Chapter

3 RANDOM VARIABLES

CHAPTER OUTLINE

One of a Swiss family producing eight distinguished scientists, Jakob (Jacques) Bernoulli (1654–1705) was forced by his father to pursue theological studies, but his love of mathematics eventually led him to a university career. He and his brother, Johann, were the most prominent champions of Leibniz's calculus on continental Europe, the two using the new theory to solve numerous problems in physics and mathematics. Bernoulli's main work in probability, Ars Conjectandi, *was published after his death by his nephew, Nikolaus, in 1713.*

3.1 INTRODUCTION

Throughout Chapter 2, probabilities were assigned to *events*—that is, to sets of sample outcomes. The events we dealt with were composed of either a finite or a countably infinite number of sample outcomes, in which case the event's probability was simply the sum of the probabilities assigned to its outcomes. One particular probability function that came up over and over again in Chapter 2 was the assignment of $\frac{1}{n}$ as the probability associated with each of the n points in a finite sample space. This is the model that typically describes games of chance (and all of our combinatorial probability problems in Chapter 2).

The first objective of this chapter is to look at several other useful ways for assigning probabilities to sample outcomes. In so doing, we confront the desirability of "redefining" sample spaces using functions known as *random variables*. How and why these are used—and what their mathematical properties are—become the focus of virtually everything covered in Chapter 3.

As a case in point, suppose a medical researcher is testing eight elderly adults for their allergic reaction (yes or no) to a new drug for controlling blood pressure. One of the $2^8 = 256$ possible sample points would be the sequence (yes, no, no, yes, no, no, yes, no), signifying that the first subject had an allergic reaction, the second did not, the third did not, and so on. Typically, in studies of this sort, the particular subjects experiencing reactions is of little interest: what does matter is the *number* who show a reaction. If that were true here, the outcome's relevant information (i.e., the number of allergic reactions) could be summarized by the number 3.[1]

Suppose X denotes the number of allergic reactions among a set of eight adults. Then X is said to be a *random variable* and the number 3 is the *value* of the random variable for the outcome (yes, no, no, yes, no, no, yes, no).

[1] By Theorem 2.6.2, of course, there would be a total of *fifty-six* (= 8!/3!5!) outcomes having exactly three yeses. All fifty-six would be equivalent in terms of what they imply about the drug's likelihood of causing allergic reactions.

In general, random variables are functions that associate numbers with some attribute of a sample outcome that is deemed to be especially important. If X denotes the random variable and s denotes a sample outcome, then $X(s) = t$, where t is a real number. For the allergy example, $s =$ (yes, no, no, yes, no, no, yes, no) and $t = 3$.

Random variables can often create a dramatically simpler sample space. That certainly is the case here—the original sample space has 256 $(= 2^8)$ outcomes, each being an ordered sequence of length eight. The random variable X, on the other hand, has only *nine* possible values, the integers from 0 to 8, inclusive.

In terms of their fundamental structure, all random variables fall into one of two broad categories, the distinction resting on the number of possible values the random variable can equal. If the latter is finite or countably infinite (which would be the case with the allergic reaction example), the random variable is said to be *discrete*; if the outcomes can be any real number in a given interval, the number of possibilities is uncountably infinite, and the random variable is said to be *continuous*. The difference between the two is critically important, as we will learn in the next several sections.

The purpose of Chapter 3 is to introduce the important definitions, concepts, and computational techniques associated with random variables, both discrete and continuous. Taken together, these ideas form the bedrock of modern probability and statistics.

3.2 Binomial and Hypergeometric Probabilities

This section looks at two specific probability scenarios that are especially important, both for their theoretical implications as well as for their ability to describe real-world problems. What we learn in developing these two models will help us understand random variables in general, the formal discussion of which begins in Section 3.3.

THE BINOMIAL PROBABILITY DISTRIBUTION

Binomial probabilities apply to situations involving a series of independent and identical trials, where each trial can have only one of two possible outcomes. Imagine three distinguishable coins being tossed, each having a probability p of coming up heads. The set of possible outcomes are the eight listed in Table 3.2.1. If the probability of any of the coins coming up heads is p, then the probability of the *sequence* (H, H, H) is p^3, since the coin tosses qualify as independent trials. Similarly, the probability of (T, H, H) is $(1 - p)p^2$. The fourth column of Table 3.2.1 shows the probabilities associated with each of the three-coin sequences.

Table 3.2.1

1st Coin	2nd Coin	3rd Coin	Probability	Number of Heads
H	H	H	p^3	3
H	H	T	$p^2(1 - p)$	2
H	T	H	$p^2(1 - p)$	2
T	H	H	$p^2(1 - p)$	2
H	T	T	$p(1 - p)^2$	1
T	H	T	$p(1 - p)^2$	1
T	T	H	$p(1 - p)^2$	1
T	T	T	$(1 - p)^3$	0

Suppose our main interest in the coin tosses is the *number* of heads that occur. Whether the actual sequence is, say, (H, H, T) or (H, T, H) is immaterial, since each outcome contains exactly two heads. The last column of Table 3.2.1 shows the number of heads in each of the eight possible outcomes. Notice that there are *three* outcomes with exactly two heads, each having an individual probability of $p^2(1-p)$. The probability, then, of the event "two heads" is the sum of those three individual probabilities—that is, $3p^2(1-p)$. Table 3.2.2 lists the probabilities of tossing k heads, where $k = 0, 1, 2$, or 3.

Table 3.2.2

Number of Heads	Probability
0	$(1-p)^3$
1	$3p(1-p)^2$
2	$3p^2(1-p)$
3	p^3

Now, more generally, suppose that n coins are tossed, in which case the number of heads can equal any integer from 0 through n. By analogy,

$$P(k \text{ heads}) = \begin{pmatrix} \text{number of} \\ \text{ways to arrange } k \\ \text{heads and } n-k \text{ tails} \end{pmatrix} \cdot \begin{pmatrix} \text{probability of} \\ \text{any particular sequence} \\ \text{having } k \text{ heads} \\ \text{and } n-k \text{ tails} \end{pmatrix}$$

$$= \begin{pmatrix} \text{number of ways} \\ \text{to arrange } k \\ \text{heads and } n-k \text{ tails} \end{pmatrix} \cdot p^k(1-p)^{n-k}$$

The number of ways to arrange k H's and $n-k$ T's, though, is $\frac{n!}{k!(n-k)!}$, or $\binom{n}{k}$ (recall Theorem 2.6.2).

Theorem 3.2.1 *Consider a series of n independent trials, each resulting in one of two possible outcomes, "success" or "failure." Let p = P (success occurs at any given trial) and assume that p remains constant from trial to trial. Then*

$$P(k \text{ successes}) = \binom{n}{k} p^k(1-p)^{n-k}, \; k = 0, 1, \ldots, n$$

Comment The probability assignment given by the equation in Theorem 3.2.1 is known as the *binomial distribution*.

Example 3.2.1 In communications of various kinds, a sent message may not be received correctly because the communications channel is "noisy." In particular, a bit transmitted might be changed with probability p. One method of coping with this problem is to send the bit five times, then use the majority of these five received bits as the intended message. Under this scheme, what is the probability that the message is received correctly?

The five bits received form a sequence of five Bernoulli trials with probability p that the bit is changed (success). Then the message is received correctly if the number of changed bits (successes) is 0, 1, or 2. The probability of this is

$$\sum_{k=0}^{2} \binom{5}{k} p^k(1-p)^{5-k}$$

The following chart gives values of this sum for some choices for p.

p	Probability
0.01	0.99999
0.05	0.99884
0.10	0.99144
0.15	0.97339

■

Example 3.2.2

Kingwest Pharmaceuticals is experimenting with a new affordable AIDS medication, PM-17, that may have the ability to strengthen a victim's immune system. Thirty monkeys infected with the HIV complex have been given the drug. Researchers intend to wait six weeks and then count the number of animals whose immunological responses show a marked improvement. Any inexpensive drug capable of being effective 60% of the time would be considered a major breakthrough; medications whose chances of success are 50% or less are not likely to have any commercial potential.

Yet to be finalized are guidelines for interpreting results. Kingwest hopes to avoid making either of two errors: (1) rejecting a drug that would ultimately prove to be marketable and (2) spending additional development dollars on a drug whose effectiveness, in the long run, would be 50% or less. As a tentative "decision rule," the project manager suggests that unless *sixteen or more* of the monkeys show improvement, research on PM-17 should be discontinued.

a. What are the chances that the "sixteen or more" rule will cause the company to reject PM-17, *even if the drug is 60% effective*?

b. How often will the "sixteen or more" rule allow a 50%-effective drug to be perceived as a major breakthrough?

a. Each of the monkeys is one of $n = 30$ independent trials, where the outcome is either a "success" (Monkey's immune system is strengthened) or a "failure" (Monkey's immune system is not strengthened). By assumption, the probability that PM-17 produces an immunological improvement in any given monkey is $p = P\,(\text{success}) = 0.60$.

By Theorem 3.2.1, the probability that exactly k monkeys (out of thirty) will show improvement after six weeks is $\binom{30}{k}(0.60)^k(0.40)^{30-k}$. The probability, then, that the "sixteen or more" rule will cause a 60%-effective drug to be discarded is the sum of "binomial" probabilities for k values ranging from 0 to 15:

$$P(60\%\text{-effective drug fails "sixteen or more" rule}) = \sum_{k=0}^{15}\binom{30}{k}(0.60)^k(0.40)^{30-k}$$
$$= 0.1754$$

Roughly 18% of the time, in other words, a "breakthrough" drug such as PM-17 will produce test results so mediocre (as measured by the "sixteen or more" rule) that the company will be misled into thinking it has no potential.

b. The other error Kingwest can make is to conclude that PM-17 warrants further study when, in fact, its value for p is below a marketable level. The chance that particular incorrect inference will be drawn here is the probability that the number of successes will be greater than or equal to sixteen when $p = 0.5$. That is,

$$\begin{aligned} &P(50\%\text{-effective PM-17 appears to be marketable}) \\ &\quad = P(\text{Sixteen or more successes occur}) \\ &\quad = \sum_{k=16}^{30} \binom{30}{k} (0.5)^k (0.5)^{30-k} \\ &\quad = 0.43 \end{aligned}$$

Thus, even if PM-17's success rate is an unacceptably low 50%, it has a 43% chance of performing sufficiently well in thirty trials to satisfy the "sixteen or more" criterion. ■

Example 3.2.3 The Stanley Cup playoff in professional hockey is a seven-game series, where the first team to win four games is declared the champion. The series, then, can last anywhere from four to seven games (just like the World Series in baseball). Calculate the likelihoods that the series will last four, five, six, or seven games. Assume that (1) each game is an independent event and (2) the two teams are evenly matched.

Consider the case where Team A wins the series in *six* games. For that to happen, they must win exactly three of the first five games *and* they must win the sixth game. Because of the independence assumption, we can write

$$\begin{aligned} P(\text{Team A wins in six games}) &= P(\text{Team A wins three of first five}) \cdot \\ &\quad\; P(\text{Team A wins sixth}) \\ &= \left[\binom{5}{3}(0.5)^3(0.5)^2\right] \cdot (0.5) = 0.15625 \end{aligned}$$

Since the probability that Team B wins the series in six games is the same (why?),

$$\begin{aligned} P(\text{Series ends in six games}) &= P(\text{Team A wins in six games} \cup \\ &\quad\; \text{Team B wins in six games}) \\ &= P(\text{A wins in six}) + P(\text{B wins in six}) \qquad \text{(why?)} \\ &= 0.15625 + 0.15625 \\ &= 0.3125 \end{aligned}$$

A similar argument allows us to calculate the probabilties of four-, five-, and seven-game series:

$$P(\text{four-game series}) = 2(0.5)^4 = 0.125$$

$$P(\text{five-game series}) = 2\left[\binom{4}{3}(0.5)^3(0.5)\right](0.5) = 0.25$$

$$P(\text{seven-game series}) = 2\left[\binom{6}{3}(0.5)^3(0.5)^3\right](0.5) = 0.3125$$

Having calculated the "theoretical" probabilities associated with the possible lengths of a Stanley Cup playoff raises an obvious question: How do those likelihoods compare with the actual distribution of playoff lengths? Between 1947 and 2015 there were sixty-eight playoffs (the 2004–05 season was cancelled). Column 2 in Table 3.2.3 shows the proportion of playoffs that have lasted four, five, six, and seven games, respectively.

Table 3.2.3

Series Length	Observed Proportion	Theoretical Probability
4	17/68 = 0.250	0.125
5	16/68 = 0.235	0.250
6	21/68 = 0.309	0.3125
7	14/68 = 0.206	0.3125

Data from: http://statshockey.homestead.com/trophies/stanleycup.html

Clearly, the agreement between the entries in columns 2 and 3 is not very good: Particularly noticeable is the excess of short playoffs (four games) and the deficit of long playoffs (seven games). What this "lack of fit" suggests is that one or more of the binomial distribution assumptions is not satisfied. Consider, for example, the parameter p, which we assumed to equal $\frac{1}{2}$. In reality, its value might be something quite different—just because the teams playing for the championship won their respective divisions, it does not necessarily follow that the two are equally good. Indeed, if the two contending teams were frequently mismatched, the consequence would be an increase in the number of short playoffs and a decrease in the number of long playoffs. It may also be the case that momentum is a factor in a team's chances of winning a given game. If so, the independence assumption implicit in the binomial model is rendered invalid. ■

Example 3.2.4

A university discovers that its colleges don't each require a full time technical support person. In fact, it estimates that any one of the ten colleges has only a 0.2 probability of needing a technical support person in any given day. They decide to create a pool of five support personnel to be called upon by the colleges. What is the probability that a college will have to wait on someone to get free?

Assume the needs of the colleges for service are independent. Then the number of support persons needed per day is a sequence of n Bernoulli trials, where a success means that a college needs a technical support person. Here $n = 10$, and $p = 0.2$. A group has to wait if the number of "successes" is strictly greater than 5. But the probability of more than five successes is $\sum_{k=6}^{10} \binom{10}{k} (0.2)^k (0.8)^{10-k} = 0.006$.

In a month, there are approximately two hundred college-days, so typically in only one of those college-days per month will there be some wait time.

Such analyses can easily be modified to observe the effects of other assumptions on n and p. (See Question 3.2.17.) ■

Questions

3.2.1. An investment analyst has tracked a certain blue-chip stock for the past six months and found that on any given day, it either goes up a point or goes down a point. Furthermore, it went up on 25% of the days and down on 75%. What is the probability that at the close of trading four days from now, the price of the stock will be the same as it is today? Assume that the daily fluctuations are independent events.

3.2.2. In a nuclear reactor, the fission process is controlled by inserting special rods into the radioactive core to absorb neutrons and slow down the nuclear chain reaction. When functioning properly, these rods serve as a first-line defense against a core meltdown. Suppose a reactor has ten control rods, each operating independently and each having an 0.80 probability of being properly inserted in the event of an "incident." Furthermore, suppose that a meltdown will be prevented if at least half the rods perform satisfactorily. What is the probability that, upon demand, the system will fail?

3.2.3. In 2009 a donor who insisted on anonymity gave seven-figure donations to twelve universities. A media report of this generous but somewhat mysterious act identified that all of the universities awarded had female presidents. It went on to say that with about 23% of U.S. college presidents being women, the probability of a dozen randomly selected institutions having female presidents is about 1/50,000,000. Is this probability approximately correct?

3.2.4. An entrepreneur owns six corporations, each with more than \$10 million in assets. The entrepreneur consults the *U.S. Internal Revenue Data Book* and discovers that the IRS audits 15.3% of businesses of that size. What is 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 at 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 (167)], 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 sixty-four series from 1950 to 2014 (there was no series in 1994).

World Series Lengths

Number of Games, k	Number of Years
4	13
5	11
6	14
7	26

Data from: www.baseball-almanac.com

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 64).

3.2.17. Redo Example 3.2.4 assuming $n = 12$ and $p = 0.3$.

3.2.18. Suppose a series of n independent trials can end in one of *three* possible outcomes. Let k_1 and k_2 denote the number of trials that result in outcomes 1 and 2, respectively. Let p_1 and p_2 denote the probabilities associated with outcomes 1 and 2. Generalize Theorem 3.2.1 to deduce a formula for the probability of getting k_1 and k_2 occurrences of outcomes 1 and 2, respectively.

3.2.19. Repair calls for central air conditioners fall into three general categories: coolant leakage, compressor failure, and electrical malfunction. Experience has shown that the probabilities associated with the three are 0.5, 0.3, and 0.2, respectively. Suppose that a dispatcher has logged in ten service requests for tomorrow morning. Use the answer to Question 3.2.18 to calculate the probability that three of those ten will involve coolant leakage and five will be compressor failures.

THE HYPERGEOMETRIC DISTRIBUTION

The second "special" distribution that we want to look at formalizes the urn problems that frequented Chapter 2. Our solutions to those earlier problems tended to be enumerations in which we listed the entire set of possible samples, and then counted the ones that satisfied the event in question. The inefficiency and redundancy of that approach should now be painfully obvious. What we are seeking here is a general formula that can be applied to any and all such problems, much like the expression in Theorem 3.2.1 can handle the full range of questions arising from the binomial model.

In the binomial model, if p is a rational number, then the experiment can be cast as an urn model. Suppose that $p = r/N$, where r and N are positive integers with $r < N$. Consider an urn with r red balls and w white balls, where $r + w = N$. Draw a ball; note its color; return it to the urn; mix the urn; draw another ball. If we continue in this way for n trials, we have a Bernoulli experiment. Then the probability of drawing k red balls is binomial.

Now let us examine a variation of the above scheme. The urn is the same, but n balls are drawn from the urn *simultaneously*. We again count the number of red balls in the sample. The result is *unordered* sampling without replacement. The probabilities associated with the number of red balls drawn are known as *hypergeometric,* and not surprisingly, rely on combinations. We formalize this discussion in the following theorem.

Theorem 3.2.2 *Suppose an urn contains r red chips and w white chips, where $r + w = N$. If n chips are drawn out at random, without replacement, and if k denotes the number of red chips selected, then*

$$P(k \text{ red chips are chosen}) = \frac{\binom{r}{k}\binom{w}{n-k}}{\binom{N}{n}} \tag{3.2.1}$$

where k varies over all the integers for which $\binom{r}{k}$ and $\binom{w}{n-k}$ are defined. The probabilities appearing on the right-hand side of Equation 3.2.1 are known as the hypergeometric distribution.

Proof Since this model concerns unordered selections, Theorem 2.6.3 applies. The number of ways to select a sample of size n from N elements is $\binom{N}{n}$. Now, ignoring the white balls, the number of ways to select k red balls from the r in the urn is $\binom{r}{k}$. Similarly, the selection of white balls can be done in $\binom{w}{n-k}$ ways. The number of ways to choose the red and white balls together is $\binom{r}{k} \cdot \binom{w}{n-k}$ by the multiplication principle. Finally, the desired probability is $\dfrac{\binom{r}{k}\binom{w}{n-k}}{\binom{N}{n}}$.

Comment A third urn model is to draw the sample in order but without replacement. In this case, the probabilities are also hypergeometric (See Question 3.2.28).

Comment The name *hypergeometric* derives from a series introduced in 1769 by the Swiss mathematician and physicist Leonhard Euler:

$$1+\frac{ab}{c}x+\frac{a(a+1)b(b+1)}{2!c(c+1)}x^2+\frac{a(a+1)(a+2)b(b+1)(b+2)}{3!c(c+1)(c+2)}x^3+\cdots$$

This is an expansion of considerable flexibility: Given appropriate values for a, b, and c, it reduces to many of the standard infinite series used in analysis. In particular, if a is set equal to 1, and b and c are set equal to each other, it reduces to the familiar *geometric* series,

$$1+x+x^2+x^3+\cdots$$

hence the name *hypergeometric*. The relationship of the probability function in Theorem 3.2.2 to Euler's series becomes apparent if we set $a=-n, b=-r, c=w-n+1$, and multiply the series by $\binom{w}{n}/\binom{N}{n}$. Then the coefficient of x^k will be

$$\frac{\binom{r}{k}\binom{w}{n-k}}{\binom{N}{n}}$$

the value the theorem gives for $P(k$ red chips are chosen).

Example 3.2.5 A hung jury is one that is unable to reach a unanimous decision. Suppose that a pool of twenty-five potential jurors is assigned to a murder case where the evidence is so overwhelming against the defendant that twenty-three of the twenty-five would return a guilty verdict. The other two potential jurors would vote to acquit regardless of the facts. What is the probability that a twelve-member panel chosen at random from the pool of twenty-five will be unable to reach a unanimous decision?

Think of the jury pool as an urn containing twenty-five chips, twenty-three of which correspond to jurors who would vote "guilty" and two of which correspond to jurors who would vote "not guilty." If either or both of the jurors who would vote "not guilty" are included in the panel of twelve, the result would be a hung jury. Applying Theorem 3.2.2 (twice) gives *0.74* as the probability that the jury impanelled would not reach a unanimous decision:

$$\begin{aligned}P(\text{Hung jury}) &= P(\text{Decision is not unanimous})\\ &= \binom{2}{1}\binom{23}{11}\bigg/\binom{25}{12}+\binom{2}{2}\binom{23}{10}\bigg/\binom{25}{12}\\ &= 0.74\end{aligned}$$

■

Example 3.2.6 The Florida Lottery features a number of games of chance, one of which is called Fantasy Five. The player chooses five numbers from a card containing the numbers 1 through 36. Each day five numbers are chosen at random, and if the player matches all five, the winnings can be as much as \$200,000 for a \$1 bet.

Lottery games like this one have spawned a mini-industry looking for biases in the selection of the winning numbers. Websites post various "analyses" claiming certain numbers are "hot" and should be played. One such examination focused on the frequency of winning numbers between 1 and 12. The probability of such occurrences fits the hypergeometric distribution, where $r = 12$, $w = 24$, $n = 5$, and $N = 36$. For example, the probability that *three* of the five numbers are 12 or less is

$$\frac{\binom{12}{3}\binom{24}{2}}{\binom{36}{5}} = \frac{60{,}720}{376{,}992} = 0.161$$

Notice how that compares to the *observed* proportion of drawings with exactly three numbers between 1 and 12. Of one leap year's daily drawings—366 of them —there were sixty-five with three numbers 12 or less, giving a relative frequency of $65/366 = 0.178$.

The full breakdown of observed and expected probabilities for winning numbers between 1 and 12 is given in Table 3.2.4.

Table 3.2.4

No. Drawn ≤ 12	Observed Proportion	Hypergeometric Probability
0	0.128	0.113
1	0.372	0.338
2	0.279	0.354
3	0.178	0.161
4	0.038	0.032
5	0.005	0.002

Data from: www.flalottery.com/exptkt/ff.html

The naive or dishonest commentator might claim that the lottery "likes" numbers ≤ 12 since the proportion of tickets drawn with three, four, or five numbers ≤ 12 is

$$0.178 + 0.038 + 0.005 = 0.221$$

This figure is in excess of the sum of the hypergeometric probabilities for $k = 3$, 4, and 5:

$$0.161 + 0.032 + 0.002 = 0.195$$

However, we shall see in Chapter 10 that such variation is well within the random fluctuations expected for truly random drawings. No bias can be inferred from these results. ■

Example 3.2.7

When a bullet is fired it becomes scored with minute striations produced by imperfections in the gun barrel. Appearing as a series of parallel lines, these striations have long been recognized as a basis for matching a bullet with a gun, since repeated firings of the same weapon will produce bullets having substantially the same configuration of markings. Until recently, deciding how close two patterns had to be before it could be concluded the bullets came from the same weapon was largely subjective. A ballistics expert would simply look at the two bullets under a microscope and make an informed judgment based on past experience. Today, however, criminologists are beginning to address the problem more quantitatively, partly with the help of the hypergeometric distribution.

Suppose a bullet is recovered from the scene of a crime, along with the suspect's gun. Under a microscope, a grid of m cells, numbered 1 to m, is superimposed over the bullet. If m is chosen large enough that the width of the cells is sufficiently small, each of that evidence bullet's n_e striations will fall into a different cell (see Figure 3.2.1a). Then the suspect's gun is fired, yielding a test bullet, which will have a total of n_t striations located in a possibly different set of cells (see Figure 3.2.1b). How might we assess the similarities in cell locations for the two striation patterns?

