{t=0} = \int_{-\infty}^{\infty} ye^{0\cdot y} f_Y(y)\,dy \\ &= \int_{-\infty}^{\infty} yf_Y(y)\,dy = E(Y) \end{aligned}$$

Questions

3.12.9. Calculate $E(Y^3)$ for a random variable whose moment-generating function is $M_Y(t) = e^{t^2/2}$.

3.12.10. Find $E(Y^4)$ if Y is an exponential random variable with $f_Y(y) = \lambda e^{-\lambda y}$, $y > 0$.

3.12.11. The form of the moment-generating function for a normal random variable is $M_Y(t) = e^{at+b^2t^2/2}$ (recall Example 3.12.4). Differentiate $M_Y(t)$ to verify that $a = E(Y)$ and $b^2 = \text{Var}(Y)$.

3.12.12. What is $E(Y^4)$ if the random variable Y has moment-generating function $M_Y(t) = (1 - \alpha t)^{-k}$?

3.12.13. Find $E(Y^2)$ if the moment-generating function for Y is given by $M_Y(t) = e^{-t+4t^2}$. Use Example 3.12.4 to find $E(Y^2)$ without taking any derivatives. (*Hint*: Recall Theorem 3.6.1.)

3.12.14. Find an expression for $E(Y^k)$ if $M_Y(t) = (1 - t/\lambda)^{-r}$, where λ is any positive real number and r is a positive integer.

3.12.15. Use $M_Y(t)$ to find the expected value of the uniform random variable described in Question 3.12.1.

3.12.16. Find the variance of Y if $M_Y(t) = e^{2t}/(1 - t^2)$.

Using Moment-Generating Functions to Identify Pdfs

Finding moments is not the only application of moment-generating functions. They are also used to identify the pdf of *sums* of random variables—that is, finding $f_W(w)$, where $W = W_1 + W_2 + \cdots + W_n$. Their assistance in the latter is particularly important for two reasons: (1) Many statistical procedures are defined in terms of sums, and (2) alternative methods for deriving $f_{W_1+W_2+\cdots+W_n}(w)$ are extremely cumbersome.

The next two theorems give the background results necessary for deriving $f_W(w)$. Theorem 3.12.2 states a key uniqueness property of moment-generating functions: If W_1 and W_2 are random variables with the same mgfs, they must necessarily have the same pdfs. In practice, applications of Theorem 3.12.2 typically rely on one or both of the algebraic properties cited in Theorem 3.12.3.

Theorem 3.12.2 *Suppose that W_1 and W_2 are random variables for which $M_{W_1}(t) = M_{W_2}(t)$ for some interval of t's containing 0. Then $f_{W_1}(w) = f_{W_2}(w)$.*

Proof See (95). □

Theorem 3.12.3

a. *Let W be a random variable with moment-generating function $M_W(t)$. Let $V = aW + b$. Then*

$$M_V(t) = e^{bt} M_W(at)$$

b. *Let $W_1, W_2, \ldots, W_n$ be independent random variables with moment-generating functions $M_{W_1}(t), M_{W_2}(t), \ldots,$ and $M_{W_n}(t)$, respectively. Let $W = W_1 + W_2 + \cdots + W_n$. Then*

$$M_W(t) = M_{W_1}(t) \cdot M_{W_2}(t) \cdots M_{W_n}(t)$$

Proof The proof is left as an exercise. □

Example 3.12.10 Suppose that X_1 and X_2 are two independent Poisson random variables with parameters λ_1 and λ_2, respectively. That is,

$$p_{X_1}(k) = P(X_1 = k) = \frac{e^{-\lambda_1}\lambda_1 k}{k!}, \quad k = 0, 1, 2, \ldots$$

But $M_Z(t) = e^{t^2/2}$ so it follows from Theorem 3.12.2 that the pdf for $\frac{Y-\mu}{\sigma}$ is the same as the pdf for $f_z(z)$. (We call $\frac{Y-\mu}{\sigma}$ a *Z transformation*. Its importance will become evident in Chapter 4.) ■

Questions

3.12.17. Use Theorem 3.12.3(a) and Question 3.12.8 to find the moment-generating function of the random variable Y, where $f_Y(y) = \lambda y e^{-\lambda y}$, $y \geq 0$.

3.12.18. Let Y_1, Y_2, and Y_3 be independent random variables, each having the pdf of Question 3.12.17. Use Theorem 3.12.3(b) to find the moment-generating function of $Y_1 + Y_2 + Y_3$. Compare your answer to the moment-generating function in Question 3.12.14.

3.12.19. Use Theorems 3.12.2 and 3.12.3 to determine which of the following statements is true:

(a) The sum of two independent Poisson random variables has a Poisson distribution.
(b) The sum of two independent exponential random variables has an exponential distribution.
(c) The sum of two independent normal random variables has a normal distribution.

3.12.20. Calculate $P(X \leq 2)$ if $M_X(t) = \left(\frac{1}{4} + \frac{3}{4}e^t\right)^5$.

3.12.21. Suppose that $Y_1, Y_2, \ldots, Y_n$ is a random sample of size n from a normal distribution with mean μ and standard deviation σ. Use moment-generating functions to deduce the pdf of $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$.

3.12.22. Suppose the moment-generating function for a random variable W is given by

$$M_W(t) = e^{-3+3e^t} \cdot \left(\frac{2}{3} + \frac{1}{3}e^t\right)^4$$

Calculate $P(W \leq 1)$. (*Hint:* Write W as a sum.)

3.12.23. Suppose that X is a Poisson random variable, where $p_X(k) = e^{-\lambda}\lambda^k/k!$, $k = 0, 1, \ldots$.

(a) Does the random variable $W = 3X$ have a Poisson distribution?
(b) Does the random variable $W = 3X + 1$ have a Poisson distribution?

3.12.24. Suppose that Y is a normal variable, where $f_Y(y) = (1/\sqrt{2\pi}\sigma)\exp\left[-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right]$, $-\infty < y < \infty$.

(a) Does the random variable $W = 3Y$ have a normal distribution?
(b) Does the random variable $W = 3Y + 1$ have a normal distribution?

3.13 Taking a Second Look at Statistics (Interpreting Means)

One of the most important ideas coming out of Chapter 3 is the notion of the *expected value* (or *mean*) of a random variable. Defined in Section 3.5 as a number that reflects the "center" of a pdf, the expected value (μ) was originally introduced for the benefit of gamblers. It spoke directly to one of their most fundamental questions—How much will I win or lose, *on the average,* if I play a certain game? (Actually, the real question they probably had in mind was "How much are *you* going to *lose,* on the average?") Despite having had such a selfish, materialistic, gambling-oriented *raison d'etre,* the expected value was quickly embraced by (respectable) scientists and researchers of all persuasions as a preeminently useful descriptor of a distribution. Today, it would not be an exaggeration to claim that the majority of *all* statistical analyses focus on either (1) the expected value of a single random variable or (2) comparing the expected values of two or more random variables.

(Case Study 4.2.1 continued)

$$
\begin{aligned}
P(X \geq 8) &= 1 - P(X \leq 7) \\
&\doteq 1 - \sum_{k=0}^{7} \frac{e^{-1.75}(1.75)^k}{k!} \\
&= 1 - 0.99953 \\
&= 0.00047
\end{aligned}
$$

How close can we expect 0.00047 to be to the "true" binomial sum? Very close. Considering the accuracy of the Poisson limit when n is as small as one hundred (recall Table 4.2.2), we should feel very confident here, where n is 7076.

Interpreting the 0.00047 probability is not nearly as easy as assessing its accuracy. The fact that the probability is so very small tends to denigrate the hypothesis that leukemia in Niles occurred at random. On the other hand, rare events, such as clusters, *do* happen by chance. The basic difficulty of putting the probability associated with a given cluster into any meaningful perspective is not knowing in how many similar communities leukemia did *not* exhibit a tendency to cluster. That there is no obvious way to do this is one reason the leukemia controversy is still with us.

About the Data Publication of the Niles cluster led to a number of research efforts on the part of biostatisticians to find quantitative methods capable of detecting clustering in space and time for diseases having low epidemicity. Several techniques were ultimately put forth, but the inherent "noise" in the data—variations in population densities, ethnicities, risk factors, and medical practices—often proved impossible to overcome.

Questions

4.2.1. If a typist averages one misspelling in every 3250 words, what are the chances that a 6000-word report is free of all such errors? Answer the question two ways—first, by using an exact binomial analysis, and second, by using a Poisson approximation. Does the similarity (or dissimilarity) of the two answers surprise you? Explain.

4.2.2. A medical study recently documented that 905 mistakes were made among the 289,411 prescriptions written during one year at a large metropolitan teaching hospital. Suppose a patient is admitted with a condition serious enough to warrant 10 different prescriptions. Approximate the probability that at least one will contain an error.

4.2.3. Five hundred people are attending the first annual "I was Hit by Lighting" Club. Approximate the probability that at most one of the five hundred was born on Poisson's birthday.

4.2.4. A chromosome mutation linked with colorblindness is known to occur, on the average, once in every ten thousand births.

(a) Approximate the probability that exactly three of the next twenty thousand babies born will have the mutation.

(b) How many babies out of the next twenty thousand would have to be born with the mutation to convince you that the "one in ten thousand" estimate is too low? [*Hint:* Calculate $P(X \geq k) = 1 - P(X \leq k-1)$ for various k. (Recall Case Study 4.2.1.)]

4.2.5. Suppose that 1% of all items in a supermarket are not priced properly. A customer buys ten items. What is the probability that she will be delayed by the cashier because one or more of her items require a price check?

Calculate both a binomial answer and a Poisson answer. Is the binomial model "exact" in this case? Explain.

4.2.6. A newly formed life insurance company has underwritten term policies on 120 women between the ages of forty and forty-four. Suppose that each woman has a 1/150 probability of dying during the next calendar year, and that each death requires the company to pay out \$50,000 in benefits. Approximate the probability that the company will have to pay at least \$150,000 in benefits next year.

4.2.7. According to an airline industry report (178), roughly 1 piece of luggage out of every 200 that are checked is lost. Suppose that a frequent-flying businesswoman will be checking 120 bags over the course of the next year. Approximate the probability that she will lose 2 of more pieces of luggage.

4.2.8. Electromagnetic fields generated by power transmission lines are suspected by some researchers to be a cause of cancer. Especially at risk would be telephone linemen because of their frequent proximity to high-voltage wires. According to one study, two cases of a rare form of cancer were detected among a group of 9500 linemen (174). In the general population, the incidence of that particular condition is on the order of one in a million. What would you conclude? (*Hint:* Recall the approach taken in Case Study 4.2.1.)

4.2.9. Astronomers estimate that as many as one hundred billion stars in the Milky Way galaxy are encircled by planets. If so, we may have a plethora of cosmic neighbors. Let p denote the probability that any such solar system contains intelligent life. How small can p be and still give a fifty-fifty chance that we are not alone?

The Poisson Distribution

The real significance of Poisson's limit theorem went unrecognized for more than fifty years. For most of the latter part of the nineteenth century, Theorem 4.2.1 was taken strictly at face value: It provides a convenient approximation for $p_X(k)$ when X is binomial, n is large, and p is small. But then in 1898 a German professor, Ladislaus von Bortkiewicz, published a monograph entitled *Das Gesetz der Kleinen Zahlen (The Law of Small Numbers)* that would quickly transform Poisson's "limit" into Poisson's "distribution."

What is best remembered about Bortkiewicz's monograph is the curious set of data described in Question 4.2.10. The measurements recorded were the numbers of Prussian cavalry soldiers who had been kicked to death by their horses. In analyzing those figures, Bortkiewicz was able to show that the function $e^{-\lambda}\lambda^k/k!$ is a useful probability model in its own right, even when (1) no explicit binomial random variable is present and (2) values for n and p are unavailable. Other researchers were quick to follow Bortkiewicz's lead, and a steady stream of Poisson distribution applications began showing up in technical journals. Today the function $p_X(k) = e^{-\lambda}\lambda^k/k!$ is universally recognized as being among the three or four most important data models in all of statistics.

Theorem 4.2.2 *The random variable X is said to have a Poisson distribution if*

$$p_X(k) = P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}, \quad k = 0, 1, 2, \ldots$$

where λ is a positive constant. Also, for any Poisson random variable, $E(X) = \lambda$ and $\text{Var}(X) = \lambda$.

Proof To show that $p_X(k)$ qualifies as a probability function, note, first of all, that $p_X(k) \geq 0$ for all nonnegative integers k. Also, $p_X(k)$ sums to 1:

$$\sum_{k=0}^{\infty} p_X(k) = \sum_{k=0}^{\infty} \frac{e^{-\lambda}\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} \cdot e^{\lambda} = 1$$

since $\sum_{k=0}^{\infty} \frac{\lambda^k}{k!}$ is the Taylor series expansion of e^{λ}. Verifying that $E(X) = \lambda$ and $\text{Var}(X) = \lambda$ has already been done in Example 3.12.9, using moment-generating functions. $\square$

is equal to k. Formulas (2) and (3) are sometimes confused because both presume to give the probability that a Poisson random variable equals k. Why are they different?

Actually, all three formulas are the same in the sense that the right-hand sides of each could be written as

4. $e^{-E(X)}\frac{[E(X)]^k}{k!}$

In formula (1), X is binomial, so $E(X) = np$. In formula (2), which comes from Theorem 4.2.2, λ is defined to be $E(X)$. Formula (3) covers all those situations where the units of X and λ are not consistent, in which case $E(X) \neq \lambda$. However, λ can always be multiplied by an appropriate constant T to make λT equal to $E(X)$.

For example, suppose a certain radioisotope is known to emit α particles at the rate of $\lambda = 1.5$ emissions/second. For whatever reason, though, the experimenter defines the Poisson random variable X to be the number of emissions counted in a given *minute*. Then $T = 60$ seconds and

$$\begin{aligned} E(X) &= 1.5 \text{ emissions/second} \times 60 \text{ seconds} \\ &= \lambda T = 90 \text{ emissions} \end{aligned}$$

Example 4.2.4

Entomologists estimate that an average person consumes almost a pound of bug parts each year (173). There are that many insect eggs, larvae, and miscellaneous body pieces in the foods we eat and the liquids we drink. The Food and Drug Administration (FDA) sets a Food Defect Action Level (FDAL) for each product: Bug-part concentrations below the FDAL are considered acceptable. The legal limit for peanut butter, for example, is thirty insect fragments per hundred grams. Suppose the crackers you just bought from a vending machine are spread with twenty grams of peanut butter. What are the chances that your snack will include at least five crunchy critters?

Let X denote the number of bug parts in twenty grams of peanut butter. Assuming the worst, suppose the contamination level equals the FDA limit—that is, thirty fragments per hundred grams (or 0.30 fragment/g). Notice that T in this case is twenty grams, making $E(X) = 6.0$:

$$\frac{0.30 \text{ fragment}}{\text{g}} \times 20 \text{ g} = 6.0 \text{ fragments}$$

It follows, then, that the probability that your snack contains five or more bug parts is a disgusting *0.71*:

$$\begin{aligned} P(X \geq 5) = 1 - P(X \leq 4) &= 1 - \sum_{k=0}^{4} \frac{e^{-6.0}(6.0)^k}{k!} \\ &= 1 - 0.29 \\ &= 0.71 \end{aligned}$$

Bon appetit! ■

Questions

4.2.10. During the latter part of the nineteenth century, Prussian officials gathered information relating to the hazards that horses posed to cavalry soldiers. A total of ten cavalry corps were monitored over a period of twenty years. Recorded for each year and each corps was X, the annual number of fatalities due to kicks. Summarized in the following table are the two hundred values recorded for X (12). Show that these data can be modeled by a Poisson pdf. Follow the procedure illustrated in Case Studies 4.2.2 and 4.2.3.

No. of Deaths, k	Observed Number of Corps-Years in Which k Fatalities Occurred
0	109
1	65
2	22
3	3
4	1
	200

4.2.11. A random sample of 356 seniors enrolled at the University of West Florida was categorized according to X, the number of times they had changed majors (110). Based on the summary of that information shown in the following table, would you conclude that X can be treated as a Poisson random variable?

Number of Major Changes	Frequency
0	237
1	90
2	22
3	7

4.2.12. Midwestern Skies books ten commuter flights each week. Passenger totals are much the same from week to week, as are the numbers of pieces of luggage that are checked. Listed in the following table are the numbers of bags that were lost during each of the first forty weeks in 2009. Do these figures support the presumption that the number of bags lost by Midwestern during a typical week is a Poisson random variable?

Week	Bags Lost	Week	Bags Lost	Week	Bags Lost
1	1	14	2	27	1
2	0	15	1	28	2
3	0	16	3	29	0
4	3	17	0	30	0
5	4	18	2	31	1
6	1	19	5	32	3
7	0	20	2	33	1
8	2	21	1	34	2
9	0	22	1	35	0
10	2	23	1	36	1
11	3	24	2	37	4
12	1	25	1	38	2
13	2	26	3	39	1
				40	0

4.2.13. In 1893, New Zealand became the first country to permit women to vote. Scattered over the ensuing 113 years, various countries joined the movement to grant this right to women. The table below (121) shows how many countries took this step in a given year. Do these data seem to follow a Poisson distribution?

Yearly Number of Countries Granting Women the Vote	Frequency
0	82
1	25
2	4
3	0
4	2

4.2.14. The following are the daily numbers of death notices for women over the age of eighty that appeared in the *London Times* over a three-year period (74).

Number of Deaths	Observed Frequency
0	162
1	267
2	271
3	185
4	111
5	61
6	27
7	8
8	3
9	1
	1096

(a) Does the Poisson pdf provide a good description of the variability pattern evident in these data?

(b) If your answer to part (a) is "no," which of the Poisson model assumptions do you think might not be holding?

