the probability that two or more of these businesses will be audited?

3.2.5. The probability is 0.10 that ball bearings in a machine component will fail under certain adverse conditions of load and temperature. If a component containing eleven ball bearings must have a least eight of them functioning to operate under the adverse conditions, what is the probability that it will break down?

3.2.6. Suppose that since the early 1950s some ten-thousand independent UFO sightings have been reported to civil authorities. If the probability that any sighting is genuine is on the order of one in one hundred thousand, what is the probability that at least one of the ten-thousand was genuine?

3.2.7. Doomsday Airlines ("Come Take the Flight of Your Life") has two dilapidated airplanes, one with two engines, and the other with four. Each plane will land safely only if at least half of its engines are working. Each engine on each aircraft operates independently and each has probability $p = 0.4$ of failing. Assuming you wish to maximize your survival probability, which plane should you fly on?

3.2.8. Two lighting systems are being proposed for an employee work area. One requires fifty bulbs, each having a probability of 0.05 of burning out within a month's time. The second has one hundred bulbs, each with a 0.02 burnout probability. Whichever system is installed will be inspected once a month for the purpose of replacing burned-out bulbs. Which system is likely to require less maintenance? Answer the question by comparing the probabilities that each will require at least one bulb to be replaced at the end of thirty days.

3.2.9. The great English diarist Samuel Pepys asked his friend Sir Isaac Newton the following question: Is it more likely to get at least one 6 when six dice are rolled, at least two 6's when twelve dice are rolled, or at least three 6's when eighteen dice are rolled? After considerable correspondence [see (158)]. Newton convinced the skeptical Pepys that the first event is the most likely. Compute the three probabilities.

3.2.10. The gunner on a small assault boat fires six missiles at an attacking plane. Each has a 20% chance of being on-target. If two or more of the shells find their mark, the plane will crash. At the same time, the pilot of the plane fires ten air-to-surface rockets, each of which has a 0.05 chance of critically disabling the boat. Would you rather be on the plane or the boat?

3.2.11. If a family has four children, is it more likely they will have two boys and two girls or three of one sex and one of the other? Assume that the probability of a child being a boy is $\frac{1}{2}$ and that the births are independent events.

3.2.12. Experience has shown that only $\frac{1}{3}$ of all patients having a certain disease will recover if given the standard treatment. A new drug is to be tested on a group of twelve volunteers. If the FDA requires that at least seven of these patients recover before it will license the new drug, what is the probability that the treatment will be discredited even if it has the potential to increase an individual's recovery rate to $\frac{1}{2}$?

3.2.13. Transportation to school for a rural county's seventy-six children is provided by a fleet of four buses. Drivers are chosen on a day-to-day basis and come from a pool of local farmers who have agreed to be "on call." What is the smallest number of drivers who need to be in the pool if the county wants to have at least a 95% probability on any given day that all the buses will run? Assume that each driver has an 80% chance of being available if contacted.

3.2.14. The captain of a Navy gunboat orders a volley of twenty-five missiles to be fired at random along a five-hundred-foot stretch of shoreline that he hopes to establish as a beachhead. Dug into the beach is a thirty-foot-long bunker serving as the enemy's first line of defense. The captain has reason to believe that the bunker will be destroyed if at least three of the missiles are on-target. What is the probability of that happening?

3.2.15. A computer has generated seven random numbers over the interval 0 to 1. Is it more likely that (a) exactly three will be in the interval $\frac{1}{2}$ to 1 or (b) fewer than three will be greater than $\frac{3}{4}$?

3.2.16. Listed in the following table is the length distribution of World Series competition for the 58 series from 1950 to 2008 (there was no series in 1994).

World Series Lengths

Number of Games, X	Number of Years
4	12
5	10
6	12
7	24
	58

Source: espn.go.com/mlb/worldseries/history/winners

Assuming that each World Series game is an independent event and that the probability of either team's winning any particular contest is 0.5, find the probability of each series length. How well does the model fit the data? (Compute the "expected" frequencies, that is, multiply the probability of a given-length series times 58).

Comment Every sampling plan invariably allows for two kinds of errors—rejecting shipments that should be accepted and accepting shipments that should be rejected. In practice, the probabilities of committing these errors can be manipulated by redefining the decision rule and/or changing the sample size. Some of these options will be explored in Chapter 6.

Questions

3.2.20. A corporate board contains twelve members. The board decides to create a five-person Committee to Hide Corporation Debt. Suppose four members of the board are accountants. What is the probability that the Committee will contain two accountants and three nonaccountants?