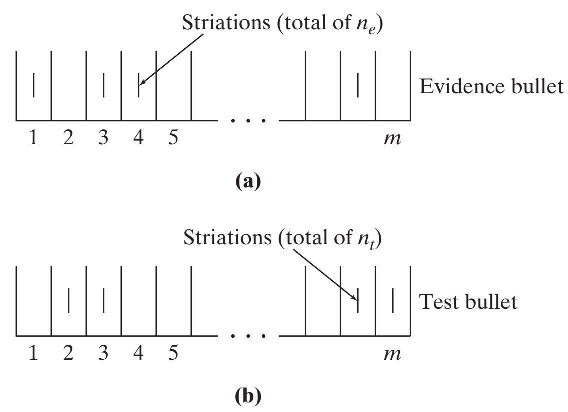


Figure 3.2.1

As a model for the striation pattern on the evidence bullet, imagine an urn containing m chips, with n_e corresponding to the striation locations. Now, think of the striation pattern on the *test* bullet as representing a sample of size n_t from the evidence urn. By Theorem 3.2.2, the probability that k of the cell locations will be shared by the two striation patterns is

$$\frac{\binom{n_e}{k}\binom{m-n_e}{n_t-k}}{\binom{m}{n_t}}$$

Suppose the bullet found at a murder scene is superimposed with a grid having $m = 25$ cells, $n_e = 4$ of which contain striations. The suspect's gun is fired and the bullet is found to have $n_t = 3$ striations, one of which matches the location of one of the striations on the evidence bullet. What do you think a ballistics expert would conclude?

Intuitively, the similarity between the two bullets would be reflected in the probability that *one or more* striations in the suspect's bullet match the evidence bullet. The smaller that probability is, the stronger would be our belief that the two bullets were fired by the same gun. Based on the values given for m, n_e, and n_t,

$$P(\text{one or more matches}) = \frac{\binom{4}{1}\binom{21}{2}}{\binom{25}{3}} + \frac{\binom{4}{2}\binom{21}{1}}{\binom{25}{3}} + \frac{\binom{4}{3}\binom{21}{0}}{\binom{25}{3}}$$
$$= 0.42$$

If P(one or more matches) had been a very small number—say, 0.001—the inference would have been clear-cut: The same gun fired both bullets. But, here with the probability of one or more matches being so large, we cannot rule out the possibility that the bullets were fired by two different guns (and, presumably, by two different people). ■

Example 3.2.8

A tax collector, finding himself short of funds, delayed depositing a large property tax payment ten different times. The money was subsequently repaid, and the whole amount deposited in the proper account. The tip-off to this behavior was the delay of the deposit. During the period of these irregularities, there was a total of 470 tax collections.

An auditing firm was preparing to do a routine annual audit of these transactions. They decided to randomly sample nineteen of the collections (approximately 4%) of the payments. The auditors would assume a pattern of malfeasance only if they saw three or more irregularities. What is the probability that three or more of the delayed deposits would be chosen in this sample?

This kind of audit sampling can be considered a hypergeometric experiment. Here, $N = 470$, $n = 19$, $r = 10$, and $w = 460$. In this case it is better to calculate the desired probability via the complement—that is,

$$1 - \frac{\binom{10}{0}\binom{460}{19}}{\binom{470}{19}} - \frac{\binom{10}{1}\binom{460}{18}}{\binom{470}{19}} - \frac{\binom{10}{2}\binom{460}{17}}{\binom{470}{19}}$$

The calculation of the first hypergeometric term is

$$\frac{\binom{10}{0}\binom{460}{19}}{\binom{470}{19}} = 1 \cdot \frac{460!}{19!441!} \cdot \frac{19!451}{470!} = \frac{451}{470} \cdot \frac{450}{469} \cdot \dots \cdot \frac{442}{461} = 0.6592$$

To compute hypergeometric probabilities where the numbers are large, a useful device is a *recursion formula*. To that end, note that the ratio of the $k+1$ term to the k term is

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

(See Question 3.2.30.)

Therefore,

$$\frac{\binom{10}{1}\binom{460}{18}}{\binom{470}{19}} = 0.6592 \cdot \frac{19+0}{1+0} \cdot \frac{10-0}{460-19+0+1} = 0.2834$$

and

$$\frac{\binom{10}{2}\binom{460}{17}}{\binom{470}{19}} = 0.2834 \cdot \frac{19-1}{1+1} \cdot \frac{10-1}{460-19+1+1} = 0.0518$$

The desired probability, then, is $1 - 0.6592 - 0.2834 - 0.0518 = 0.0056$, which shows that a larger audit sample would be necessary to have a reasonable chance of detecting this sort of impropriety. ■

CASE STUDY 3.2.1

Biting into a plump, juicy apple is one of the innocent pleasures of autumn. Critical to that enjoyment is the *firmness* of the apple, a property that growers and shippers monitor closely. The apple industry goes so far as to set a lowest acceptable limit for firmness, which is measured (in lbs) by inserting a probe into the apple. For the Red Delicious variety, for example, firmness is supposed to be at least 12 lbs; in the state of Washington, wholesalers are not allowed to sell apples if more than 10% of their shipment falls below that 12-lb limit.

All of this raises an obvious question: How can shippers demonstrate that their apples meet the 10% standard? Testing each one is not an option—the probe that measures firmness renders an apple unfit for sale. That leaves *sampling* as the only viable strategy.

(Continued on next page)

(Case Study 3.2.1 continued)

Suppose, for example, a shipper has a supply of one hundred forty-four apples. She decides to select fifteen at random and measure each one's firmness, with the intention of selling the remaining apples if two or fewer in the sample are substandard. What are the consequences of her plan? More specifically, does it have a good chance of "accepting" a shipment that meets the 10% rule and "rejecting" one that does not? (If either or both of those objectives are not met, the plan is inappropriate.)

For example, suppose there are actually ten defective apples among the original one hundred forty-four. Since $\frac{10}{144} \times 100 = 6.9\%$, that shipment would be suitable for sale because fewer than 10% failed to meet the firmness standard. The question is, how likely is it that a sample of fifteen chosen at random from that shipment will pass inspection?

Notice, here, that the number of substandard apples in the sample has a hypergeometric distribution with $r = 10$, $w = 134$, $n = 15$, and $N = 144$. Therefore,

$$\begin{aligned}
P(\text{Sample passes inspection}) &= P(\text{Two or fewer substandard apples are found}) \\
&= \frac{\binom{10}{0}\binom{134}{15}}{\binom{144}{15}} + \frac{\binom{10}{1}\binom{134}{14}}{\binom{144}{15}} + \frac{\binom{10}{2}\binom{134}{13}}{\binom{144}{15}} \\
&= 0.320 + 0.401 + 0.208 = 0.929
\end{aligned}$$

So, the probability is reassuringly high that a supply of apples this good would, in fact, be judged acceptable to ship. Of course, it also follows from this calculation that roughly 7% of the time, the number of substandard apples found will be *greater* than two, in which case the apples would be (incorrectly) assumed to be unsuitable for sale (earning them an undeserved one-way ticket to the applesauce factory . . .).

How good is the proposed sampling plan at recognizing apples that would, in fact, be inappropriate to ship? Suppose, for example, that *30*, or *21*%, of the one hundred forty-four apples would fall below the 12-lb limit. Ideally, the probability here that a sample passes inspection should be small. The number of substandard apples found in this case would be hypergeometric with $r = 30$, $w = 114$, $n = 15$, and $N = 144$, so

$$\begin{aligned}
P(\text{Sample passes inspection}) &= \frac{\binom{30}{0}\binom{114}{15}}{\binom{144}{15}} + \frac{\binom{30}{1}\binom{114}{14}}{\binom{144}{15}} + \frac{\binom{30}{2}\binom{114}{13}}{\binom{144}{15}} \\
&= 0.024 + 0.110 + 0.221 = 0.355
\end{aligned}$$

Here the bad news is that the sampling plan will allow a 21% defective supply to be shipped 36% of the time. The good news is that 64% of the time, the number of substandard apples in the sample will exceed two, meaning that the correct decision "not to ship" will be made.

Figure 3.2.2 shows P(Sample passes) plotted against the percentage of defectives in the entire supply. Graphs of this sort are called *operating characteristic* (or *OC*) *curves:* They summarize how a sampling plan will respond to all possible levels of quality.

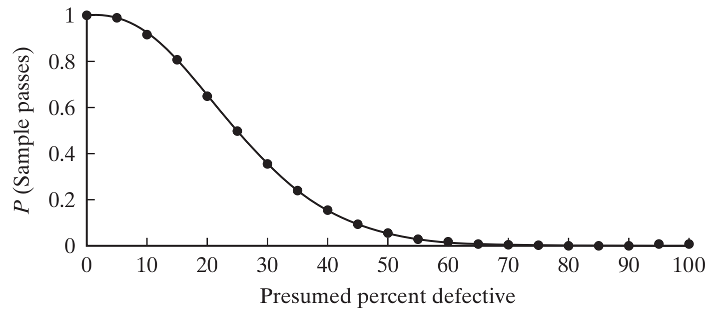


Figure 3.2.2

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. Consider an urn with r red balls and w white balls, where $r + w = N$. Draw n balls in order without replacement. Show that the probability of k red balls is hypergeometric.

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.

3.2.31. Urn I contains five red chips and four white chips; urn II contains four red and five white chips. Two chips are drawn simultaneously from urn I and placed into urn II. Then a single chip is drawn from urn II. What is the probability that the chip drawn from urn II is white? (*Hint:* Use Theorem 2.4.1.)

3.2.32. As the owner of a chain of sporting goods stores, you have just been offered a "deal" on a shipment of one hundred robot table tennis machines. The price is right, but the prospect of picking up the merchandise at midnight from an unmarked van parked on the side of the New Jersey Turnpike is a bit disconcerting. Being of low repute yourself, you do not consider the legality of the transaction to be an issue, but you do have concerns about being cheated. If too many of the machines are in poor working order, the offer ceases to be a bargain. Suppose you decide to close the deal only if a sample of ten machines contains no more than one defective. Construct the corresponding operating characteristic curve. For approximately what incoming quality will you accept a shipment 50% of the time?

3.2.33. Suppose that r of N chips are red. Divide the chips into three groups of sizes n_1, n_2, and n_3, where $n_1 + n_2 + n_3 = N$. Generalize the hypergeometric distribution to find the probability that the first group contains r_1 red chips, the second group r_2 red chips, and the third group r_3 red chips, where $r_1 + r_2 + r_3 = r$.

3.2.34. Some nomadic tribes, when faced with a life-threatening contagious disease, try to improve their chances of survival by dispersing into smaller groups. Suppose a tribe of twenty-one people, of whom four are carriers of the disease, split into three groups of seven each. What is the probability that at least one group is free of the disease? (*Hint:* Find the probability of the complement.)

3.2.35. Suppose a population contains n_1 objects of one kind, n_2 objects of a second kind, ..., and n_t objects of a tth kind, where $n_1 + n_2 + \cdots + n_t = N$. A sample of size n is drawn at random and without replacement. Deduce an expression for the probability of drawing k_1 objects of the first kind, k_2 objects of the second kind, ..., and k_t objects of the tth kind by generalizing Theorem 3.2.2.

3.2.36. Sixteen students—five freshmen, four sophomores, four juniors, and three seniors—have applied for membership in their school's Communications Board, a group that oversees the college's newspaper, literary magazine, and radio show. Eight positions are open. If the selection is done at random, what is the probability that each class gets two representatives? (*Hint:* Use the generalized hypergeometric model asked for in Question 3.2.35.)

3.3 Discrete Random Variables

The binomial and hypergeometric distributions described in Section 3.2 are special cases of some important general concepts that we want to explore more fully in this section. Previously in Chapter 2, we studied in depth the situation where every point in a sample space is equally likely to occur (recall Section 2.6). The sample space of independent trials that ultimately led to the binomial distribution presented a quite different scenario: specifically, individual points in S had different probabilities. For example, if $n = 4$ and $p = \frac{1}{3}$, the probabilities assigned to the sample points (s, f, s, f) and (f, f, f, f) are $(1/3)^2(2/3)^2 = \frac{4}{81}$ and $(2/3)^4 = \frac{16}{81}$, respectively. Allowing for the possibility that different outcomes may have different probabilities will obviously broaden enormously the range of real-world problems that probability models can address.

How to assign probabilities to outcomes that are not binomial or hypergeometric is one of the major questions investigated in this chapter. A second critical issue is the nature of the sample space itself and whether it makes sense to redefine the outcomes and create, in effect, an alternative sample space. Why we would want to do that has already come up in our discussion of independent trials. The "original" sample space in such cases is a set of ordered sequences, where the ith member of a sequence is either an "s" or an "f," depending on whether the ith trial ended in success or failure, respectively. However, knowing which particular trials ended in success is typically less important than knowing the *number* that did (recall the medical researcher discussion on p. 102). That being the case, it often makes sense to replace each ordered sequence with the number of successes that sequence contains. Doing so collapses the original set of 2^n ordered sequences (i.e., outcomes) in S to the set of $n+1$ integers ranging from 0 to n. The probabilities assigned to those integers, of course, are given by the binomial formula in Theorem 3.2.1.

In general, a function that assigns numbers to outcomes is called a *random variable*. The purpose of such functions in practice is to define a new sample space whose outcomes speak more directly to the objectives of the experiment. That was the rationale that ultimately motivated both the binomial and hypergeometric distributions.

The purpose of this section is to (1) outline the general conditions under which probabilities can be assigned to sample spaces and (2) explore the ways and means of redefining sample spaces through the use of random variables. The notation introduced in this section is especially important and will be used throughout the remainder of the book.

ASSIGNING PROBABILITIES: THE DISCRETE CASE

We begin with the general problem of assigning probabilities to sample outcomes, the simplest version of which occurs when the number of points in S is either finite or countably infinite. The probability functions, $p(s)$, that we are looking for in those cases satisfy the conditions in Definition 3.3.1.

Definition 3.3.1
Suppose that S is a finite or countably infinite sample space. Let p be a real-valued function defined for each element of S such that

a. $0 \leq p(s)$ for each $s \in S$
b. $\sum_{s \in S} p(s) = 1$

Then p is said to be a *discrete probability function*.

Comment Once $p(s)$ is defined for all s, it follows that the probability of any event A—that is, $P(A)$—is the sum of the probabilities of the outcomes comprising A:

$$P(A) = \sum_{s \in A} p(s) \tag{3.3.1}$$

Defined in this way, the function $P(A)$ satisfies the probability axioms given in Section 2.3. The next several examples illustrate some of the specific forms that $p(s)$ can have and how $P(A)$ is calculated.

Example 3.3.1

Ace-six flats are a type of crooked dice where the cube is foreshortened in the one-six direction, the effect being that 1's and 6's are more likely to occur than any of the other four faces. Let $p(s)$ denote the probability that the face showing is s. For many ace-six flats, the "cube" is asymmetric to the extent that $p(1) = p(6) = \frac{1}{4}$, while $p(2) = p(3) = p(4) = p(5) = \frac{1}{8}$. Notice that $p(s)$ here qualifies as a discrete probability function because each $p(s)$ is greater than or equal to 0 and the sum of $p(s)$, over all s, is $1\left[= 2\left(\frac{1}{4}\right) + 4\left(\frac{1}{8}\right)\right]$.

Suppose A is the event that an even number occurs. It follows from Equation 3.3.1 that $P(A) = P(2) + P(4) + P(6) = \frac{1}{8} + \frac{1}{8} + \frac{1}{4} = \frac{1}{2}$. ■

Comment If two ace-six flats are rolled, the probability of getting a sum equal to 7 is equal to $2p(1)p(6) + 2p(2)p(5) + 2p(3)p(4) = 2\left(\frac{1}{4}\right)^2 + 4\left(\frac{1}{8}\right)^2 = \frac{3}{16}$. If two *fair* dice are rolled, the probability of getting a sum equal to 7 is $2p(1)p(6)+2p(2)p(5)+2p(3)p(4) = 6\left(\frac{1}{6}\right)^2 = \frac{1}{6}$, which is less than $\frac{3}{16}$. Gamblers cheat with ace-six flats by switching back and forth between fair dice and ace-six flats, depending on whether or not they want a sum of 7 to be rolled.

Example 3.3.2 Suppose a fair coin is tossed until a head comes up for the first time. What are the chances of that happening on an odd-numbered toss?

Note that the sample space here is countably infinite and so is the set of outcomes making up the event whose probability we are trying to find. The $P(A)$ that we are looking for, then, will be the sum of an infinite number of terms.

Let $p(s)$ be the probability that the first head appears on the sth toss. Since the coin is presumed to be fair, $p(1) = \frac{1}{2}$. Furthermore, we would expect that half the time, when a tail appears, the next toss would be a head, so $p(2) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$. In general, $p(s) = \left(\frac{1}{2}\right)^s$, $s = 1, 2, \ldots$.

Does $p(s)$ satisfy the conditions stated in Definition 3.3.1? Yes. Clearly, $p(s) \geq 0$ for all s. To see that the sum of the probabilities is 1, recall the formula for the sum of a geometric series: If $0 < r < 1$,

$$\sum_{s=0}^{\infty} r^s = \frac{1}{1-r} \tag{3.3.2}$$

Applying Equation 3.3.2 to the sample space here confirms that $P(S) = 1$:

$$P(S) = \sum_{s=1}^{\infty} p(s) = \sum_{s=1}^{\infty} \left(\frac{1}{2}\right)^s = \sum_{s=0}^{\infty} \left(\frac{1}{2}\right)^s - \left(\frac{1}{2}\right)^0 = 1 \Big/ \left(1 - \frac{1}{2}\right) - 1 = 1$$

Now, let A be the event that the first head appears on an odd-numbered toss. Then $P(A) = p(1) + p(3) + p(5) + \cdots$ But

$$\begin{aligned} p(1) + p(3) + p(5) + \cdots &= \sum_{s=0}^{\infty} p(2s+1) = \sum_{s=0}^{\infty} \left(\frac{1}{2}\right)^{2s+1} = \left(\frac{1}{2}\right) \sum_{s=0}^{\infty} \left(\frac{1}{4}\right)^s \\ &= \left(\frac{1}{2}\right) \left[1 \Big/ \left(1 - \frac{1}{4}\right)\right] = \frac{2}{3} \end{aligned}$$

■

CASE STUDY 3.3.1

For good pedagogical reasons, the principles of probability are always introduced by considering events defined on familiar sample spaces generated by simple experiments. To that end, we toss coins, deal cards, roll dice, and draw chips from urns. It would be a serious error, though, to infer that the importance of probability extends no further than the nearest casino. In its infancy, gambling and probability were, indeed, intimately related: Questions arising from games of chance were often the catalyst that motivated mathematicians to study random phenomena in earnest. But more than three hundred forty years have passed since Huygens published *De Ratiociniis*. Today, the application of probability to gambling is relatively insignificant (the NCAA March basketball tournament notwithstanding) compared to the depth and breadth of uses the subject finds in business, medicine, engineering, and science.

Probability functions—properly chosen—can "model" complex real-world phenomena every bit as well as $P(\text{heads}) = \frac{1}{2}$ describes the behavior of a fair coin. The following set of actuarial data is a case in point. Over a period of three years (= 1096 days) in London, records showed that a total of nine hundred three deaths occurred among males eighty-five years of age and older (191). Columns 1 and 2 of Table 3.3.1 give the breakdown of those 903 deaths according to the number occurring on a given day. Column 3 gives the *proportion* of days for which exactly s elderly men died.

Table 3.3.1

(1) Number of Deaths, s	(2) Number of Days	(3) Proportion [= Col.(2)/1096]	(4) $p(s)$
0	484	0.442	0.440
1	391	0.357	0.361
2	164	0.150	0.148
3	45	0.041	0.040
4	11	0.010	0.008
5	1	0.001	0.003
6+	0	0.000	0.000
	1096	1	1

For reasons that we will go into at length in Chapter 4, the probability function that describes the behavior of this particular phenomenon is

$$\begin{aligned} p(s) &= P(s \text{ elderly men die on a given day}) \\ &= \frac{e^{-0.82}(0.82)^s}{s!}, \qquad s = 0, 1, 2, \ldots \end{aligned} \tag{3.3.3}$$

How do we know that the $p(s)$ in Equation 3.3.3 is an appropriate way to assign probabilities to the "experiment" of elderly men dying? Because it accurately predicts what happened. Column 4 of Table 3.3.1 shows $p(s)$ evaluated for $s = 0, 1, 2, \ldots$. To two decimal places, the agreement between the entries in Column 3 and Column 4 is perfect.

Example 3.3.3

Consider the following experiment: Every day for the next month you copy down each number that appears in the stories on the front pages of your hometown newspaper. Those numbers would necessarily be extremely diverse: One might be the age of a celebrity who had just died, another might report the interest rate currently paid on government Treasury bills, and still another might give the number of square feet of retail space recently added to a local shopping mall.

Suppose you then calculated the proportion of those numbers whose leading digit was a 1, the proportion whose leading digit was a 2, and so on. What relationship would you expect those proportions to have? Would numbers starting with a 2, for example, occur as often as numbers starting with a 6?

Let $p(s)$ denote the probability that the first significant digit of a "newspaper number" is s, $s = 1, 2, \dots, 9$. Our intuition is likely to tell us that the nine first digits should be equally probable—that is, $p(1) = p(2) = \cdots = p(9) = \frac{1}{9}$. Given the diversity and the randomness of the numbers, there is no obvious reason why one digit should be more common than another. Our intuition, though, would be wrong—first digits are *not* equally likely. Indeed, they are not even close to being equally likely!

Credit for making this remarkable discovery goes to Simon Newcomb, a mathematician who observed more than a hundred years ago that some portions of logarithm tables are used more than others (85). Specifically, pages at the beginning of such tables are more dog-eared than pages at the end, suggesting that users have more occasion to look up logs of numbers starting with small digits than they do numbers starting with large digits.

Almost fifty years later, a physicist, Frank Benford, reexamined Newcomb's claim in more detail and looked for a mathematical explanation. What is now known as *Benford's law* asserts that the first digits of many different types of measurements, or combinations of measurements, often follow the discrete probability model:

$$p(s) = P(\text{First significant digit is } s) = \log\left(1 + \frac{1}{s}\right), \quad s = 1, 2, \dots, 9$$

Table 3.3.2 compares Benford's law to the uniform assumption that $p(s) = \frac{1}{9}$, for all s. The differences are striking. According to Benford's law, for example, 1's are the most frequently occurring first digit, appearing 6.5 times ($= 0.301/0.046$) as often as 9's.

Table 3.3.2

s	"Uniform" Law	Benford's Law
1	0.111	0.301
2	0.111	0.176
3	0.111	0.125
4	0.111	0.097
5	0.111	0.079
6	0.111	0.067
7	0.111	0.058
8	0.111	0.051
9	0.111	0.046

Comment A key to *why* Benford's law is true is the differences in proportional changes associated with each leading digit. To go from one thousand to two thousand, for example, represents a 100% increase; to go from eight thousand to nine thousand, on the other hand, is only a 12.5% increase. That would suggest that evolutionary phenomena such as stock prices would be more likely to start with 1's and 2's than with 8's and 9's—and they are. Still, the precise conditions under which $p(s) = \log\left(1 + \frac{1}{s}\right), s = 1, 2, \dots, 9$, are not fully understood and remain a topic of research. ■

Example 3.3.4

Is $p(s)$ as defined below a discrete probability function? Why or why not?

$$p(s) = \frac{1}{1+\lambda}\left(\frac{\lambda}{1+\lambda}\right)^s, \; s = 0, 1, 2, \ldots; \quad \lambda > 0$$

To qualify as a discrete probability function, a given $p(s)$ needs to satisfy parts (a) and (b) of Definition 3.3.1. A simple inspection shows that part (a) is satisfied. Since $\lambda > 0$, $p(s)$ is, in fact, greater than or equal to 0 for all $s = 0, 1, 2, \ldots$. Part (b) is satisfied if the sum of all the probabilities defined on the outcomes in S is 1. But

$$\begin{aligned}
\sum_{\text{all } s \in S} p(s) &= \sum_{s=0}^{\infty} \frac{1}{1+\lambda}\left(\frac{\lambda}{1+\lambda}\right)^s \\
&= \frac{1}{1+\lambda}\left(\frac{1}{1-\frac{\lambda}{1+\lambda}}\right) \qquad \text{(why?)} \\
&= \frac{1}{1+\lambda} \cdot \frac{1+\lambda}{1} \\
&= 1
\end{aligned}$$

The answer, then, is "yes"—$p(s) = \frac{1}{1+\lambda}\left(\frac{\lambda}{1+\lambda}\right)^s$, $s = 0, 1, 2, \ldots; \lambda > 0$ *does* qualify as a discrete probability function. Of course, whether it has any practical value depends on whether the set of values for $p(s)$ actually do describe the behavior of real-world phenomena. ■

DEFINING "NEW" SAMPLE SPACES

We have seen how the function $p(s)$ associates a probability with each outcome, s, in a sample space. Related is the key idea that outcomes can often be grouped or reconfigured in ways that may facilitate problem solving. Recall the sample space associated with a series of n independent trials, where each s is an ordered sequence of successes and failures. The most relevant information in such outcomes is often the *number* of successes that occur, not a detailed listing of which trials ended in success and which ended in failure. That being the case, it makes sense to define a "new" sample space by grouping the original outcomes according to the number of successes they contained. The outcome $(f, f, \ldots, f)$, for example, had 0 successes. On the other hand, there were n outcomes that yielded 1 success—$(s, f, f, \ldots, f)$, $(f, s, f, \ldots, f)$, $\ldots$, and $(f, f, \ldots, s)$. As we saw earlier in this chapter, that particular regrouping of outcomes ultimately led to the binomial distribution.