4.2.15. A certain species of European mite is capable of damaging the bark on orange trees. The following are the results of inspections done on one hundred saplings chosen at random from a large orchard. The measurement recorded, X, is the number of mite infestations found on the trunk of each tree. Is it reasonable to assume that X is a Poisson random variable? If not, which of the Poisson model assumptions is likely not to be true?

No. of Infestations, k	No. of Trees
0	55
1	20
2	21
3	1
4	1
5	1
6	0
7	1

4.2.16. A tool and die press that stamps out cams used in small gasoline engines tends to break down once every five hours. The machine can be repaired and put back on line quickly, but each such incident costs \$50. What is the probability that maintenance expenses for the press will be no more than \$100 on a typical eight-hour workday?

4.2.17. In a new fiber-optic communication system, transmission errors occur at the rate of 1.5 per ten seconds. What is the probability that more than two errors will occur during the next half-minute?

4.2.18. Assume that the number of hits, X, that a baseball team makes in a nine-inning game has a Poisson distribution. If the probability that a team makes zero hits is $\frac{1}{3}$, what are their chances of getting two or more hits?

4.2.19. Flaws in metal sheeting produced by a high-temperature roller occur at the rate of one per ten square feet. What is the probability that three or more flaws will appear in a five-by-eight-foot panel?

4.2.20. Suppose a radioactive source is metered for two hours, during which time the total number of alpha particles counted is 482. What is the probability that exactly three particles will be counted in the next two minutes? Answer the question two ways—first, by defining X to be the number of particles counted in two minutes, and second, by defining X to be the number of particles counted in one minute.

4.2.21. Suppose that on-the-job injuries in a textile mill occur at the rate of 0.1 per day.

(a) What is the probability that two accidents will occur during the next (five-day) workweek?
(b) Is the probability that four accidents will occur over the next two workweeks the square of your answer to part (a)? Explain.

4.2.22. Find $P(X = 4)$ if the random variable X has a Poisson distribution such that $P(X = 1) = P(X = 2)$.

4.2.23. Let X be a Poisson random variable with parameter λ. Show that the probability that X is even is $\frac{1}{2}(1 + e^{-2\lambda})$.

4.2.24. Let X and Y be independent Poisson random variables with parameters λ and μ, respectively. Example 3.12.10 established that $X + Y$ is also Poisson with parameter $\lambda + \mu$. Prove that same result using Theorem 3.8.3.

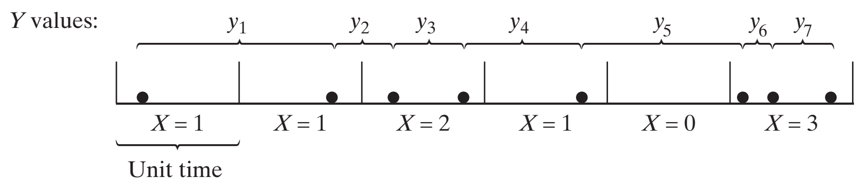
4.2.25. If X_1 is a Poisson random variable for which $E(X_1) = \lambda$ and if the conditional pdf of X_2 given that $X_1 = x_1$ is binomial with parameters x_1 and p, show that the marginal pdf of X_2 is Poisson with $E(X_2) = \lambda p$.

Intervals Between Events: The Poisson/Exponential Relationship

Situations sometimes arise where the time interval between consecutively occurring events is an important random variable. Imagine being responsible for the maintenance on a network of computers. Clearly, the number of technicians you would need to employ in order to be capable of responding to service calls in a timely fashion would be a function of the "waiting time" from one breakdown to another.

Figure 4.2.3 shows the relationship between the random variables X and Y, where X denotes the number of occurrences in a unit of time and Y denotes the interval between consecutive occurrences. Pictured are six intervals: $X = 0$ on one occasion, $X = 1$ on three occasions, $X = 2$ once, and $X = 3$ once. Resulting from those eight occurrences are seven measurements on the random variable Y. Obviously, the pdf for Y will depend on the pdf for X. One particularly important special case of that dependence is the Poisson/exponential relationship outlined in Theorem 4.2.3.

Figure 4.2.3



Theorem 4.2.3 *Suppose a series of events satisfying the Poisson model are occurring at the rate of λ per unit time. Let the random variable Y denote the interval between consecutive events. Then Y has the exponential distribution*

$$f_Y(y) = \lambda e^{-\lambda y}, \quad y > 0$$

they should come one month apart. But that is simply not the way random events behave, as Theorem 4.2.3 clearly shows.

Look at the entries in Table 4.2.5. The average of those thirty-six (randomly occurring) eruption separations was 37.7 months, yet seven of the separations were extremely short (less than or equal to six months). If two of those extremely short separations happened to occur consecutively, it would be tempting (but wrong) to conclude that the eruptions (since they came so close together) were "occurring in threes" for some supernatural reason.

Using the combinatorial techniques discussed in Section 2.6, we can calculate the probability that two extremely short intervals would occur consecutively. Think of the thirty-six intervals as being either "normal" or "extremely short." There are twenty-nine in the first group and seven in the second. Using the method described in Example 2.6.21, the probability that two extremely short separations would occur consecutively at least once is 61%, which hardly qualifies as a rare event:

$$P(\text{Two extremely short separations occur consecutively at least once})$$
$$= \frac{\binom{30}{6}\cdot\binom{6}{1}+\binom{30}{5}\cdot\binom{5}{2}+\binom{30}{4}\cdot\binom{4}{3}}{\binom{36}{29}} = 0.61$$

So, despite what our intuitions might tell us, the phenomenon of bad things coming in threes is neither mysterious nor uncommon or unexpected.

Example 4.2.5 Among the most famous of all meteor showers are the Perseids, which occur each year in early August. In some areas the frequency of visible Perseids can be as high as forty per hour. Given that such sightings are Poisson events, calculate the probability that an observer who has just seen a meteor will have to wait at least five minutes before seeing another one.

Let the random variable Y denote the interval (in minutes) between consecutive sightings. Expressed in the units of Y, the *forty-per-hour* rate of visible Perseids becomes *0.67 per minute*. A straightforward integration, then, shows that the probability is *0.035* that an observer will have to wait five minutes or more to see another meteor:

$$\begin{aligned} P(Y > 5) &= \int_5^\infty 0.67e^{-0.67y}\,dy \\ &= \int_{3.35}^\infty e^{-u}\,du \quad (\text{where } u = 0.67y) \\ &= -e^{-u}\Big|_{3.35}^{\infty} = e^{-3.35} \\ &= 0.035 \end{aligned}$$

■

Questions

4.2.26. Suppose that commercial airplane crashes in a certain country occur at the rate of 2.5 per year.

(a) Is it reasonable to assume that such crashes are Poisson events? Explain.
(b) What is the probability that four or more crashes will occur next year?
(c) What is the probability that the next two crashes will occur within three months of one another?

4.2.27. Records show that deaths occur at the rate of 0.1 per day among patients residing in a large nursing home. If someone dies today, what are the chances that a week or more will elapse before another death occurs?

4.2.28. Suppose that Y_1 and Y_2 are independent exponential random variables, each having pdf $f_Y(y) = \lambda e^{-\lambda y}$, $y > 0$. If $Y = Y_1 + Y_2$, it can be shown that

$$f_{Y_1+Y_2}(y) = \lambda^2 y e^{-\lambda y}, \quad y > 0$$

Recall Case Study 4.2.4. What is the probability that the next three eruptions of Mauna Loa will be less than forty months apart?

4.2.29. Fifty spotlights have just been installed in an outdoor security system. According to the manufacturer's specifications, these particular lights are expected to burn out at the rate of 1.1 per one hundred hours. What is the expected number of bulbs that will fail to last for at least seventy-five hours?

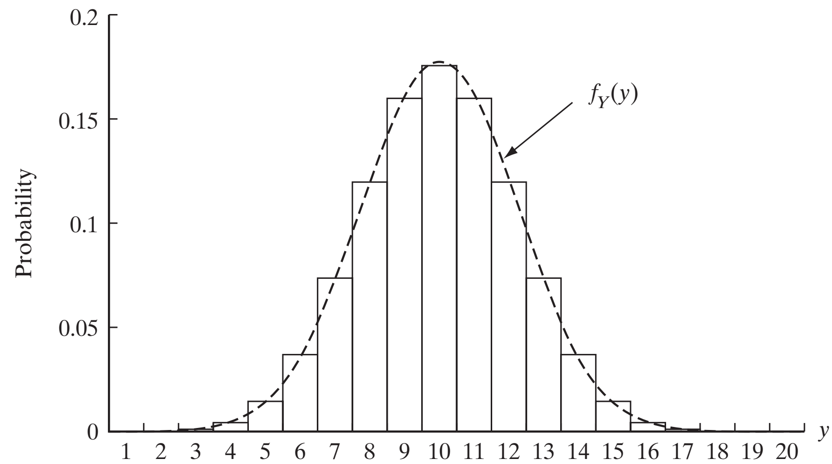
4.2.30. Suppose you want to invent a new superstition that "Bad things come in fours." Using the data given in Case Study 4.2.4 and the type of analysis described on p. 238, calculate the probability that your superstition would appear to be true.

4.3 The Normal Distribution

The Poisson limit described in Section 4.2 was not the only, or the first, approximation developed for the purpose of facilitating the calculation of binomial probabilities. Early in the eighteenth century, Abraham DeMoivre proved that areas under the curve $f_z(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$, $-\infty < z < \infty$, can be used to estimate $P\left[a \le \frac{X - n\left(\frac{1}{2}\right)}{\sqrt{n\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)}} \le b\right]$, where X is a binomial random variable with a large n and $p = \frac{1}{2}$.

Figure 4.3.1 illustrates the central idea in DeMoivre's discovery. Pictured is a probability histogram of the binomial distribution with $n = 20$ and $p = \frac{1}{2}$. Superimposed over the histogram is the function $f_Y(y) = \frac{1}{\sqrt{2\pi} \cdot \sqrt{5}} e^{-\frac{1}{2}\frac{(y-10)^2}{5}}$. Notice how closely the area under the curve approximates the area of the bar, even for this relatively small value of n. The French mathematician Pierre-Simon Laplace generalized DeMoivre's original idea to binomial approximations for arbitrary p and brought this theorem to the full attention of the mathematical community by including it in his influential 1812 book, *Theorie Analytique des Probabilities.*

Figure 4.3.1



Theorem 4.3.1 *Let X be a binomial random variable defined on n independent trials for which $p = P(\text{success})$. For any numbers a and b,*

$$\lim_{n\to\infty} P\left[a \le \frac{X - np}{\sqrt{np(1-p)}} \le b\right] = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-z^2/2}\, dz$$

About the Data This is a good set of data for illustrating why we need formal mathematical methods for interpreting data. As we have seen on other occasions, our intuitions, when left unsupported by probability calculations, can often be deceived. A typical first reaction to the Pratt-Woodruff results is to dismiss as inconsequential the 489 additional correct answers. To many, it seems entirely believable that sixty thousand guesses could produce, by chance, an extra 489 correct responses. Only after making the $P(X \geq 12{,}489)$ computation do we see the utter implausibility of that conclusion. What statistics is doing here is what we would like it to do in general—rule out hypotheses that are not supported by the data and point us in the direction of inferences that are more likely to be true.

Questions

4.3.1. Use Appendix Table A.1 to evaluate the following integrals. In each case, draw a diagram of $f_Z(z)$ and shade the area that corresponds to the integral.

(a) $\int_{-0.44}^{1.33} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\, dz$
(b) $\int_{-\infty}^{0.94} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\, dz$
(c) $\int_{-1.48}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\, dz$
(d) $\int_{-\infty}^{-4.32} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\, dz$

4.3.2. Let Z be a standard normal random variable. Use Appendix Table A.1 to find the numerical value for each of the following probabilities. Show each of your answers as an area under $f_Z(z)$.

(a) $P(0 \leq Z \leq 2.07)$
(b) $P(-0.64 \leq Z < -0.11)$
(c) $P(Z > -1.06)$
(d) $P(Z < -2.33)$
(e) $P(Z \geq 4.61)$

4.3.3.

(a) Let $0 < a < b$. Which number is larger?

$$\int_a^b \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\, dz \quad \text{or} \quad \int_{-b}^{-a} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\, dz$$

(b) Let $a > 0$. Which number is larger?

$$\int_a^{a+1} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\, dz \quad \text{or} \quad \int_{a-1/2}^{a+1/2} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\, dz$$

4.3.4.

(a) Evaluate $\int_0^{1.24} e^{-z^2/2}\, dz$.
(b) Evaluate $\int_{-\infty}^{\infty} 6e^{-z^2/2}\, dz$.

4.3.5. Assume that the random variable Z is described by a standard normal curve $f_Z(z)$. For what values of z are the following statements true?

(a) $P(Z \leq z) = 0.33$
(b) $P(Z \geq z) = 0.2236$
(c) $P(-1.00 \leq Z \leq z) = 0.5004$
(d) $P(-z < Z < z) = 0.80$
(e) $P(z \leq Z \leq 2.03) = 0.15$

4.3.6. Let z_α denote the value of Z for which $P(Z \geq z_\alpha) = \alpha$. By definition, the *interquartile range*, Q, for the standard normal curve is the difference

$$Q = z_{.25} - z_{.75}$$

Find Q.

4.3.7. Oak Hill has 74,806 registered automobiles. A city ordinance requires each to display a bumper decal showing that the owner paid an annual wheel tax of \$50. By law, new decals need to be purchased during the month of the owner's birthday. This year's budget assumes that at least \$306,000 in decal revenue will be collected in November. What is the probability that the wheel taxes reported in that month will be less than anticipated and produce a budget shortfall?

4.3.8. Hertz Brothers, a small, family-owned radio manufacturer, produces electronic components domestically but subcontracts the cabinets to a foreign supplier. Although inexpensive, the foreign supplier has a quality-control program that leaves much to be desired. On the average, only 80% of the standard 1600-unit shipment that Hertz receives is usable. Currently, Hertz has back orders for 1260 radios but storage space for no more than 1310 cabinets. What are the chances that the number of usable units in Hertz's latest shipment will be large enough to allow Hertz to fill all the orders already on hand, yet small enough to avoid causing any inventory problems?

4.3.9. Fifty-five percent of the registered voters in Sheridanville favor their incumbent mayor in her bid for re-election. If four hundred voters go to the polls, approximate the probability that

(a) the race ends in a tie.
(b) the challenger scores an upset victory.

4.3.10. State Tech's basketball team, the Fighting Logarithms, have a 70% foul-shooting percentage.

(a) Write a formula for the exact probability that out of their next one hundred free throws, they will make between seventy-five and eighty, inclusive.
(b) Approximate the probability asked for in part (a).

4.3.11. A random sample of 747 obituaries published recently in Salt Lake City newspapers revealed that 344 (or 46%) of the decedents died in the three-month period following their birthdays (123). Assess the statistical significance of that finding by approximating the probability that 46% or more would die in that particular interval if deaths occurred randomly throughout the year. What would you conclude on the basis of your answer?

4.3.12. There is a theory embraced by certain parapsychologists that hypnosis can enhance a person's ESP ability. To test that hypothesis, an experiment was set up with fifteen hypnotized subjects (21). Each was asked to make 100 guesses using the same sort of ESP cards and protocol that were described in Case Study 4.3.1. A total of 326 correct identifications were made. Can it be argued on the basis of those results that hypnosis does have an effect on a person's ESP ability? Explain.

4.3.13. If $p_X(k) = \binom{10}{k}(0.7)^k(0.3)^{10-k}$, $k = 0, 1, \ldots, 10$, is it appropriate to approximate $P(4 \leq X \leq 8)$ by computing the following?

$$P\left[\frac{3.5 - 10(0.7)}{\sqrt{10(0.7)(0.3)}} \leq Z \leq \frac{8.5 - 10(0.7)}{\sqrt{10(0.7)(0.3)}}\right]$$

Explain.

4.3.14. A sell-out crowd of 42,200 is expected at Cleveland's Jacobs Field for next Tuesday's game against the Baltimore Orioles, the last before a long road trip. The ballpark's concession manager is trying to decide how much food to have on hand. Looking at records from games played earlier in the season, she knows that, on the average, 38% of all those in attendance will buy a hot dog. How large an order should she place if she wants to have no more that a 20% chance of demand exceeding supply?

Central Limit Theorem

It was pointed out in Example 3.9.3 that every binomial random variable X can be written as the sum of n independent Bernoulli random variables $X_1, X_2, \ldots, X_n$, where

$$X_i = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability } 1-p \end{cases}$$

But if $X = X_1 + X_2 + \cdots + X_n$, Theorem 4.3.1 can be reexpressed as

$$\lim_{n\to\infty} P\left[a \leq \frac{X_1 + X_2 + \cdots + X_n - np}{\sqrt{np(1-p)}} \leq b\right] = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-z^2/2}\, dz \qquad (4.3.2)$$

Implicit in Equation 4.3.2 is an obvious question: Does the DeMoivre-Laplace limit apply to sums of other types of random variables as well? Remarkably, the answer is "yes." Efforts to extend Equation 4.3.2 have continued for more than 150 years. Russian probabilists—A. M. Lyapunov, in particular—made many of the key advances. In 1920, George Polya gave these new generalizations a name that has been associated with the result ever since: He called it the *central limit theorem* (136).