3.2.21. One of the popular tourist attractions in Alaska is watching black bears catch salmon swimming upstream to spawn. Not all "black" bears are black, though—some are tan-colored. Suppose that six black bears and three tan-colored bears are working the rapids of a salmon stream. Over the course of an hour, six different bears are sighted. What is the probability that those six include at least twice as many black bears as tan-colored bears?

3.2.22. A city has 4050 children under the age of ten, including 514 who have not been vaccinated for measles. Sixty-five of the city's children are enrolled in the ABC Day Care Center. Suppose the municipal health department sends a doctor and a nurse to ABC to immunize any child who has not already been vaccinated. Find a formula for the probability that exactly k of the children at ABC have not been vaccinated.

3.2.23. Country A inadvertently launches ten guided missiles—six armed with nuclear warheads—at Country B. In response, Country B fires seven antiballistic missiles, each of which will destroy exactly one of the incoming rockets. The antiballistic missiles have no way of detecting, though, which of the ten rockets are carrying nuclear warheads. What are the chances that Country B will be hit by at least one nuclear missile?

3.2.24. Anne is studying for a history exam covering the French Revolution that will consist of five essay questions selected at random from a list of ten the professor has handed out to the class in advance. Not exactly a Napoleon buff, Anne would like to avoid researching all ten questions but still be reasonably assured of getting a fairly good grade. Specifically, she wants to have at least an 85% chance of getting at least four of the five questions right. Will it be sufficient if she studies eight of the ten questions?

3.2.25. Each year a college awards five merit-based scholarships to members of the entering freshman class who have exceptional high school records. The initial pool of applicants for the upcoming academic year has been reduced to a "short list" of eight men and ten women, all of whom seem equally deserving. If the awards are made at random from among the eighteen finalists, what are the chances that both men and women will be represented?

3.2.26. Keno is a casino game in which the player has a card with the numbers 1 through 80 on it. The player selects a set of k numbers from the card, where k can range from one to fifteen. The "caller" announces twenty winning numbers, chosen at random from the eighty. The amount won depends on how many of the called numbers match those the player chose. Suppose the player picks ten numbers. What is the probability that among those ten are six winning numbers?

3.2.27. A display case contains thirty-five gems, of which ten are real diamonds and twenty-five are fake diamonds. A burglar removes four gems at random, one at a time and without replacement. What is the probability that the last gem she steals is the second real diamond in the set of four?

3.2.28. A bleary-eyed student awakens one morning, late for an 8:00 class, and pulls two socks out of a drawer that contains two black, six brown, and two blue socks, all randomly arranged. What is the probability that the two he draws are a matched pair?

3.2.29. Show directly that the set of probabilities associated with the hypergeometric distribution sum to 1. (*Hint:* Expand the identity

$$(1+\mu)^N = (1+\mu)^r(1+\mu)^{N-r}$$

and equate coefficients.)

3.2.30. Show that the ratio of two successive hypergeometric probability terms satisfies the following equation,

$$\frac{\binom{r}{k+1}\binom{w}{n-k-1}}{\binom{N}{n}} \div \frac{\binom{r}{k}\binom{w}{n-k}}{\binom{N}{n}} = \frac{n-k}{k+1}\cdot\frac{r-k}{w-n+k+1}$$

for any k where both numerators are defined.

As it turns out, values of the cdf for a binomial random variable are widely available, both in books and in computer software. Here, for example, $F_X(40) = 0.9992$ and $F_X(20) = 0.0034$, so

$$\begin{aligned} P(21 \le X \le 40) &= 0.9992 - 0.0034 \\ &= 0.9958 \end{aligned}$$

Example 3.3.10 Suppose that two fair dice are rolled. Let the random variable X denote the larger of the two faces showing: (a) Find $F_X(t)$ for $t = 1, 2, \ldots, 6$ and (b) Find $F_X(2.5)$.

a. The sample space associated with the experiment of rolling two fair dice is the set of ordered pairs $s = (i, j)$, where the face showing on the first die is i and the face showing on the second die is j. By assumption, all thirty-six possible outcomes are equally likely. Now, suppose t is some integer from 1 to 6, inclusive. Then

$$\begin{aligned} F_X(t) &= P(X \le t) \\ &= P[\text{Max}\,(i, j) \le t] \\ &= P(i \le t \quad \text{and} \quad j \le t) \qquad \text{(why?)} \\ &= P(i \le t) \cdot P(j \le t) \qquad \text{(why?)} \\ &= \frac{t}{6} \cdot \frac{t}{6} \\ &= \frac{t^2}{36}, \qquad t = 1, 2, 3, 4, 5, 6 \end{aligned}$$

b. Even though the random variable X has nonzero probability only for the integers 1 through 6, the cdf is defined for *any* real number from $-\infty$ to $+\infty$. By definition, $F_X(2.5) = P(X \le 2.5)$. But

$$\begin{aligned} P(X \le 2.5) &= P(X \le 2) + P(2 < X \le 2.5) \\ &= F_X(2) + 0 \end{aligned}$$

so

$$F_X(2.5) = F_X(2) = \frac{2^2}{36} = \frac{1}{9}$$

What would the graph of $F_X(t)$ as a function of t look like?