The function that replaces the outcome $(s, f, f, \ldots, f)$ with the numerical value 1 is called a *random variable*. We conclude this section with a discussion of some of the concepts, terminology, and applications associated with random variables.

Definition 3.3.2
A function whose domain is a sample space S and whose values form a finite or countably infinite set of real numbers is called a *discrete random variable.* We denote random variables by uppercase letters, often X or Y.

Example 3.3.5 Consider tossing two dice, an experiment for which the sample space is a set of ordered pairs, $S = \{(i, j) \mid i = 1, 2, \ldots, 6;\ j = 1, 2, \ldots, 6\}$. For a variety of games ranging from Monopoly to craps, the *sum* of the numbers showing is what matters on a given turn. That being the case, the original sample space S of thirty-six ordered pairs would not provide a particularly convenient backdrop for discussing the rules of those games. It would be better to work directly with the sums. Of course, the eleven possible sums (from 2 to 12) are simply the different values of the random variable X, where $X(i, j) = i + j$.

Comment In the above example, suppose we define a random variable X_1 that gives the result on the first die and a random variable X_2 that gives the result on the second die. Then $X = X_1 + X_2$. Note how easily we could extend this idea to the toss of *three* dice, or *ten* dice. The ability to conveniently express complex events in terms of simpler ones is an advantage of the random variable concept that we will see playing out over and over again. ■

THE PROBABILITY DENSITY FUNCTION

We began this section discussing the function $p(s)$, which assigns a probability to each outcome s in S. Now, having introduced the notion of a random variable X as a real-valued function defined on S, that is, $X(s) = k$, we need to find a mapping analogous to $p(s)$ that assigns probabilities to the different values of k.

Definition 3.3.3
Associated with every discrete random variable X is a *probability density function* (or *pdf*), denoted $p_X(k)$, where

$$p_X(k) = P(\{s \in S \mid X(s) = k\})$$

Note that $p_X(k) = 0$ for any k not in the range of X. For notational simplicity, we will usually delete all references to s and S and write $p_X(k) = P(X = k)$.

Comment We have already discussed at length two examples of the function $p_X(k)$. Recall the binomial distribution derived in Section 3.2. If we let the random variable X denote the number of successes in n independent trials, then Theorem 3.2.1 states that

$$P(X = k) = p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \ldots, n$$

A similar result was given in that same section in connection with the hypergeometric distribution. If a sample of size n is drawn without replacement from an urn containing r red chips and w white chips, and if we let the random variable X denote the number of red chips included in the sample, then (according to Theorem 3.2.2),

$$P(X = k) = p_X(k) = \binom{r}{k}\binom{w}{n-k} \Big/ \binom{r+w}{n}$$

where k ranges over the values for which the numerator terms are defined.

Example 3.3.6

Consider again the rolling of two dice as described in Example 3.3.5. Let i and j denote the faces showing on the first and second die, respectively, and define the random variable X to be the sum of the two faces: $X(i, j) = i + j$. Find $p_X(k)$.

According to Definition 3.3.3, each value of $p_X(k)$ is the sum of the probabilities of the outcomes that get mapped by X onto the value k. For example,

$$\begin{aligned}
P(X = 5) = p_X(5) &= P(\{s \in S \mid X(s) = 5\}) \\
&= P[(1, 4), (4, 1), (2, 3), (3, 2)] \\
&= P(1, 4) + P(4, 1) + P(2, 3) + P(3, 2) \\
&= \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} \\
&= \frac{4}{36}
\end{aligned}$$

assuming the dice are fair. Values of $p_X(k)$ for other k are calculated similarly. Table 3.3.3 shows the random variable's entire pdf.

Table 3.3.3

k	$p_X(k)$	k	$p_X(k)$
2	1/36	8	5/36
3	2/36	9	4/36
4	3/36	10	3/36
5	4/36	11	2/36
6	5/36	12	1/36
7	6/36		

■

Example 3.3.7

Acme Industries typically produces three electric power generators a day; some pass the company's quality-control inspection on their first try and are ready to be shipped; others need to be retooled. The probability of a generator needing further work is 0.05. If a generator is ready to ship, the firm earns a profit of \$10,000. If it needs to be retooled, it ultimately costs the firm \$2,000. Let X be the random variable quantifying the company's daily profit. Find $p_X(k)$.

The underlying sample space here is a set of $n = 3$ independent trials, where $p = P(\text{Generator passes inspection}) = 0.95$. If the random variable X is to measure the company's daily profit, then

$$\begin{aligned}
X = &\ \$10{,}000 \times (\text{no. of generators passing inspection}) \\
&- \$2{,}000 \times (\text{no. of generators needing retooling})
\end{aligned}$$

For instance, $X(s, f, s) = 2(\$10{,}000) - 1(\$2{,}000) = \$18{,}000$. Moreover, the random variable X equals \$18,000 whenever the day's output consists of two successes and one failure. That is, $X(s, f, s) = X(s, s, f) = X(f, s, s)$. It follows that

$$P(X = \$18{,}000) = p_X(18{,}000) = \binom{3}{2}(0.95)^2(0.05)^1 = 0.135375$$

Table 3.3.4 shows $p_X(k)$ for the four possible values of k (\$30,000, \$18,000, \$6,000, and $-$\$6,000).

Table 3.3.4

No. Defectives	k = Profit	$p_X(k)$
0	\$30,000	0.857375
1	\$18,000	0.135375
2	\$6,000	0.007125
3	−\$6,000	0.000125

■

Example 3.3.8

As part of her warm-up drill, each player on State's basketball team is required to shoot free throws until two baskets are made. If Rhonda has a 65% success rate at the foul line, what is the pdf of the random variable X that describes the number of throws it takes her to complete the drill? Assume that individual throws constitute independent events.

Figure 3.3.1 illustrates what must occur if the drill is to end on the kth toss, $k = 2, 3, 4, \ldots$ First, Rhonda needs to make exactly one basket sometime during the first $k-1$ attempts, and, second, she needs to make a basket on the kth toss. Written formally,

$$\begin{aligned} p_X(k) &= P(X = k) = P(\text{Drill ends on } k\text{th throw}) \\ &= P[(1 \text{ basket and } k-2 \text{ misses in first } k-1 \text{ throws}) \cap (\text{basket on } k\text{th throw})] \\ &= P(1 \text{ basket and } k-2 \text{ misses}) \cdot P(\text{basket}) \end{aligned}$$

$$\overbrace{\underset{1}{\underline{\text{Miss}}} \quad \underset{2}{\underline{\text{Basket}}} \quad \underset{3}{\underline{\text{Miss}}} \quad \cdots \quad \underset{k-1}{\underline{\text{Miss}}}}^{\text{Exactly one basket}} \quad \overbrace{\underset{k}{\underline{\text{Basket}}}}^{\text{Second basket}}$$

Attempts

Figure 3.3.1

Notice that $k-1$ different sequences have the property that exactly one of the first $k-1$ throws results in a basket:

$$\begin{array}{c} k-1 \\ \text{sequences} \end{array} \left\{ \begin{array}{cccccc} \frac{B}{1} & \frac{M}{2} & \frac{M}{3} & \frac{M}{4} & \cdots & \frac{M}{k-1} \\ \frac{M}{1} & \frac{B}{2} & \frac{M}{3} & \frac{M}{4} & \cdots & \frac{M}{k-1} \\ & & \vdots & & & \\ \frac{M}{1} & \frac{M}{2} & \frac{M}{3} & \frac{M}{4} & \cdots & \frac{B}{k-1} \end{array} \right.$$

Since each sequence has probability $(0.35)^{k-2}(0.65)$,

$$P(1 \text{ basket and } k-2 \text{ misses}) = (k-1)(0.35)^{k-2}(0.65)$$

Therefore,

$$\begin{aligned} p_X(k) &= (k-1)(0.35)^{k-2}(0.65) \cdot (0.65) \\ &= (k-1)(0.35)^{k-2}(0.65)^2, \quad k = 2, 3, 4, \ldots \end{aligned} \tag{3.3.4}$$

Table 3.3.5 shows the pdf evaluated for specific values of k. Although the range of k is infinite, the bulk of the probability associated with X is concentrated in the values 2 through 7: It is highly unlikely, for example, that Rhonda would need more than seven shots to complete the drill.

Table 3.3.5

k	$p_X(k)$
2	0.4225
3	0.2958
4	0.1553
5	0.0725
6	0.0317
7	0.0133
8+	0.0089

■

THE CUMULATIVE DISTRIBUTION FUNCTION

In working with random variables, we frequently need to calculate the probability that the value of a random variable is somewhere between two numbers. For example, suppose we have an integer-valued random variable. We might want to calculate an expression like $P(s \le X \le t)$. If we know the pdf for X, then

$$P(s \le X \le t) = \sum_{k=s}^{t} p_X(k)$$

But depending on the nature of $p_X(k)$ and the number of terms that need to be added, calculating the sum of $p_X(k)$ from $k = s$ to $k = t$ may be quite difficult. An alternate strategy is to use the fact that

$$P(s \le X \le t) = P(X \le t) - P(X \le s - 1)$$

where the two probabilities on the right represent *cumulative* probabilities of the random variable X. If the latter were available (and they often are), then evaluating $P(s \le X \le t)$ by one simple subtraction would clearly be easier than doing all the calculations implicit in $\sum_{k=s}^{t} p_X(k)$.

Definition 3.3.4
Let X be a discrete random variable. For any real number t, the probability that X takes on a value $\le t$ is the *cumulative distribution function* (*cdf*) of X [written $F_X(t)$]. In formal notation, $F_X(t) = P(\{s \in S \mid X(s) \le t\})$. As was the case with pdfs, references to s and S are typically deleted, and the cdf is written $F_X(t) = P(X \le t)$.

Example 3.3.9

Suppose we wish to compute $P(21 \le X \le 40)$ for a binomial random variable X with $n = 50$ and $p = 0.6$. From Theorem 3.2.1, we know the formula for $p_X(k)$, so $P(21 \le X \le 40)$ can be written as a simple, although computationally cumbersome, sum:

$$P(21 \le X \le 40) = \sum_{k=21}^{40} \binom{50}{k} (0.6)^k (0.4)^{50-k}$$

Equivalently, the probability we are looking for can be expressed as the difference between two cdfs:

$$P(21 \le X \le 40) = P(X \le 40) - P(X \le 20) = F_X(40) - F_X(20)$$

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

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

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 infinite-valued discrete random variable X whose cdf at the points $x = 1, 2, 3, \ldots$ is given by $F_X(x) = 1 - (1-p)^x$, where $0 < p < 1$.

3.3.16. Recall the game of Fantasy Five from Example 3.2.6. For any Fantasy Five draw of five balls, let the random variable X be the largest number drawn.

(a) Find $F_X(k)$

(b) Find $P(X = 36)$

(c) In one hundred eight plays of Fantasy Five, thirty-six was the largest number drawn fifteen times. Compare this observation to the theoretical probability in part (b).

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 given interval will be numerically similar to the integral of the exponential curve above that same interval.

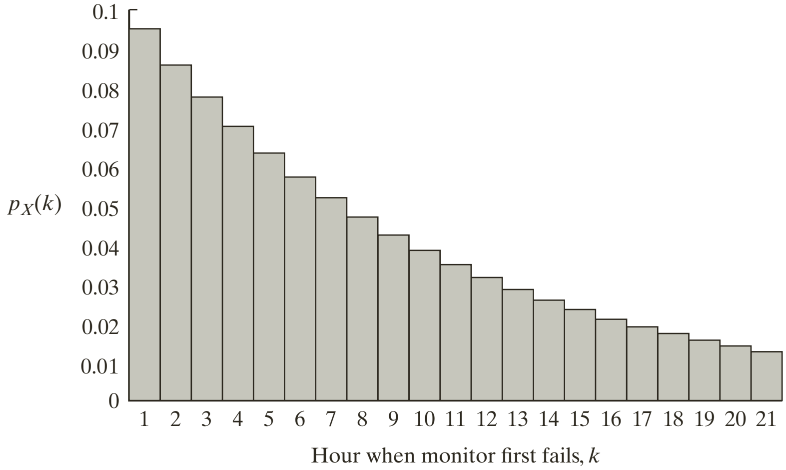


Figure 3.4.1

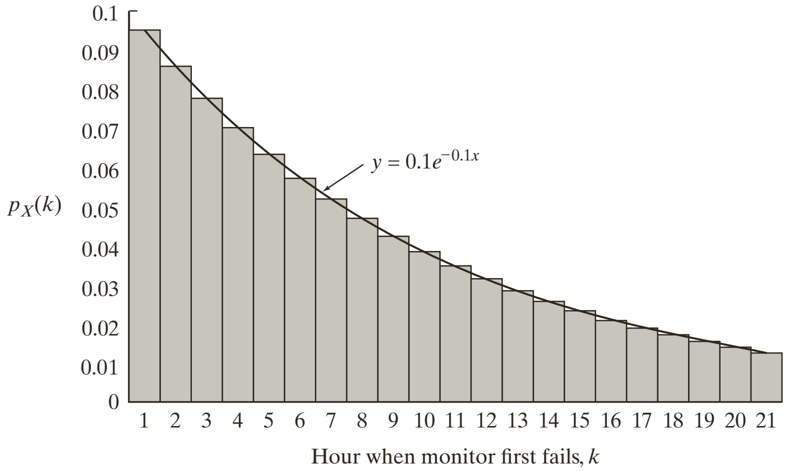


Figure 3.4.2

For example, the probability that the monitor fails sometime during the first four hours would be the sum

$$\begin{aligned} P(1 \le X \le 4) &= \sum_{k=1}^{4} p_X(k) \\ &= \sum_{k=1}^{4} (0.905)^{k-1}(0.095) \\ &= 0.3297 \end{aligned}$$

To four decimal places, the corresponding area under the exponential curve is the same:

$$\int_0^4 0.1e^{-0.1x}\,dx = 0.3297$$

Implicit in the similarity here between $p_X(k)$ and the exponential curve $y = 0.1e^{-0.1x}$ is our sought-after alternative to $p(s)$ for continuous sample spaces. Instead of defining probabilities for individual points, we will define probabilities for *intervals* of points, and those probabilities will be areas under the graph of some function (such as $y = 0.1e^{-0.1x}$), where the shape of the function will reflect the desired probability "measure" to be associated with the sample space.

Definition 3.4.1
A probability function P on a set of real numbers S is called *continuous* if there exists a function $f(t)$ such that for any closed interval $[a, b] \subset S$, $P([a, b]) = \int_a^b f(t)\,dt$. The function $f(t)$ must have the following two properties:

a. $f(t) \geq 0$ for all t.
b. $\int_{-\infty}^{\infty} f(t)dt = 1$.

Then the probability $P(A) = \int_A f(t)dt$ for any set A where the integral is defined.

Comment Using the $-\infty$ and ∞ limits on the integral is simply a convention to mean the function should be integrated over its entire domain. The examples below will make this clear.

Comment If a probability function P satisfies Definition 3.4.1, then it will satisfy the probability axioms given in Section 2.3.

Comment Replacing a sum for discrete probability with an integral is not without its own logic. One of the founders of calculus, Gottfried Wilhelm Leibniz, regarded the integral as an infinite sum of infinitesimal summands. The integral sign was based on a version of the German letter long s (for the German Summe).

CHOOSING THE FUNCTION $f(t)$

We have seen that the probability structure of any sample space with a finite or countably infinite number of outcomes is defined by the function $p(s) = P(\text{Outcome is } s)$. For sample spaces having an uncountably infinite number of possible outcomes, the function $f(t)$ serves an analogous purpose. Specifically, $f(t)$ defines the probability structure of S in the sense that the probability of any *interval* in the sample space is the *integral* of $f(t)$. The next set of examples illustrate several different choices for $f(t)$.

Example 3.4.1 The continuous equivalent of the equiprobable probability model on a discrete sample space is the function $f(t)$ defined by $f(t) = 1/(b - a)$ for all t in the interval $[a, b]$ (and $f(t) = 0$, otherwise). This particular $f(t)$ places equal probability weighting on every closed interval of the same length contained in the interval $[a, b]$. For example, suppose $a = 0$ and $b = 10$, and let $A = [1, 3]$ and $B = [6, 8]$. Then $f(t) = \frac{1}{10}$, and

$$P(A) = \int_1^3 \left(\frac{1}{10}\right) dt = \frac{2}{10} = P(B) = \int_6^8 \left(\frac{1}{10}\right) dt$$

(See Figure 3.4.3.)

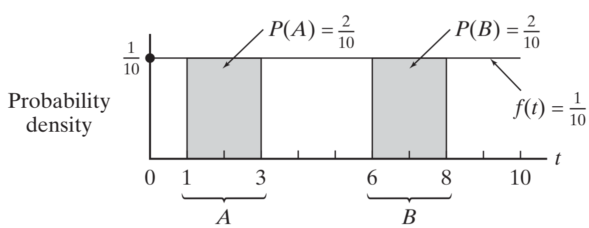


Figure 3.4.3 ■

Example 3.4.2

Could $f(t) = 3t^2$, $0 \le t \le 1$, be used to define the probability function for a continuous sample space whose outcomes consist of all the real numbers in the interval $[0, 1]$? Yes, because (1) $f(t) \ge 0$ for all t, and (2) $\int_0^1 f(t)\,dt = \int_0^1 3t^2\,dt = t^3\big|_0^1 = 1$.

Notice that the shape of $f(t)$ (see Figure 3.4.4) implies that outcomes close to 1 are more likely to occur than are outcomes close to 0. For example, $P\left(\left[0, \frac{1}{3}\right]\right) = \int_0^{1/3} 3t^2\,dt = t^3\big|_0^{1/3} = \frac{1}{27}$, while $P\left(\left[\frac{2}{3}, 1\right]\right) = \int_{2/3}^1 3t^2\,dt = t^3\big|_{2/3}^1 = 1 - \frac{8}{27} = \frac{19}{27}$.

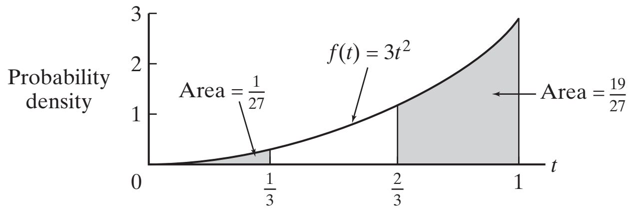


Figure 3.4.4 ■

Example 3.4.3

By far the most important of all continuous probability functions is the "bell-shaped" curve, known more formally as the *normal* (or *Gaussian*) *distribution*. The sample space for the normal distribution is the entire real line; its probability function is given by

$$f(t) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2\right], \quad -\infty < t < \infty, \quad -\infty < \mu < \infty, \quad \sigma > 0$$

Depending on the values assigned to the parameters μ and σ, $f(t)$ can take on a variety of shapes and locations; three are illustrated in Figure 3.4.5.

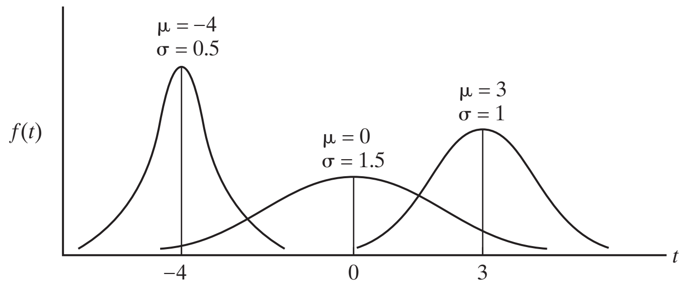


Figure 3.4.5 ■

FITTING $f(t)$ TO DATA: THE DENSITY-SCALED HISTOGRAM

The notion of using a continuous probability function to approximate an integer-valued discrete probability model has already been discussed (recall Figure 3.4.2). The "trick" there was to replace the spikes that define $p_X(k)$ with rectangles whose heights are $p_X(k)$ and whose widths are 1. Doing that makes the sum of the areas of the rectangles corresponding to $p_X(k)$ equal to 1, which is the same as the total area under the approximating continuous probability function. Because of the equality of those two areas, it makes sense to superimpose (and compare) the "histogram" of $p_X(k)$ and the continuous probability function on the same set of axes.

Now, consider the related, but slightly more general, problem of using a continuous probability function to model the distribution of a set of n measurements,

$y_1, y_2, \ldots, y_n$. Following the approach taken in Figure 3.4.2, we would start by making a histogram of the n observations. The problem, though, is that the sum of the areas of the bars comprising that histogram would not necessarily equal 1.

As a case in point, Table 3.4.1 shows a set of forty observations. Grouping those y_i's into five classes, each of width 10, produces the distribution and histogram pictured in Figure 3.4.6. Furthermore, suppose we have reason to believe that these forty y_i's may be a random sample from a uniform probability function defined over the interval [20, 70], that is,

$$f(t) = \frac{1}{70 - 20} = \frac{1}{50}, \quad 20 \le t \le 70$$

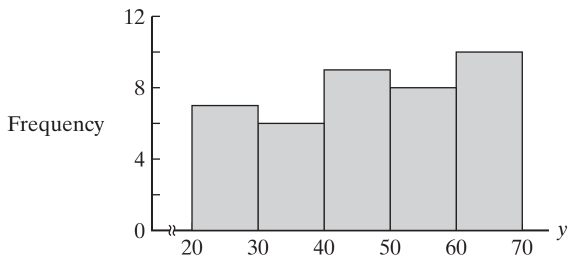
Table 3.4.1

33.8	62.6	42.3	62.9	32.9	58.9	60.8	49.1	42.6	59.8
41.6	54.5	40.5	30.3	22.4	25.0	59.2	67.5	64.1	59.3
24.9	22.3	69.7	41.2	64.5	33.4	39.0	53.1	21.6	46.0
28.1	68.7	27.6	57.6	54.8	48.9	68.4	38.4	69.0	46.6

Recall Example 3.4.1. How can we appropriately draw the distribution of the y_i's and the uniform probability model on the same graph?

Figure 3.4.6

Class	Frequency
$20 \le y < 30$	7
$30 \le y < 40$	6
$40 \le y < 50$	9
$50 \le y < 60$	8
$60 \le y < 70$	10
	40



Note, first, that $f(t)$ and the histogram are not compatible in the sense that the area under $f(t)$ is (necessarily) 1 $(= 50 \times \frac{1}{50})$, but the sum of the areas of the bars making up the histogram is 400:

$$\begin{aligned} \text{histogram area} &= 10(7) + 10(6) + 10(9) + 10(8) + 10(10) \\ &= 400 \end{aligned}$$

Nevertheless, we can "force" the total area of the five bars to match the area under $f(t)$ by redefining the scale of the vertical axis on the histogram. Specifically, *frequency* needs to be replaced with the analog of *probability density,* which would be the scale used on the vertical axis of any graph of $f(t)$. Intuitively, the density associated with, say, the interval [20, 30) would be defined as the quotient

$$\frac{7}{40 \times 10}$$

because integrating that constant over the interval [20, 30) would give $\frac{7}{40}$, and the latter does represent the estimated probability that an observation belongs to the interval [20, 30).