Theorem 4.3.2 *(Central Limit Theorem) Let $W_1, W_2, \ldots$ be an infinite sequence of independent random variables, each with the same distribution. Suppose that the mean μ and the variance σ^2 of $f_W(w)$ are both finite. For any numbers a and b,* and

$$\text{Var}(Y_i) = E\left(Y_i^2\right) = \int_{-50}^{50} \frac{1}{100} y^2 \, dy$$
$$= \frac{2500}{3}$$

Therefore,

$$E(S_{100}) = E(Y_1 + Y_2 + \cdots + Y_{100}) = 0$$

and

$$\text{Var}(S_{100}) = \text{Var}(Y_1 + Y_2 + \cdots + Y_{100}) = 100\left(\frac{2500}{3}\right)$$
$$= \frac{250{,}000}{3}$$

Applying Theorem 4.3.2, then, shows that her strategy has roughly an *8%* chance of being in error by more than \$500:

$$P(|S_{100}| > \$500) = 1 - P(-500 \leq S_{100} \leq 500)$$
$$= 1 - P\left(\frac{-500-0}{500/\sqrt{3}} \leq \frac{S_{100}-0}{500/\sqrt{3}} \leq \frac{500-0}{500/\sqrt{3}}\right)$$
$$= 1 - P(-1.73 < Z < 1.73)$$
$$= 0.0836$$

■

Questions

4.3.15. A fair coin is tossed two hundred times. Let $X_i = 1$ if the ith toss comes up heads and $X_i = 0$ otherwise, $i = 1, 2, \ldots, 200$; $X = \sum_{i=1}^{200} X_i$. Calculate the central limit theorem approximation for $P(|X - E(X)| \leq 5)$. How does this differ from the DeMoivre-Laplace approximation?

4.3.16. Suppose that one hundred fair dice are tossed. Estimate the probability that the sum of the faces showing exceeds 370. Include a continuity correction in your analysis.

4.3.17. Let X be the amount won or lost in betting \$5 on red in roulette. Then $p_x(5) = \frac{18}{38}$ and $p_x(-5) = \frac{20}{38}$. If a gambler bets on red one hundred times, use the central limit theorem to estimate the probability that those wagers result in less than \$50 in losses.

4.3.18. If $X_1, X_2, \ldots, X_n$ are independent Poisson random variables with parameters $\lambda_1, \lambda_2, \ldots, \lambda_n$, respectively, and if $X = X_1 + X_2 + \cdots + X_n$, then X is a Poisson random variable with parameter $\lambda = \sum_{i=1}^{n} \lambda_i$ (recall Example 3.12.10). What specific form does the ratio in Theorem 4.3.2 take if the X_i's are Poisson random variables?

4.3.19. An electronics firm receives, on the average, fifty orders per week for a particular silicon chip. If the company has sixty chips on hand, use the central limit theorem to approximate the probability that they will be unable to fill all their orders for the upcoming week. Assume that weekly demands follow a Poisson distribution. (*Hint:* See Question 4.3.18.)

4.3.20. Considerable controversy has arisen over the possible aftereffects of a nuclear weapons test conducted in Nevada in 1957. Included as part of the test were some three thousand military and civilian "observers." Now, more than fifty years later, eight cases of leukemia have been diagnosed among those three thousand. The expected number of cases, based on the demographic characteristics of the observers, was three. Assess the statistical significance of those findings. Calculate both an exact answer using the Poisson distribution as well as an approximation based on the central limit theorem.

Corollary 4.3.2 *Let $Y_1, Y_2, \ldots, Y_n$ be any set of independent normal random variables with means $\mu_1, \mu_2, \ldots, \mu_n$ and variances $\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2$, respectively. Let $a_1, a_2, \ldots, a_n$ be any set of constants. Then $Y = a_1Y_1 + a_2Y_2 + \cdots + a_nY_n$ is normally distributed with mean $\mu = \sum_{i=1}^{n} a_i\mu_i$ and variance $\sigma^2 = \sum_{i=1}^{n} a_i^2\sigma_i^2$.* ◀

Example 4.3.8 The elevator in the athletic dorm at Swampwater Tech has a maximum capacity of twenty-four hundred pounds. Suppose that ten football players get on at the twentieth floor. If the weights of Tech's players are normally distributed with a mean of two hundred twenty pounds and a standard deviation of twenty pounds, what is the probability that there will be ten fewer Muskrats at tomorrow's practice?

Let the random variables $Y_1, Y_2, \ldots, Y_{10}$ denote the weights of the ten players. At issue is the probability that $Y = \sum_{i=1}^{10} Y_i$ exceeds twenty-four hundred pounds. But

$$P\left(\sum_{i=1}^{10} Y_i > 2400\right) = P\left(\frac{1}{10}\sum_{i=1}^{10} Y_i > \frac{1}{10}\cdot 2400\right) = P(\bar{Y} > 240.0)$$

A Z transformation can be applied to the latter expression using the corollary on p. 257:

$$P(\bar{Y} > 240.0) = P\left(\frac{\bar{Y} - 220}{20/\sqrt{10}} > \frac{240.0 - 220}{20/\sqrt{10}}\right) = P(Z > 3.16)$$
$$= 0.0008$$

Clearly, the chances of a Muskrat splat are minimal. (How much would the probability change if eleven players squeezed onto the elevator?) ■

Questions

4.3.21. Econo-Tire is planning an advertising campaign for its newest product, an inexpensive radial. Preliminary road tests conducted by the firm's quality-control department have suggested that the lifetimes of these tires will be normally distributed with an average of thirty thousand miles and a standard deviation of five thousand miles. The marketing division would like to run a commercial that makes the claim that at least nine out of ten drivers will get at least twenty-five thousand miles on a set of Econo-Tires. Based on the road test data, is the company justified in making that assertion?

4.3.22. A large computer chip manufacturing plant under construction in Westbank is expected to result in an additional fourteen hundred children in the county's public school system once the permament workforce arrives. Any child with an IQ under 80 or over 135 will require individualized instruction that will cost the city an additional \$1750 per year. How much money should Westbank anticipate spending next year to meet the needs of its new special ed students? Assume that IQ scores are normally distributed with a mean (μ) of 100 and a standard deviation (σ) of 16.

4.3.23. Records for the past several years show that the amount of money collected daily by a prominent televangelist is normally distributed with a mean (μ) of \$20,000 and a standard deviation (σ) of \$5000. What are the chances that tomorrow's donations will exceed \$30,000?

4.3.24. The following letter was written to a well-known dispenser of advice to the lovelorn (171):

> Dear Abby: You wrote in your column that a woman is pregnant for 266 days. Who said so? I carried my baby for ten months and five days, and there is no doubt about it because I know the exact date my baby was conceived. My husband is in the Navy and it couldn't have possibly been conceived any other time because I saw him only once for an hour, and I didn't see him again until the day before the baby was born.
>
> I don't drink or run around, and there is no way this baby isn't his, so please print a retraction about the 266-day carrying time because otherwise I am in a lot of trouble.
>
> San Diego Reader

Whether or not San Diego Reader is telling the truth is a judgment that lies beyond the scope of any statistical analysis, but quantifying the plausibility of her story does not. According to the collective experience of generations of pediatricians, pregnancy durations, Y, tend to be normally distributed with $\mu = 266$ days and $\sigma = 16$ days. Do a probability calculation that addresses San Diego Reader's credibility. What would you conclude?

4.3.25. A criminologist has developed a questionnaire for predicting whether a teenager will become a delinquent. Scores on the questionnaire can range from 0 to 100, with higher values reflecting a presumably greater criminal tendency. As a rule of thumb, the criminologist decides to classify a teenager as a potential delinquent if his or her score exceeds 75. The questionnaire has already been tested on a large sample of teenagers, both delinquent and nondelinquent. Among those considered nondelinquent, scores were normally distributed with a mean (μ) of 60 and a standard deviation (σ) of 10. Among those considered delinquent, scores were normally distributed with a mean of 80 and a standard deviation of 5.

- **(a)** What proportion of the time will the criminologist misclassify a nondelinquent as a delinquent? A delinquent as a nondelinquent?
- **(b)** On the same set of axes, draw the normal curves that represent the distributions of scores made by delinquents and nondelinquents. Shade the two areas that correspond to the probabilities asked for in part (a).

4.3.26. The cross-sectional area of plastic tubing for use in pulmonary resuscitators is normally distributed with $\mu = 12.5$ mm^2 and $\sigma = 0.2$ mm^2. When the area is less than 12.0 mm^2 or greater than 13.0 mm^2, the tube does not fit properly. If the tubes are shipped in boxes of one thousand, how many wrong-sized tubes per box can doctors expect to find?

4.3.27. At State University, the average score of the entering class on the verbal portion of the SAT is 565, with a standard deviation of 75. Marian scored a 660. How many of State's other 4250 freshmen did better? Assume that the scores are normally distributed.

4.3.28. A college professor teaches Chemistry 101 each fall to a large class of freshmen. For tests, she uses standardized exams that she knows from past experience produce bell-shaped grade distributions with a mean of 70 and a standard deviation of 12. Her philosophy of grading is to impose standards that will yield, in the long run, 20% A's, 26% B's, 38% C's, 12% D's, and 4% F's. Where should the cutoff be between the A's and the B's? Between the B's and the C's?

4.3.29. Suppose the random variable Y can be described by a normal curve with $\mu = 40$. For what value of σ is

$$P(20 \leq Y \leq 60) = 0.50$$

4.3.30. It is estimated that 80% of all eighteen-year-old women have weights ranging from 103.5 to 144.5 lb. Assuming the weight distribution can be adequately modeled by a normal curve and that 103.5 and 144.5 are equidistant from the average weight μ, calculate σ.

4.3.31. Recall the breath analyzer problem described in Example 4.3.5. Suppose the driver's blood alcohol concentration is actually 0.09% rather than 0.075%. What is the probability that the breath analyzer will make an error in his favor and indicate that he is *not* legally drunk? Suppose the police offer the driver a choice—either take the sobriety test once or take it twice and average the readings. Which option should a "0.075%" driver take? Which option should a "0.09%" driver take? Explain.

4.3.32. If a random variable Y is normally distributed with mean μ and standard deviation σ, the Z ratio $\frac{Y-\mu}{\sigma}$ is often referred to as a *normed* score: It indicates the magnitude of y *relative to the distribution from which it came.* "Norming" is sometimes used as an affirmative-action mechanism in hiring decisions. Suppose a cosmetics company is seeking a new sales manager. The aptitude test they have traditionally given for that position shows a distinct gender bias: Scores for men are normally distributed with $\mu = 62.0$ and $\sigma = 7.6$, while scores for women are normally distributed with $\mu = 76.3$ and $\sigma = 10.8$. Laura and Michael are the two candidates vying for the position: Laura has scored 92 on the test and Michael 75. If the company agrees to norm the scores for gender bias, whom should they hire?

4.3.33. The IQs of nine randomly selected people are recorded. Let $\overline{Y}$ denote their average. Assuming the distribution from which the Y_i's were drawn is normal with a mean of 100 and a standard deviation of 16, what is the probability that $\overline{Y}$ will exceed 103? What is the probability that any arbitrary Y_i will exceed 103? What is the probability that exactly three of the Y_i's will exceed 103?

4.3.34. Let $Y_1, Y_2, \ldots, Y_n$ be a random sample from a normal distribution where the mean is 2 and the variance is 4. How large must n be in order that

$$P(1.9 \leq \overline{Y} \leq 2.1) \geq 0.99$$

4.3.35. A circuit contains three resistors wired in series. Each is rated at 6 ohms. Suppose, however, that the true resistance of each one is a normally distributed random variable with a mean of 6 ohms and a standard deviation of 0.3 ohm. What is the probability that the combined resistance will exceed 19 ohms? How "precise" would the manufacturing process have to be to make the probability less than 0.005 that the combined resistance of the circuit would exceed 19 ohms?

4.3.36. The cylinders and pistons for a certain internal combustion engine are manufactured by a process that gives a normal distribution of cylinder diameters with a mean of 41.5 cm and a standard deviation of 0.4 cm. Similarly, the distribution of piston diameters is normal with a mean of 40.5 cm and a standard deviation of 0.3 cm. If the piston diameter is greater than the cylinder diameter, the former can be reworked until the two "fit." What proportion of cylinder-piston pairs will need to be reworked?

4.3.37. Use moment-generating functions to prove the two corollaries to Theorem 4.3.3.

4.3.38. Let $Y_1, Y_2, \ldots, Y_9$ be a random sample of size 9 from a normal distribution where $\mu = 2$ and $\sigma = 2$. Let $Y_1^*, Y_2^*, \ldots, Y_9^*$ be an independent random sample from a normal distribution having $\mu = 1$ and $\sigma = 1$. Find $P(\bar{Y} \geq \bar{Y}^*)$.

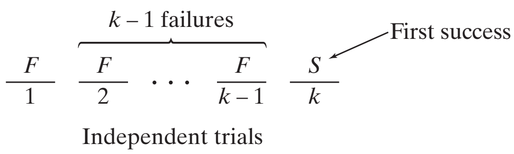
4.4 The Geometric Distribution

Consider a series of independent trials, each having one of two possible outcomes, success or failure. Let $p = P(\text{Trial ends in success})$. Define the random variable X to be the trial *at which the first success occurs*. Figure 4.4.1 suggests a formula for the pdf of X:

$$\begin{aligned} p_X(k) = P(X=k) &= P(\text{First success occurs on } k\text{th trial}) \\ &= P(\text{First } k-1 \text{ trials end in failure and } k\text{th trial ends in success}) \\ &= P(\text{First } k-1 \text{ trials end in failure}) \cdot P(k\text{th trial ends in success}) \\ &= (1-p)^{k-1}p, \quad k = 1, 2, \ldots \end{aligned} \tag{4.4.1}$$

We call the probability model in Equation 4.4.1 a *geometric distribution* (with parameter p).

Figure 4.4.1



Comment Even without its association with independent trials and Figure 4.4.1, the function $p_X(k) = (1-p)^{k-1}p$, $k = 1, 2, \ldots$ qualifies as a discrete pdf because (1) $p_X(k) \geq 0$ for all k and (2) $\sum_{\text{all } k} p_X(k) = 1$:

$$\begin{aligned} \sum_{k=1}^{\infty} (1-p)^{k-1} p = p \sum_{j=0}^{\infty} (1-p)^j & \\ &= p \cdot \left[\frac{1}{1-(1-p)} \right] \\ &= 1 \end{aligned}$$

Example 4.4.1 A pair of fair dice are tossed until a sum of 7 appears for the first time. What is the probability that more than four rolls will be required for that to happen?

Each throw of the dice here is an independent trial for which

$$p = P(\text{sum} = 7) = \frac{6}{36} = \frac{1}{6}$$

and $E(X_3) = \frac{6}{4}$. Continuing in this fashion, we can find the remaining $E(X_i)$'s. It follows that a customer will have to make *14.7* trips to the store, on the average, to collect a complete set of six letters:

$$\begin{aligned} E(X) &= \sum_{i=1}^{6} E(X_i) \\ &= 1 + \frac{6}{5} + \frac{6}{4} + \frac{6}{3} + \frac{6}{2} + \frac{6}{1} \\ &= 14.7 \end{aligned}$$

■

Questions

4.4.1. Because of her past convictions for mail fraud and forgery, Jody has a 30% chance each year of having her tax returns audited. What is the probability that she will escape detection for at least three years? Assume that she exaggerates, distorts, misrepresents, lies, and cheats every year.

4.4.2. A teenager is trying to get a driver's license. Write out the formula for the pdf $p_x(k)$, where the random variable X is the number of tries that he needs to pass the road test. Assume that his probability of passing the exam on any given attempt is 0.10. On the average, how many attempts is he likely to require before he gets his license?

4.4.3. Is the following set of data likely to have come from the geometric pdf $p_X(k) = \left(\frac{3}{4}\right)^{k-1} \cdot \left(\frac{1}{4}\right)$, $k = 1, 2, \ldots$? Explain.

2	8	1	2	2	5	1	2	8	3
5	4	2	4	7	2	2	8	4	7
2	6	2	3	5	1	3	3	2	5
4	2	2	3	6	3	6	4	9	3
3	7	5	1	3	4	3	4	6	2

4.4.4. Recently married, a young couple plans to continue having children until they have their first girl. Suppose the probability that a child is a girl is $\frac{1}{2}$, the outcome of each birth is an independent event, and the birth at which the first girl appears has a geometric distribution. What is the couple's expected family size? Is the geometric pdf a reasonable model here? Discuss.

4.4.5. Show that the cdf for a geometric random variable is given by $F_X(t) = P(X \leq t) = 1 - (1 - p)^{[t]}$, where $[t]$ denotes the greatest integer in t, $t \geq 0$.

4.4.6. Suppose three fair dice are tossed repeatedly. Let the random variable X denote the roll on which a sum of 4 appears for the first time. Use the expression for $F_x(t)$ given in Question 4.4.5 to evaluate $P(65 \leq X \leq 75)$.

4.4.7. Let Y be an exponential random variable, where $f_Y(y) = \lambda e^{-\lambda y}$, $0 \leq y$. For any positive integer n, show that $P(n \leq Y \leq n+1) = e^{-\lambda n}(1 - e^{-\lambda})$. Note that if $p = 1 - e^{-\lambda}$, the "discrete" version of the exponential pdf is the geometric pdf.

4.4.8. Sometimes the geometric random variable is defined to be the number of trials, *X, preceding* the first success. Write down the corresponding pdf and derive the moment-generating function for X two ways—(1) by evaluating $E(e^{tX})$ directly and (2) by using Theorem 3.12.3.

4.4.9. Differentiate the moment-generating function for a geometric random variable and verify the expressions given for $E(X)$ and $\text{Var}(X)$ in Theorem 4.4.1.

4.4.10. Suppose that the random variables X_1 and X_2 have mgfs $M_{X_1}(t) = \frac{\frac{1}{2}e^t}{1-\left(1-\frac{1}{2}\right)e^t}$ and $M_{X_2}(t) = \frac{\frac{1}{4}e^t}{1-\left(1-\frac{1}{4}t\right)e^t}$, respectively. Let $X = X_1 + X_2$. Does X have a geometric distribution? Assume that X_1 and X_2 are independent.

4.4.11. The *factorial moment-generating function* for any random variable W is the expected value of t^w. Moreover $\frac{d^r}{dt^r}E(t^W)\,|_{t=1} = E[W(W-1)\cdots(W-r+1)]$. Find the factorial moment-generating function for a geometric random variable and use it to verify the expected value and variance formulas given in Theorem 4.4.1.