Questions

3.3.1. An urn contains five balls numbered 1 to 5. Two balls are drawn simultaneously.

(a) Let X be the larger of the two numbers drawn. Find $p_X(k)$.

(b) Let V be the sum of the two numbers drawn. Find $p_V(k)$.

3.3.2. Repeat Question 3.3.1 for the case where the two balls are drawn *with replacement*.

3.3.3. Suppose a fair die is tossed three times. Let X be the largest of the three faces that appear. Find $p_X(k)$.

3.3.4. Suppose a fair die is tossed three times. Let X be the number of different faces that appear (so $X = 1, 2,$ or 3). Find $p_X(k)$.

3.3.5. A fair coin is tossed three times. Let X be the number of heads in the tosses minus the number of tails. Find $p_X(k)$.

3.3.6. Suppose die one has spots $1, 2, 2, 3, 3, 4$ and die two has spots $1, 3, 4, 5, 6, 8$. If both dice are rolled, what is the sample space? Let $X =$ total spots showing. Show that the pdf for X is the same as for normal dice.

3.3.7. Suppose a particle moves along the x-axis beginning at 0. It moves one integer step to the left or right with equal probability. What is the pdf of its position after four steps?

3.3.8. How would the pdf asked for in Question 3.3.7 be affected if the particle was twice as likely to move to the right as to the left?

3.3.9. Suppose that five people, including you and a friend, line up at random. Let the random variable X denote the number of people standing between you and your friend. What is $p_X(k)$?

3.3.10. Urn I and urn II each have two red chips and two white chips. Two chips are drawn simultaneously from each urn. Let X_1 be the number of red chips in the first sample and X_2 the number of red chips in the second sample. Find the pdf of $X_1 + X_2$.

3.3.11. Suppose X is a binomial random variable with $n = 4$ and $p = \frac{2}{3}$. What is the pdf of $2X + 1$?

3.3.12. Find the cdf for the random variable X in Question 3.3.3.

3.3.13. A fair die is rolled four times. Let the random variable X denote the number of 6's that appear. Find and graph the cdf for X.

3.3.14. At the points $x = 0, 1, \ldots, 6$, the cdf for the discrete random variable X has the value $F_X(x) = x(x+1)/42$. Find the pdf for X.

3.3.15. Find the pdf for the discrete random variable X whose cdf at the points $x = 0, 1, \ldots, 6$ is given by $F_X(x) = x^3/216$.

3.4 Continuous Random Variables

The statement was made in Chapter 2 that all sample spaces belong to one of two generic types—*discrete* sample spaces are ones that contain a finite or a countably infinite number of outcomes and *continuous* sample spaces are those that contain an uncountably infinite number of outcomes. Rolling a pair of dice and recording the faces that appear is an experiment with a discrete sample space; choosing a number at random from the interval [0, 1] would have a continuous sample space.

How we assign probabilities to these two types of sample spaces is different. Section 3.3 focused on discrete sample spaces. Each outcome s is assigned a probability by the discrete probability function $p(s)$. If a random variable X is defined on the sample space, the probabilities associated with its outcomes are assigned by the probability density function $p_X(k)$. Applying those same definitions, though, to the outcomes in a continuous sample space will not work. The fact that a continuous sample space has an uncountably infinite number of outcomes eliminates the option of assigning a probability to each point as we did in the discrete case with the function $p(s)$. We begin this section with a particular pdf defined on a discrete sample space that suggests how we might define probabilities, in general, on a continuous sample space.

Suppose an electronic surveillance monitor is turned on briefly at the beginning of every hour and has a 0.905 probability of working properly, regardless of how long it has remained in service. If we let the random variable X denote the hour at which the monitor first fails, then $p_X(k)$ is the product of k individual probabilities:

$$\begin{aligned} p_X(k) &= P(X = k) = P(\text{Monitor fails for the first time at the } k\text{th hour}) \\ &= P(\text{Monitor functions properly for first } k-1 \text{ hours} \cap \text{Monitor fails at the } k\text{th hour}) \\ &= (0.905)^{k-1}(0.095), \quad k = 1, 2, 3, \ldots \end{aligned}$$

Figure 3.4.1 shows a probability histogram of $p_X(k)$ for k values ranging from 1 to 21. Here the height of the kth bar is $p_X(k)$, and since the width of each bar is 1, the *area* of the kth bar is also $p_X(k)$.