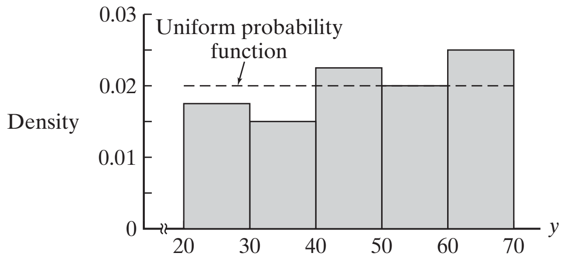
Figure 3.4.7 shows a histogram of the data in Table 3.4.1, where the height of each bar has been converted to a *density,* according to the formula

$$\text{density (of a class)} = \frac{\text{class frequency}}{\text{total no. of observations} \times \text{class width}}$$

Superimposed is the uniform probability model, $f(t) = \frac{1}{50}$, $20 \le t \le 70$. Scaled in this fashion, areas under both $f(t)$ and the histogram are 1.

Figure 3.4.7

Class	Density
$20 \le y < 30$	$7/[40(10)] = 0.0175$
$30 \le y < 40$	$6/[40(10)] = 0.0150$
$40 \le y < 50$	$9/[40(10)] = 0.0225$
$50 \le y < 60$	$8/[40(10)] = 0.0200$
$60 \le y < 70$	$10/[40(10)] = 0.0250$



In practice, density-scaled histograms offer a simple, but effective, format for examining the "fit" between a set of data and a presumed continuous model. We will use it often in the chapters ahead. Applied statisticians have especially embraced this particular graphical technique. Indeed, computer software packages that include *Histograms* on their menus routinely give users the choice of putting either *frequency* or *density* on the vertical axis.

CASE STUDY 3.4.1

Years ago, the V805 transmitter tube was standard equipment on many aircraft radar systems. Table 3.4.2 summarizes part of a reliability study done on the V805; listed are the lifetimes (in hrs) recorded for nine hundred three tubes (38). Grouped into intervals of width 80, the densities for the nine classes are shown in the last column.

Table 3.4.2

Lifetime (hrs)	Number of Tubes	Density
0–80	317	0.0044
80–160	230	0.0032
160–240	118	0.0016
240–320	93	0.0013
320–400	49	0.0007
400–480	33	0.0005
480–560	17	0.0002
560–700	26	0.0002
700+	20	0.0002
	903	

Experience has shown that lifetimes of electrical equipment can often be nicely modeled by the exponential probability function,

$$f(t) = \lambda e^{-\lambda t}, \quad t > 0$$

where the value of λ (for reasons explained in Chapter 5) is set equal to the reciprocal of the average lifetime of the tubes in the sample. Can the distribution of these data also be described by the exponential model?

One way to answer such a question is to superimpose the proposed model on a graph of the density-scaled histogram. The extent to which the two graphs are similar then becomes an obvious measure of the appropriateness of the model.

For these data, λ would be 0.0056. Figure 3.4.8 shows the function

$$f(t) = 0.0056e^{-0.0056t}$$

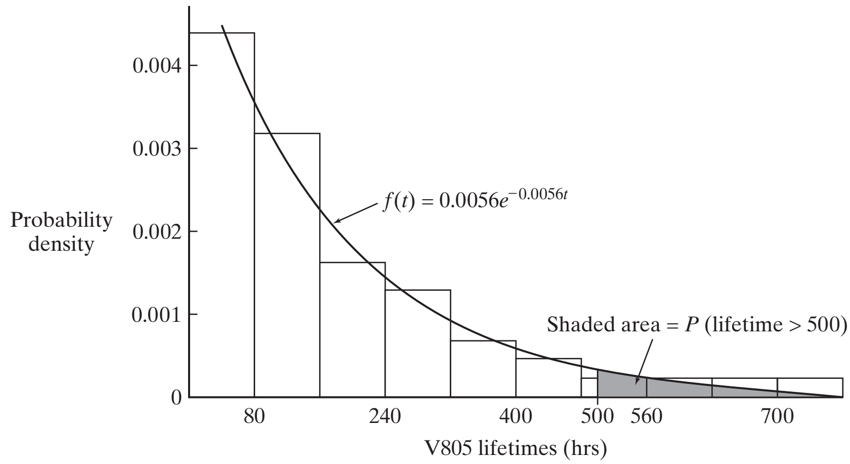


Figure 3.4.8

plotted on the same axes as the density-scaled histogram. Clearly, the agreement is excellent, and we would have no reservations about using areas under $f(t)$ to estimate lifetime probabilities. How likely is it, for example, that a V805 tube will last longer than five hundred hrs? Based on the exponential model, that probability would be 0.0608:

$$P(\text{V805 lifetime exceeds 500 hrs}) = \int_{500}^{\infty} 0.0056e^{-0.0056y}\,dy$$

$$= -e^{-0.0056y}\Big|_{500}^{\infty} = e^{-0.0056(500)} = e^{-2.8} = 0.0608$$

CONTINUOUS PROBABILITY DENSITY FUNCTIONS

We saw in Section 3.3 how the introduction of discrete random variables facilitates the solution of certain problems. The same sort of function can also be defined on sample spaces with an uncountably infinite number of outcomes. Usually, the sample space is an interval of real numbers—finite or infinite. The notation and techniques for this type of random variable replace sums with integrals.

Definition 3.4.2
Let Y be a function from a sample space S to the real numbers. The function Y is called a *continuous random variable* if there exists a function $f_Y(y)$ such that for any real numbers a and b with $a < b$

$$P(a \leq Y \leq b) = \int_a^b f_Y(y)dy$$

The function $f_Y(y)$ is the *probability density function* (*pdf*) for Y.

As in the discrete case, the *cumulative distribution function* (*cdf*) is defined by

$$F_Y(y) = P(Y \le y)$$

The cdf in the continuous case is just an integral of $f_Y(y)$, that is,

$$F_Y(y) = \int_{-\infty}^{y} f_Y(t)\,dt$$

Let $f(y)$ be an arbitrary real-valued function defined on some subset S of the real numbers. If $f(y)$ satisfies Definition 3.4.1, then $f(y) = f_Y(y)$ for all y, where the random variable Y is the identity mapping.

Example 3.4.4 We saw in Case Study 3.4.1 that lifetimes of V805 radar tubes can be nicely modeled by the exponential probability function

$$f(t) = 0.0056e^{-0.0056t}, \quad t > 0$$

To couch that statement in random variable notation would simply require that we define Y to be the life of a V805 radar tube. Then Y would be the identity mapping, and the pdf for the random variable Y would be the same as the probability function, $f(t)$. That is, we would write

$$f_Y(y) = 0.0056e^{-0.0056y}, \quad y \ge 0$$

Similarly, when we work with the bell-shaped normal distribution in later chapters, we will write the model in random variable notation as

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

■

Example 3.4.5 Suppose we would like a continuous random variable Y to "select" a number between 0 and 1 in such a way that intervals near the middle of the range would be more likely to be represented than intervals near either 0 or 1. One pdf having that property is the function $f_Y(y) = 6y(1-y)$, $0 \le y \le 1$ (see Figure 3.4.9). Do we know for certain that the function pictured in Figure 3.4.9 is a "legitimate" pdf? Yes, because $f_Y(y) \ge 0$ for all y between 0 and 1, and $\int_0^1 6y(1-y)\,dy = 6[y^2/2 - y^3/3]\big|_0^1 = 1$.

Comment To simplify the way pdfs are written, it will be assumed that $f_Y(y) = 0$ for all y outside the range actually specified in the function's definition. In Example 3.4.5,

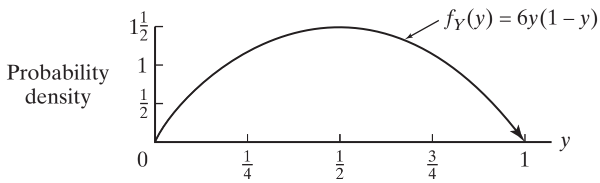


Figure 3.4.9

for instance, the statement $f_Y(y) = 6y(1-y)$, $0 \le y \le 1$, is to be interpreted as an abbreviation for

$$f_Y(y) = \begin{cases} 0, & y < 0 \\ 6y(1-y), & 0 \le y \le 1 \\ 0, & y > 1 \end{cases}$$

■

CONTINUOUS CUMULATIVE DISTRIBUTION FUNCTIONS

Associated with every random variable, discrete or continuous, is a cumulative distribution function. For discrete random variables (recall Definition 3.3.4), the cdf is a nondecreasing step function, where the "jumps" occur at the values of t for which the pdf has positive probability. For continuous random variables, the cdf is a monotonically nondecreasing continuous function. In both cases, the cdf can be helpful in calculating the probability that a random variable takes on a value in a given interval. As we will see in later chapters, there are also several important relationships that hold for continuous cdfs and pdfs. One such relationship is cited in Theorem 3.4.1.

Definition 3.4.3
The cdf for a continuous random variable Y is an indefinite integral of its pdf:

$$F_Y(y) = \int_{-\infty}^{y} f_Y(t)\,dt = P(\{s \in S \mid Y(s) \le y\}) = P(Y \le y)$$

Theorem 3.4.1 *Let $F_Y(y)$ be the cdf of a continuous random variable Y. Then*

$$\frac{d}{dy}F_Y(y) = f_Y(y)$$

Proof The statement of Theorem 3.4.1 follows immediately from the Fundamental Theorem of Calculus.

Theorem 3.4.2 *Let Y be a continuous random variable with cdf $F_Y(y)$. Then*

a. $P(Y > s) = 1 - F_Y(s)$

b. $P(r < Y \le s) = F_Y(s) - F_Y(r)$

c. $\lim_{y\to\infty} F_Y(y) = 1$

d. $\lim_{y\to-\infty} F_Y(y) = 0$

Proof

a. $P(Y > s) = 1 - P(Y \le s)$ since $(Y > s)$ and $(Y \le s)$ are complementary events. But $P(Y \le s) = F_Y(s)$, and the conclusion follows.

b. Since the set $(r < Y \le s) = (Y \le s) - (Y \le r)$, $P(r < Y \le s) = P(Y \le s) - P(Y \le r) = F_Y(s) - F_Y(r)$.

c. Let $\{y_n\}$ be a set of values of Y, $n = 1, 2, 3, \ldots$, where $y_n < y_{n+1}$ for all n, and $\lim_{n\to\infty} y_n = \infty$. If $\lim_{n\to\infty} F_Y(y_n) = 1$ for every such sequence $\{y_n\}$, then $\lim_{y\to\infty} F_Y(y) = 1$. To that end, set $A_1 = (Y \le y_1)$, and let $A_n = (y_{n-1} < Y \le y_n)$

(Continued on next page)

(Theorem 3.4.2 continued)

for $n=2,3,\ldots$. Then $F_Y(y_n)=P(\cup_{k=1}^{n}A_k)=\sum_{k=1}^{n}P(A_k)$, 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 $1=\sum_{k=1}^{\infty}P(A_k)=\lim_{n\to\infty}\sum_{k=1}^{n}P(A_k)=\lim_{n\to\infty}F_Y(y_n)$.

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

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. For a high-risk driver, the time in days between the beginning of a year and an accident has an exponential pdf. Suppose an insurance company believes the probability that such a driver will be involved in an accident in the first forty days is 0.25. What is the probability that such a driver will be involved in an accident during the first seventy-five days of the year?

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 $W = 2Y + 1$. Find $f_W(w)$. For which values of w is $f_W(w) \neq 0$?

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 $P(-a \leq Y \leq a) = 2F_Y(a) - 1$.

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,

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

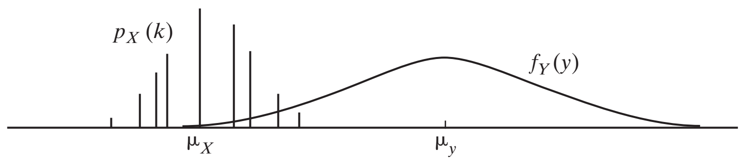
Find $h(y)$ if Y has an exponential pdf (see Question 3.4.8).

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, $P(Y \in A) = \int_A f_Y(y)\,dy$. 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,

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

and

$$p_X(-1) = P(X = -1) = \frac{20}{38} = \frac{10}{19}$$

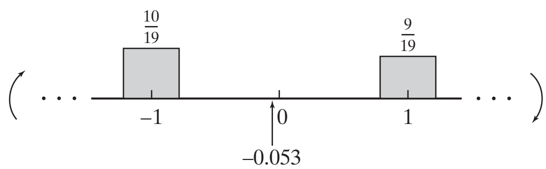
Then $\frac{9}{19}$ of the time we will win \$1 and $\frac{10}{19}$ of the time we will lose \$1. Intuitively, then, if we persist in this foolishness, we stand to *lose*, on the average, a little more than 5¢ each time we play the game:

$$\begin{aligned}\text{"expected" winnings} &= \$1 \cdot \frac{9}{19} + (-\$1) \cdot \frac{10}{19} \\ &= -\$0.053 \doteq -5¢\end{aligned}$$

The number -0.053 is called the *expected value of* X.

Physically, an expected value can be thought of as a center of gravity. Here, for example, imagine two bars of height $\frac{10}{19}$ and $\frac{9}{19}$ positioned along a weightless X-axis at the points -1 and $+1$, respectively (see Figure 3.5.2). If a fulcrum were placed at the point -0.053, the system would be in balance, implying that we can think of that point as marking off the center of the random variable's distribution.

Figure 3.5.2



If X is a discrete random variable taking on each of its values with the same probability, the expected value of X is simply the everyday notion of an arithmetic average or mean:

$$\text{expected value of } X = \sum_{\text{all } k} k \cdot \frac{1}{n} = \frac{1}{n} \sum_{\text{all } k} k$$

Extending this idea to a discrete X described by an arbitrary pdf, $p_X(k)$, gives

$$\text{expected value of } X = \sum_{\text{all } k} k \cdot p_X(k) \tag{3.5.1}$$

For a continuous random variable Y, the summation in Equation 3.5.1 is replaced by an integration and $k \cdot p_X(k)$ becomes $y \cdot f_Y(y)$.

Definition 3.5.1
Let X be a discrete random variable with probability function $p_X(k)$. The *expected value of* X is denoted $E(X)$ (or sometimes μ or μ_X) and is given by

$$E(X) = \mu = \mu_X = \sum_{\text{all } k} k \cdot p_X(k)$$

Similarly, if Y is a continuous random variable with pdf $f_Y(y)$,

$$E(Y) = \mu = \mu_Y = \int_{-\infty}^{\infty} y \cdot f_Y(y)\, dy$$

Comment We assume that both the sum and the integral in Definition 3.5.1 converge absolutely:

$$\sum_{\text{all } k} |k| p_X(k) < \infty \qquad \int_{-\infty}^{\infty} |y| f_Y(y)\, dy < \infty$$

If not, we say that the random variable has no finite expected value. One immediate reason for requiring *absolute* convergence is that a convergent sum that is not absolutely convergent depends on the order in which the terms are added, and order should obviously not be a consideration when defining an average.

Example 3.5.1 Suppose X is a binomial random variable with $p = \frac{5}{9}$ and $n = 3$. Then $p_X(k) = P(X = k) = \binom{3}{k}\left(\frac{5}{9}\right)^k\left(\frac{4}{9}\right)^{3-k}$, $k = 0, 1, 2, 3$. What is the expected value of X?

Applying Definition 3.5.1 gives

$$E(X) = \sum_{k=0}^{3} k \cdot \binom{3}{k}\left(\frac{5}{9}\right)^k\left(\frac{4}{9}\right)^{3-k}$$

$$= (0)\left(\frac{64}{729}\right) + (1)\left(\frac{240}{729}\right) + (2)\left(\frac{300}{729}\right) + (3)\left(\frac{125}{729}\right) = \frac{1215}{729} = \frac{5}{3} = 3\left(\frac{5}{9}\right)$$

Comment Notice that the expected value here reduces to five-thirds, which can be written as three times five-ninths, the latter two factors being n and p, respectively. As the next theorem proves, that relationship is not a coincidence. ■

Theorem 3.5.1 *Suppose X is a binomial random variable with parameters n and p. Then $E(X) = np$.*

Proof According to Definition 3.5.1, $E(X)$ for a binomial random variable is the sum

$$E(X) = \sum_{k=0}^{n} k \cdot p_X(k) = \sum_{k=0}^{n} k\binom{n}{k} p^k (1-p)^{n-k}$$

$$= \sum_{k=0}^{n} \frac{k \cdot n!}{k!(n-k)!} p^k (1-p)^{n-k}$$

$$= \sum_{k=1}^{n} \frac{n!}{(k-1)!(n-k)!} p^k (1-p)^{n-k} \tag{3.5.2}$$

At this point, a trick is called for. If $E(X) = \sum_{\text{all } k} g(k)$ can be factored in such a way that $E(X) = h \sum_{\text{all } k} p_{X^*}(k)$, where $p_{X^*}(k)$ is the pdf for some random variable X^*, then $E(X) = h$, since the sum of a pdf over its entire range is 1. Here, suppose that np is factored out of Equation 3.5.2. Then

$$E(X) = np \sum_{k=1}^{n} \frac{(n-1)!}{(k-1)!(n-k)!} p^{k-1} (1-p)^{n-k}$$

$$= np \sum_{k=1}^{n} \binom{n-1}{k-1} p^{k-1} (1-p)^{n-k}$$

(Continued on next page)

(Theorem 3.5.1 continued)

Now, let $j = k - 1$. It follows that

$$E(X) = np \sum_{j=0}^{n-1} \binom{n-1}{j} p^j (1-p)^{n-j-1}$$

Finally, letting $m = n - 1$ gives

$$E(X) = np \sum_{j=0}^{m} \binom{m}{j} p^j (1-p)^{m-j}$$

and, since the value of the sum is 1 (why?),

$$E(X) = np \tag{3.5.3}$$

Comment The statement of Theorem 3.5.1 should come as no surprise. If a multiple-choice test, for example, has one hundred questions, each with five possible answers, we would "expect" to get twenty correct, just by guessing. But if the random variable X denotes the number of correct answers (out of one hundred), $20 = E(X) = 100\left(\frac{1}{5}\right) = np$.

Example 3.5.2 An urn contains nine chips, five red and four white. Three are drawn out at random without replacement. Let X denote the number of red chips in the sample. Find $E(X)$.

From Section 3.2, we recognize X to be a hypergeometric random variable, where

$$P(X = k) = p_X(k) = \frac{\binom{5}{k}\binom{4}{3-k}}{\binom{9}{3}}, \quad k = 0, 1, 2, 3$$

Therefore,

$$\begin{aligned} E(X) &= \sum_{k=0}^{3} k \cdot \frac{\binom{5}{k}\binom{4}{3-k}}{\binom{9}{3}} \\ &= (0)\left(\frac{4}{84}\right) + (1)\left(\frac{30}{84}\right) + (2)\left(\frac{40}{84}\right) + (3)\left(\frac{10}{84}\right) \\ &= \frac{5}{3} \end{aligned}$$

■

Comment As was true in Example 3.5.1, the value found here for $E(X)$ suggests a general formula—in this case, for the expected value of a hypergeometric random variable.

Theorem 3.5.2 *Suppose X is a hypergeometric random variable with parameters r, w, and n. That is, suppose an urn contains r red balls and w white balls. A sample of size n is drawn simultaneously from the urn. Let X be the number of red balls in the sample. Then $E(X) = \frac{rn}{r+w}$.*

Proof See Question 3.5.25.

Comment Let p represent the proportion of red balls in an urn, that is, $p = \frac{r}{r+w}$. The formula, then, for the expected value of a hypergeometric random variable has the same structure as the formula for the expected value of a binomial random variable:

$$E(X) = \frac{rn}{r+w} = n\frac{r}{r+w} = np$$

Example 3.5.3

Among the more common versions of the "numbers" racket is a game called D.J., its name deriving from the fact that the winning ticket is determined from Dow Jones averages. Three sets of stocks are used: Industrials, Transportations, and Utilities. Traditionally, the three are quoted at two different times, 11 A.M. and noon. The last digits of the earlier quotation are arranged to form a three-digit number; the noon quotation generates a second three-digit number, formed the same way. Those two numbers are then added together and the last three digits of that sum become the winning pick. Figure 3.5.3 shows a set of quotations for which 906 would be declared the winner.

11 A.M. quotation	
Industrials	845.61
Transportation	375.27
Utilities	110.63

Noon quotation	
Industrials	848.17
Transportation	376.73
Utilities	110.63

173 + 733

=

906 = Winning number

Figure 3.5.3

The payoff in D.J. is 700 to 1. Suppose that we bet \$5. How much do we stand to win, or lose, *on the average*?

Let p denote the probability of our number being the winner and let X denote our earnings. Then

$$X = \begin{cases} \$3500 & \text{with probability } p \\ -\$5 & \text{with probability } 1-p \end{cases}$$

and

$$E(X) = \$3500 \cdot p - \$5 \cdot (1-p)$$

Our intuition would suggest (and this time it would be correct!) that each of the possible winning numbers, 000 through 999, is equally likely. That being the case, $p = 1/1000$ and

$$E(X) = \$3500 \cdot \left(\frac{1}{1000}\right) - \$5 \cdot \left(\frac{999}{1000}\right) = -\$1.50$$

On the average, then, we lose \$1.50 on a \$5.00 bet. ■

Example 3.5.4

Suppose that fifty people are to be given a blood test to see who has a certain disease. The obvious laboratory procedure is to examine each person's blood individually, meaning that fifty tests would eventually be run. An alternative strategy is to divide each person's blood sample into two parts—say, A and B. All of the A's would then be mixed together and treated as one sample. If that "pooled" sample proved to be negative for the disease, all fifty individuals must necessarily be free of the infection, and no further testing would need to be done. If the pooled sample gave a positive reading, of course, all fifty B samples would have to be analyzed separately. Under what conditions would it make sense for a laboratory to consider pooling the fifty samples?

In principle, the pooling strategy is preferable (i.e., more economical) if it can substantially reduce the number of tests that need to be performed. Whether or not it can do so depends ultimately on the probability p that a person is infected with the disease.

Let the random variable X denote the number of tests that will have to be performed if the samples are pooled. Clearly,

$$X = \begin{cases} 1 & \text{if none of the fifty is infected} \\ 51 & \text{if at least one of the fifty is infected} \end{cases}$$

But

$$P(X = 1) = p_X(1) = P(\text{None of the fifty is infected}) = (1 - p)^{50}$$

(assuming independence), and

$$P(X = 51) = p_X(51) = 1 - P(X = 1) = 1 - (1 - p)^{50}$$

Therefore,

$$E(X) = 1 \cdot (1 - p)^{50} + 51 \cdot [1 - (1 - p)^{50}]$$

Table 3.5.1 shows $E(X)$ as a function of p. As our intuition would suggest, the pooling strategy becomes increasingly feasible as the prevalence of the disease diminishes. If the chance of a person being infected is 1 in 1000, for example, the pooling strategy requires an average of only 3.4 tests, a dramatic improvement over the fifty tests that would be needed if the samples were tested one by one. On the other hand, if 1 in 10 individuals is infected, pooling would be clearly inappropriate, requiring *more* than fifty tests [$E(X) = 50.7$].

Table 3.5.1

p	$E(X)$
0.5	51.0
0.1	50.7
0.01	20.8
0.001	3.4
0.0001	1.2

■

Example 3.5.5

Consider the following game. A fair coin is flipped until the first tail appears; we win \$2 if it appears on the first toss, \$4 if it appears on the second toss, and, in general, \$$2^k$ if it first occurs on the kth toss. Let the random variable X denote our winnings. How much should we have to pay in order for this to be a fair game? (*Note:* A fair game is one where the difference between the ante and $E(X)$ is 0.)

Known as the St. Petersburg paradox, this problem has a rather unusual answer. First, note that

$$p_X(2^k) = P(X = 2^k) = \frac{1}{2^k}, \quad k = 1, 2, \ldots$$

Therefore,

$$E(X) = \sum_{\text{all } k} 2^k p_X(2^k) = \sum_{k=1}^{\infty} 2^k \cdot \frac{1}{2^k} = 1 + 1 + 1 + \cdots$$

which is a divergent sum. That is, X does not have a finite expected value, so in order for this game to be fair, our ante would have to be an infinite amount of money! ■

Comment Mathematicians have been trying to "explain" the St. Petersburg paradox for almost two hundred years (61). The answer seems clearly absurd—no gambler would consider paying even \$25 to play such a game, much less an infinite amount—yet the computations involved in showing that X has no finite expected value are unassailably correct. Where the difficulty lies, according to one common theory, is with our inability to put in perspective the very small probabilities of winning very large payoffs. Furthermore, the problem assumes that our opponent has infinite capital, which is an impossible state of affairs. We get a much more reasonable answer for $E(X)$ if the stipulation is added that our winnings can be at most, say, \$1000 (see Question 3.5.19) or if the payoffs are assigned according to some formula other than 2^k (see Question 3.5.20).