4.5 The Negative Binomial Distribution

The geometric distribution introduced in Section 4.4 can be generalized in a very straightforward fashion. Imagine waiting for the rth (instead of the first) success in a series of independent trials, where each trial has a probability of p of ending in success (see Figure 4.5.1).

is a set of uniform random variables as defined earlier, then $Y_i = -(1/\lambda)\ln U_i$, $i = 1, 2, \ldots$, will be the desired set of exponential observations. Why that should be so is an exercise in differentiating the cdf of Y. By definition,

$$F_Y(y) = P(Y \le y) = P(\ln U > -\lambda y) = P(U > e^{-\lambda y})$$
$$= \int_{e^{-\lambda y}}^{1} 1\, du = 1 - e^{-\lambda y}$$

which implies that

$$f_Y(y) = F_Y'(y) = \lambda e^{-\lambda y},\ y \ge 0$$

Questions

4.5.1. A door-to-door encyclopedia salesperson is required to document five in-home visits each day. Suppose that she has a 30% chance of being invited into any given home, with each address representing an independent trial. What is the probability that she requires fewer than eight houses to achieve her fifth success?

4.5.2. An underground military installation is fortified to the extent that it can withstand up to three direct hits from air-to-surface missiles and still function. Suppose an enemy aircraft is armed with missiles, each having a 30% chance of scoring a direct hit. What is the probability that the installation will be destroyed with the seventh missile fired?

4.5.3. Darryl's statistics homework last night was to flip a fair coin and record the toss, X, when heads appeared for the second time. The experiment was to be repeated a total of one hundred times. The following are the one hundred values for X that Darryl turned in this morning. Do you think that he actually did the assignment? Explain.

3	7	3	2	9	3	4	3	3	2
7	3	8	4	3	3	3	4	3	3
4	3	2	2	4	5	2	2	2	4
2	5	6	4	2	6	2	8	3	2
8	2	3	2	4	3	2	6	3	3
3	2	5	3	6	4	5	6	5	6
3	5	2	7	2	10	4	3	2	2
4	2	4	5	5	5	6	2	4	3
3	4	4	6	3	4	2	5	5	2
5	7	5	3	2	7	4	4	4	3

4.5.4. When a machine is improperly adjusted, it has probability 0.15 of producing a defective item. Each day, the machine is run until three defective items are produced. When this occurs, it is stopped and checked for adjustment. What is the probability that an improperly adjusted machine will produce five or more items before being stopped? What is the average number of items an improperly adjusted machine will produce before being stopped?

4.5.5. For a negative binomial random variable whose pdf is given by Equation 4.5.1, find $E(X)$ directly by evaluating $\sum_{k=r}^{\infty} k \binom{k-1}{r-1} p^r (1-p)^{k-r}$. (*Hint:* Reduce the sum to one involving negative binomial probabilities with parameters $r+1$ and p.)

4.5.6. Let the random variable X denote the number of trials in *excess of* r that are required to achieve the rth success in a series of independent trials, where p is the probability of success at any given trial. Show that

$$p_X(k) = \binom{k+r-1}{k} p^r (1-p)^k, \quad k = 0, 1, 2, \ldots$$

[*Note:* This particular formula for $p_X(k)$ is often used in place of Equation 4.5.1 as the definition of the pdf for a negative binomial random variable.]

4.5.7. Calculate the mean, variance, and moment-generating function for a negative binomial random variable X whose pdf is given by the expression

$$p_X(k) = \binom{k+r-1}{k} p^r (1-p)^k, \quad k = 0, 1, 2, \ldots$$

(See Question 4.5.6.)

4.5.8. Let X_1, X_2, and X_3 be three independent negative binomial random variables with pdfs

$$p_{X_i}(k) = \binom{k-1}{2} \left(\frac{4}{5}\right)^3 \left(\frac{1}{5}\right)^{k-3}, \quad k = 3, 4, 5, \ldots$$

for $i = 1, 2, 3$. Define $X = X_1 + X_2 + X_3$. Find $P(10 \le X \le 12)$. (*Hint:* Use the moment-generating functions of X_1, X_2, and X_3 to deduce the pdf of X.)

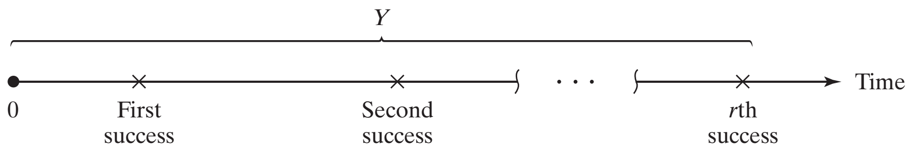
4.5.9. Differentiate the moment-generating function $M_X(t) = \left[\frac{pe^t}{1-(1-p)e^t}\right]^r$ to verify the formula given in Theorem 4.5.1 for $E(X)$.

4.5.10. Suppose that $X_1, X_2, \ldots, X_k$ are independent negative binomial random variables with parameters r_1 and p, r_2 and $p, \ldots$, and r_k and p, respectively. Let $X = X_1 + X_2 + \cdots + X_k$. Find $M_X(t)$, $p_X(t)$, $E(X)$, and $\text{Var}(X)$.

4.6 The Gamma Distribution

Suppose a series of independent events are occurring at the constant rate of λ per unit time. If the random variable Y denotes the interval between consecutive occurrences, we know from Theorem 4.2.3 that $f_Y(y) = \lambda e^{-\lambda y}$, $y > 0$. Equivalently, Y can be interpreted as the "waiting time" for the first occurrence. This section generalizes the Poisson/exponential relationship and focuses on the interval, or waiting time, required for the rth event to occur (see Figure 4.6.1).

Figure 4.6.1



Theorem 4.6.1 *Suppose that Poisson events are occurring at the constant rate of λ per unit time. Let the random variable Y denote the waiting time for the rth event. Then Y has pdf $f_Y(y)$, where*

$$f_Y(y) = \frac{\lambda^r}{(r-1)!} y^{r-1} e^{-\lambda y}, \quad y > 0$$

Proof We will establish the formula for $f_Y(y)$ by deriving and differentiating its cdf, $F_Y(y)$. Let Y denote the waiting time to the rth occurrence. Then

$$\begin{aligned} F_Y(y) = P(Y \le y) = 1 - P(Y > y) \\ &= 1 - P(\text{Fewer than } r \text{ events occur in } [0, y]) \\ &= 1 - \sum_{k=0}^{r-1} e^{-\lambda y} \frac{(\lambda y)^k}{k!} \end{aligned}$$

since the number of events that occur in the interval $[0, y]$ is a Poisson random variable with parameter λy.

From Theorem 3.4.1,

$$\begin{aligned} f_Y(y) = F'_Y(y) &= \frac{d}{dy}\left[1 - \sum_{k=0}^{r-1} e^{-\lambda y} \frac{(\lambda y)^k}{k!}\right] \\ &= \sum_{k=0}^{r-1} \lambda e^{-\lambda y} \frac{(\lambda y)^k}{k!} - \sum_{k=1}^{r-1} \lambda e^{-\lambda y} \frac{(\lambda y)^{k-1}}{(k-1)!} \\ &= \sum_{k=0}^{r-1} \lambda e^{-\lambda y} \frac{(\lambda y)^k}{k!} - \sum_{k=0}^{r-2} \lambda e^{-\lambda y} \frac{(\lambda y)^k}{k!} \\ &= \frac{\lambda^r}{(r-1)!} y^{r-1} e^{-\lambda y}, \quad y > 0 \end{aligned}$$

$\square$

Questions

4.6.1. An Arctic weather station has three electronic wind gauges. Only one is used at any given time. The lifetime of each gauge is exponentially distributed with a mean of one thousand hours. What is the pdf of Y, the random variable measuring the time until the last gauge wears out?

4.6.2. A service contact on a new university computer system provides twenty-four free repair calls from a technician. Suppose the technician is required, on the average, three times a month. What is the average time it will take for the service contract to be fulfilled?

4.6.3. Suppose a set of measurements $Y_1, Y_2, \ldots, Y_{100}$ is taken from a gamma pdf for which $E(Y) = 1.5$ and $\text{Var}(Y) = 0.75$. How many Y_i's would you expect to find in the interval [1.0, 2.5]?

4.6.4. Demonstrate that λ plays the role of a scale parameter by showing that if Y is gamma with parameters r and λ, then λY is gamma with parameters r and 1.

4.6.5. Show that a gamma pdf has the unique mode $\frac{r-1}{\lambda}$; that is, show that the function $f_Y(y) = \frac{\lambda^r}{\Gamma(r)} y^{r-1} e^{-\lambda y}$ takes its maximum value at $y_{\text{mode}} = \frac{r-1}{\lambda}$ and at no other point.

4.6.6. Prove that $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$. [*Hint:* Consider $E(Z^2)$, where Z is a standard normal random variable.]

4.6.7. Show that $\Gamma\left(\frac{7}{2}\right) = \frac{15}{8}\sqrt{\pi}$.

4.6.8. If the random variable Y has the gamma pdf with integer parameter r and arbitrary $\lambda > 0$, show that

$$E(Y^m) = \frac{(m+r-1)!}{(r-1)!\lambda^m}$$

[*Hint:* Use the fact that $\int_0^\infty y^{r-1} e^{-y}\, dy = (r-1)!$ when r is a positive integer.]

4.6.9. Differentiate the gamma moment-generating function to verify the formulas for $E(Y)$ and $\text{Var}(Y)$ given in Theorem 4.6.3.

4.6.10. Differentiate the gamma moment-generating function to show that the formula for $E(Y^m)$ given in Question 4.6.8 holds for arbitrary $r > 0$.

4.7 Taking a Second Look at Statistics (Monte Carlo Simulations)

Calculating probabilities associated with (1) single random variables and (2) functions of sets of random variables has been the overarching theme of Chapters 3 and 4. Facilitating those computations has been a variety of transformations, summation properties, and mathematical relationships linking one pdf with another. Collectively, these results are enormously effective. Sometimes, though, the intrinsic complexity of a random variable overwhelms our ability to model its probabilistic behavior in any formal or precise way. An alternative in those situations is to use a computer to draw random samples from one or more distributions that model portions of the random variable's behavior. If a large enough number of such samples is generated, a histogram (or density-scaled histogram) can be constructed that will accurately reflect the random variable's true (but unknown) distribution. Sampling "experiments" of this sort are known as *Monte Carlo studies*.

Real-life situations where a Monte Carlo analysis could be helpful are not difficult to imagine. Suppose, for instance, you just bought a state-of-the-art, high-definition, plasma screen television. In addition to the pricey initial cost, an optional warranty is available that covers all repairs made during the first two years. According to an independent laboratory's reliability study, this particular television is likely to require 0.75 service call per year, on the average. Moreover, the costs of service calls are expected to be normally distributed with a mean (μ) of \$100 and a standard deviation (σ) of \$20. If the warranty sells for \$200, should you buy it?

Moreover,

$$\frac{\partial \ln L(\mu, \sigma^2)}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=1}^{n} (y_i - \mu)$$

and

$$\frac{\partial \ln L(\mu, \sigma^2)}{\partial \sigma^2} = -\frac{n}{2} \cdot \frac{1}{\sigma^2} + \frac{1}{2}\left(\frac{1}{\sigma^2}\right)^2 \sum_{i=1}^{n} (y_i - \mu)^2$$

Setting the two derivatives equal to zero gives the equations

$$\sum_{i=1}^{n} (y_i - \mu) = 0 \tag{5.2.1}$$

and

$$-n\sigma^2 + \sum_{i=1}^{n} (y_i - \mu)^2 = 0 \tag{5.2.2}$$

Equation 5.2.1 simplifies to

$$\sum_{i=1}^{n} y_i = n\mu$$

which implies that $\mu_e = \frac{1}{n} \sum_{i=1}^{n} y_i = \overline{y}$. Substituting μ_e, then, into Equation 5.2.2 gives

$$-n\sigma^2 + \sum_{i=1}^{n} (y_i - \overline{y})^2 = 0$$

or

$$\sigma_e^2 = \frac{1}{n} \sum_{i=1}^{n} (y_i - \overline{y})^2$$

Comment The method of maximum likelihood has a long history: Daniel Bernoulli was using it as early as 1777 (130). It was Ronald Fisher, though, in the early years of the twentieth century, who first studied the mathematical properties of likelihood estimation in any detail, and the procedure is often credited to him. ■

Questions

5.2.1. A random sample of size 8—$X_1 = 1$, $X_2 = 0$, $X_3 = 1$, $X_4 = 1$, $X_5 = 0$, $X_6 = 1$, $X_7 = 1$, and $X_8 = 0$—is taken from the probability function

$$p_X(k; \theta) = \theta^k (1 - \theta)^{1-k}, \quad k = 0, 1; \quad 0 < \theta < 1$$

Find the maximum likelihood estimate for θ.

5.2.2. The number of red chips and white chips in an urn is unknown, but the *proportion*, p, of reds is either $\frac{1}{3}$ or $\frac{1}{2}$. A sample of size 5, drawn with replacement, yields the sequence red, white, white, red, and white. What is the maximum likelihood estimate for p?

5.2.3. Use the sample $Y_1 = 8.2$, $Y_2 = 9.1$, $Y_3 = 10.6$, and $Y_4 = 4.9$ to calculate the maximum likelihood estimate for λ in the exponential pdf

$$f_Y(y; \lambda) = \lambda e^{-\lambda y}, \quad y \geq 0$$

5.2.4. Suppose a random sample of size n is drawn from the probability model

$$p_X(k; \theta) = \frac{\theta^{2k} e^{-\theta^2}}{k!}, \quad k = 0, 1, 2, \ldots$$

Find a formula for the maximum likelihood estimator, $\hat{\theta}$.

5.2.5. Given that $Y_1 = 2.3, Y_2 = 1.9,$ and $Y_3 = 4.6$ is a random sample from

$$f_Y(y;\theta) = \frac{y^3 e^{-y/\theta}}{6\theta^4}, \quad y \geq 0$$

calculate the maximum likelihood estimate for θ.

5.2.6. Use the method of maximum likelihood to estimate θ in the pdf

$$f_Y(y;\theta) = \frac{\theta}{2\sqrt{y}} e^{-\theta\sqrt{y}}, \quad y \geq 0$$

Evaluate θ_e for the following random sample of size 4: $Y_1 = 6.2$, $Y_2 = 7.0$, $Y_3 = 2.5$, and $Y_4 = 4.2$.

5.2.7. An engineer is creating a project scheduling program and recognizes that the tasks making up the project are not always completed on time. However, the completion proportion tends to be fairly high. To reflect this condition, he uses the pdf

$$f_Y(y;\theta) = \theta y^{\theta-1}, \quad 0 \leq y \leq 1, \quad \text{and} \quad 0 < \theta$$

where y is the proportion of the task completed. Suppose that in his previous project, the proportions of tasks completed were 0.77, 0.82, 0.92, 0.94, and 0.98. Estimate θ.

5.2.8. The following data show the number of occupants in passenger cars observed during one hour at a busy intersection in Los Angeles (69). Suppose it can be assumed that these data follow a geometric distribution, $p_X(k;p) = (1-p)^{k-1}p, k = 1, 2, \ldots$. Estimate p and compare the observed and expected frequencies for each value of X.

Number of Occupants	Frequency
1	678
2	227
3	56
4	28
5	8
6+	14
	1011

5.2.9. For the Major League Baseball seasons from 1950 through 2008, there were fifty-nine nine-inning games in which one of the teams did not manage to get a hit. The data in the table give the number of no-hitters *per season* over this period. Assume that the data follow a Poisson distribution,

$$p_X(k;\ \lambda) = e^{-\lambda}\frac{\lambda^k}{k!}, k = 0, 1, 2, \ldots$$

(a) Estimate λ and compare the observed and expected frequencies.
(b) Does the agreement (or lack of agreement) in part (a) come as a surprise? Explain.

No. of No-Hitters	Frequency
0	6
1	19
2	12
3	13
4+	9

Source: en.wikipedia.org/wiki/List_of_Major_League_Baseball_no-hitlers.

5.2.10. (a) Based on the random sample $Y_1 = 6.3$, $Y_2 = 1.8$, $Y_3 = 14.2$, and $Y_4 = 7.6$, use the method of maximum likelihood to estimate the parameter θ in the uniform pdf

$$f_Y(y;\theta) = \frac{1}{\theta}, \quad 0 \leq y \leq \theta$$

(b) Suppose the random sample in part (a) represents the two-parameter uniform pdf

$$f_Y(y;\theta_1,\theta_2) = \frac{1}{\theta_2 - \theta_1}, \quad \theta_1 \leq y \leq \theta_2$$

Find the maximum likelihood estimates for θ_1 and θ_2.

5.2.11. Find the maximum likelihood estimate for θ in the pdf

$$f_Y(y;\theta) = \frac{2y}{1-\theta^2}, \quad \theta \leq y \leq 1$$

if a random sample of size 6 yielded the measurements 0.70, 0.63, 0.92, 0.86, 0.43, and 0.21.

5.2.12. A random sample of size n is taken from the pdf

$$f_Y(y;\theta) = \frac{2y}{\theta^2}, \quad 0 \leq y \leq \theta$$

Find an expression for $\hat{\theta}$, the maximum likelihood estimator for θ.

5.2.13. If the random variable Y denotes an individual's income, Pareto's law claims that $P(Y \geq y) = \left(\frac{k}{y}\right)^\theta$, where k is the entire population's minimum income. It follows that $F_Y(y) = 1 - \left(\frac{k}{y}\right)^\theta$, and, by differentiation,

$$f_Y(y;\theta) = \theta k^\theta \left(\frac{1}{y}\right)^{\theta+1}, \quad y \geq k; \quad \theta \geq 1$$

Assume k is known. Find the maximum likelihood estimator for θ if income information has been collected on a random sample of 25 individuals.