Now, look at Figure 3.4.2, where the exponential curve $y = 0.1e^{-0.1x}$ is superimposed on the graph of $p_X(k)$. Notice how closely the area under the curve approximates the area of the bars. It follows that the probability that X lies in some

$F_Y(y_n) = P(\cup_{k=1}^{n} A_k) = \sum_{k=1}^{n} P(A_k)$, since the A_k's are disjoint. Also, the sample space $S = \cup_{k=1}^{\infty} A_k$, and by Axiom 4, $1 = P(S) = P(\cup_{k=1}^{\infty} A_k) = \sum_{k=1}^{\infty} P(A_k)$. Putting these equalities together gives $1 = \sum_{k=0}^{\infty} P(A_k) = \lim_{n\to\infty} \sum_{k=0}^{n} P(A_k) = \lim_{n\to\infty} F_Y(y_n)$.

d.
$$\begin{aligned}\lim_{y\to-\infty} F_Y(y) &= \lim_{y\to-\infty} P(Y \le y) = \lim_{y\to-\infty} P(-Y \ge -y) = \lim_{y\to-\infty} [1 - P(-Y \le -y)] \\ &= 1 - \lim_{y\to-\infty} P(-Y \le -y) = 1 - \lim_{y\to\infty} P(-Y \le y) \\ &= 1 - \lim_{y\to\infty} F_{-Y}(y) = 0\end{aligned}$$
□

Questions

3.4.1. Suppose $f_Y(y) = 4y^3$, $0 \le y \le 1$. Find $P\left(0 \le Y \le \frac{1}{2}\right)$.

3.4.2. For the random variable Y with pdf $f_Y(y) = \frac{2}{3} + \frac{2}{3}y$, $0 \le y \le 1$, find $P\left(\frac{3}{4} \le Y \le 1\right)$.

3.4.3. Let $f_Y(y) = \frac{3}{2}y^2$, $-1 \le y \le 1$. Find $P\left(|Y - \frac{1}{2}| < \frac{1}{4}\right)$. Draw a graph of $f_Y(y)$ and show the area representing the desired probability.

3.4.4. For persons infected with a certain form of malaria, the length of time spent in remission is described by the continuous pdf $f_Y(y) = \frac{1}{9}y^2$, $0 \le y \le 3$, where Y is measured in years. What is the probability that a malaria patient's remission lasts longer than one year?

3.4.5. The length of time, Y, that a customer spends in line at a bank teller's window before being served is described by the exponential pdf $f_Y(y) = 0.2e^{-0.2y}$, $y \ge 0$.

(a) What is the probability that a customer will wait more than ten minutes?
(b) Suppose the customer will leave if the wait is more than ten minutes. Assume that the customer goes to the bank twice next month. Let the random variable X be the number of times the customer leaves without being served. Calculate $p_X(1)$.

3.4.6. Let n be a positive integer. Show that $f_Y(y) = (n+2)(n+1)y^n(1-y)$, $0 \le y \le 1$, is a pdf.

3.4.7. Find the cdf for the random variable Y given in Question 3.4.1. Calculate $P\left(0 \le Y \le \frac{1}{2}\right)$ using $F_Y(y)$.

3.4.8. If Y is an exponential random variable, $f_Y(y) = \lambda e^{-\lambda y}$, $y \ge 0$, find $F_Y(y)$.

3.4.9. If the pdf for Y is

$$f_Y(y) = \begin{cases} 0, & |y| > 1 \\ 1 - |y|, & |y| \le 1 \end{cases}$$

find and graph $F_Y(y)$.

3.4.10. A continuous random variable Y has a cdf given by

$$F_Y(y) = \begin{cases} 0 & y < 0 \\ y^2 & 0 \le y < 1 \\ 1 & y \ge 1 \end{cases}$$

Find $P\left(\frac{1}{2} < Y \le \frac{3}{4}\right)$ two ways—first, by using the cdf and second, by using the pdf.