Comment There are two important lessons to be learned from the St. Petersburg paradox. First is the realization that $E(X)$ is not necessarily a meaningful characterization of the "location" of a distribution. Question 3.5.24 shows another situation where the formal computation of $E(X)$ gives a similarly inappropriate answer. Second, we need to be aware that the notion of expected value is not necessarily synonymous with the concept of *worth*. Just because a game, for example, has a positive expected value—even a very *large* positive expected value—does not imply that someone would want to play it. Suppose, for example, that you had the opportunity to spend your last \$10,000 on a sweepstakes ticket where the prize was \$1 billion but the probability of winning was only one in ten thousand. The expected value of such a bet would be over \$90,000,

$$\begin{aligned} E(X) &= \$1{,}000{,}000{,}000\left(\frac{1}{10{,}000}\right) + (-\$10{,}000)\left(\frac{9{,}999}{10{,}000}\right) \\ &= \$90{,}001 \end{aligned}$$

but it is doubtful that many people would rush out to buy a ticket. (Economists have long recognized the distinction between a payoff's numerical value and its perceived desirability. They refer to the latter as *utility*.)

Example 3.5.6 The distance, Y, that a molecule in a gas travels before colliding with another molecule can be modeled by the exponential pdf

$$f_Y(y) = \frac{1}{\mu} e^{-y/\mu}, \quad y \geq 0$$

where μ is a positive constant known as the *mean free path*. Find $E(Y)$.

Since the random variable here is continuous, its expected value is an integral:

$$E(Y) = \int_0^{\infty} y \frac{1}{\mu} e^{-y/\mu}\, dy$$

Let $w = y/\mu$, so that $dw = 1/\mu\ dy$. Then $E(Y) = \mu \int_0^\infty we^{-w}dw$. Setting $u = w$ and $dv = e^{-w}dw$ and integrating by parts gives

$$E(Y) = \mu[-we^{-w} - e^{-w}]\Big|_0^\infty = \mu \tag{3.5.4}$$

Equation 3.5.4 shows that μ is aptly named—it does, in fact, represent the average distance a molecule travels, free of any collisions. Nitrogen (N_2), for example, at room temperature and standard atmospheric pressure has $\mu = 0.00005$ cm. An N_2 molecule, then, travels that far before colliding with another N_2 molecule, *on the average*. ■

Example 3.5.7 One continuous pdf that has a number of interesting applications in physics is the *Rayleigh distribution,* where the pdf is given by

$$f_Y(y) = \frac{y}{a^2}\, e^{-y^2/2a^2}, \quad a > 0; \quad 0 \le y < \infty \tag{3.5.5}$$

Calculate the expected value for a random variable having a Rayleigh distribution.

From Definition 3.5.1,

$$E(Y) = \int_0^\infty y \cdot \frac{y}{a^2} e^{-y^2/2a^2}\, dy$$

Let $v = y/(\sqrt{2}a)$. Then

$$E(Y) = 2\sqrt{2}a \int_0^\infty v^2 e^{-v^2}\, dv$$

The integrand here is a special case of the general form $v^{2k}e^{-v^2}$. For $k = 1$,

$$\int_0^\infty v^{2k}e^{-v^2}\, dv = \int_0^\infty v^2 e^{-v^2}\, dv = \frac{1}{4}\sqrt{\pi}$$

Therefore,

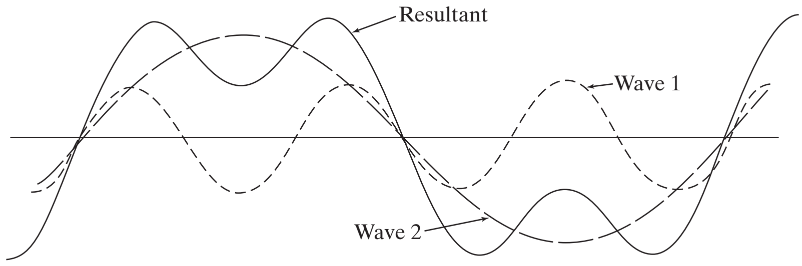
$$\begin{aligned} E(Y) &= 2\sqrt{2}a \cdot \frac{1}{4}\sqrt{\pi} \\ &= a\sqrt{\pi/2} \end{aligned}$$ ■

Comment The pdf here is named for John William Strutt, Baron Rayleigh, the nineteenth- and twentieth-century British physicist who showed that Equation 3.5.5 is the solution to a problem arising in the study of wave motion. If two waves are superimposed, it is well known that the height of the resultant at any time t is simply the algebraic sum of the corresponding heights of the waves being added (see Figure 3.5.4). Seeking to extend that notion, Rayleigh posed the following question: If n waves, each having the same amplitude h and the same wavelength, are superimposed randomly with respect to phase, what can we say about the amplitude R of the resultant? Clearly, R is a random variable, its value depending on the particular collection of phase angles represented by the sample. What Rayleigh was able to show in his 1880 paper (177) is that when n is large, the probabilistic behavior of R is described by the pdf

$$f_R(r) = \frac{2r}{nh^2} \cdot e^{-r^2/nh^2}, \quad r > 0$$

which is just a special case of Equation 3.5.5 with $a = \sqrt{2/nh^2}$.

Figure 3.5.4



A SECOND MEASURE OF CENTRAL TENDENCY: THE MEDIAN

While the expected value is the most frequently used measure of a random variable's central tendency, it does have a weakness that sometimes makes it misleading and inappropriate. Specifically, if one or several possible values of a random variable are either much smaller or much larger than all the others, the value of μ can be distorted in the sense that it no longer reflects the center of the distribution in any meaningful way. For example, suppose a small community consists of a homogeneous group of middle-range salary earners, and then Ms. Rich moves to town. Obviously, the town's average salary before and after the multibillionaire arrives will be quite different, even though she represents only one new value of the "salary" random variable.

It would be helpful to have a measure of central tendency that is not so sensitive to "outliers" or to probability distributions that are markedly skewed. One such measure is the *median*, which, in effect, divides the area under a pdf into two equal areas.

Definition 3.5.2
If X is a discrete random variable, the median, m, is that point for which $P(X < m) = P(X > m)$. In the event that $P(X \leq m) = 0.5$ and $P(X \geq m') = 0.5$, the median is defined to be the arithmetic average, $(m + m')/2$.

If Y is a continuous random variable, its median is the solution to the integral equation $\int_{-\infty}^{m} f_Y(y)\,dy = 0.5$.

Example 3.5.8

If a random variable's pdf is symmetric, μ and m will be equal. Should $p_X(k)$ or $f_Y(y)$ not be symmetric, though, the difference between the expected value and the median can be considerable, especially if the asymmetry takes the form of extreme skewness. The situation described here is a case in point.

Soft-glow makes a 60-watt light bulb that is advertised to have an average life of one thousand hours. Assuming that the performance claim is valid, is it reasonable for consumers to conclude that the Soft-glow bulbs they buy will last for approximately one thousand hours?

No! If the average life of a bulb is one thousand hours, the (continuous) pdf, $f_Y(y)$, modeling the length of time, Y, that it remains lit before burning out is likely to have the form

$$f_Y(y) = 0.001e^{-0.001y}, \quad y > 0 \tag{3.5.6}$$

(for reasons explained in Chapter 4). But Equation 3.5.6 is a very skewed pdf, having a shape much like the curve drawn in Figure 3.4.8. The median for such a distribution will lie considerably to the left of the mean.

More specifically, the median lifetime for these bulbs—according to Definition 3.5.2—is the value m for which

$$\int_0^m 0.001e^{-0.001y}\,dy = 0.5$$

But $\int_0^m 0.001e^{-0.001y}\,dy = 1 - e^{-0.001m}$. Setting the latter equal to 0.5 implies that

$$m = (1/-0.001)\ln(0.5) = 693$$

So, even though the *average* life of one of these bulbs is one thousand hours, there is a 50% chance that the one you buy will last less than six hundred ninety-three hours. ■

Questions

3.5.1. Recall the game of Keno described in Question 3.2.26. The following are all the payoffs on a \$1 wager where the player has bet on ten numbers. Calculate $E(X)$, where the random variable X denotes the amount of money won.

Number of Correct Guesses	Payoff	Probability
< 5	$-\$1$	.935
5	2	.0514
6	18	.0115
7	180	.0016
8	1,300	1.35×10^{-4}
9	2,600	6.12×10^{-6}
10	10,000	1.12×10^{-7}

3.5.2. The roulette wheels in Monte Carlo typically have a 0 but not a 00. What is the expected value of betting on red in this case? If a trip to Monte Carlo costs \$3000, how much would a player have to bet to justify gambling there rather than Las Vegas?

3.5.3. The pdf describing the daily profit, X, earned by Acme Industries was derived in Example 3.3.7. Find the company's *average* daily profit.

3.5.4. In the game of redball, two drawings are made without replacement from a bowl that has four white ping-pong balls and two red ping-pong balls. The amount won is determined by how many of the red balls are selected. For a \$5 bet, a player can opt to be paid under either Rule A or Rule B, as shown. If you were playing the game, which would you choose? Why?

A		B	
No. of Red Balls Drawn	Payoff	No. of Red Balls Drawn	Payoff
0	0	0	0
1	\$2	1	\$1
2	\$10	2	\$20

3.5.5. Suppose a life insurance company sells a \$50,000, five-year term policy to a twenty-five-year-old woman. At the beginning of each year the woman is alive, the company collects a premium of \$$P$. The probability that the woman dies and the company pays the \$50,000 is given in the table below. So, for example, in Year 3, the company loses \$50,000 – \$$P$ with probability 0.00054 and gains \$$P$ with probability $1 - 0.00054 = 0.99946$. If the company expects to make \$1000 on this policy, what should P be?

Year	Probability of Payoff
1	0.00051
2	0.00052
3	0.00054
4	0.00056
5	0.00059

3.5.6. A manufacturer has one hundred memory chips in stock, 4% of which are likely to be defective (based on past experience). A random sample of twenty chips is selected and shipped to a factory that assembles laptops. Let X denote the number of computers that receive faulty memory chips. Find $E(X)$.

3.5.7. Records show that 642 new students have just entered a certain Florida school district. Of those 642, a total of 125 are not adequately vaccinated. The district's physician has scheduled a day for students to receive whatever shots they might need. On any given day, though, 12% of the district's students are likely to be absent. How many new students, then, can be expected to remain inadequately vaccinated?

3.5.8. Calculate $E(Y)$ for the following pdfs:

(a) $f_Y(y) = 3(1-y)^2,\ 0 \le y \le 1$

(b) $f_Y(y) = 4ye^{-2y},\ y \ge 0$

(c) $f_Y(y) = \begin{cases} \frac{3}{4}, & 0 \le y \le 1 \\ \frac{1}{4}, & 2 \le y \le 3 \\ 0, & \text{elsewhere} \end{cases}$

(d) $f_Y(y) = \sin y, \quad 0 \le y \le \frac{\pi}{2}$

3.5.9. Recall Question 3.4.4, where the length of time Y (in years) that a malaria patient spends in remission has pdf $f_Y(y) = \frac{1}{9}y^2, 0 \le y \le 3$. What is the average length of time that such a patient spends in remission?

3.5.10. Let the random variable Y have the uniform distribution over $[a, b]$; that is, $f_Y(y) = \frac{1}{b-a}$ for $a \le y \le b$. Find $E(Y)$ using Definition 3.5.1. Also, deduce the value of $E(Y)$, knowing that the expected value is the center of gravity of $f_Y(y)$.

3.5.11. Show that the expected value associated with the exponential distribution, $f_Y(y) = \lambda e^{-\lambda y}, y > 0$, is $1/\lambda$, where λ is a positive constant.

3.5.12. Show that

$$f_Y(y) = \frac{1}{y^2}, \quad y \ge 1$$

is a valid pdf but that Y does not have a finite expected value.

3.5.13. Based on recent experience, ten-year-old passenger cars going through a motor vehicle inspection station have an 80% chance of passing the emissions test. Suppose that two hundred such cars will be checked out next week. Write two formulas that show the number of cars that are expected to pass.

3.5.14. Suppose that fifteen observations are chosen at random from the pdf $f_Y(y) = 3y^2, 0 \le y \le 1$. Let X denote the number that lie in the interval $\left(\frac{1}{2}, 1\right)$. Find $E(X)$.

3.5.15. A city has 74,806 registered automobiles. Each is required 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. How much wheel tax revenue can the city expect to receive in November?

3.5.16. Regulators have found that twenty-three of the sixty-eight investment companies that filed for bankruptcy in the past five years failed because of fraud, not for reasons related to the economy. Suppose that nine additional firms will be added to the bankruptcy rolls during the next quarter. How many of those failures are likely to be attributed to fraud?

3.5.17. An urn contains four chips numbered 1 through 4. Two are drawn without replacement. Let the random variable X denote the larger of the two. Find $E(X)$.

3.5.18. A fair coin is tossed three times. Let the random variable X denote the total number of heads that appear times the number of heads that appear on the first and third tosses. Find $E(X)$.

3.5.19. How much would you have to ante to make the St. Petersburg game "fair" (recall Example 3.5.5) if the most you could win was \$1000? That is, the payoffs are $\$2^k$ for $1 \le k \le 9$, and \$1000 for $k \ge 10$.

3.5.20. For the St. Petersburg problem (Example 3.5.5), find the expected payoff if

(a) the amounts won are c^k instead of 2^k, where $0 < c < 2$.

(b) the amounts won are $\log 2^k$. [This was a modification suggested by D. Bernoulli (a nephew of James Bernoulli) to take into account the decreasing marginal utility of money—the more you have, the less useful a bit more is.]

3.5.21. A fair die is rolled three times. Let X denote the number of different faces showing, $X = 1, 2, 3$. Find $E(X)$.

3.5.22. Two distinct integers are chosen at random from the first five positive integers. Compute the expected value of the absolute value of the difference of the two numbers.

3.5.23. Suppose that two evenly matched teams are playing in the World Series. On the average, how many games will be played? (The winner is the first team to get four victories.) Assume that each game is an independent event.

3.5.24. An urn contains one white chip and one black chip. A chip is drawn at random. If it is white, the "game" is over; if it is black, that chip and another black one are put into the urn. Then another chip is drawn at random from the "new" urn and the same rules for ending or continuing the game are followed (i.e., if the chip is white, the game is over; if the chip is black, it is placed back in the urn, together with another chip of the same color). The drawings continue until a white chip is selected. Show that the expected number of drawings necessary to get a white chip is not finite.

3.5.25. A random sample of size n is drawn without replacement from an urn containing r red chips and w white chips. Define the random variable X to be the number of red chips in the sample. Use the summation technique described in Theorem 3.5.1 to prove that $E(X) = rn/(r + w)$.

3.5.26. Given that X is a nonnegative, integer-valued random variable, show that

$$E(X) = \sum_{k=1}^{\infty} P(X \ge k)$$

3.5.27. Find the median for each of the following pdfs:

(a) $f_Y(y) = (\theta + 1)y^\theta, \ 0 \le y \le 1$, where $\theta > 0$

(b) $f_Y(y) = y + \frac{1}{2}, \ 0 \le y \le 1$

THE EXPECTED VALUE OF A FUNCTION OF A RANDOM VARIABLE

There are many situations that call for finding the expected value of a *function* of a random variable—say, $Y = g(X)$. One common example would be change of scale problems, where $g(X) = aX + b$ for constants a and b. Sometimes the pdf of the new random variable Y can be easily determined, in which case $E(Y)$ can be calculated by simply applying Definition 3.5.1. Often, though, $f_Y(y)$ can be difficult to derive, depending on the complexity of $g(X)$. Fortunately, Theorem 3.5.3 allows us to calculate the expected value of Y without knowing the pdf for Y.

Theorem 3.5.3 *Suppose X is a discrete random variable with pdf $p_X(k)$. Let $g(X)$ be a function of X. Then the expected value of the random variable $g(X)$ is given by*

$$E[g(X)] = \sum_{\text{all } k} g(k) \cdot p_X(k)$$

provided that $\sum_{\text{all } k} |g(k)| p_X(k) < \infty$.

If Y is a continuous random variable with pdf $f_Y(y)$, and if $g(Y)$ is a continuous function, then the expected value of the random variable $g(Y)$ is

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

provided that $\int_{-\infty}^{\infty} |g(y)| f_Y(y)\, dy < \infty$.

Proof We will prove the result for the discrete case. See (155) for details showing how the argument is modified when the pdf is continuous. Let $W = g(X)$. The set of all possible k values, $k_1, k_2, \ldots$, of X will give rise to a set of w values, $w_1, w_2, \ldots$, where, in general, more than one k may be associated with a given w. Let S_j be the set of k's for which $g(k) = w_j$ [so $\cup_j S_j$ is the entire set of k values for which $p_X(k)$ is defined]. We obviously have that $P(W = w_j) = P(X \in S_j)$, and we can write

$$\begin{aligned}
E(W) = \sum_j w_j \cdot P(W = w_j) &= \sum_j w_j \cdot P(X \in S_j) \\
&= \sum_j w_j \sum_{k \in S_j} p_X(k) \\
&= \sum_j \sum_{k \in S_j} w_j \cdot p_X(k) \\
&= \sum_j \sum_{k \in S_j} g(k) p_X(k) \quad \text{(why?)} \\
&= \sum_{\text{all } k} g(k) p_X(k)
\end{aligned}$$

Since it is being assumed that $\sum_{\text{all } k} |g(k)| p_X(k) < \infty$, the statement of the theorem holds.

Corollary 3.5.1 *For any random variable W, $E(aW + b) = aE(W) + b$, where a and b are constants.*

Proof Suppose W is continuous; the proof for the discrete case is similar. By Theorem 3.5.3, $E(aW + b) = \int_{-\infty}^{\infty} (aw + b) f_W(w)\, dw$, but the latter can be written $a \int_{-\infty}^{\infty} w \cdot f_W(w)\, dw + b \int_{-\infty}^{\infty} f_W(w)\, dw = aE(W) + b \cdot 1 = aE(W) + b$.

Example 3.5.9

Suppose that X is a random variable whose pdf is nonzero only for the three values -2, 1, and $+2$:

k	$p_X(k)$
-2	$\frac{5}{8}$
1	$\frac{1}{8}$
2	$\frac{2}{8}$
	1

Let $W = g(X) = X^2$. Verify the statement of Theorem 3.5.3 by computing $E(W)$ two ways—first, by finding $p_W(w)$ and summing $w \cdot p_W(w)$ over w and, second, by summing $g(k) \cdot p_X(k)$ over k.

By inspection, the pdf for W is defined for only two values, 1 and 4:

$w\ (= k^2)$	$p_W(w)$
1	$\frac{1}{8}$
4	$\frac{7}{8}$
	1

Taking the first approach to find $E(W)$ gives

$$E(W) = \sum_w w \cdot p_W(w) = 1 \cdot \left(\frac{1}{8}\right) + 4 \cdot \left(\frac{7}{8}\right)$$
$$= \frac{29}{8}$$

To find the expected value via Theorem 3.5.3, we take

$$E[g(X)] = \sum_k k^2 \cdot p_X(k) = (-2)^2 \cdot \frac{5}{8} + (1)^2 \cdot \frac{1}{8} + (2)^2 \cdot \frac{2}{8}$$

with the sum here reducing to the answer we already found, $\frac{29}{8}$.

For this particular situation, neither approach was easier than the other. In general, that will not be the case. Finding $p_W(w)$ is often quite difficult, and on those occasions Theorem 3.5.3 can be of great benefit. ■

Example 3.5.10

Suppose the amount of propellant, Y, put into a can of spray paint is a random variable with pdf

$$f_Y(y) = 3y^2, \quad 0 < y < 1$$

Experience has shown that the largest surface area that can be painted by a can having Y amount of propellant is twenty times the area of a circle generated by a radius of Y ft. If the Purple Dominoes, a newly formed urban gang, have just stolen their first can of spray paint, can they expect to have enough to cover a $5' \times 8'$ subway panel with grafitti?

No. By assumption, the maximum area (in ft^2) that can be covered by a can of paint is described by the function

$$g(Y) = 20\pi Y^2$$

According to the second statement in Theorem 3.5.3, though, the average value for $g(Y)$ is slightly less than the desired 40 ft^2:

$$\begin{aligned} E[g(Y)] &= \int_0^1 20\pi y^2 \cdot 3y^2\,dy \\ &= \left.\frac{60\pi y^5}{5}\right|_0^1 \\ &= 12\pi \\ &= 37.7\ \text{ft}^2 \end{aligned}$$

■

Example 3.5.11 A fair coin is tossed until a head appears. You will be given $\left(\frac{1}{2}\right)^k$ dollars if that first head occurs on the kth toss. How much money can you expect to be paid?

Let the random variable X denote the toss at which the first head appears. Then

$$\begin{aligned} p_X(k) = P(X = k) &= P(\text{First } k-1 \text{ tosses are tails and } k\text{th toss is a head}) \\ &= \left(\frac{1}{2}\right)^{k-1} \cdot \frac{1}{2} \\ &= \left(\frac{1}{2}\right)^k, \quad k = 1, 2, \ldots \end{aligned}$$

Moreover,

$$\begin{aligned} E(\text{amount won}) = E\left[\left(\frac{1}{2}\right)^X\right] = E[g(X)] &= \sum_{\text{all } k} g(k) \cdot p_X(k) \\ &= \sum_{k=1}^{\infty} \left(\frac{1}{2}\right)^k \cdot \left(\frac{1}{2}\right)^k \\ &= \sum_{k=1}^{\infty} \left(\frac{1}{2}\right)^{2k} = \sum_{k=1}^{\infty} \left(\frac{1}{4}\right)^k \\ &= \sum_{k=0}^{\infty} \left(\frac{1}{4}\right)^k - \left(\frac{1}{4}\right)^0 \\ &= \frac{1}{1-\frac{1}{4}} - 1 \\ &= \$0.33 \end{aligned}$$

■

Example 3.5.12 In one of the early applications of probability to physics, James Clerk Maxwell (1831–1879) showed that the speed S of a molecule in a perfect gas has a density function given by

$$f_S(s) = 4\sqrt{\frac{a^3}{\pi}}\, s^2 e^{-as^2}, \quad s > 0$$

where a is a constant depending on the temperature of the gas and the mass of the particle. What is the average *energy* of a molecule in a perfect gas?

Let m denote the molecule's mass. Recall from physics that energy (W), mass (m), and speed (S) are related through the equation

$$W = \frac{1}{2}mS^2 = g(S)$$

To find $E(W)$ we appeal to the second part of Theorem 3.5.3:

$$\begin{aligned} E(W) &= \int_0^\infty g(s) f_S(s)\, ds \\ &= \int_0^\infty \frac{1}{2}ms^2 \cdot 4\sqrt{\frac{a^3}{\pi}} s^2 e^{-as^2}\, ds \\ &= 2m\sqrt{\frac{a^3}{\pi}} \int_0^\infty s^4 e^{-as^2}\, ds \end{aligned}$$

We make the substitution $t = as^2$. Then

$$E(W) = \frac{m}{a\sqrt{\pi}} \int_0^\infty t^{3/2} e^{-t}\, dt$$

But

$$\int_0^\infty t^{3/2} e^{-t}\, dt = \left(\frac{3}{2}\right)\left(\frac{1}{2}\right)\sqrt{\pi} \quad \text{(see Section 4.4.6)}$$

so

$$\begin{aligned} E(\text{energy}) = E(W) &= \frac{m}{a\sqrt{\pi}}\left(\frac{3}{2}\right)\left(\frac{1}{2}\right)\sqrt{\pi} \\ &= \frac{3m}{4a} \end{aligned}$$

■

Example 3.5.13

Consolidated Industries is planning to market a new product and they are trying to decide how many to manufacture. They estimate that each item sold will return a profit of m dollars; each one not sold represents an n-dollar loss. Furthermore, they suspect the demand for the product, V, will have an exponential distribution,

$$f_V(v) = \left(\frac{1}{\lambda}\right) e^{-v/\lambda}, \quad v > 0$$

How many items should the company produce if they want to maximize their expected profit? (Assume that n, m, and λ are known.)