5.2.14. The exponential pdf is a measure of lifetimes of devices that do not age (see Question 3.11.11). However, the exponential pdf is a special case of the *Weibull distribution*, which measures time to failure of devices where the probability of failure increases as time does. A Weibull random variable Y has pdf $f_Y(y;\alpha,\beta) = \alpha\beta y^{\beta-1}e^{-\alpha y^\beta}, 0 \leq y$, $0 < \alpha, 0 < \beta$.

(a) Find the maximum likelihood estimator for α assuming that β is known.
(b) Suppose α and β are both unknown. Write down the equations that would be solved simultaneously to find the maximum likelihood estimators of α and β.

5.2.15. Suppose a random sample of size n is drawn from a normal pdf where the mean μ is known but the variance σ^2 is unknown. Use the method of maximum likelihood to find a formula for $\hat{\sigma}^2$. Compare your answer to the maximum likelihood estimator found in Example 5.2.4.

The Method of Moments

A second procedure for estimating parameters is the *method of moments*. Proposed near the turn of the twentieth century by the great British statistician, Karl Pearson, the method of moments is often more tractable than the method of maximum likelihood in situations where the underlying probability model has multiple parameters.

Suppose that Y is a continuous random variable and that its pdf is a function of s unknown parameters, $\theta_1, \theta_2, \ldots, \theta_s$. The first s moments of Y, if they exist, are given by the integrals

$$E(Y^j) = \int_{-\infty}^{\infty} y^j \cdot f_Y(y; \theta_1, \theta_2, \ldots, \theta_s)\, dy, \quad j = 1, 2, \ldots, s$$

In general, each $E(Y^j)$ will be a different function of the s parameters. That is,

$$\begin{aligned} E(Y^1) &= g_1(\theta_1, \theta_2, \ldots, \theta_s) \\ E(Y^2) &= g_2(\theta_1, \theta_2, \ldots, \theta_s) \\ &\vdots \\ E(Y^s) &= g_s(\theta_1, \theta_2, \ldots, \theta_s) \end{aligned}$$

Corresponding to each *theoretical* moment, $E(Y^j)$, is a *sample* moment, $\frac{1}{n}\sum_{i=1}^{n} y_i^j$. Intuitively, the jth sample moment is an approximation to the jth theoretical moment. Setting the two equal *for each j* produces a system of s simultaneous equations, the solutions to which are the desired set of estimates, $\theta_{1e}, \theta_{2e}, \ldots$, and θ_{se}.

Definition 5.2.3. Let $y_1, y_2, \ldots, y_n$ be a random sample from the continuous pdf $f_Y(y; \theta_1, \theta_2, \ldots, \theta_s)$. The *method of moments* estimates, $\theta_{1e}, \theta_{2e}, \ldots$, and θ_{se}, for the model's unknown parameters are the solutions of the s simultaneous equations

$$\begin{aligned} \int_{-\infty}^{\infty} y\, f_Y(y; \theta_1, \theta_2, \ldots, \theta_s)\, dy &= \left(\frac{1}{n}\right)\sum_{i=1}^{n} y_i \\ \int_{-\infty}^{\infty} y^2 f_Y(y; \theta_1, \theta_2, \ldots, \theta_s)\, dy &= \left(\frac{1}{n}\right)\sum_{i=1}^{n} y_i^2 \\ \vdots \qquad & \qquad \vdots \\ \int_{-\infty}^{\infty} y^s f_Y(y; \theta_1, \theta_2, \ldots, \theta_s)\, dy &= \left(\frac{1}{n}\right)\sum_{i=1}^{n} y_i^s \end{aligned}$$

Questions

5.2.16. Let $y_1, y_2, \ldots, y_n$ be a random sample of size n from the pdf $f_Y(y;\theta) = \frac{2y}{\theta^2}, 0 \le y \le \theta$. Find a formula for the method of moments estimate for θ. Compare the values of the method of moments estimate and the maximum likelihood estimate if a random sample of size 5 consists of the numbers 17, 92, 46, 39, and 56 (recall Question 5.2.12).

5.2.17. Use the method of moments to estimate θ in the pdf

$$f_Y(y;\theta) = (\theta^2 + \theta)y^{\theta-1}(1-y), \quad 0 \le y \le 1$$

Assume that a random sample of size n has been collected.

5.2.18. A criminologist is searching through FBI files to document the prevalence of a rare double-whorl fingerprint. Among six consecutive sets of 100,000 prints scanned by a computer, the numbers of persons having the abnormality are 3, 0, 3, 4, 2, and 1, respectively. Assume that double whorls are Poisson events. Use the method of moments to estimate their occurrence rate, λ. How would your answer change if λ were estimated using the method of maximum likelihood?

5.2.19. Find the method of moments estimate for λ if a random sample of size n is taken from the exponential pdf, $f_Y(y;\lambda) = \lambda e^{-\lambda y}, y \ge 0$.

5.2.20. Suppose that $Y_1 = 8.3$, $Y_2 = 4.9$, $Y_3 = 2.6$, and $Y_4 = 6.5$ is a random sample of size 4 from the two-parameter uniform pdf,

$$f_Y(y;\theta_1,\theta_2) = \frac{1}{2\theta_2}, \quad \theta_1 - \theta_2 \le y \le \theta_1 + \theta_2$$

Use the method of moments to calculate θ_{1e} and θ_{2e}.

5.2.21. Find a formula for the method of moments estimate for the parameter θ in the Pareto pdf,

$$f_Y(y;\theta) = \theta k^\theta \left(\frac{1}{y}\right)^{\theta+1}, \quad y \ge k; \quad \theta \ge 1$$

Assume that k is known and that the data consist of a random sample of size n. Compare your answer to the maximum likelihood estimator found in Question 5.2.13.

5.2.22. Calculate the method of moments estimate for the parameter θ in the probability function

$$p_X(k;\theta) = \theta^k(1-\theta)^{1-k}, \quad k = 0, 1$$

if a sample of size 5 is the set of numbers 0, 0, 1, 0, 1.

5.2.23. Find the method of moments estimates for μ and σ^2, based on a random sample of size n drawn from a normal pdf, where $\mu = E(Y)$ and $\sigma^2 = \text{Var}(Y)$. Compare your answers with the maximum likelihood estimates derived in Example 5.2.4.

5.2.24. Use the method of moments to derive formulas for estimating the parameters r and p in the negative binomial pdf,

$$p_X(k;r,p) = \binom{k-1}{r-1} p^r (1-p)^{k-r}, \quad k = r, r+1, \ldots$$

5.2.25. Bird songs can be characterized by the number of clusters of "syllables" that are strung together in rapid succession. If the last cluster is defined as a "success," it may be reasonable to treat the *number* of clusters in a song as a geometric random variable. Does the model $p_X(k) = (1-p)^{k-1}p, k = 1, 2, \ldots$, adequately describe the following distribution of 250 song lengths (100)? Begin by finding the method of moments estimate for p. Then calculate the set of "expected" frequencies.

No. of Clusters/Song	Frequency
1	132
2	52
3	34
4	9
5	7
6	5
7	5
8	6
	250

5.2.26. Let $y_1, y_2, \ldots, y_n$ be a random sample from the continuous pdf $f_Y(y;\theta_1,\theta_2)$. Let $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{y})^2$. Show that the solutions of the equations

$$E(Y) = \overline{y} \text{ and } \text{Var}(Y) = \hat{\sigma}^2$$

for θ_1 and θ_2 give the same results as using the equations in Definition 5.2.3.

5.3 Interval Estimation

Point estimates, no matter how they are determined, share the same fundamental weakness: They provide no indication of their inherent precision. We know, for instance, that $\hat{\lambda} = \overline{X}$ is both the maximum likelihood and the method of moments estimator for the Poisson parameter, λ. But suppose a sample of size 6 is taken from though, seems to contradict that assumption: Samples used in opinion surveys are invariably drawn *without replacement*, in which case X is hypergeometric, not binomial. The consequences of that particular "error," however, are easily corrected and frequently negligible.

It can be shown mathematically that the expected value of $\frac{X}{n}$ is the same regardless of whether X is binomial or hypergeometric; its variance, though, is different. If X is binomial,

$$\text{Var}\left(\frac{X}{n}\right) = \frac{p(1-p)}{n}$$

If X is hypergeometric,

$$\text{Var}\left(\frac{X}{n}\right) = \frac{p(1-p)}{n}\left(\frac{N-n}{N-1}\right)$$

where N is the total number of subjects in the population.

Since $\frac{N-n}{N-1} < 1$, the actual variance of $\frac{X}{n}$ is somewhat smaller than the (binomial) variance we have been assuming, $\frac{p(1-p)}{n}$. The ratio $\frac{N-n}{N-1}$ is called the *finite correction factor*. If N is much larger than n, which is typically the case, then the magnitude of $\frac{N-n}{N-1}$ will be so close to 1 that the variance of $\frac{X}{n}$ is equal to $\frac{p(1-p)}{n}$ for all practical purposes. Thus the "binomial" assumption in those situations is more than adequate. Only when the sample is a sizeable fraction of the population do we need to include the finite correction factor in any calculations that involve the variance of $\frac{X}{n}$. ■

Questions

5.3.1. A commonly used IQ test is scaled to have a mean of 100 and a standard deviation of $\sigma = 15$. A school counselor was curious about the average IQ of the students in her school and took a random sample of fifty students' IQ scores. The average of these was $\overline{y} = 107.9$. Find a 95% confidence interval for the student IQ in the school.

5.3.2. The production of a nationally marketed detergent results in certain workers receiving prolonged exposures to a *Bacillus subtilis* enzyme. Nineteen workers were tested to determine the effects of those exposures, if any, on various respiratory functions. One such function, air-flow rate, is measured by computing the ratio of a person's forced expiratory volume (FEV_1) to his or her vital capacity (VC). (Vital capacity is the maximum volume of air a person can exhale after taking as deep a breath as possible; FEV_1 is the maximum volume of air a person can exhale in one second.) In persons with no lung dysfunction, the "norm" for FEV_1/VC ratios is 0.80. Based on the following data (164), is it believable that exposure to the *Bacillus subtilis* enzyme has no effect on the FEV_1/VC ratio? Answer the question by constructing a 95% confidence interval. Assume that FEV_1/VC ratios are normally distributed with $\sigma = 0.09$.

Subject	FEV_1/VC	Subject	FEV_1/VC
RH	0.61	WS	0.78
RB	0.70	RV	0.84
MB	0.63	EN	0.83
DM	0.76	WD	0.82
WB	0.67	FR	0.74
RB	0.72	PD	0.85
BF	0.64	EB	0.73
JT	0.82	PC	0.85
PS	0.88	RW	0.87
RB	0.82		

5.3.3. Mercury pollution is widely recognized as a serious ecological problem. Much of the mercury released into the environment originates as a byproduct of coal burning and other industrial processes. It does not become dangerous until it falls into large bodies of water, where microorganisms convert it to methylmercury (CH_3^{203}), an organic form that is particularly toxic. Fish are the intermediaries: They ingest and absorb the methylmercury and are then eaten by humans. Men and women, however, may not metabolize CH_3^{203} at the same rate. In one study investigating that issue, six women were given a known amount of protein-bound methylmercury. Shown in the following table are the half-lives of the methylmercury in their systems (114). For men, the average CH_3^{203} half-life is believed to be eighty days. Assume that for both genders, CH_3^{203} half-lives are normally distributed with a standard deviation (σ) of eight days. Construct a 95% confidence interval for the true female CH_3^{203} half-life. Based on these data, is it believable that males and females metabolize methylmercury at the same rate? Explain.

Females	CH_3^{203} Half-Life
AE	52
EH	69
LJ	73
AN	88
KR	87
LU	56

5.3.4. A physician who has a group of thirty-eight female patients aged 18 to 24 on a special diet wishes to estimate the effect of the diet on total serum cholesterol. For this group, their average serum cholesterol is 188.4 (measured in mg/100mL). Because of a large-scale government study, the physician is willing to assume that the total serum cholesterol measurements are normally distributed with standard deviation of $\sigma = 40.7$. Find a 95% confidence interval of the mean serum cholesterol of patients on the special diet. Does the diet seem to have any effect on their serum cholesterol, given that the national average for women aged 18 to 24 is 192.0?

5.3.5. Suppose a sample of size n is to be drawn from a normal distribution where σ is known to be 14.3. How large does n have to be to guarantee that the length of the 95% confidence interval for μ will be less than 3.06?

5.3.6. What "confidence" would be associated with each of the following intervals? Assume that the random variable Y is normally distributed and that σ is known.

(a) $\left(\bar{y} - 1.64 \cdot \frac{\sigma}{\sqrt{n}}, \bar{y} + 2.33 \cdot \frac{\sigma}{\sqrt{n}}\right)$

(b) $\left(-\infty, \bar{y} + 2.58 \cdot \frac{\sigma}{\sqrt{n}}\right)$

(c) $\left(\bar{y} - 1.64 \cdot \frac{\sigma}{\sqrt{n}}, \bar{y}\right)$

5.3.7. Five independent samples, each of size n, are to be drawn from a normal distribution where σ is known. For each sample, the interval $\left(\bar{y} - 0.96 \cdot \frac{\sigma}{\sqrt{n}}, \bar{y} + 1.06 \cdot \frac{\sigma}{\sqrt{n}}\right)$ will be constructed. What is the probability that at least four of the intervals will contain the unknown μ?

5.3.8. Suppose that $y_1, y_2, \ldots, y_n$ is a random sample of size n from a normal distribution where σ is known. Depending on how the tail-area probabilities are split up, an infinite number of random intervals having a 95% probability of containing μ can be constructed. What is unique about the particular interval $\left(\bar{y} - 1.96 \cdot \frac{\sigma}{\sqrt{n}}, \bar{y} + 1.96 \cdot \frac{\sigma}{\sqrt{n}}\right)$?

5.3.9. If the standard deviation (σ) associated with the pdf that produced the following sample is 3.6, would it be correct to claim that

$$\left(2.61 - 1.96 \cdot \frac{3.6}{\sqrt{20}}, 2.61 + 1.96 \cdot \frac{3.6}{\sqrt{20}}\right) = (1.03, 4.19)$$

is a 95% confidence interval for μ? Explain.

2.5	0.1	0.2	1.3
3.2	0.1	0.1	1.4
0.5	0.2	0.4	11.2
0.4	7.4	1.8	2.1
0.3	8.6	0.3	10.1

5.3.10. In 1927, the year he hit sixty home runs, Babe Ruth batted .356, having collected 192 hits in 540 official at-bats (140). Based on his performance that season, construct a 95% confidence interval for Ruth's probability of getting a hit in a future at-bat.

5.3.11. To buy a thirty-second commercial break during the telecast of Super Bowl XXIX cost approximately \$1,000,000. Not surprisingly, potential sponsors wanted to know how many people might be watching. In a survey of 1015 potential viewers, 281 said they expected to see less than a quarter of the advertisements aired during the game. Define the relevant parameter and estimate it using a 90% confidence interval.

5.3.12. During one of the first "beer wars" in the early 1980s, a taste test between Schlitz and Budweiser was the focus of a nationally broadcast TV commercial. One hundred people agreed to drink from two unmarked mugs and indicate which of the two beers they liked better; fifty-four said, "Bud." Construct and interpret the corresponding 95% confidence interval for p, the true proportion of beer drinkers who prefered Budweiser to Schlitz. How would Budweiser and Schlitz executives each have put these results in the best possible light for their respective companies?

5.3.13. The Pew Research Center did a survey of 2253 adults and discovered that 63% of them had broadband Internet connections in their homes. The survey report noted that this figure represented a "significant jump" from the similar figure of 54% from two years earlier. One way to define "significant jump" is to show that the earlier number does not lie in the 95% confidence interval. Was the increase significant by this definition?

Source: http://www.pewinternet.org/Reports/2009/10-Home-Broadband-Adoption-2009.aspx.

5.3.14. If (0.57, 0.63) is a 50% confidence interval for p, what does $\frac{k}{n}$ equal and how many observations were taken?

5.3.15. Suppose a coin is to be tossed n times for the purpose of estimating p, where $p = P(\text{heads})$. How large must n be to guarantee that the length of the 99% confidence interval for p will be less than 0.02?

5.3.16. On the morning of November 9, 1994—the day after the electoral landslide that had returned Republicans to power in both branches of Congress—several key races were still in doubt. The most prominent was the Washington contest involving Democrat Tom Foley, the reigning speaker of the house. An Associated Press story showed how narrow the margin had become (120):

> With 99 percent of precincts reporting, Foley trailed Republican challenger George Nethercutt by just 2,174 votes, or 50.6 percent to 49.4 percent. About 14,000 absentee ballots remained uncounted, making the race too close to call.

Let $p = P(\text{Absentee voter prefers Foley})$. How small could p have been and still have given Foley a 20% chance of overcoming Nethercutt's lead and winning the election?

5.3.17. Which of the following two intervals has the greater probability of containing the binomial parameter p?

$$\left[\frac{X}{n} - 0.67\sqrt{\frac{(X/n)(1-X/n)}{n}}, \frac{X}{n} + 0.67\sqrt{\frac{(X/n)(1-X/n)}{n}}\right]$$

$$\text{or} \quad \left(\frac{X}{n}, \infty\right)$$

5.3.18. Examine the first two derivatives of the function $g(p) = p(1-p)$ to verify the claim on p. 305 that $p(1-p) \leq \frac{1}{4}$ for $0 < p < 1$.

5.3.19. The financial crisis of 2008 highlighted the issue of excessive compensation for business CEOs. In a Gallup poll in the summer of 2009, 998 adults were asked, "Do you favor or oppose the federal government taking steps to limit the pay of executives at major companies?", with 59% responding in favor. The report of the poll noted a margin of error of ± 3 percentage points. Verify the margin of error and construct a 95% confidence interval.