3.4.11. A random variable Y has cdf

$$F_Y(y) = \begin{cases} 0 & y < 1 \\ \ln y & 1 \le y \le e \\ 1 & e < y \end{cases}$$

Find

(a) $P(Y < 2)$
(b) $P\left(2 < Y \le 2\frac{1}{2}\right)$
(c) $P\left(2 < Y < 2\frac{1}{2}\right)$
(d) $f_Y(y)$

3.4.12. The cdf for a random variable Y is defined by $F_Y(y) = 0$ for $y < 0$; $F_Y(y) = 4y^3 - 3y^4$ for $0 \le y \le 1$; and $F_Y(y) = 1$ for $y > 1$. Find $P\left(\frac{1}{4} \le Y \le \frac{3}{4}\right)$ by integrating $f_Y(y)$.

3.4.13. Suppose $F_Y(y) = \frac{1}{12}(y^2 + y^3)$, $0 \le y \le 2$. Find $f_Y(y)$.

3.4.14. In a certain country, the distribution of a family's disposable income, Y, is described by the pdf $f_Y(y) = ye^{-y}$, $y \ge 0$. Find $F_Y(y)$.

3.4.15. The logistic curve $F(y) = \frac{1}{1+e^{-y}}$, $-\infty < y < \infty$, can represent a cdf since it is increasing, $\lim_{y\to-\infty} \frac{1}{1+e^{-y}} = 0$, and $\lim_{y\to+\infty} \frac{1}{1+e^{-y}} = 1$. Verify these three assertions and also find the associated pdf.

3.4.16. Let Y be the random variable described in Question 3.4.1. Define $W = 2Y$. Find $f_W(w)$. For which values of w is $f_W(w) \ne 0$?

3.4.17. Suppose that $f_Y(y)$ is a continuous and symmetric pdf, where *symmetry* is the property that $f_Y(y) = f_Y(-y)$ for all y. Show that $P(-a \le Y \le a) = 2F_Y(a) - 1$.

3.4.18. Let Y be a random variable denoting the age at which a piece of equipment fails. In reliability theory, the probability that an item fails at time y given that it has survived until time y is called the *hazard rate*, $h(y)$. In terms of the pdf and cdf,

$$h(y) = \frac{f_Y(y)}{1 - F_Y(y)}$$

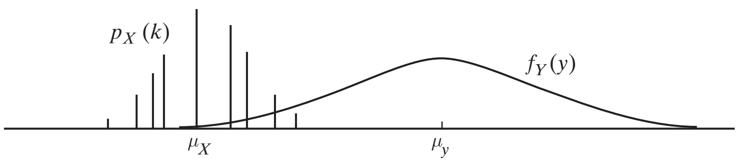
Find $h(y)$ if Y has an exponential pdf (see Question 3.4.8).

3.5 Expected Values

Probability density functions, as we have already seen, provide a global overview of a random variable's behavior. If X is discrete, $p_X(k)$ gives $P(X = k)$ for all k; if Y is continuous, and A is any interval or a countable union of intervals, $P(Y \epsilon A) = \int_A f_Y(y)\,dy$. Detail that explicit, though, is not always necessary—or even helpful. There are times when a more prudent strategy is to focus the information contained in a pdf by summarizing certain of its features with single numbers.

The first such feature that we will examine is *central tendency,* a term referring to the "average" value of a random variable. Consider the pdfs $p_X(k)$ and $f_Y(y)$ pictured in Figure 3.5.1. Although we obviously cannot predict with certainty what values any future X's and Y's will take on, it seems clear that X values will tend to lie somewhere near μ_X, and Y values somewhere near μ_Y. In some sense, then, we can characterize $p_X(k)$ by μ_X, and $f_Y(y)$ by μ_Y.

Figure 3.5.1



The most frequently used measure for describing central tendency—that is, for quantifying μ_X and μ_Y—is the *expected value.* Discussed at some length in this section and in Section 3.9, the expected value of a random variable is a slightly more abstract formulation of what we are already familiar with in simple discrete settings as the arithmetic average. Here, though, the values included in the average are "weighted" by the pdf.

Gambling affords a familiar illustration of the notion of an expected value. Consider the game of roulette. After bets are placed, the croupier spins the wheel and declares one of thirty-eight numbers, 00, 0, 1, 2, …, 36, to be the winner. Disregarding what seems to be a perverse tendency of many roulette wheels to land on numbers for which no money has been wagered, we will assume that each of these thirty-eight numbers is equally likely (although only the eighteen numbers 1, 3, 5, …, 35 are considered to be odd and only the eighteen numbers 2, 6, 4, …, 36 are considered to be even). Suppose that our particular bet (at "even money") is \$1 on odds. If the random variable X denotes our winnings, then X takes on the value *1* if an odd number occurs, and -1 otherwise. Therefore,

$$p_X(1) = P(X = 1) = \frac{18}{38} = \frac{9}{19}$$