If a total of x items are made, the company's profit can be expressed as a function $Q(v)$, where

$$Q(v) = \begin{cases} mv - n(x - v) & \text{if } v < x \\ mx & \text{if } v \geq x \end{cases}$$

and v is the number of items sold. It follows that their *expected* profit is

$$\begin{aligned} E[Q(V)] &= \int_0^\infty Q(v) \cdot f_V(v)\, dv \\ &= \int_0^x [(m+n)v - nx]\left(\frac{1}{\lambda}\right) e^{-v/\lambda}\, dv + \int_x^\infty mx \cdot \left(\frac{1}{\lambda}\right) e^{-v/\lambda}\, dv \qquad (3.5.7) \end{aligned}$$

The integration here is straightforward, though a bit tedious. Equation 3.5.7 eventually simplifies to

$$E[Q(V)] = \lambda \cdot (m+n) - \lambda \cdot (m+n)e^{-x/\lambda} - nx$$

To find the optimal production level, we need to solve $dE[Q(V)]/dx = 0$ for x. But

$$\frac{dE[Q(V)]}{dx} = (m+n)e^{-x/\lambda} - n$$

and the latter equals zero when

$$x = -\lambda \cdot \ln\left(\frac{n}{m+n}\right)$$

■

Example 3.5.14 A point, y, is selected at random from the interval [0, 1], dividing the line into two segments (see Figure 3.5.5). What is the expected value of the ratio of the shorter segment to the longer segment?

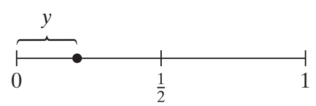


Figure 3.5.5

Notice, first, that the function

$$g(Y) = \frac{\text{shorter segment}}{\text{longer segment}}$$

has two expressions, depending on the location of the chosen point:

$$g(Y) = \begin{cases} y/(1-y), & 0 \le y \le \frac{1}{2} \\ (1-y)/y, & \frac{1}{2} < y \le 1 \end{cases}$$

By assumption, $f_Y(y) = 1, 0 \le y \le 1$, so

$$E[g(Y)] = \int_0^{\frac{1}{2}} \frac{y}{1-y} \cdot 1\, dy + \int_{\frac{1}{2}}^1 \frac{1-y}{y} \cdot 1\, dy$$

Writing the second integrand as $(1/y - 1)$ gives

$$\int_{\frac{1}{2}}^1 \frac{1-y}{y} \cdot 1\, dy = \int_{\frac{1}{2}}^1 \left(\frac{1}{y} - 1\right) dy = (\ln y - y)\Big|_{\frac{1}{2}}^1$$

$$= \ln 2 - \frac{1}{2}$$

By symmetry, though, the two integrals are the same, so

$$E\left(\frac{\text{shorter segment}}{\text{longer segment}}\right) = 2\ln 2 - 1$$

$$= 0.39$$

On the average, then, the longer segment will be a little more than $2\frac{1}{2}$ times the length of the shorter segment. ■

Questions

3.5.28. Suppose X is a binomial random variable with $n = 10$ and $p = \frac{2}{5}$. What is the expected value of $3X - 4$?

3.5.29. A typical day's production of a certain electronic component is twelve. The probability that one of these components needs rework is 0.11. Each component needing rework costs \$100. What is the average daily cost for defective components?

3.5.30. Let Y have probability density function

$$f_Y(y) = 2(1 - y),\ 0 \le y \le 1$$

Suppose that $W = Y^2$, in which case

$$f_W(w) = \frac{1}{\sqrt{w}} - 1,\ 0 \le w \le 1$$

Find $E(W)$ in two different ways.

3.5.31. A tool and die company makes castings for steel stress-monitoring gauges. Their annual profit, Q, in hundreds of thousands of dollars, can be expressed as a function of product demand, y:

$$Q(y) = 2(1 - e^{-2y})$$

Suppose that the demand (in thousands) for their castings follows an exponential pdf, $f_Y(y) = 6e^{-6y}$, $y > 0$. Find the company's expected profit.

3.5.32. A box is to be constructed so that its height is five inches and its base is Y inches by Y inches, where Y is a random variable described by the pdf, $f_Y(y) = 6y(1 - y)$, $0 < y < 1$. Find the expected volume of the box.

3.5.33. Grades on the last Economics 301 exam were not very good. Graphed, their distribution had a shape similar to the pdf

$$f_Y(y) = \frac{1}{5000}(100 - y), \quad 0 \le y \le 100$$

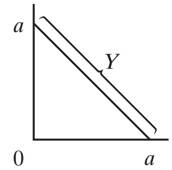
As a way of "curving" the results, the professor announces that he will replace each person's grade, Y, with a new grade, $g(Y)$, where $g(Y) = 10\sqrt{Y}$. Will the professor's strategy be successful in raising the class average above 60?

3.5.34. If Y has probability density function

$$f_Y(y) = 2y,\ 0 \le y \le 1$$

then $E(Y) = \frac{2}{3}$. Define the random variable W to be the squared deviation of Y from its mean, that is, $W = \left(Y - \frac{2}{3}\right)^2$. Find $E(W)$.

3.5.35. The hypotenuse, Y, of the isosceles right triangle shown is a random variable having a uniform pdf over the interval [6, 10]. Calculate the expected value of the triangle's area. Do not leave the answer as a function of a.



3.5.36. An urn contains n chips numbered 1 through n. Assume that the probability of choosing chip i is equal to ki, $i = 1, 2, \ldots, n$. If one chip is drawn, calculate $E\left(\frac{1}{X}\right)$, where the random variable X denotes the number showing on the chip selected. [*Hint:* Recall that the sum of the first n integers is $n(n + 1)/2$.]

3.6 The Variance

We saw in Section 3.5 that the location of a distribution is an important characteristic and that it can be effectively measured by calculating either the mean or the median. A second feature of a distribution that warrants further scrutiny is its *dispersion*—that is, the extent to which its values are spread out. The two properties are totally different: Knowing a pdf's location tells us absolutely nothing about its dispersion. Table 3.6.1, for example, shows two simple discrete pdfs with the same expected value (equal to zero), but with vastly different dispersions.

Table 3.6.1

k	$p_{X_1}(k)$	k	$p_{X_2}(k)$
-1	$\frac{1}{2}$	$-1{,}000{,}000$	$\frac{1}{2}$
1	$\frac{1}{2}$	1,000,000	$\frac{1}{2}$

It is not immediately obvious how the dispersion in a pdf should be quantified. Suppose that X is any discrete random variable. One seemingly reasonable approach would be to average the deviations of X from their mean—that is, calculate the expected value of $X - \mu$. As it happens, that strategy will not work because the negative deviations will exactly cancel the positive deviations, making the numerical value of such an average always zero, regardless of the amount of spread present in $p_X(k)$:

$$E(X - \mu) = E(X) - \mu = \mu - \mu = 0 \tag{3.6.1}$$

Another possibility would be to modify Equation 3.6.1 by making all the deviations positive, that is, to replace $E(X - \mu)$ with $E(|X - \mu|)$. This does work, and it *is* sometimes used to measure dispersion, but the absolute value is somewhat troublesome mathematically: It does not have a simple arithmetic formula, nor is it a differentiable function. *Squaring* the deviations proves to be a much better approach.

Definition 3.6.1
The *variance* of a random variable is the expected value of its squared deviations from μ. If X is discrete, with pdf $p_X(k)$,

$$\text{Var}(X) = \sigma^2 = E[(X - \mu)^2] = \sum_{\text{all } k} (k - \mu)^2 \cdot p_X(k)$$

If Y is continuous, with pdf $f_Y(y)$,

$$\text{Var}(Y) = \sigma^2 = E[(Y - \mu)^2] = \int_{-\infty}^{\infty} (y - \mu)^2 \cdot f_Y(y)\, dy$$

[If $E(X^2)$ or $E(Y^2)$ is not finite, the variance is not defined.]

Comment One unfortunate consequence of Definition 3.6.1 is that the units for the variance are the square of the units for the random variable: If Y is measured in inches, for example, the units for $\text{Var}(Y)$ are inches squared. This causes obvious problems in relating the variance back to the sample values. For that reason, in applied statistics, where unit compatibility is especially important, dispersion is measured not by the variance but by the *standard deviation*, which is defined to be the square root of the variance. That is,

$$\sigma = \text{standard deviation} = \begin{cases} \sqrt{\displaystyle\sum_{\text{all } k} (k - \mu)^2 \cdot p_X(k)} & \text{if } X \text{ is discrete} \\ \sqrt{\displaystyle\int_{-\infty}^{\infty} (y - \mu)^2 \cdot f_Y(y)\, dy} & \text{if } Y \text{ is continuous} \end{cases}$$

Comment The analogy between the expected value of a random variable and the center of gravity of a physical system was pointed out in Section 3.5. A similar equivalency holds between the variance and what engineers call a *moment of inertia*. If a set of weights having masses $m_1, m_2, \ldots$ are positioned along a (weightless) rigid bar at distances $r_1, r_2, \ldots$ from an axis of rotation (see Figure 3.6.1), the moment of inertia of the system is defined to be value $\sum_i m_i r_i^2$. Notice, though, that if the masses were the probabilities associated with a discrete random variable and if the axis of rotation were actually μ, then $r_1, r_2, \ldots$ could be written $(k_1 - \mu), (k_2 - \mu), \ldots$ and $\sum_i m_i r_i^2$ would be the same as the variance, $\sum_{\text{all } k} (k-\mu)^2 \cdot p_X(k)$.

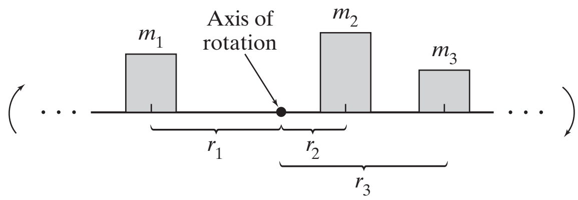


Figure 3.6.1

Definition 3.6.1 gives a formula for calculating σ^2 in both the discrete and the continuous cases. An equivalent, but easier-to-use, formula is given in Theorem 3.6.1.

Theorem 3.6.1 *Let W be any random variable, discrete or continuous, having mean μ and for which $E(W^2)$ is finite. Then*

$$\text{Var}(W) = \sigma^2 = E(W^2) - \mu^2$$

Proof We will prove the theorem for the continuous case. The argument for discrete W is similar. In Theorem 3.5.3, let $g(W) = (W - \mu)^2$. Then

$$\text{Var}(W) = E[(W-\mu)^2] = \int_{-\infty}^{\infty} g(w) f_W(w)\, dw = \int_{-\infty}^{\infty} (w-\mu)^2 f_W(w)\, dw$$

Squaring out the term $(w - \mu)^2$ that appears in the integrand and using the additive property of integrals gives

$$\begin{aligned}
\int_{-\infty}^{\infty} (w-\mu)^2 f_W(w)\, dw &= \int_{-\infty}^{\infty} (w^2 - 2\mu w + \mu^2) f_W(w)\, dw \\
&= \int_{-\infty}^{\infty} w^2 f_W(w)\, dw - 2\mu \int_{-\infty}^{\infty} w f_W(w)\, dw + \int_{-\infty}^{\infty} \mu^2 f_W(w)\, dw \\
&= E(W^2) - 2\mu^2 + \mu^2 = E(W^2) - \mu^2
\end{aligned}$$

Note that the equality $\int_{-\infty}^{\infty} w^2 f_W(w)\, dw = E(W^2)$ also follows from Theorem 3.5.3.

Example 3.6.1 An urn contains five chips, two red and three white. Suppose that two are drawn out simultaneously. Let X denote the number of red chips in the sample. Find $\text{Var}(X)$.

Note, first, that since the chips are not being replaced from drawing to drawing, X is a hypergeometric random variable. Moreover, we need to find μ, regardless of which formula is used to calculate σ^2. In the notation of Theorem 3.5.2, $r = 2$, $w = 3$, and $n = 2$, so

$$\mu = rn/(r + w) = 2 \cdot 2/(2 + 3) = 0.8$$

To find Var(X) using Definition 3.6.1, we write

$$\begin{aligned}
\text{Var}(X) = E[(X-\mu)^2] &= \sum_{\text{all } x}(x-\mu)^2 \cdot f_X(x) \\
&= (0-0.8)^2 \cdot \frac{\binom{2}{0}\binom{3}{2}}{\binom{5}{2}} + (1-0.8)^2 \cdot \frac{\binom{2}{1}\binom{3}{1}}{\binom{5}{2}} + (2-0.8)^2 \cdot \frac{\binom{2}{2}\binom{3}{0}}{\binom{5}{2}} \\
&= 0.36
\end{aligned}$$

To use Theorem 3.6.1, we would first find $E(X^2)$. From Theorem 3.5.3,

$$\begin{aligned}
E(X^2) = \sum_{\text{all } x} x^2 \cdot f_X(x) &= 0^2 \cdot \frac{\binom{2}{0}\binom{3}{2}}{\binom{5}{2}} + 1^2 \cdot \frac{\binom{2}{1}\binom{3}{1}}{\binom{5}{2}} + 2^2 \cdot \frac{\binom{2}{2}\binom{3}{0}}{\binom{5}{2}} \\
&= 1.00
\end{aligned}$$

Then

$$\begin{aligned}
\text{Var}(X) = E(X^2) - \mu^2 &= 1.00 - (0.8)^2 \\
&= 0.36
\end{aligned}$$

confirming what we calculated earlier. ■

In Section 3.5 we encountered a change of scale formula that applied to expected values. For any constants a and b and any random variable W, $E(aW+b) = aE(W)+b$. A similar issue arises in connection with the *variance* of a linear transformation: If Var$(W) = \sigma^2$, what is the variance of $aW+b$?

Theorem 3.6.2 *Let W be any random variable having mean μ and where $E(W^2)$ is finite. Then $Var(aW+b) = a^2 Var(W)$.*

Proof Using the same approach taken in the proof of Theorem 3.6.1, it can be shown that $E[(aW+b)^2] = a^2E(W^2) + 2ab\mu + b^2$. We also know from the corollary to Theorem 3.5.3 that $E(aW+b) = a\mu + b$. Using Theorem 3.6.1, then, we can write

$$\begin{aligned}
\text{Var}(aW+b) &= E[(aW+b)^2] - [E(aW+b)]^2 \\
&= [a^2E(W^2) + 2ab\mu + b^2] - [a\mu + b]^2 \\
&= [a^2E(W^2) + 2ab\mu + b^2] - [a^2\mu^2 + 2ab\mu + b^2] \\
&= a^2[E(W^2) - \mu^2] = a^2\text{Var}(W)
\end{aligned}$$

Example 3.6.2 A random variable Y is described by the pdf

$$f_Y(y) = 2y, \quad 0 \le y \le 1$$

What is the standard deviation of $3Y+2$?

First, we need to find the variance of Y. But

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

and

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

so

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

Then, by Theorem 3.6.2,

$$\text{Var}(3Y + 2) = (3)^2 \cdot \text{Var}(Y) = 9 \cdot \frac{1}{18}$$
$$= \frac{1}{2}$$

which makes the standard deviation of $3Y + 2$ equal to $\sqrt{\frac{1}{2}}$ or *0.71*. ■

Comment When the median is used as the measure of central tendency, the usual measure of dispersion is called the *interquartile range*. It marks off the middle 50% of the data. In other words, let Q_1 represent the first quartile; that is, $P(Y \leq Q_1) = 0.25$ and Q_3 represent the third quartile, that is $P(Y \leq Q_3) = 0.75$. Then $Q_3 - Q_1$ is the interquartile range.

Example 3.6.3

For $f_Y(y) = 3y^2$, $0 \leq y \leq 1$, Q_1 is the solution to the equation $0.25 = \int_0^{Q_1} 3y^2 dy$, or $0.25 = Q_1^3$, giving $Q_1 = \sqrt[3]{0.25} = 0.630$. Similarly $Q_3 = \sqrt[3]{0.75} = 0.909$, so the interquartile range is $0.909 - 0.630 = 0.279$. ■

Questions

3.6.1. Find Var(X) for the urn problem of Example 3.6.1 if the sampling is done *with* replacement.

3.6.2. Find the variance of Y if

$$f_Y(y) = \begin{cases} \frac{3}{4}, & 0 \leq y \leq 1 \\ \frac{1}{4}, & 2 \leq y \leq 3 \\ 0, & \text{elsewhere} \end{cases}$$

3.6.3. Ten equally qualified applicants, six men and four women, apply for three lab technician positions. Unable to justify choosing any of the applicants over all the others, the personnel director decides to select the three at random. Let X denote the number of men hired. Compute the standard deviation of X.

3.6.4. A certain hospitalization policy pays a cash benefit for up to five days in the hospital. It pays \$250 per day for the first three days and \$150 per day for the next two. The number of days of hospitalization, X, is a discrete random variable with probability function $P(X = k) = \frac{1}{15}(6 - k)$ for $k = 1, 2, 3, 4, 5$. Find Var(X).

3.6.5. Use Theorem 3.6.1 to find the variance of the random variable Y, where

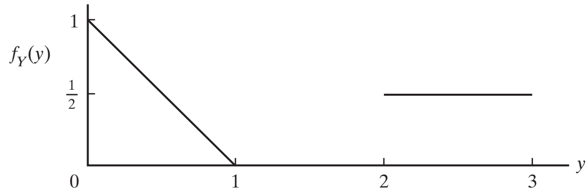
$$f_Y(y) = 3(1 - y)^2, \quad 0 < y < 1$$

3.6.6. If

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

for what value of k does Var(Y) = 2?

3.6.7. Calculate the standard deviation, σ, for the random variable Y whose pdf has the graph shown below:



3.6.8. Consider the pdf defined by

$$f_Y(y) = \frac{2}{y^3}, \quad y \geq 1$$

Show that (a) $\int_1^\infty f_Y(y)\ dy = 1$, (b) $E(Y) = 2$, and (c) Var(Y) is not finite.

3.6.9. Frankie and Johnny play the following game. Frankie selects a number at random from the interval $[a, b]$. Johnny, not knowing Frankie's number, is to pick a second number from that same interval and pay Frankie an amount, W, equal to the squared difference between the two [so $0 \le W \le (b-a)^2$]. What should be Johnny's strategy if he wants to minimize his expected loss?

3.6.10. Let Y be a random variable whose pdf is given by $f_Y(y) = 5y^4, 0 \le y \le 1$. Use Theorem 3.6.1 to find $\text{Var}(Y)$.

3.6.11. Suppose that Y is an exponential random variable, so $f_Y(y) = \lambda e^{-\lambda y}, y \ge 0$. Show that the variance of Y is $1/\lambda^2$.

3.6.12. Suppose that Y is an exponential random variable with $\lambda = 2$ (recall Question 3.6.11). Find $P[Y > E(Y) + 2\sqrt{\text{Var}(Y)}]$.

3.6.13. Let X be a random variable with finite mean μ. Define for every real number a, $g(a) = E[(X-a)^2]$. Show that

$$g(a) = E[(X-\mu)^2] + (\mu - a)^2.$$

What is another name for $\min g(a)$?

3.6.14. Suppose the charge for repairing an automobile averages \$200 with a standard deviation of \$16. If a 10% tax is added to the charge and then a \$15 flat fee for environmental impact, what is the standard deviation of the charge to the car owner?

3.6.15. If Y denotes a temperature recorded in degrees Fahrenheit, then $\frac{5}{9}(Y-32)$ is the corresponding temperature in degrees Celsius. If the standard deviation for a set of temperatures is 15.7°F, what is the standard deviation of the equivalent Celsius temperatures?

3.6.16. If $E(W) = \mu$ and $\text{Var}(W) = \sigma^2$, show that

$$E\left(\frac{W-\mu}{\sigma}\right) = 0 \quad \text{and} \quad \text{Var}\left(\frac{W-\mu}{\sigma}\right) = 1$$

3.6.17. Suppose U is a uniform random variable over $[0, 1]$.

(a) Show that $Y = (b-a)U + a$ is uniform over $[a, b]$.

(b) Use part (a) and Theorem 3.6.2 to find the variance of Y.

3.6.18. Recovering small quantities of calcium in the presence of magnesium can be a difficult problem for an analytical chemist. Suppose the amount of calcium Y to be recovered is uniformly distributed between 4 and 7 mg. The amount of calcium recovered by one method is the random variable

$$W_1 = 0.2281 + (0.9948)Y + E_1$$

where the error term E_1 has mean 0 and variance 0.0427 and is independent of Y.

A second procedure has random variable

$$W_2 = -0.0748 + (1.0024)Y + E_2$$

where the error term E_2 has mean 0 and variance 0.0159 and is independent of Y.

The better technique should have a mean as close as possible to the mean of $Y(=5.5)$, and a variance as small as possible. Compare the two methods on the basis of mean and variance.

HIGHER MOMENTS

The quantities we have identified as the mean and the variance are actually special cases of what are referred to more generally as the *moments* of a random variable. More precisely, $E(W)$ is the *first moment about the origin* and σ^2 is the *second moment about the mean*. As the terminology suggests, we will have occasion to define higher moments of W. Just as $E(W)$ and σ^2 reflect a random variable's location and dispersion, so it is possible to characterize other aspects of a distribution in terms of other moments. We will see, for example, that the skewness of a distribution—that is, the extent to which it is not symmetric around μ—can be effectively measured in terms of a *third* moment. Likewise, there are issues that arise in certain applied statistics problems that require a knowledge of the flatness of a pdf, a property that can be quantified by the *fourth* moment.

Definition 3.6.2
Let W be any random variable with pdf $f_W(w)$. For any positive integer r,

a. The rth *moment of W about the origin*, μ_r, is given by

$$\mu_r = E(W^r)$$

provided $\int_{-\infty}^{\infty} |w|^r \cdot f_W(w)\, dw < \infty$ (or provided the analogous condition on the *summation* of $|w|^r$ holds, if W is discrete). When $r = 1$, we usually delete the subscript and write $E(W)$ as μ rather than μ_1.

b. The rth *moment of W about the mean*, μ'_r, is given by

$$\mu'_r = E[(W - \mu)^r]$$

provided the finiteness conditions of part 1 hold.

Comment We can express μ'_r in terms of μ_j, $j = 1, 2, \ldots, r$, by simply writing out the binomial expansion of $(W - \mu)^r$:

$$\mu'_r = E[(W - \mu)^r] = \sum_{j=0}^{r} \binom{r}{j} E(W^j)(-\mu)^{r-j}$$

Thus,

$$\mu'_2 = E[(W - \mu)^2] = \sigma^2 = \mu_2 - \mu_1^2$$
$$\mu'_3 = E[(W - \mu)^3] = \mu_3 - 3\mu_1\mu_2 + 2\mu_1^3$$
$$\mu'_4 = E[(W - \mu)^4] = \mu_4 - 4\mu_1\mu_3 + 6\mu_1^2\mu_2 - 3\mu_1^4$$

and so on.

Example 3.6.4 The *skewness* of a pdf can be measured in terms of its third moment about the mean. If a pdf is symmetric, $E[(W - \mu)^3]$ will obviously be zero; for pdfs not symmetric, $E[(W - \mu)^3]$ will not be zero. In practice, the symmetry (or lack of symmetry) of a pdf is often measured by the *coefficient of skewness*, γ_1, where

$$\gamma_1 = \frac{E[(W - \mu)^3]}{\sigma^3}$$

Dividing μ'_3 by σ^3 makes γ_1 dimensionless.

A second "shape" parameter in common use is the *coefficient of kurtosis*, γ_2, which involves the *fourth* moment about the mean. Specifically,

$$\gamma_2 = \frac{E[(W - \mu)^4]}{\sigma^4} - 3$$

For certain pdfs, γ_2 is a useful measure of the probability of outliers ("fat tails"). Relatively flat pdfs are said to be *platykurtic*; more peaked pdfs are called *leptokurtic*. ■

Earlier in this chapter we encountered random variables whose means do not exist—recall, for example, the St. Petersburg paradox. More generally, there are random variables having certain of their higher moments finite and certain others, not finite. Addressing the question of whether or not a given $E(W^j)$ is finite is the following existence theorem.