Source: http://www.gallup.com/poll/120872/Americans-Favor-Gov-Action-Limit-Executive-Pay.aspx.

5.3.20. Viral infections contracted early during a woman's pregnancy can be very harmful to the fetus. One study found a total of 86 deaths and birth defects among 202 pregnancies complicated by a first-trimester German measles infection (45). Is it believable that the true proportion of abnormal births under similar circumstances could be as high as 50%? Answer the question by calculating the margin of error for the sample proportion, 86/202.

5.3.21. Rewrite Definition 5.3.1 to cover the case where a finite correction factor needs to be included (i.e., situations where the sample size n is not negligible relative to the population size N).

5.3.22. A public health official is planning for the supply of influenza vaccine needed for the upcoming flu season. She took a poll of 350 local citizens and found that only 126 said they would be vaccinated.

(a) Find the 90% confidence interval for the true proportion of people who plan to get the vaccine.
(b) Find the confidence interval, including the finite correction factor, assuming the town's population is 3000.

5.3.23. Given that n observations will produce a binomial parameter estimator, $\frac{X}{n}$, having a margin of error equal to 0.06, how many observations are required for the proportion to have a margin of error half that size?

5.3.24. Given that a political poll shows that 52% of the sample favors Candidate A, whereas 48% would vote for Candidate B, and given that the margin of error associated with the survey is 0.05, does it make sense to claim that the two candidates are tied? Explain.

5.3.25. Assume that the binomial parameter p is to be estimated with the function $\frac{X}{n}$, where X is the number of successes in n independent trials. Which demands the larger sample size: requiring that $\frac{X}{n}$ have a 96% probability of being within 0.05 of p, or requiring that $\frac{X}{n}$ have a 92% probability of being within 0.04 of p?

5.3.26. Suppose that p is to be estimated by $\frac{X}{n}$ and we are willing to assume that the true p will not be greater than 0.4. What is the smallest n for which $\frac{X}{n}$ will have a 99% probability of being within 0.05 of p?

5.3.27. Let p denote the true proportion of college students who support the movement to colorize classic films. Let the random variable X denote the number of students (out of n) who prefer colorized versions to black and white. What is the smallest sample size for which the probability is 80% that the difference between $\frac{X}{n}$ and p is less than 0.02?

5.3.28. University officials are planning to audit 1586 new appointments to estimate the proportion p who have been incorrectly processed by the payroll department.

(a) How large does the sample size need to be in order for $\frac{X}{n}$, the sample proportion, to have an 85% chance of lying within 0.03 of p?
(b) Past audits suggest that p will not be larger than 0.10. Using that information, recalculate the sample size asked for in part (a).

To "unbias" the maximum likelihood estimator in this case, we need simply multiply $\hat{\sigma}^2$ by $\frac{n}{n-1}$. By convention, the unbiased version of the maximum likelihood estimator for σ^2 in a normal distribution is denoted S^2 and is referred to as the *sample variance*:

$$S^2 = \text{sample variance} = \frac{n}{n-1} \cdot \frac{1}{n} \sum_{i=1}^{n} (Y_i - \overline{Y})^2$$

$$= \frac{1}{n-1} \sum_{i=1}^{n} (Y_i - \overline{Y})^2$$

Comment The square root of the sample variance is called the *sample standard deviation*:

$$S = \text{sample standard deviation} = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (Y_i - \overline{Y})^2}$$

In practice, S is the most commonly used estimator for σ *even though* $E(S) \neq \sigma$ [despite the fact that $E(S^2) = \sigma^2$]. ■

Questions

5.4.1. Two chips are drawn without replacement from an urn containing five chips, numbered 1 through 5. The average of the two drawn is to be used as an estimator, $\hat{\theta}$, for the true average of all the chips ($\theta = 3$). Calculate $P(|\hat{\theta} - 3| > 1.0)$.

5.4.2. Suppose a random sample of size $n = 6$ is drawn from the uniform pdf $f_Y(y;\theta) = 1/\theta, 0 \leq y \leq \theta$, for the purpose of using $\hat{\theta} = Y_{\max}$ to estimate θ.

(a) Calculate the probability that $\hat{\theta}$ falls within 0.2 of θ given that the parameter's true value is 3.0.
(b) Calculate the probability of the event asked for in part (a), assuming the sample size is 3 instead of 6.

5.4.3. Five hundred adults are asked whether they favor a bipartisan campaign finance reform bill. If the true proportion of the electorate in favor of the legislation is 52%, what are the chances that fewer than half of those in the sample support the proposal? Use a Z transformation to approximate the answer.

5.4.4. A sample of size $n = 16$ is drawn from a normal distribution where $\sigma = 10$ but μ is unknown. If $\mu = 20$, what is the probability that the estimator $\hat{\mu} = \overline{Y}$ will lie between 19.0 and 21.0?

5.4.5. Suppose $X_1, X_2, \ldots, X_n$ is a random sample of size n drawn from a Poisson pdf where λ is an unknown parameter. Show that $\hat{\lambda} = \overline{X}$ is unbiased for λ. For what type of parameter, in general, will the sample mean necessarily be an unbiased estimator? (*Hint:* The answer is implicit in the derivation showing that $\overline{X}$ is unbiased for the Poisson λ.)

5.4.6. Let $Y_{\min}$ be the smallest order statistic in a random sample of size n drawn from the uniform pdf, $f_Y(y;\theta) = 1/\theta, 0 \leq y \leq \theta$. Find an unbiased estimator for θ based on $Y_{\min}$.

5.4.7. Let Y be the random variable described in Example 5.2.3, where $f_Y(y,\theta) = e^{-(y-\theta)}$, $y \geq \theta$, $\theta > 0$. Show that $Y_{\min} - \frac{1}{n}$ is an unbiased estimator of θ.

5.4.8. Suppose that 14, 10, 18, and 21 constitute a random sample of size 4 drawn from a uniform pdf defined over the interval $[0, \theta]$, where θ is unknown. Find an unbiased estimator for θ based on Y_3', the third order statistic. What numerical value does the estimator have for these particular observations? Is it possible that we would know that an estimate for θ based on Y_3' was incorrect, even if we had no idea what the true value of θ might be? Explain.

5.4.9. A random sample of size 2, Y_1 and Y_2, is drawn from the pdf

$$f_Y(y;\theta) = 2y\theta^2, \quad 0 < y < \frac{1}{\theta}$$

What must c equal if the statistic $c(Y_1 + 2Y_2)$ is to be an unbiased estimator for $\frac{1}{\theta}$?

5.4.10. A sample of size 1 is drawn from the uniform pdf defined over the interval $[0, \theta]$. Find an unbiased estimator for θ^2. (*Hint:* Is $\hat{\theta} = Y^2$ unbiased?)

5.4.11. Suppose that W is an unbiased estimator for θ. Can W^2 be an unbiased estimator for θ^2?

5.4.12. We showed in Example 5.4.4 that $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \overline{Y})^2$ is biased for σ^2. Suppose μ is known and does not have to be estimated by $\overline{Y}$. Show that $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \mu)^2$ is unbiased for σ^2.

5.4.13. As an alternative to imposing unbiasedness, an estimator's distribution can be "centered" by requiring that its median be equal to the unknown parameter θ. If it is, $\hat{\theta}$ is said to be *median unbiased*. Let $Y_1, Y_2, \ldots, Y_n$ be a random sample of size n from the uniform pdf, $f_Y(y;\theta) = 1/\theta, 0 \le y \le \theta$. For arbitrary n, is $\hat{\theta} = \frac{n+1}{n} \cdot Y_{\max}$ median unbiased? Is it median unbiased for any value of n?

5.4.14. Let $Y_1, Y_2, \ldots, Y_n$ be a random sample of size n from the pdf $f_Y(y;\theta) = \frac{1}{\theta}e^{-y/\theta}$, $y > 0$. Let $\hat{\theta} = n \cdot Y_{\min}$. Is $\hat{\theta}$ unbiased for θ? Is $\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} Y_i$ unbiased for θ?

5.4.15. An estimator $\hat{\theta}_n = h(W_1, \ldots, W_n)$ is said to be *asymptotically unbiased for* θ if $\lim_{n\to\infty} E(\hat{\theta}_n) = \theta$. Suppose W is a random variable with $E(W) = \mu$ and with variance σ^2. Show that $\overline{W}^2$ is an asymptotically unbiased estimator for μ^2.

5.4.16. Is the maximum likelihood estimator for σ^2 in a normal pdf, where both μ and σ^2 are unknown, asymptotically unbiased?

Efficiency

As we have seen, unknown parameters can have a multiplicity of unbiased estimators. For samples drawn from the uniform pdf, $f_Y(y;\theta) = 1/\theta, 0 \le y \le \theta$, for example, both $\hat{\theta} = \frac{n+1}{n} \cdot Y_{\max}$ and $\hat{\theta} = \frac{2}{n}\sum_{i=1}^{n} Y_i$ have expected values equal to θ. Does it matter which we choose?

Yes. Unbiasedness is not the only property we would like an estimator to have; also important is its *precision*. Figure 5.4.3 shows the pdfs associated with two hypothetical estimators, $\hat{\theta}_1$ and $\hat{\theta}_2$. Both are unbiased for θ, but $\hat{\theta}_2$ is clearly the better of the two because of its smaller variance. For any value r,

$$P(\theta - r \le \hat{\theta}_2 \le \theta + r) > P(\theta - r \le \hat{\theta}_1 \le \theta + r)$$

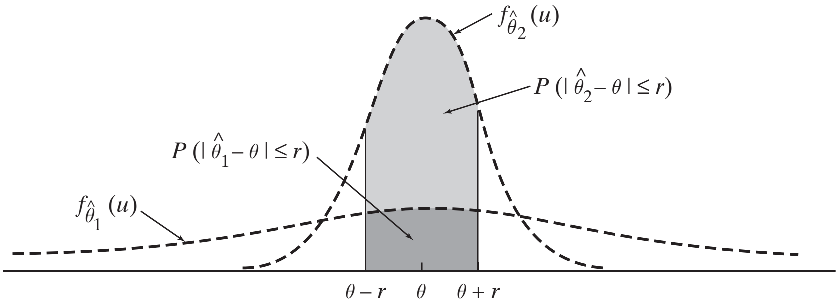
That is, $\hat{\theta}_2$ has a greater chance of being within a distance r of the unknown θ than does $\hat{\theta}_1$.

Definition 5.4.2. Let $\hat{\theta}_1$ and $\hat{\theta}_2$ be two unbiased estimators for a parameter θ. If

$$\text{Var}(\hat{\theta}_1) < \text{Var}(\hat{\theta}_2)$$

we say that $\hat{\theta}_1$ is *more efficient* than $\hat{\theta}_2$. Also, the *relative efficiency of* $\hat{\theta}_1$ *with respect to* $\hat{\theta}_2$ is the ratio $\text{Var}(\hat{\theta}_2)/\text{Var}(\hat{\theta}_1)$.

Figure 5.4.3



Example 5.4.6 Let $Y_1, \ldots, Y_n$ be a random sample from the pdf $f_Y(y;\theta) = \frac{2y}{\theta^2}$, $0 \le y \le \theta$. We know from Example 5.4.2 that $\hat{\theta}_1 = \frac{3}{2}\overline{Y}$ and $\hat{\theta}_2 = \frac{2n+1}{2n} Y_{\max}$ are both unbiased for θ. Which estimator is more efficient?

First, let us calculate the variance of $\hat{\theta}_1 = \frac{3}{2}\overline{Y}$. To do so, we need the variance of Y. To that end, note that

$$E(Y^2) = \int_0^\theta y^2 \cdot \frac{2y}{\theta^2}\, dy = \frac{2}{\theta^2}\int_0^\theta y^3\, dy = \frac{2}{\theta^2}\cdot\frac{\theta^4}{4} = \frac{1}{2}\theta^2$$

and

$$\mathrm{Var}(Y) = E(Y^2) - E(Y)^2 = \frac{1}{2}\theta^2 - \left(\frac{2}{3}\theta\right)^2 = \frac{\theta^2}{18}$$

Then

$$\mathrm{Var}(\hat{\theta}_1) = \mathrm{Var}\left(\frac{3}{2}\overline{Y}\right) = \frac{9}{4}\mathrm{Var}(\overline{Y}) = \frac{9}{4}\frac{\mathrm{Var}(Y)}{n} = \frac{9}{4n}\cdot\frac{\theta^2}{18} = \frac{\theta^2}{8n}$$

To address the variance of $\hat{\theta}_2 = \frac{2n+1}{2n} Y_{\max}$, we start with finding the variance of $Y_{\max}$. Recall that its pdf is

$$nF_Y(y)^{n-1} f_Y(y) = \frac{2n}{\theta^{2n}} y^{2n-1}, 0 \le y \le \theta$$

From that expression, we obtain

$$E(Y_{\max}^2) = \int_0^\theta y^2 \cdot \frac{2n}{\theta^{2n}} y^{2n-1}\, dy = \frac{2n}{\theta^{2n}}\int_0^\theta y^{2n+1}\, dy = \frac{2n}{\theta^{2n}}\cdot\frac{\theta^{2n+2}}{2n+2} = \frac{n}{n+1}\theta^2$$

and then

$$\mathrm{Var}(Y_{\max}) = E(Y_{\max}^2) - E(Y_{\max})^2 = \frac{n}{n+1}\theta^2 - \left(\frac{2n}{2n+1}\theta\right)^2 = \frac{n}{(n+1)(2n+1)^2}\theta^2$$

Finally,

$$\mathrm{Var}(\hat{\theta}_2) = \mathrm{Var}\left(\frac{2n+1}{2n}Y_{\max}\right) = \frac{(2n+1)^2}{4n^2}\mathrm{Var}(Y_{\max}) = \frac{(2n+1)^2}{4n^2}\cdot\frac{n}{(n+1)(2n+1)^2}\theta^2$$

$$= \frac{1}{4n(n+1)}\theta^2$$

Note that $\mathrm{Var}(\hat{\theta}_2) = \frac{1}{4n(n+1)}\theta^2 < \frac{1}{8n}\theta^2 = \mathrm{Var}(\hat{\theta}_1)$ for $n > 1$, so we say that $\hat{\theta}_2$ is *more efficient* than $\hat{\theta}_1$. The *relative efficiency* of $\hat{\theta}_2$ with respect to $\hat{\theta}_1$ is the ratio of their variances:

$$\frac{\mathrm{Var}(\hat{\theta}_1)}{\mathrm{Var}(\hat{\theta}_2)} = \frac{1}{8n}\theta^2 \div \frac{1}{4n(n+1)}\theta^2 = \frac{4n(n+1)}{8n} = \frac{(n+1)}{2}$$

■

Questions

5.4.17. Let $X_1, X_2, \ldots, X_n$ denote the outcomes of a series of n independent trials, where

$$X_i = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability } 1-p \end{cases}$$

for $i = 1, 2, \ldots, n$. Let $X = X_1 + X_2 + \cdots + X_n$.

(a) Show that $\hat{p}_1 = X_1$ and $\hat{p}_2 = \frac{X}{n}$ are unbiased estimators for p.

(b) Intuitively, $\hat{p}_2$ is a better estimator than $\hat{p}_1$ because $\hat{p}_1$ fails to include any of the information about the parameter contained in trials 2 through n. Verify that speculation by comparing the variances of $\hat{p}_1$ and $\hat{p}_2$.

5.4.18. Suppose that $n = 5$ observations are taken from the uniform pdf, $f_Y(y;\theta) = 1/\theta, 0 \le y \le \theta$, where θ is unknown. Two unbiased estimators for θ are

$$\hat{\theta}_1 = \frac{6}{5} \cdot Y_{\max} \quad \text{and} \quad \hat{\theta}_2 = 6 \cdot Y_{\min}$$

Which estimator would be better to use? [*Hint:* What must be true of $\text{Var}(Y_{\max})$ and $\text{Var}(Y_{\min})$ given that $f_Y(y;\theta)$ is symmetric?] Does your answer as to which estimator is better make sense on intuitive grounds? Explain.

5.4.19. Let $Y_1, Y_2, \ldots, Y_n$ be a random sample of size n from the pdf $f_Y(y;\theta) = \frac{1}{\theta}e^{-y/\theta}, y > 0$.

(a) Show that $\hat{\theta}_1 = Y_1, \hat{\theta}_2 = \overline{Y}$, and $\hat{\theta}_3 = n \cdot Y_{\min}$ are all unbiased estimators for θ.
(b) Find the variances of $\hat{\theta}_1, \hat{\theta}_2$, and $\hat{\theta}_3$.
(c) Calculate the relative efficiencies of $\hat{\theta}_1$ to $\hat{\theta}_3$ and $\hat{\theta}_2$ to $\hat{\theta}_3$.

5.4.20. Given a random sample of size n from a Poisson distribution, $\hat{\lambda}_1 = X_1$ and $\hat{\lambda}_2 = \overline{X}$ are two unbiased estimators for λ. Calculate the relative efficiency of $\hat{\lambda}_1$ to $\hat{\lambda}_2$.

5.4.21. If $Y_1, Y_2, \ldots, Y_n$ are random observations from a uniform pdf over $[0, \theta]$, both $\hat{\theta}_1 = \left(\frac{n+1}{n}\right) \cdot Y_{\max}$ and $\hat{\theta}_2 = (n + 1) \cdot Y_{\min}$ are unbiased estimators for θ. Show that $\text{Var}(\hat{\theta}_2)/\text{Var}(\hat{\theta}_1) = n^2$.

5.4.22. Suppose that W_1 is a random variable with mean μ and variance σ_1^2 and W_2 is a random variable with mean μ and variance σ_2^2. From Example 5.4.3, we know that $cW_1 + (1 - c)W_2$ is an unbiased estimator of μ for any constant $c > 0$. If W_1 and W_2 are independent, for what value of c is the estimator $cW_1 + (1 - c)W_2$ most efficient?