Theorem 3.6.3 *If the kth moment of a random variable exists, all moments of order less than k exist.*

Proof Let $f_Y(y)$ be the pdf of a continuous random variable Y. By Definition 3.6.2, $E(Y^k)$ exists if and only if

$$\int_{-\infty}^{\infty} |y|^k \cdot f_Y(y)\, dy < \infty \tag{3.6.2}$$

(Continued on next page)

(Theorem 3.6.3 continued)

Let $1 \leq j < k$. To prove the theorem we must show that

$$\int_{-\infty}^{\infty} |y|^j \cdot f_Y(y)\, dy < \infty$$

is implied by Inequality 3.6.2. But

$$\begin{aligned}\int_{-\infty}^{\infty} |y|^j \cdot f_Y(y)\, dy &= \int_{|y|\leq 1} |y|^j \cdot f_Y(y)\, dy + \int_{|y|>1} |y|^j \cdot f_Y(y)\, dy \\ &\leq \int_{|y|\leq 1} f_Y(y)\, dy + \int_{|y|>1} |y|^j \cdot f_Y(y)\, dy \\ &\leq 1 + \int_{|y|>1} |y|^j \cdot f_Y(y)\, dy \\ &\leq 1 + \int_{|y|>1} |y|^k \cdot f_Y(y)\, dy < \infty\end{aligned}$$

Therefore, $E(Y^j)$ exists, $j = 1, 2, \ldots, k - 1$. The proof for discrete random variables is similar.

Questions

3.6.19. Let Y be a uniform random variable defined over the interval $(0, 2)$. Find an expression for the rth moment of Y about the origin. Also, use the binomial expansion as described in the Comment to find $E[(Y - \mu)^6]$.

3.6.20. Find the coefficient of skewness for an exponential random variable having the pdf

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

3.6.21. Calculate the coefficient of kurtosis for a uniform random variable defined over the unit interval, $f_Y(y) = 1$, for $0 \leq y \leq 1$.

3.6.22. Suppose that W is a random variable for which $E[(W - \mu)^3] = 10$ and $E(W^3) = 4$. Is it possible that $\mu = 2$?

3.6.23. If $Y = aX + b$, $a > 0$, show that Y has the same coefficients of skewness and kurtosis as X.

3.6.24. Let Y be the random variable of Question 3.4.6, where for a positive integer n, $f_Y(y) = (n + 2)(n + 1)y^n(1 - y)$, $0 \leq y \leq 1$.

(a) Find $\text{Var}(Y)$.

(b) For any positive integer k, find the kth moment around the origin.

3.6.25. Suppose that the random variable Y is described by the pdf

$$f_Y(y) = c \cdot y^{-6}, \quad y > 1$$

(a) Find c.

(b) What is the highest moment of Y that exists?

3.7 Joint Densities

Sections 3.3 and 3.4 introduced the basic terminology for describing the probabilistic behavior of a *single* random variable. Such information, while adequate for many problems, is insufficient when more than one variable are of interest to the experimenter. Medical researchers, for example, continue to explore the relationship between blood cholesterol and heart disease, and, more recently, between "good" cholesterol and "bad" cholesterol. And more than a little attention—both political and pedagogical—is given to the role played by K–12 funding in the performance of would-be high school graduates on exit exams. On a smaller scale, electronic equipment and systems are often designed to have built-in redundancy: Whether or not that equipment functions properly ultimately depends on the reliability of two different components.

The point is, there are many situations where two relevant random variables, say, X and Y, are defined on the same sample space.[2] Knowing only $f_X(x)$ and $f_Y(y)$, though, does not necessarily provide enough information to characterize the all-important *simultaneous* behavior of X and Y. The purpose of this section is to introduce the concepts, definitions, and mathematical techniques associated with distributions based on two (or more) random variables.

DISCRETE JOINT PDFS

As we saw in the single-variable case, the pdf is defined differently depending on whether the random variable is discrete or continuous. The same distinction applies to joint pdfs. We begin with a discussion of joint pdfs as they apply to two discrete random variables.

Definition 3.7.1
Suppose S is a discrete sample space on which two random variables, X and Y, are defined. The *joint probability density function of X and Y (or joint pdf)* is denoted $p_{X,Y}(x, y)$, where

$$p_{X,Y}(x, y) = P(\{s|X(s) = x \text{ and } Y(s) = y\})$$

Comment A convenient shorthand notation for the meaning of $p_{X,Y}(x, y)$, consistent with what we used earlier for pdfs of single discrete random variables, is to write $p_{X,Y}(x, y) = P(X = x, Y = y)$.

Example 3.7.1 A supermarket has two express lines. Let X and Y denote the number of customers in the first and in the second, respectively, at any given time. During nonrush hours, the joint pdf of X and Y is summarized by the following table:

		X			
		0	1	2	3
Y	0	0.1	0.2	0	0
	1	0.2	0.25	0.05	0
	2	0	0.05	0.05	0.025
	3	0	0	0.025	0.05

Find $P(|X - Y| = 1)$, the probability that X and Y differ by exactly 1.

By definition,

$$\begin{aligned} P(|X - Y| = 1) &= \sum_{|x-y|=1} \sum p_{X,Y}(x, y) \\ &= p_{X,Y}(0, 1) + p_{X,Y}(1, 0) + p_{X,Y}(1, 2) \\ &\quad + p_{X,Y}(2, 1) + p_{X,Y}(2, 3) + p_{X,Y}(3, 2) \\ &= 0.2 + 0.2 + 0.05 + 0.05 + 0.025 + 0.025 \\ &= 0.55 \end{aligned}$$

[Would you expect $p_{X,Y}(x, y)$ to be symmetric? Would you expect the event $|X - Y| \geq 2$ to have zero probability?] ■

[2] For the next several sections we will suspend our earlier practice of using X to denote a discrete random variable and Y to denote a continuous random variable. The category of the random variables will need to be determined from the context of the problem. Typically, though, X and Y will either be both discrete or both continuous.

Example 3.7.2 Suppose two fair dice are rolled. Let X be the sum of the numbers showing, and let Y be the larger of the two. So, for example,

$$p_{X,Y}(2,3) = P(X = 2, Y = 3) = P(\emptyset) = 0$$

$$p_{X,Y}(4,3) = P(X = 4, Y = 3) = P(\{(1,3)(3,1)\}) = \frac{2}{36}$$

and

$$p_{X,Y}(6,3) = P(X = 6, Y = 3) = P(\{(3,3)\}) = \frac{1}{36}$$

The entire joint pdf is given in Table 3.7.1.

Table 3.7.1

x \ y	1	2	3	4	5	6	Row totals
2	1/36	0	0	0	0	0	1/36
3	0	2/36	0	0	0	0	2/36
4	0	1/36	2/36	0	0	0	3/36
5	0	0	2/36	2/36	0	0	4/36
6	0	0	1/36	2/36	2/36	0	5/36
7	0	0	0	2/36	2/36	2/36	6/36
8	0	0	0	1/36	2/36	2/36	5/36
9	0	0	0	0	2/36	2/36	4/36
10	0	0	0	0	1/36	2/36	3/36
11	0	0	0	0	0	2/36	2/36
12	0	0	0	0	0	1/36	1/36
Col. totals	1/36	3/36	5/36	7/36	9/36	11/36	

Notice that the row totals in the right-hand margin of the table give the pdf for X. Similarly, the column totals along the bottom detail the pdf for Y. Those are not coincidences. Theorem 3.7.1 gives a formal statement of the relationship between the joint pdf and the individual pdfs. ■

Theorem 3.7.1 *Suppose that $p_{X,Y}(x, y)$ is the joint pdf of the discrete random variables X and Y. Then*

$$p_X(x) = \sum_{\text{all } y} p_{X,Y}(x, y) \quad \textit{and} \quad p_Y(y) = \sum_{\text{all } x} p_{X,Y}(x, y)$$

Proof We will prove the first statement. Note that the collection of sets $(Y = y)$ for all y forms a partition of S; that is, they are disjoint and $\bigcup_{\text{all } y}(Y = y) = S$. The set $(X = x) = (X = x) \cap S = (X = x) \cap \bigcup_{\text{all } y}(Y = y) = \bigcup_{\text{all } y}[(X = x) \cap (Y = y)]$, so

$$p_X(x) = P(X = x) = P\left(\bigcup_{\text{all } y}[(X = x) \cap (Y = y)]\right)$$

$$= \sum_{\text{all } y} P(X = x, Y = y) = \sum_{\text{all } y} p_{X,Y}(x, y)$$

Definition 3.7.2
An individual pdf obtained by summing a joint pdf over all values of the other random variable is called a *marginal pdf*.

CONTINUOUS JOINT PDFS

If X and Y are both continuous random variables, Definition 3.7.1 does not apply because $P(X = x, Y = y)$ will be identically 0 for all (x, y). As was the case in single-variable situations, the joint pdf for two continuous random variables will be defined as a function that when integrated yields the probability that (X, Y) lies in a specified region of the xy-plane.

Definition 3.7.3
Two random variables defined on the same set of real numbers are *jointly continuous* if there exists a function $f_{X,Y}(x, y)$ such that for any region R in the xy-plane, $P[(X, Y) \in R] = \int\int_R f_{X,Y}(x, y)\,dx\,dy$. The function $f_{X,Y}(x, y)$ is the *joint pdf of X and Y*.

Comment Any function $f_{X,Y}(x, y)$ for which

1. $f_{X,Y}(x, y) \geq 0$ for all x and y
2. $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x, y)\,dx\,dy = 1$

qualifies as a joint pdf. We shall employ the convention of naming the domain only where the joint pdf is nonzero; everywhere else it will be assumed to be zero. This is analogous, of course, to the notation used earlier in describing the domain of single random variables. Also, for the functions used, the order of integration does not matter.

Example 3.7.3 Suppose that the variation in two continuous random variables, X and Y, can be modeled by the joint pdf $f_{X,Y}(x, y) = cxy$, for $0 < y < x < 1$. Find c.

By inspection, $f_{X,Y}(x, y)$ will be nonnegative as long as $c \geq 0$. The particular c that qualifies $f_{X,Y}(x, y)$ as a joint pdf, though, is the one that makes the volume under $f_{X,Y}(x, y)$ equal to 1. But

$$\int\int_S cxy\,dy\,dx = 1 = c\int_0^1\left[\int_0^x (xy)\,dy\right]dx = c\int_0^1 x\left(\frac{y^2}{2}\Big|_0^x\right)dx$$
$$= c\int_0^1\left(\frac{x^3}{2}\right)dx = c\frac{x^4}{8}\Big|_0^1 = \left(\frac{1}{8}\right)c$$

Therefore, $c = 8$. ■

Example 3.7.4 A study claims that the daily number of hours, X, a teenager watches television and the daily number of hours, Y, he works on his homework are approximated by the joint pdf

$$f_{X,Y}(x, y) = xye^{-(x+y)}, \quad x > 0, \quad y > 0$$

What is the probability that a teenager chosen at random spends at least twice as much time watching television as he does working on his homework?

The region, R, in the xy-plane corresponding to the event "$X \geq 2Y$" is shown in Figure 3.7.1. It follows that $P(X \geq 2Y)$ is the volume under $f_{X,Y}(x, y)$ above the region R:

$$P(X \geq 2Y) = \int_0^\infty \int_0^{x/2} xye^{-(x+y)}\, dy\, dx$$

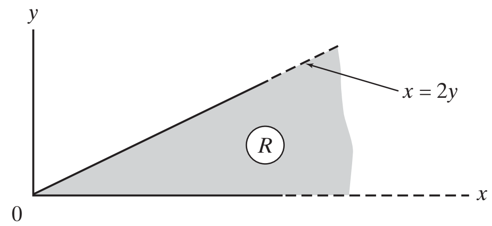


Figure 3.7.1

Separating variables, we can write

$$P(X \geq 2Y) = \int_0^\infty xe^{-x} \left[\int_0^{x/2} ye^{-y} dy \right] dx$$

and the double integral reduces to $\frac{7}{27}$:

$$\begin{aligned} P(X \geq 2Y) &= \int_0^\infty xe^{-x} \left[1 - \left(\frac{x}{2} + 1 \right) e^{-x/2} \right] dx \\ &= \int_0^\infty xe^{-x} dx - \int_0^\infty \frac{x^2}{2} e^{-3x/2}\, dx - \int_0^\infty xe^{-3x/2} dx \\ &= 1 - \frac{16}{54} - \frac{4}{9} \\ &= \frac{7}{27} \end{aligned}$$

■

GEOMETRIC PROBABILITY

One particularly important special case of Definition 3.7.3 is the *joint uniform pdf*, which is represented by a surface having a constant height everywhere above a specified rectangle in the xy-plane. That is,

$$f_{X,Y}(x, y) = \frac{1}{(b-a)(d-c)}, \quad a \leq x \leq b, c \leq y \leq d$$

If R is some region in the rectangle where X and Y are defined, $P((X, Y) \in R)$ reduces to a simple ratio of areas:

$$P((X, Y) \in R) = \frac{\text{area of } R}{(b-a)(d-c)} \tag{3.7.1}$$

Calculations based on Equation 3.7.1 are referred to as *geometric probabilities*.

Example 3.7.5

Two friends agree to meet on the University Commons "sometime around 12:30." But neither of them is particularly punctual—or patient. What will actually happen is that each will arrive at random sometime in the interval from 12:00 to 1:00. If one arrives and the other is not there, the first person will wait fifteen minutes or until 1:00, whichever comes first, and then leave. What is the probability that the two will get together?

To simplify notation, we can represent the time period from 12:00 to 1:00 as the interval from zero to sixty minutes. Then if x and y denote the two arrival times, the sample space is the 60×60 square shown in Figure 3.7.2. Furthermore, the event M, "The two friends meet," will occur if and only if $|x - y| \leq 15$ or, equivalently, if and only if $-15 \leq x - y \leq 15$. These inequalities appear as the shaded region in Figure 3.7.2.

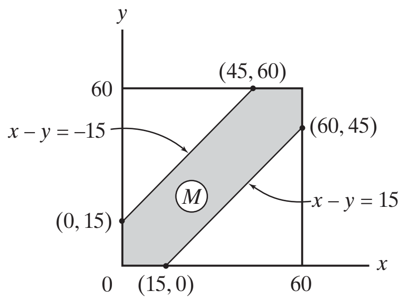


Figure 3.7.2

Notice that the areas of the triangles above and below M are each equal to $\frac{1}{2}(45)(45)$. It follows that the two friends have a 44% chance of meeting:

$$\begin{aligned} P(M) &= \frac{\text{area of } M}{\text{area of } S} \\ &= \frac{(60)^2 - 2\left[\frac{1}{2}(45)(45)\right]}{(60)^2} \\ &= 0.44 \end{aligned}$$

■

Example 3.7.6

A carnival operator wants to set up a ringtoss game. Players will throw a ring of diameter d onto a grid of squares, the side of each square being of length s (see Figure 3.7.3). If the ring lands entirely inside a square, the player wins a prize. To ensure a profit, the operator must keep the player's chances of winning down to something less than one in five. How small can the operator make the ratio d/s?

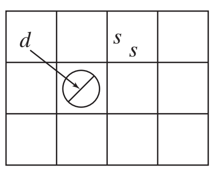


Figure 3.7.3

First, assume that the player is required to stand far enough away that no skill is involved and the ring is falling at random on the grid. From Figure 3.7.4, we see that in order for the ring not to touch any side of the square, the ring's center must be somewhere in the interior of a smaller square, each side of which is a distance $d/2$ from one of the grid lines.

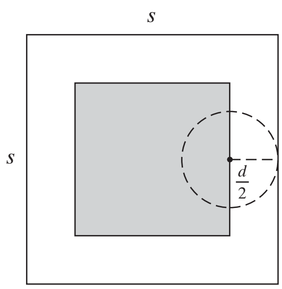


Figure 3.7.4

Since the area of a grid square is s^2 and the area of an interior square is $(s-d)^2$, the probability of a winning toss can be written as the ratio:

$$P(\text{Ring touches no lines}) = \frac{(s-d)^2}{s^2}$$

But the operator requires that

$$\frac{(s-d)^2}{s^2} \leq 0.20$$

Solving for d/s gives

$$\frac{d}{s} \geq 1 - \sqrt{0.20} = 0.55$$

That is, if the diameter of the ring is at least 55% as long as the side of one of the squares, the player will have no more than a 20% chance of winning. ■

Questions

3.7.1. If $p_{X,Y}(x, y) = cxy$ at the points $(1, 1)$, $(2, 1)$, $(2, 2)$, and $(3, 1)$, and equals 0 elsewhere, find c.

3.7.2. Let X and Y be two continuous random variables defined over the unit square. What does c equal if $f_{X,Y}(x, y) = c(x^2 + y^2)$?

3.7.3. Suppose that random variables X and Y vary in accordance with the joint pdf, $f_{X,Y}(x, y) = c(x+y)$, $0 < x < y < 1$. Find c.

3.7.4. Find c if $f_{X,Y}(x, y) = cxy$ for X and Y defined over the triangle whose vertices are the points $(0, 0)$, $(0, 1)$, and $(1, 1)$.

3.7.5. An urn contains four red chips, three white chips, and two blue chips. A random sample of size 3 is drawn without replacement. Let X denote the number of white chips in the sample and Y the number of blue chips. Write a formula for the joint pdf of X and Y.

3.7.6. Four cards are drawn from a standard poker deck. Let X be the number of kings drawn and Y the number of queens. Find $p_{X,Y}(x, y)$.

3.7.7. An advisor looks over the schedules of his fifty students to see how many math and science courses each has registered for in the coming semester. He summarizes his results in a table. What is the probability that a student selected at random will have signed up for more math courses than science courses?

		Number of math courses, X		
		0	1	2
Number of science courses, Y	0	11	6	4
	1	9	10	3
	2	5	0	2

3.7.8. Consider the experiment of tossing a fair coin three times. Let X denote the number of heads on the last flip, and let Y denote the total number of heads on the three flips. Find $p_{X,Y}(x, y)$.

3.7.9. Suppose that two fair dice are tossed one time. Let X denote the number of 2's that appear, and Y the number of 3's. Write the matrix giving the joint probability density function for X and Y. Suppose a third random variable, Z, is defined, where $Z = X + Y$. Use $p_{X,Y}(x, y)$ to find $p_Z(z)$.

3.7.10. Let X be the time in days between a car accident and reporting a claim to the insurance company. Let Y be the time in days between the report and payment of the claim. Suppose that $f_{X,Y}(x, y) = c, 0 \le x \le 7, 0 \le y \le 7$, and zero otherwise.

(a) Find c.

(b) Find $P(0 \le X \le 2, 0 \le Y \le 4)$.

3.7.11. Let X and Y have the joint pdf

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

Find $P(Y < 3X)$.

3.7.12. A point is chosen at random from the interior of a circle whose equation is $x^2 + y^2 \le 4$. Let the random variables X and Y denote the x- and y-coordinates of the sampled point. Find $f_{X,Y}(x, y)$.

3.7.13. Find $P(X < 2Y)$ if $f_{X,Y}(x, y) = x + y$ for X and Y each defined over the unit interval.

3.7.14. Suppose that five independent observations are drawn from the continuous pdf $f_T(t) = 2t, 0 \le t \le 1$. Let X denote the number of t's that fall in the interval $0 \le t < \frac{1}{3}$ and let Y denote the number of t's that fall in the interval $\frac{1}{3} \le t < \frac{2}{3}$. Find $p_{X,Y}(1, 2)$.

3.7.15. A point is chosen at random from the interior of a right triangle with base b and height h. What is the probability that the y value is between 0 and $h/2$?

MARGINAL PDFS FOR CONTINUOUS RANDOM VARIABLES

The notion of marginal pdfs in connection with discrete random variables was introduced in Theorem 3.7.1 and Definition 3.7.2. An analogous relationship holds in the continuous case—*integration*, though, replaces the summation that appears in Theorem 3.7.1.

Theorem 3.7.2 *Suppose X and Y are jointly continuous with joint pdf $f_{X,Y}(x, y)$. Then the marginal pdfs, $f_X(x)$ and $f_Y(y)$, are given by*

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dy \quad \text{and} \quad f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dx$$

Proof It suffices to verify the first of the theorem's two equalities. As is often the case with proofs for continuous random variables, we begin with the cdf:

$$F_X(x) = P(X \le x) = \int_{-\infty}^{\infty} \int_{-\infty}^{x} f_{X,Y}(t, y)\, dt\, dy = \int_{-\infty}^{x} \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dy\, dt$$

Differentiating both ends of the equation above gives

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

Recall Theorem 3.4.1.

Example 3.7.7 Suppose that two continuous random variables, X and Y, have the joint uniform pdf

$$f_{X,Y}(x, y) = \frac{1}{6}, \quad 0 \le x \le 3, \quad 0 \le y \le 2$$

Find $f_X(x)$.

Applying Theorem 3.7.2 gives

$$f_X(x) = \int_0^2 f_{X,Y}(x, y)\, dy = \int_0^2 \frac{1}{6}\, dy = \frac{1}{3}, \quad 0 \le x \le 3$$

Notice that X, by itself, is a uniform random variable defined over the interval $[0, 3]$; similarly, we would find that $f_Y(y)$ is a uniform pdf over the interval $[0, 2]$. ■

Example 3.7.8 Consider the case where X and Y are two continuous random variables, jointly distributed over the first quadrant of the xy-plane according to the joint pdf,

$$f_{X,Y}(x, y) = \begin{cases} y^2 e^{-y(x+1)} & x \geq 0, \quad y \geq 0 \\ 0 & \text{elsewhere} \end{cases}$$

Find the two marginal pdfs.

First, consider $f_X(x)$. By Theorem 3.7.2,

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dy = \int_0^{\infty} y^2 e^{-y(x+1)}\, dy$$

In the integrand, substitute

$$u = y(x+1)$$

making $du = (x+1)\, dy$. This gives

$$f_X(x) = \frac{1}{x+1} \int_0^{\infty} \frac{u^2}{(x+1)^2} e^{-u}\, du = \frac{1}{(x+1)^3} \int_0^{\infty} u^2 e^{-u}\, du$$

After applying integration by parts (twice) to $\int_0^{\infty} u^2 e^{-u}\, du$, we get

$$\begin{aligned} f_X(x) &= \frac{1}{(x+1)^3} \left[-u^2 e^{-u} - 2ue^{-u} - 2e^{-u}\right]\Big|_0^{\infty} \\ &= \frac{1}{(x+1)^3} \left[2 - \lim_{u \to \infty} \left(\frac{u^2}{e^u} + \frac{2u}{e^u} + \frac{2}{e^u}\right)\right] \\ &= \frac{2}{(x+1)^3}, \quad x \geq 0 \end{aligned}$$

Finding $f_Y(y)$ is a bit easier:

$$\begin{aligned} f_Y(y) &= \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dx = \int_0^{\infty} y^2 e^{-y(x+1)}\, dx \\ &= y^2 e^{-y} \int_0^{\infty} e^{-yx}\, dx = y^2 e^{-y} \left(\frac{1}{y}\right) \left(-e^{-yx}\Big|_0^{\infty}\right) \\ &= ye^{-y}, \quad y \geq 0 \end{aligned}$$

■

Questions

3.7.16. Find the marginal pdf of X for the joint pdf derived in Question 3.7.5.

3.7.17. Find the marginal pdfs of X and Y for the joint pdf derived in Question 3.7.8.

3.7.18. The campus recruiter for an international conglomerate classifies the large number of students she interviews into three categories—the lower quarter, the middle half, and the upper quarter. If she meets six students on a given morning, what is the probability that they will be evenly divided among the three categories? What is the marginal probability that exactly two will belong to the middle half?

3.7.19. For each of the following joint pdfs, find $f_X(x)$ and $f_Y(y)$.

(a) $f_{X,Y}(x, y) = \frac{1}{2}$, $0 \leq x \leq 2$, $0 \leq y \leq 1$

(b) $f_{X,Y}(x, y) = \frac{3}{2}y^2$, $0 \leq x \leq 2$, $0 \leq y \leq 1$

(c) $f_{X,Y}(x, y) = \frac{2}{3}(x + 2y)$, $0 \leq x \leq 1$, $0 \leq y \leq 1$

(d) $f_{X,Y}(x, y) = c(x + y)$, $0 \leq x \leq 1$, $0 \leq y \leq 1$

(e) $f_{X,Y}(x, y) = 4xy$, $0 \leq x \leq 1$, $0 \leq y \leq 1$

(f) $f_{X,Y}(x, y) = xye^{-(x+y)}$, $0 \leq x$, $0 \leq y$

(g) $f_{X,Y}(x, y) = ye^{-xy-y}$, $0 \leq x$, $0 \leq y$

3.7.20. For each of the following joint pdfs, find $f_X(x)$ and $f_Y(y)$.

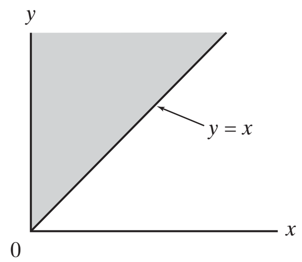
(a) $f_{X,Y}(x, y) = \frac{1}{2}, 0 \leq x \leq y \leq 2$

(b) $f_{X,Y}(x, y) = \frac{1}{x}, 0 \leq y \leq x \leq 1$

(c) $f_{X,Y}(x, y) = 6x, 0 \leq x \leq 1, 0 \leq y \leq 1 - x$

3.7.21. Suppose that $f_{X,Y}(x, y) = 6(1 - x - y)$ for x and y defined over the unit square, subject to the restriction that $0 \leq x + y \leq 1$. Find the marginal pdf for X.