5.5 Minimum-Variance Estimators: The Cramér-Rao Lower Bound

Given two estimators, $\hat{\theta}_1$ and $\hat{\theta}_2$, each unbiased for the parameter θ, we know from Section 5.4 which is "better"—the one with the smaller variance. But nothing in that section speaks to the more fundamental question of how good $\hat{\theta}_1$ and $\hat{\theta}_2$ are *relative to the infinitely many other unbiased estimators for* θ. Is there a $\hat{\theta}_3$, for example, that has a smaller variance than either $\hat{\theta}_1$ or $\hat{\theta}_2$ has? Can we identify the unbiased estimator having the *smallest* variance? Addressing those concerns is one of the most elegant, yet practical, theorems in all of mathematical statistics, a result known as the *Cramér-Rao lower bound*.

Suppose a random sample of size n is taken from, say, a continuous probability distribution $f_Y(y;\theta)$, where θ is an unknown parameter. Associated with $f_Y(y;\theta)$ is a theoretical limit below which the variance of any unbiased estimator for θ cannot fall. That limit is the Cramér-Rao lower bound. If the variance of a given $\hat{\theta}$ is *equal* to the Cramér-Rao lower bound, we know that estimator is *optimal* in the sense that no unbiased $\hat{\theta}$ can estimate θ with greater precision.

Theorem 5.5.1 *(Cramér-Rao Inequality.) Let $f_Y(y;\theta)$ be a continuous pdf with continuous first-order and second-order derivatives. Also, suppose that the set of y values, where $f_Y(y;\theta) \neq 0$, does not depend on θ.*

Let $Y_1, Y_2, \ldots, Y_n$ be a random sample from $f_Y(y;\theta)$, and let $\hat{\theta} = h(Y_1, Y_2, \ldots, Y_n)$ be any unbiased estimator of θ. Then

$$\text{Var}(\hat{\theta}) \ge \left\{ nE\left[\left(\frac{\partial \ln f_Y(Y;\theta)}{\partial \theta} \right)^2 \right] \right\}^{-1} = \left\{ -nE\left[\frac{\partial^2 \ln f_Y(Y;\theta)}{\partial \theta^2} \right] \right\}^{-1}$$

[A similar statement holds if the n observations come from a discrete pdf, $p_X(k;\theta)$].

Proof See (93). □

b. Let W be any random variable whose mean is μ and whose variance is finite, and let b be any constant. Then

$$\begin{aligned} E[(W-b)^2] &= E[(W-\mu)+(\mu-b)]^2 \\ &= E[(W-\mu)^2]+2(\mu-b)E(W-\mu)+(\mu-b)^2 \\ &= \mathrm{Var}(W)+0+(\mu-b)^2 \end{aligned}$$

implying that $E[(W-b)]^2$ is minimized when $b=\mu$. It follows that the Bayes estimate for θ, given a quadratic loss function, is the *mean* of the posterior distribution. □

Example 5.8.6 Recall Example 5.8.3, where the parameter in a Poisson distribution was assumed to have a gamma prior distribution. For a random sample of size n, where $W=\sum_{i=1}^{n} X_i$,

$$p_W(w|\theta)=e^{-n\theta}(n\theta)^w/w!, \quad w=0,1,2,\ldots$$

$$f_\Theta(\theta)=\frac{\mu^s}{\Gamma(s)}\theta^{s-1}e^{-\mu\theta}$$

which resulted in the posterior distribution being a gamma pdf with parameters $w+s$ and $\mu+n$.

Suppose the loss function associated with $\hat{\theta}$ is quadratic, $L(\hat{\theta},\theta)=(\hat{\theta}-\theta)^2$. By part (b) of Theorem 5.8.1, the Bayes estimate for θ is the mean of the posterior distribution. From Theorem 4.6.3, though, the mean of $g_\Theta(\theta \mid W=w)$ is $(w+s)/(\mu+n)$.

Notice that

$$\frac{w+s}{\mu+n}=\frac{n}{\mu+n}\left(\frac{w}{n}\right)+\frac{\mu}{\mu+n}\left(\frac{s}{\mu}\right)$$

which shows that the Bayes estimate is a weighted average of $\frac{w}{n}$, the maximum likelihood estimate for θ and $\frac{s}{\mu}$, the mean of the prior distribution. Moreover, as n gets large, the Bayes estimate converges to the maximum likelihood estimate. ■

Questions

5.8.1. Suppose that X is a geometric random variable, where $p_X(k|\theta)=(1-\theta)^{k-1}\theta, k=1,2,\ldots$. Assume that the prior distribution for θ is the beta pdf with parameters r and s. Find the posterior distribution for θ.

5.8.2. Find the squared-error loss $[L(\hat{\theta},\theta)=(\hat{\theta}-\theta)^2]$ Bayes estimate for θ in Example 5.8.2 and express it as a weighted average of the maximum likelihood estimate for θ and the mean of the prior pdf.

5.8.3. Suppose the binomial pdf described in Example 5.8.2 refers to the number of votes a candidate might receive in a poll conducted before the general election. Moreover, suppose a beta prior distribution has been assigned to θ, and every indicator suggests the election will be close. The pollster, then, has good reason for concentrating the bulk of the prior distribution around the value $\theta=\frac{1}{2}$. Setting the two beta parameters r and s both equal to 135 will accomplish that objective (in the event $r=s=135$, the probability of θ being between 0.45 and 0.55 is approximately 0.90).

(a) Find the corresponding posterior distribution.

(b) Find the squared-error loss Bayes estimate for θ and express it as a weighted average of the maximum likelihood estimate for θ and the mean of the prior pdf.

5.8.4. What is the squared-error loss Bayes estimate for the parameter θ in a binomial pdf, where θ has a uniform distribution—that is, a noninformative prior? (Recall that a uniform prior is a beta pdf for which $r=s=1$.)

5.8.5. In Questions 5.8.2–5.8.4, is the Bayes estimate unbiased? Is it asymptotically unbiased?

5.8.6. Suppose that Y is a gamma random variable with parameters r and θ and the prior is also gamma with parameters s and μ. Show that the posterior pdf is gamma with parameters $r+s$ and $y+\mu$.

5.8.7. Let $Y_1, Y_2, \ldots, Y_n$ be a random sample from a gamma pdf with parameters r and θ, where the prior distribution assigned to θ is the gamma pdf with parameters s and μ. Let $W = Y_1 + Y_2 + \cdots + Y_n$. Find the posterior pdf for θ.

5.8.8. Find the squared-error loss Bayes estimate for θ in Question 5.8.7.

5.8.9. Consider, again, the scenario described in Example 5.8.2—a binomial random variable X has parameters n and θ, where the latter has a beta prior with integer parameters r and s. Integrate the joint pdf $p_X(k\,|\,\theta)f_\Theta(\theta)$ with respect to θ to show that the marginal pdf of X is given by

$$p_X(k) = \frac{\binom{k+r-1}{k}\binom{n-k+s-1}{n-k}}{\binom{n+r+s-1}{n}}, \quad k = 0, 1, \ldots, n$$

5.9 Taking a Second Look at Statistics (Beyond Classical Estimation)

The theory of estimation presented in this chapter can properly be called *classical*. It is a legacy of the late nineteenth and early twentieth centuries, culminating in the work of R.A. Fisher, especially his foundational paper published in 1922 (47).

This chapter covers the historical, yet still vibrant, theory and technique of estimation. This material is the basis for many of the modern advances in statistics. And, these approaches still provide useful methods for estimating parameters and building models.

But statistics, like every other branch of knowledge, progresses. As is the case for most sciences, the computer has dramatically changed the landscape. Classical problems—such as finding maximum likelihood estimators—that were difficult, if not impossible, to solve in Fisher's day can now be attacked through computer approximations.

However, modern computers not only give new methods for old problems, but they also provide new avenues of approach. One such set of new methods goes under the general name of *resampling*. One part of resampling is known as *bootstrapping*. This technique is useful when classical inference is impossible.

A general explication of bootstrapping is not possible in this section, but an example of its application to estimating the *standard* error should provide a sense of the idea.

The *standard error* of an estimator $\hat{\theta}$ is just its standard deviation; that is, $\sqrt{\text{Var}(\hat{\theta})}$. The standard error, or an approximation of it, is an essential part of the construction of confidence intervals. For the normal case, $\overline{Y}$ is the basis of the confidence interval, and its standard error is $\sigma/\sqrt{n}$. If X is a binomial random variable with parameters n and p, then the standard error $\sqrt{\frac{p(1-p)}{n}}$ is readily approximated by $\sqrt{\frac{\frac{k}{n}(1-\frac{k}{n})}{n}}$, where k is the observed number of successes.

In general, though, estimating the standard error may not be so straightforward. As a case in point, consider the gamma pdf with $r = 2$ and unknown parameter θ, $f_Y(y;\,\theta) = \frac{1}{\theta^2}ye^{-y/\theta}$. Recall from Example 5.2.2 that the maximum likelihood estimator for θ is $\frac{1}{2}\overline{Y}$. Then its variance is

$$\text{Var}(\frac{1}{2}\overline{Y}) = \frac{1}{4}\text{Var}(\overline{Y}) = \frac{1}{4}\frac{\text{Var}(\overline{Y})}{n} = \frac{1}{4n}2\theta^2 = \frac{\theta^2}{2n}$$

and the standard error is the square root of the variance, or $\frac{\theta}{\sqrt{2n}}$.

Comment Test statistics that yield small P-values should be interpreted as evidence *against* H_0. More specifically, if the P-value calculated for a test statistic is less than or equal to α, the null hypothesis can be rejected at the α level of significance. Or, put another way, the P-value is the smallest α at which we can reject H_0.

Example 6.2.2 Recall Example 6.2.1. Given that $H_0: \mu = 494$ is being tested against $H_1: \mu \neq 494$, what P-value is associated with the calculated test statistic, $z = 0.60$, and how should it be interpreted?

If $H_0: \mu = 494$ is true, the random variable $Z = \frac{\overline{Y}-494}{124/\sqrt{86}}$ has a standard normal pdf. Relative to the two-sided H_1, any value of Z greater than or equal to 0.60 *or* less than or equal to -0.60 qualifies as being "as extreme as or more extreme than" the observed z. Therefore, by Definition 6.2.4,

$$\begin{aligned} P\text{-value} &= P(Z \geq 0.60) + P(Z \leq -0.60) \\ &= 0.2743 + 0.2743 \\ &= 0.5486 \end{aligned}$$

(see Figure 6.2.6).

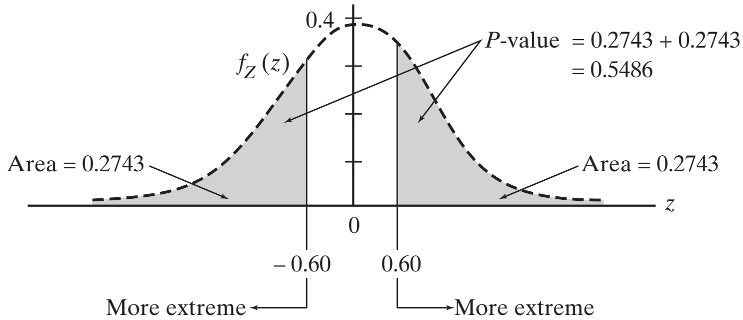


Figure 6.2.6

As noted in the preceding comment, P-values can be used as decision rules. In Example 6.2.1, 0.05 was the stated level of significance. Having determined here that the P-value associated with $z = 0.60$ is 0.5486, we know that $H_0: \mu = 494$ would *not* be rejected at the given α. Indeed, the null hypothesis would not be rejected for any value of α up to and including 0.5486.

Notice that the P-value would have been halved had H_1 been one-sided. Suppose we were confident that the new algebra and geometry classes would not *lower* a student's math SAT. The appropriate hypothesis test in that case would be $H_0: \mu = 494$ versus $H_1: \mu > 494$. Moreover, only values in the right-hand tail of $f_Z(z)$ would be considered more extreme than the observed $z = 0.60$, so

$$P\text{-value} = P(Z \geq 0.60) = 0.2743$$

■

Questions

6.2.1. State the decision rule that would be used to test the following hypotheses. Evaluate the appropriate test statistic and state your conclusion.

(a) $H_0: \mu = 120$ versus $H_1: \mu < 120$; $\overline{y} = 114.2$, $n = 25$, $\sigma = 18$, $\alpha = 0.08$

(b) $H_0: \mu = 42.9$ versus $H_1: \mu \neq 42.9$; $\overline{y} = 45.1, n = 16, \sigma = 3.2, \alpha = 0.01$

(c) $H_0: \mu = 14.2$ versus $H_1: \mu > 14.2$; $\overline{y} = 15.8, n = 9, \sigma = 4.1, \alpha = 0.13$

6.2.2. An herbalist is experimenting with juices extracted from berries and roots that may have the ability to affect the Stanford-Binet IQ scores of students afflicted with mild cases of attention deficit disorder (ADD). A random sample of twenty-two children diagnosed with the condition have been drinking Brain-Blaster daily for two months. Past experience suggests that children with ADD score an average of 95 on the IQ test with a standard deviation of 15. If the data are to be analyzed using the $\alpha = 0.06$ level of significance, what values of $\overline{y}$ would cause H_0 to be rejected? Assume that H_1 is two-sided.

6.2.3. (a) Suppose $H_0: \mu = \mu_o$ is rejected in favor of $H_1: \mu \neq \mu_o$ at the $\alpha = 0.05$ level of significance. Would H_0 necessarily be rejected at the $\alpha = 0.01$ level of significance?

(b) Suppose $H_0: \mu = \mu_o$ is rejected in favor of $H_1: \mu \neq \mu_o$ at the $\alpha = 0.01$ level of significance. Would H_0 necessarily be rejected at the $\alpha = 0.05$ level of significance?

6.2.4. Company records show that drivers get an average of 32,500 miles on a set of Road Hugger All-Weather radial tires. Hoping to improve that figure, the company has added a new polymer to the rubber that should help protect the tires from deterioration caused by extreme temperatures. Fifteen drivers who tested the new tires have reported getting an average of 33,800 miles. Can the company claim that the polymer has produced a statistically significant increase in tire mileage? Test $H_0: \mu = 32{,}500$ against a one-sided alternative at the $\alpha = 0.05$ level. Assume that the standard deviation (σ) of the tire mileages has not been affected by the addition of the polymer and is still 4000 miles.

6.2.5. If $H_0: \mu = \mu_o$ is rejected in favor of $H_1: \mu > \mu_o$, will it necessarily be rejected in favor of $H_1: \mu \neq \mu_o$? Assume that α remains the same.

6.2.6. A random sample of size 16 is drawn from a normal distribution having $\sigma = 6.0$ for the purpose of testing $H_0: \mu = 30$ versus $H_1: \mu \neq 30$. The experimenter chooses to define the critical region C to be the set of sample means lying in the interval (29.9, 30.1). What level of significance does the test have? Why is (29.9, 30.1) a poor choice for the critical region? What range of $\overline{y}$ values should comprise C, assuming the same α is to be used?

6.2.7. Recall the breath analyzers described in Example 4.3.5. The following are thirty blood alcohol determinations made by Analyzer GTE-10, a three-year-old unit that may be in need of recalibration. All thirty measurements were made using a test sample on which a properly adjusted machine would give a reading of 12.6%.

12.3	12.7	13.6	12.7	12.9	12.6
12.6	13.1	12.6	13.1	12.7	12.5
13.2	12.8	12.4	12.6	12.4	12.4
13.1	12.9	13.3	12.6	12.6	12.7
13.1	12.4	12.4	13.1	12.4	12.9

(a) If μ denotes the true average reading that Analyzer GTE-10 would give for a person whose blood alcohol concentration is 12.6%, test

$$H_0: \mu = 12.6$$

versus

$$H_1: \mu \neq 12.6$$

at the $\alpha = 0.05$ level of significance. Assume that $\sigma = 0.4$. Would you recommend that the machine be readjusted?

(b) What statistical assumptions are implicit in the hypothesis test done in part (a)? Is there any reason to suspect that those assumptions may not be satisfied?

6.2.8. Calculate the P-values for the hypothesis tests indicated in Question 6.2.1. Do they agree with your decisions on whether or not to reject H_0?

6.2.9. Suppose $H_0: \mu = 120$ is tested against $H_1: \mu \neq 120$. If $\sigma = 10$ and $n = 16$, what P-value is associated with the sample mean $\overline{y} = 122.3$? Under what circumstances would H_0 be rejected?

6.2.10 As a class research project, Rosaura wants to see whether the stress of final exams elevates the blood pressures of freshmen women. When they are not under any untoward duress, healthy eighteen-year-old women have systolic blood pressures that average 120 mm Hg with a standard deviation of 12 mm Hg. If Rosaura finds that the average blood pressure for the fifty women in Statistics 101 on the day of the final exam is 125.2, what should she conclude? Set up and test an appropriate hypothesis.

6.2.11. As input for a new inflation model, economists predicted that the average cost of a hypothetical "food basket" in east Tennessee in July would be \$145.75. The standard deviation (σ) of basket prices was assumed to be \$9.50, a figure that has held fairly constant over the years. To check their prediction, a sample of twenty-five baskets representing different parts of the region were checked in late July, and the average cost was \$149.75. Let $\alpha = 0.05$. Is the difference between the economists' prediction and the sample mean statistically significant?

where

$$k_i = \begin{cases} 0 & \text{if the new drug fails to relieve } i\text{th patient's pain} \\ 1 & \text{if the new drug does relieve } i\text{th patient's pain} \end{cases}$$

What should the decision rule be if the intention is to keep α somewhere near 10%? [Note that Theorem 6.3.1 does not apply here because Inequality 6.3.1 is not satisfied—specifically, $np_o + 3\sqrt{np_o(1 - p_o)} = 19(0.85) + 3\sqrt{19(0.85)(0.15)} = 20.8$ is not less than $n(= 19)$.]

If the null hypothesis is true, the expected number of successes would be $np_o = 19(0.85)$. or *16.2*. It follows that values of k to the extreme right or extreme left of 16.2 should constitute the critical region.

```
MTB > pdf;
SUBC > binomial 19   0.85.
```

Probability Density Function

Binomial with n = 19 and p = 0.85

x	P(X = x)	
6	0.000000	
7	0.000002	
8	0.000018	
9	0.000123	$\rightarrow P(X \leq 13) = 0.053696$
10	0.000699	
11	0.003242	
12	0.012246	
13	0.037366	
14	0.090746	
15	0.171409	
16	0.242829	
17	0.242829	
18	0.152892	
19	0.045599	$\rightarrow P(X = 19) = 0.045599$

Figure 6.3.1

Figure 6.3.1 is a Minitab printout of $p_X(k) = \binom{19}{k}(0.85)^k(0.15)^{19-k}$. By inspection, we can see that the critical region

$$C = \{k: k \leq 13 \quad \text{or} \quad k = 19\}$$

would produce an α close to the desired 0.10 (and would keep the probabilities associated with the two sides of the rejection region roughly the same). In random variable notation,

$$\begin{aligned} P(X \in C \mid H_0 \text{ is true}) &= P(X \leq 13 \mid p = 0.85) + P(X = 19 \mid p = 0.85) \\ &= 0.053696 + 0.045599 \\ &= 0.099295 \\ &\doteq 0.10 \end{aligned}$$

■

Questions

6.3.1. Commercial fishermen working certain parts of the Atlantic Ocean sometimes find their efforts hindered by the presence of whales. Ideally, they would like to scare away the whales without frightening the fish. One of the strategies being experimented with is to transmit underwater the sounds of a killer whale. On the fifty-two occasions that technique has been tried, it worked twenty-four times (that is, the whales immediately left the area). Experience has shown, though, that 40% of all whales sighted near fishing boats leave of their own accord, probably just to get away from the noise of the boat.

(a) Let $p = P$(Whale leaves area after hearing sounds of killer whale). Test H_0: $p = 0.40$ versus H_1: $p > 0.40$ at the $\alpha = 0.05$ level of significance. Can it be argued on the basis of these data that transmitting underwater predator sounds is an effective technique for clearing fishing waters of unwanted whales?

(b) Calculate the P-value for these data. For what values of α would H_0 be rejected?

6.3.2. Efforts to find a genetic explanation for why certain people are right-handed and others left-handed have been largely unsuccessful. Reliable data are difficult to find because of environmental factors that also influence a child's "handedness." To avoid that complication, researchers often study the analogous problem of "pawedness" in animals, where both genotypes and the environment can be partially controlled. In one such experiment (27), mice were put into a cage having a feeding tube that was equally accessible from the right or the left. Each mouse was then carefully watched over a number of feedings. If it used its right paw more than half the time to activate the tube, it was defined to be "right-pawed." Observations of this sort showed that 67% of mice belonging to strain A/J are right-pawed. A similar protocol was followed on a sample of thirty-five mice belonging to strain A/HeJ. Of those thirty-five, a total of eighteen were eventually classified as right-pawed. Test whether the proportion of right-pawed mice found in the A/HeJ sample was significantly different from what was known about the A/J strain. Use a two-sided alternative and let 0.05 be the probability associated with the critical region.

6.3.3. Defeated in his most recent attempt to win a congressional seat because of a sizeable gender gap, a politician has spent the last two years speaking out in favor of women's rights issues. A newly released poll claims to have contacted a random sample of 120 of the politician's current supporters and found that 72 were men. In the election that he lost, exit polls indicated that 65% of those who voted for him were men. Using an $\alpha = 0.05$ level of significance, test the null hypothesis that the proportion of his male supporters has remained the same. Make the alternative hypothesis one-sided.

6.3.4. Suppose H_0: $p = 0.45$ is to be tested against H_1: $p > 0.45$ at the $\alpha = 0.14$ level of significance, where $p = P(i\text{th}$ trial ends in success). If the sample size is 200, what is the smallest number of successes that will cause H_0 to be rejected?

6.3.5. Recall the median test described in Example 5.3.2. Reformulate that analysis as a hypothesis test rather than a confidence interval. What P-value is associated with the outcomes listed in Table 5.3.3?

6.3.6. Among the early attempts to revisit the death postponement theory introduced in Case Study 6.3.2 was an examination of the birth dates and death dates of 348 U.S. celebrities (134). It was found that 16 of those individuals had died in the month preceding their birth month. Set up and test the appropriate H_0 against a one-sided H_1. Use the 0.05 level of significance.

6.3.7. What α levels are possible with a decision rule of the form "Reject H_0 if $k \geq k^*$" when H_0: $p = 0.5$ is to be tested against H_1: $p > 0.5$ using a random sample of size $n = 7$?

6.3.8. The following is a Minitab printout of the binomial pdf $p_X(k) = \binom{9}{k}(0.6)^k(0.4)^{9-k}$, $k = 0, 1, \ldots, 9$. Suppose H_0: $p = 0.6$ is to be tested against H_1: $p > 0.6$ and we wish the level of significance to be *exactly* 0.05. Use Theorem 2.4.1 to combine two different critical regions into a single *randomized decision rule* for which $\alpha = 0.05$.

MTB > pdf;
SUBC > binomial 9 0.6.
Probability Density Function
Binomial with n = 9 and p = 0.6

x	P(X = x)
0	0.000262
1	0.003539
2	0.021234
3	0.074318
4	0.167215
5	0.250823
6	0.250823
7	0.161243
8	0.060466
9	0.010078

6.3.9. Suppose H_0: $p = 0.75$ is to be tested against H_1: $p < 0.75$ using a random sample of size $n = 7$ and the decision rule "Reject H_0 if $k \leq 3$."

(a) What is the test's level of significance?

(b) Graph the probability that H_0 will be rejected *as a function of p*.

6.4 Type I and Type II Errors

The possibility of drawing incorrect conclusions is an inevitable byproduct of hypothesis testing. No matter what sort of mathematical facade is laid atop the decision-making process, there is no way to guarantee that what the test tells us is the truth. One kind of error—rejecting H_0 when H_0 is true—figured prominently in Section 6.3: It was argued that critical regions should be defined so as to keep the probability of making such errors small, often on the order of 0.05.

By inspection, the decision rule "Reject $H_0: \lambda = 0.8$ if $\sum_{i=1}^{4} k_i \geq 6$" gives an α close to the desired 0.10.

If H_1 is true and $\lambda = 1.2$, $\sum_{i=1}^{4} X_i$ will have a Poisson distribution with a parameter equal to 4.8. According to Figure 6.4.10, the probability that the sum of a random sample of size 4 from such a distribution would equal or exceed 6 (i.e., $1 - \beta$ when $\lambda = 1.2$) is *0.348993*. ■

Example 6.4.4

Suppose a random sample of seven observations is taken from the pdf $f_Y(y;\theta) = (\theta+1)y^{\theta}, 0 \leq y \leq 1$, to test

$$H_0: \theta = 2$$

versus

$$H_1: \theta > 2$$

As a decision rule, the experimenter plans to record X, the number of y_i's that exceed 0.9, and reject H_0 if $X \geq 4$. What proportion of the time would such a decision rule lead to a Type I error?

To evaluate $\alpha = P(\text{Reject } H_0 \mid H_0 \text{ is true})$, we first need to recognize that X is a binomial random variable where $n = 7$ and the parameter p is an area under $f_Y(y; \theta = 2)$:

$$\begin{aligned} p = P(Y \geq 0.9 \mid H_0 \text{ is true}) &= P[Y \geq 0.9 \mid f_Y(y;2) = 3y^2] \\ &= \int_{0.9}^{1} 3y^2\, dy \\ &= 0.271 \end{aligned}$$

It follows, then, that H_0 will be incorrectly rejected *9.2%* of the time:

$$\begin{aligned} \alpha = P(X \geq 4 \mid \theta = 2) &= \sum_{k=4}^{7} \binom{7}{k} (0.271)^k (0.729)^{7-k} \\ &= 0.092 \end{aligned}$$

Comment The basic notions of Type I and Type II errors first arose in a quality-control context. The pioneering work was done at the Bell Telephone Laboratories: There the terms *producer's risk* and *consumer's risk* were introduced for what we now call α and β. Eventually, these ideas were generalized by Neyman and Pearson in the 1930s and evolved into the theory of hypothesis testing as we know it today. ■

Questions

6.4.1. Recall the "Math for the Twenty-First Century" hypothesis test done in Example 6.2.1. Calculate the power of that test when the true mean is 500.

6.4.2. Carry out the details to verify the decision rule change cited on p. 371 in connection with Figure 6.4.6.

6.4.3. For the decision rule found in Question 6.2.2 to test $H_0: \mu = 95$ versus $H_1: \mu \neq 95$ at the $\alpha = 0.06$ level of significance, calculate $1 - \beta$ when $\mu = 90$.

6.4.4. Construct a power curve for the $\alpha = 0.05$ test of $H_0: \mu = 60$ versus $H_1: \mu \neq 60$ if the data consist of a random sample of size 16 from a normal distribution having $\sigma = 4$.

6.4.5. If $H_0: \mu = 240$ is tested against $H_1: \mu < 240$ at the $\alpha = 0.01$ level of significance with a random sample of twenty-five normally distributed observations, what proportion of the time will the procedure fail to recognize that μ has dropped to 220? Assume that $\sigma = 50$.

6.4.6. Suppose $n = 36$ observations are taken from a normal distribution where $\sigma = 8.0$ for the purpose of testing $H_0: \mu = 60$ versus $H_1: \mu \neq 60$ at the $\alpha = 0.07$ level of significance. The lead investigator skipped statistics class the day decision rules were being discussed and intends to reject H_0 if $\overline{y}$ falls in the region $(60 - \overline{y}^*, 60 + \overline{y}^*)$.

(a) Find $\overline{y}^*$.
(b) What is the power of the test when $\mu = 62$?
(c) What would the power of the test be when $\mu = 62$ if the critical region had been defined the correct way?

6.4.7. If $H_0: \mu = 200$ is to be tested against $H_1: \mu < 200$ at the $\alpha = 0.10$ level of significance based on a random sample of size n from a normal distribution where $\sigma = 15.0$, what is the smallest value for n that will make the power equal to at least 0.75 when $\mu = 197$?

6.4.8. Will $n = 45$ be a sufficiently large sample to test $H_0: \mu = 10$ versus $H_1: \mu \neq 10$ at the $\alpha = 0.05$ level of significance if the experimenter wants the Type II error probability to be no greater than 0.20 when $\mu = 12$? Assume that $\sigma = 4$.

6.4.9. If $H_0: \mu = 30$ is tested against $H_1: \mu > 30$ using $n = 16$ observations (normally distributed) and if $1 - \beta = 0.85$ when $\mu = 34$, what does α equal? Assume that $\sigma = 9$.

6.4.10. Suppose a sample of size 1 is taken from the pdf $f_Y(y) = (1/\lambda)e^{-y/\lambda}$, $y > 0$, for the purpose of testing

$$H_0: \lambda = 1$$
$$\text{versus}$$
$$H_1: \lambda > 1$$

The null hypothesis will be rejected if $y \geq 3.20$.

(a) Calculate the probability of committing a Type I error.
(b) Calculate the probability of committing a Type II error when $\lambda = \frac{4}{3}$.
(c) Draw a diagram that shows the α and β calculated in parts (a) and (b) as areas.

6.4.11. Polygraphs used in criminal investigations typically measure five bodily functions: (1) thoracic respiration, (2) abdominal respiration, (3) blood pressure and pulse rate, (4) muscular movement and pressure, and (5) galvanic skin response. In principle, the magnitude of these responses when the subject is asked a relevant question ("Did you murder your wife?") indicate whether he is lying or telling the truth. The procedure, of course, is not infallible, as a recent study bore out (82). Seven experienced polygraph examiners were given a set of forty records—twenty were from innocent suspects and twenty from guilty suspects. The subjects had been asked eleven questions, on the basis of which each examiner was to make an overall judgment: "Innocent" or "Guilty." The results are as follows:

		Suspect's True Status	
		Innocent	Guilty
Examiner's	"Innocent"	131	15
Decision	"Guilty"	9	125

What would be the numerical values of α and β in this context? In a judicial setting, should Type I and Type II errors carry equal weight? Explain.

6.4.12. An urn contains ten chips. An unknown number of the chips are white; the others are red. We wish to test

$$H_0: \text{exactly half the chips are white}$$
$$\text{versus}$$
$$H_1: \text{more than half the chips are white}$$

We will draw, without replacement, three chips and reject H_0 if two or more are white. Find α. Also, find β when the urn is (a) 60% white and (b) 70% white.

6.4.13. Suppose that a random sample of size 5 is drawn from a uniform pdf:

$$f_Y(y;\theta) = \begin{cases} \frac{1}{\theta}, & 0 < y < \theta \\ 0, & \text{elsewhere} \end{cases}$$

We wish to test

$$H_0: \theta = 2$$
$$\text{versus}$$
$$H_1: \theta > 2$$

by rejecting the null hypothesis if $y_{\max} \geq k$. Find the value of k that makes the probability of committing a Type I error equal to 0.05.

6.4.14. A sample of size 1 is taken from the pdf

$$f_Y(y) = (\theta + 1)y^{\theta}, \quad 0 \leq y \leq 1$$

The hypothesis $H_0: \theta = 1$ is to be rejected in favor of $H_1: \theta > 1$ if $y \geq 0.90$. What is the test's level of significance?

6.4.15. A series of n Bernoulli trials is to be observed as data for testing

$$H_0: p = \tfrac{1}{2}$$
$$\text{versus}$$
$$H_1: p > \tfrac{1}{2}$$

The null hypothesis will be rejected if k, the observed number of successes, equals n. For what value of p will the probability of committing a Type II error equal 0.05?

6.4.16. Let X_1 be a binomial random variable with $n = 2$ and $p_{X_1} = P(\text{success})$. Let X_2 be an independent binomial random variable with $n = 4$ and $p_{X_2} = P(\text{success})$. Let $X = X_1 + X_2$. Calculate α if

$$H_0\colon p_{X_1} = p_{X_2} = \tfrac{1}{2}$$
$$\text{versus}$$
$$H_1\colon p_{X_1} = p_{X_2} > \tfrac{1}{2}$$

is to be tested by rejecting the null hypothesis when $k \geq 5$.

6.4.17. A sample of size 1 from the pdf $f_Y(y) = (1+\theta)y^\theta$, $0 \leq y \leq 1$, is to be the basis for testing

$$H_0\colon \theta = 1$$
$$\text{versus}$$
$$H_1\colon \theta < 1$$

The critical region will be the interval $y \leq \frac{1}{2}$. Find an expression for $1-\beta$ as a function of θ.

6.4.18. An experimenter takes a sample of size 1 from the Poisson probability model, $p_X(k) = e^{-\lambda}\lambda^k/k!$, $k = 0, 1, 2, \ldots$, and wishes to test

$$H_0\colon \lambda = 6$$
$$\text{versus}$$
$$H_1\colon \lambda < 6$$

by rejecting H_0 if $k \leq 2$.

(a) Calculate the probability of committing a Type I error.

(b) Calculate the probability of committing a Type II error when $\lambda = 4$.

6.4.19. A sample of size 1 is taken from the geometric probability model, $p_X(k) = (1-p)^{k-1}p$, $k = 1, 2, 3, \ldots$, to test $H_0\colon p = \frac{1}{3}$ versus $H_1\colon p > \frac{1}{3}$. The null hypothesis is to be rejected if $k \geq 4$. What is the probability that a Type II error will be committed when $p = \frac{1}{2}$?

6.4.20. Suppose that one observation from the exponential pdf, $f_Y(y) = \lambda e^{-\lambda y}$, $y > 0$, is to be used to test $H_0\colon \lambda = 1$ versus $H_1\colon \lambda < 1$. The decision rule calls for the null hypothesis to be rejected if $y \geq \ln 10$. Find β as a function of λ.

6.4.21. A random sample of size 2 is drawn from a uniform pdf defined over the interval $[0, \theta]$. We wish to test

$$H_0\colon \theta = 2$$
$$\text{versus}$$
$$H_1\colon \theta < 2$$

by rejecting H_0 when $y_1 + y_2 \leq k$. Find the value for k that gives a level of significance of 0.05.

6.4.22. Suppose that the hypotheses of Question 6.4.21 are to be tested with a decision rule of the form "Reject $H_0\colon \theta = 2$ if $y_1 y_2 \leq k^*$." Find the value of k^* that gives a level of significance of 0.05 (see Theorem 3.8.5).

6.5 A Notion of Optimality: The Generalized Likelihood Ratio

In the next several chapters we will be studying some of the particular hypothesis tests that statisticians most often use in dealing with real-world problems. All of these have the same conceptual heritage—a fundamental notion known as the *generalized likelihood ratio,* or *GLR*. More than just a principle, the generalized likelihood ratio is a working criterion for actually *suggesting* test procedures.

As a first look at this important idea, we will conclude Chapter 6 with an application of the generalized likelihood ratio to the problem of testing the parameter θ in a uniform pdf. Notice the relationship here between the likelihood ratio and the definition of an "optimal" hypothesis test.

Suppose $y_1, y_2, \ldots, y_n$ is a random sample from a uniform pdf over the interval $[0, \theta]$, where θ is unknown, and our objective is to test

$$H_0\colon \theta = \theta_o$$
$$\text{versus}$$
$$H_1\colon \theta < \theta_o$$

at a specified level of significance α. What is the "best" decision rule for choosing between H_0 and H_1, and by what criterion is it considered optimal?