3.7.22. Find $f_Y(y)$ if $f_{X,Y}(x, y) = 2e^{-x}e^{-y}$ for x and y defined over the shaded region pictured.



3.7.23. Suppose that X and Y are discrete random variables with

$$p_{X,Y}(x, y) = \frac{4!}{x!y!(4-x-y)!}\left(\frac{1}{2}\right)^x\left(\frac{1}{3}\right)^y\left(\frac{1}{6}\right)^{4-x-y},$$
$$0 \leq x + y \leq 4$$

Find $p_X(x)$ and $p_Y(x)$.

3.7.24. A generalization of the binomial model occurs when there is a sequence of n independent trials with *three* outcomes, where $p_1 = P(\text{outcome 1})$ and $p_2 = P(\text{outcome 2})$. Let X and Y denote the number of trials (out of n) resulting in outcome 1 and outcome 2, respectively.

(a) Show that $p_{X,Y}(x, y) = \frac{n!}{x!y!(n-x-y)!}p_1^x p_2^y (1 - p_1 - p_2)^{n-x-y}, 0 \leq x + y \leq n$

(b) Find $p_X(x)$ and $p_Y(x)$.

(*Hint:* See Question 3.7.23.)

JOINT CDFS

For a single random variable X, the cdf of X evaluated at some point x—that is, $F_X(x)$—is the probability that the random variable X takes on a value less than or equal to x. Extended to two variables, a *joint cdf* [evaluated at the point (x, y)] is the probability that $X \leq x$ and, simultaneously, that $Y \leq y$.

Definition 3.7.4
Let X and Y be any two random variables. The *joint cumulative distribution function of X and Y* (or *joint cdf*) is denoted $F_{X,Y}(x, y)$, where

$$F_{X,Y}(x, y) = P(X \leq x \text{ and } Y \leq y)$$

Example 3.7.9

Find the joint cdf, $F_{X,Y}(x, y)$, for the two random variables X and Y with joint pdf $f_{X,Y}(x, y) = \frac{4}{3}(x + xy), 0 \leq x \leq 1, 0 \leq y \leq 1$.

If Definition 3.7.4 is applied, the probability that $X \leq x$ and $Y \leq y$ becomes a double integral of the pdf. In order to keep the cdf a function of x and y, use u and v as the variables of integration.

$$F_{X,Y}(x, y) = \frac{4}{3}\int_0^y\int_0^x (u + uv)\,du\,dv = \frac{4}{3}\int_0^y\left(\int_0^x (u + uv)\,du\right)dv$$
$$= \frac{4}{3}\int_0^y\left(\frac{u^2}{2}(1+v)\Big|_0^x\right)dv = \frac{4}{3}\int_0^y \frac{x^2}{2}(1+v)\,dv$$
$$= \frac{4}{3}\frac{x^2}{2}\left(v + \frac{v^2}{2}\right)\Big|_0^y = \frac{4}{3}\frac{x^2}{2}\left(y + \frac{y^2}{2}\right),$$

which simplifies to

$$F_{X,Y}(x, y) = \frac{1}{3}x^2(2y + y^2).$$

[For what values of x and y is $F_{X,Y}(x, y)$ defined?] ■

Theorem 3.7.3 *Let $F_{X,Y}(x, y)$ be the joint cdf associated with the continuous random variables X and Y. Then the joint pdf of X and Y, $f_{X,Y}(x, y)$, is a second partial derivative of the joint cdf, that is, $f_{X,Y}(x, y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x, y)$, provided $F_{X,Y}(x, y)$ has continuous second partial derivatives.*

Example 3.7.10 What is the joint pdf of the random variables X and Y whose joint cdf is $F_{X,Y}(x, y) = \frac{1}{3}x^2(2y + y^2)$, $0 \le x \le 1$, $0 \le y \le 1$?

By Theorem 3.7.3,

$$f_{X,Y}(x, y) = \frac{\partial^2}{\partial x\, \partial y} F_{X,Y}(x, y) = \frac{\partial^2}{\partial x\, \partial y} \frac{1}{3}x^2(2y + y^2)$$

$$= \frac{\partial}{\partial y} \frac{2}{3}x(2y + y^2) = \frac{2}{3}x(2 + 2y) = \frac{4}{3}(x + xy),\ 0 \le x \le 1,\ 0 \le y \le 1$$

Notice the similarity between Examples 3.7.9 and 3.7.10. $f_{X,Y}(x, y)$ is the same in both examples; so is $F_{X,Y}(x, y)$. ■

MULTIVARIATE DENSITIES

The definitions and theorems in this section extend in a very straightforward way to situations involving more than two variables. The joint pdf for n discrete random variables, for example, is denoted $p_{X_1,\ldots,X_n}(x_1, \ldots, x_n)$ where

$$p_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = P(X_1 = x_1, \ldots, X_n = x_n)$$

For n continuous random variables, the joint pdf is that function $f_{X_1,\ldots,X_n}(x_1, \ldots, x_n)$ having the property that for any region R in n-space,

$$P[(X_1, \ldots, X_n) \in R] = \iint_R \cdots \int f_{X_1,\ldots,X_n}(x_1, \ldots, x_n)\, dx_1 \cdots dx_n$$

And if $F_{X_1,\ldots,X_n}(x_1, \ldots, x_n)$ is the joint *cdf* of continuous random variables $X_1, \ldots, X_n$—that is, $F_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = P(X_1 \le x_1, \ldots, X_n \le x_n)$—then

$$f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = \frac{\partial^n}{\partial x_1 \cdots \partial x_n} F_{X_1,\ldots,X_n}(x_1, \ldots, x_n)$$

The notion of a marginal pdf also extends readily, although in the n-variate case, a marginal pdf can, itself, be a joint pdf. Given $X_1, \ldots, X_n$, the marginal pdf of any subset of r of those variables $(X_{i_1}, X_{i_2}, \ldots, X_{i_r})$ is derived by integrating (or summing) the joint pdf with respect to the remaining $n - r$ variables $(X_{j_1}, X_{j_2}, \ldots, X_{j_{n-r}})$. If the X_i's are all continuous, for example,

$$f_{X_{i_1},\ldots,X_{i_r}}(x_{i_1}, \ldots, x_{i_r}) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1,\ldots,X_n}(x_1, \ldots, x_n)\, dx_{j_1} \cdots dx_{j_{n-r}}$$

Questions

3.7.25. Consider the experiment of simultaneously tossing a fair coin and rolling a fair die. Let X denote the number of heads showing on the coin and Y the number of spots showing on the die.

(a) List the outcomes in S.

(b) Find $F_{X,Y}(1, 2)$.

3.7.26. An urn contains twelve chips—four red, three black, and five white. A sample of size 4 is to be drawn without replacement. Let X denote the number of white chips in the sample, Y the number of red. Find $F_{X,Y}(1, 2)$.

3.7.27. For each of the following joint pdfs, find $F_{X,Y}(x, y)$.

(a) $f_{X,Y}(x, y) = \frac{3}{2}y^2, 0 \le x \le 2, 0 \le y \le 1$

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

(c) $f_{X,Y}(x, y) = 4xy, 0 \le x \le 1, 0 \le y \le 1$

3.7.28. For each of the following joint pdfs, find $F_{X,Y}(x, y)$.

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

(b) $f_{X,Y}(x, y) = \frac{1}{x}, 0 \le y \le x \le 1$

(c) $f_{X,Y}(x, y) = 6x, 0 \le x \le 1, 0 \le y \le 1 - x$

3.7.29. Find and graph $f_{X,Y}(x, y)$ if the joint cdf for random variables X and Y is

$$F_{X,Y}(x, y) = xy, \qquad 0 \le x \le 1, \quad 0 \le y \le 1$$

3.7.30. Find the joint pdf associated with two random variables X and Y whose joint cdf is

$$F_{X,Y}(x, y) = (1 - e^{-\lambda y})(1 - e^{-\lambda x}), \qquad x > 0, \quad y > 0$$

3.7.31. Given that $F_{X,Y}(x, y) = k(4x^2y^2 + 5xy^4), 0 < x < 1, 0 < y < 1$, find the corresponding pdf and use it to calculate $P(0 < X < \frac{1}{2}, \frac{1}{2} < Y < 1)$.

3.7.32. Prove that

$$P(a < X \le b, c < Y \le d) = F_{X,Y}(b, d) - F_{X,Y}(a, d) - F_{X,Y}(b, c) + F_{X,Y}(a, c)$$

3.7.33. A certain brand of fluorescent bulbs will last, on the average, one thousand hours. Suppose that four of these bulbs are installed in an office. What is the probability that all four are still functioning after one thousand fifty hours? If X_i denotes the ith bulb's life, assume that

$$f_{X_1,X_2,X_3,X_4}(x_1, x_2, x_3, x_4) = \prod_{i=1}^{4} \left(\frac{1}{1000}\right) e^{-x/1000}$$

for $x_i > 0, i = 1, 2, 3, 4$.

3.7.34. A hand of six cards is dealt from a standard poker deck. Let X denote the number of aces, Y the number of kings, and Z the number of queens.

(a) Write a formula for $p_{X,Y,Z}(x, y, z)$.

(b) Find $p_{X,Y}(x, y)$ and $p_{X,Z}(x, z)$.

3.7.35. Calculate $p_{X,Y}(0, 1)$ if $p_{X,Y,Z}(x, y, z) = \frac{3!}{x!y!z!(3-x-y-z)!}\left(\frac{1}{2}\right)^x\left(\frac{1}{12}\right)^y\left(\frac{1}{6}\right)^z \cdot \left(\frac{1}{4}\right)^{3-x-y-z}$ for $x, y, z = 0, 1, 2, 3$ and $0 \le x + y + z \le 3$.

3.7.36. Suppose that the random variables X, Y, and Z have the multivariate pdf

$$f_{X,Y,Z}(x, y, z) = (x + y)e^{-z}$$

for $0 < x < 1, 0 < y < 1$, and $z > 0$. Find (a) $f_{X,Y}(x, y)$, (b) $f_{Y,Z}(y, z)$, and (c) $f_Z(z)$.

3.7.37. The four random variables W, X, Y, and Z have the multivariate pdf

$$f_{W,X,Y,Z}(w, x, y, z) = 16wxyz$$

for $0 \le w \le 1, 0 \le x \le 1, 0 \le y \le 1$, and $0 \le z \le 1$. Find the marginal pdf, $f_{W,X}(w, x)$, and use it to compute $P(0 \le W \le \frac{1}{2}, \frac{1}{2} \le X \le 1)$.

INDEPENDENCE OF TWO RANDOM VARIABLES

The concept of independent events that was introduced in Section 2.5 leads quite naturally to a similar definition for independent random variables.

Definition 3.7.5
Two discrete random variables X and Y are said to be *independent* if for every points a and b, $P(X = a \text{ and } Y = b) = P(X = a)P(Y = b)$. Two continuous random variables X and Y are said to be *independent* if for every interval A and every interval B, $P(X \in A \text{ and } Y \in B) = P(X \in A)P(Y \in B)$.

Theorem 3.7.4 *The continuous random variables X and Y are independent if and only if there are functions $g(x)$ and $h(y)$ such that*

$$f_{X,Y}(x, y) = g(x)h(y) \quad \textit{for all } x \textit{ and } y \tag{3.7.2}$$

If Equation 3.7.2 holds, there is a constant k such that $f_X(x) = kg(x)$ and $f_Y(y) = (1/k)h(y)$.

Proof First, suppose that X and Y are independent. Then $F_{X,Y}(x, y) = P(X \leq x$ and $Y \leq y) = P(X \leq x)P(Y \leq y) = F_X(x)F_Y(y)$, and we can write

$$f_{X,Y}(x, y) = \frac{\partial^2}{\partial x\, \partial y} F_{X,Y}(x, y) = \frac{\partial^2}{\partial x\, \partial y} F_X(x)F_Y(y) = \frac{d}{dx} F_X(x) \frac{d}{dy} F_Y(y) = f_X(x) f_Y(y)$$

Next we need to show that Equation 3.7.2 implies that X and Y are independent. To begin, note that

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dy = \int_{-\infty}^{\infty} g(x)h(y)\, dy = g(x) \int_{-\infty}^{\infty} h(y)\, dy$$

Set $k = \int_{-\infty}^{\infty} h(y)\, dy$, so $f_X(x) = kg(x)$. Similarly, it can be shown that $f_Y(y) = (1/k)h(y)$. Therefore,

$$\begin{aligned} P(X \in A \text{ and } Y \in B) &= \int_A \int_B f_{X,Y}(x, y)\, dx\, dy = \int_A \int_B g(x)h(y)\, dx\, dy \\ &= \int_A \int_B kg(x)(1/k)h(y)\, dx\, dy = \int_A f_X(x)\, dx \int_B f_Y(y)\, dy \\ &= P(X \in A)P(Y \in B) \end{aligned}$$

and the theorem is proved.

Comment Theorem 3.7.4 can be adapted to the case that X and Y are discrete.

Example 3.7.11 Suppose that the probabilistic behavior of two random variables X and Y is described by the joint pdf $f_{X,Y}(x, y) = 12xy(1 - y)$, $0 \leq x \leq 1$, $0 \leq y \leq 1$. Are X and Y independent? If they are, find $f_X(x)$ and $f_Y(y)$.

According to Theorem 3.7.4, the answer to the independence question is "yes" if $f_{X,Y}(x, y)$ can be factored into a function of x times a function of y. There are such functions. Let $g(x) = 12x$ and $h(y) = y(1 - y)$.

To find $f_X(x)$ and $f_Y(y)$ requires that the "12" appearing in $f_{X,Y}(x, y)$ be factored in such a way that $g(x) \cdot h(y) = f_X(x) \cdot f_Y(y)$. Let

$$k = \int_{-\infty}^{\infty} h(y)\, dy = \int_0^1 y(1 - y)\, dy = [y^2/2 - y^3/3]\Big|_0^1 = \frac{1}{6}$$

Therefore, $f_X(x) = kg(x) = \frac{1}{6}(12x) = 2x$, $0 \leq x \leq 1$ and $f_Y(y) = (1/k)h(y) = 6y(1 - y)$, $0 \leq y \leq 1$. ■

INDEPENDENCE OF n (>2) RANDOM VARIABLES

In Chapter 2, extending the notion of independence from *two* events to n events proved to be something of a problem. The independence of each subset of the n events had to be checked separately (recall Definition 2.5.2). This is not necessary in the case of n random variables. We simply use the extension of Theorem 3.7.4 to n random variables as the definition of independence in the multidimensional case.

The theorem that independence is equivalent to the factorization of the joint pdf holds in the multidimensional case.

Definition 3.7.6
The n random variables $X_1, X_2, \ldots, X_n$ are said to be *independent* if there are functions $g_1(x_1), g_2(x_2), \ldots, g_n(x_n)$ such that for every $x_1, x_2, \ldots, x_n$

$$f_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) = g_1(x_1)g_2(x_2)\cdots g_n(x_n)$$

A similar statement holds for discrete random variables, in which case f is replaced with p.

Comment Analogous to the result for $n = 2$ random variables, the expression on the right-hand side of the equation in Definition 3.7.6 can also be written as the product of the marginal pdfs of $X_1, X_2, \ldots,$ and X_n.

Example 3.7.12 Consider k urns, each holding n chips numbered 1 through n. A chip is to be drawn at random from each urn. What is the probability that all k chips will bear the same number?

If $X_1, X_2, \ldots, X_k$ denote the numbers on the 1st, 2nd, ..., and kth chips, respectively, we are looking for the probability that $X_1 = X_2 = \cdots = X_k$. In terms of the joint pdf,

$$P(X_1 = X_2 = \cdots = X_k) = \sum_{x_1=x_2=\cdots=x_k} p_{X_1,X_2,\ldots,X_k}(x_1, x_2, \ldots, x_k)$$

Each of the selections here is obviously independent of all the others, so the joint pdf factors according to Definition 3.7.6, and we can write

$$\begin{aligned} P(X_1 = X_2 = \cdots = X_k) &= \sum_{i=1}^{n} p_{X_1}(x_i) \cdot p_{X_2}(x_i) \cdots p_{X_k}(x_i) \\ &= n \cdot \left(\frac{1}{n} \cdot \frac{1}{n} \cdot \cdots \cdot \frac{1}{n}\right) \\ &= \frac{1}{n^{k-1}} \end{aligned}$$

■

RANDOM SAMPLES

Definition 3.7.6 addresses the question of independence as it applies to n random variables having marginal pdfs—say, $f_{X_1}(x_1), f_{X_2}(x_2), \ldots, f_{X_n}(x_n)$—that might be quite different. A special case of that definition occurs for virtually every set of data collected for statistical analysis. Suppose an experimenter takes a set of n measurements, $x_1, x_2, \ldots, x_n$, under the same conditions. Those X_i's, then, qualify as a set of independent random variables—moreover, each represents the *same* pdf. The special—but familiar—notation for that scenario is given in Definition 3.7.7. We will encounter it often in the chapters ahead.

Definition 3.7.7
Let $W_1, W_2, \ldots, W_n$ be a set of n independent random variables, all having the same pdf. Then $W_1, W_2, \ldots, W_n$ are said to be a *random sample of size n*.

Questions

3.7.38. Two fair dice are tossed. Let X denote the number appearing on the first die and Y the number on the second. Show that X and Y are independent.

3.7.39. Let $f_{X,Y}(x, y) = \lambda^2 e^{-\lambda(x+y)}$, $0 \le x$, $0 \le y$. Show that X and Y are independent. What are the marginal pdfs in this case?

3.7.40. Suppose that each of two urns has four chips, numbered 1 through 4. A chip is drawn from the first urn and bears the number X. That chip is added to the second urn. A chip is then drawn from the second urn. Call its number Y.

(a) Find $p_{X,Y}(x, y)$.

(b) Show that $p_X(k) = p_Y(k) = \frac{1}{4}$, $k = 1, 2, 3, 4$.

(c) Show that X and Y are not independent.

3.7.41. Let X and Y be random variables with joint pdf

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

Give a geometric argument to show that X and Y are not independent.

3.7.42. Are the random variables X and Y independent if $f_{X,Y}(x, y) = \frac{2}{3}(x + 2y)$, $0 \le x \le 1$, $0 \le y \le 1$?

3.7.43. Suppose that random variables X and Y are independent with marginal pdfs $f_X(x) = 2x$, $0 \le x \le 1$, and $f_Y(y) = 3y^2$, $0 \le y \le 1$. Find $P(Y < X)$.

3.7.44. Find the joint cdf of the independent random variables X and Y, where $f_X(x) = \dfrac{x}{2}$, $0 \le x \le 2$, and $f_Y(y) = 2y$, $0 \le y \le 1$.

3.7.45. If two random variables X and Y are independent with marginal pdfs $f_X(x) = 2x$, $0 \le x \le 1$, and $f_Y(y) = 1$, $0 \le y \le 1$, calculate $P\left(\frac{Y}{X} > 2\right)$.

3.7.46. Suppose $f_{X,Y}(x, y) = xye^{-(x+y)}$, $x > 0$, $y > 0$. Prove for any real numbers a, b, c, and d that

$$P(a < X < b, c < Y < d) = P(a < X < b) \cdot P(c < Y < d)$$

thereby establishing the independence of X and Y.

3.7.47. Given the joint pdf $f_{X,Y}(x, y) = 2x + y - 2xy$, $0 < x < 1$, $0 < y < 1$, find numbers a, b, c, and d such that

$$P(a < X < b, c < Y < d) \ne P(a < X < b) \cdot P(c < Y < d)$$

thus demonstrating that X and Y are not independent.

3.7.48. Prove that if X and Y are two independent random variables, then $U = g(X)$ and $V = h(Y)$ are also independent.

3.7.49. If two random variables X and Y are defined over a region in the XY-plane that is *not* a rectangle (possibly infinite) with sides parallel to the coordinate axes, can X and Y be independent?

3.7.50. Write down the joint probability density function for a random sample of size n drawn from the exponential pdf, $f_X(x) = (1/\lambda)e^{-x/\lambda}$, $x \ge 0$.

3.7.51. Suppose that X_1, X_2, X_3, and X_4 are independent random variables, each with pdf $f_{X_i}(x_i) = 4x_i^3$, $0 \le x_i \le 1$. Find

(a) $P\left(X_1 < \frac{1}{2}\right)$.

(b) $P\left(\text{exactly one } X_i < \frac{1}{2}\right)$.

(c) $f_{X_1,X_2,X_3,X_4}(x_1, x_2, x_3, x_4)$.

(d) $F_{X_2,X_3}(x_2, x_3)$.

3.7.52. A random sample of size $n = 2k$ is taken from a uniform pdf defined over the unit interval. Calculate $P\left(X_1 < \frac{1}{2}, X_2 > \frac{1}{2}, X_3 < \frac{1}{2}, X_4 > \frac{1}{2}, \ldots, X_{2k} > \frac{1}{2}\right)$.

3.8 Transforming and Combining Random Variables

TRANSFORMATIONS

Transforming a variable from one scale to another is a problem that is comfortably familiar. If a thermometer says the temperature outside is 83°F, we know that the temperature *in degrees Celsius* is 28:

$$^\circ\text{C} = \left(\frac{5}{9}\right)(^\circ\text{F} - 32) = \left(\frac{5}{9}\right)(83 - 32) = 28$$

An analogous question arises in connection with random variables. Suppose that X is a discrete random variable with pdf $p_X(k)$. If a second random variable, Y, is defined to be $aX + b$, where a and b are constants, what can be said about the pdf for Y? Recall Questions 3.3.11 and 3.4.16 as examples of such transformations.

Theorem 3.8.1 *Suppose X is a discrete random variable. Let $Y = aX + b$, where a and b are constants. Then $p_Y(y) = p_X\left(\frac{y-b}{a}\right)$.*

Proof $p_Y(y) = P(Y = y) = P(aX + b = y) = P\left(X = \dfrac{y-b}{a}\right) = p_X\left(\dfrac{y-b}{a}\right)$

Example 3.8.1 Let X be a random variable for which $p_X(k) = \frac{1}{10}$, for $k = 1, 2, \ldots, 10$. What is the probability distribution associated with the random variable Y, where $Y = 4X - 1$? That is, find $p_Y(y)$.

From Theorem 3.8.1, $P(Y = y) = P(4X - 1 = y) = P[X = (y+1)/4] = p_X\left(\frac{y+1}{4}\right)$, which implies that $p_Y(y) = \frac{1}{10}$ for the ten values of $(y+1)/4$ that equal 1, $2, \ldots, 10$. But $(y+1)/4 = 1$ when $y = 3$, $(y+1)/4 = 2$ when $y = 7, \ldots, (y+1)/4 = 10$ when $y = 39$. Therefore, $p_Y(y) = \frac{1}{10}$, for $y = 3, 7, \ldots, 39$. ■

Next we give the analogous result for a linear transformation of a *continuous* random variable.

Theorem 3.8.2 *Suppose X is a continuous random variable. Let $Y = aX + b$, where $a \neq 0$ and b is a constant. Then*

$$f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right)$$

Proof We begin by writing an expression for the cdf of Y:

$$F_Y(y) = P(Y \leq y) = P(aX + b \leq y) = P(aX \leq y - b)$$

At this point we need to consider two cases, the distinction being the sign of a. Suppose, first, that $a > 0$. Then

$$F_Y(y) = P(aX \leq y - b) = P\left(X \leq \frac{y-b}{a}\right)$$

and differentiating $F_Y(y)$ yields $f_Y(y)$:

$$f_Y(y) = \frac{d}{dy}F_Y(y) = \frac{d}{dy}F_X\left(\frac{y-b}{a}\right) = \frac{1}{a} f_X\left(\frac{y-b}{a}\right) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right)$$

If $a < 0$,

$$F_Y(y) = P(aX \leq y - b) = P\left(X > \frac{y-b}{a}\right) = 1 - P\left(X \leq \frac{y-b}{a}\right)$$

Differentiation in this case gives

$$f_Y(y) = \frac{d}{dy}F_Y(y) = \frac{d}{dy}\left[1 - F_X\left(\frac{y-b}{a}\right)\right] = -\frac{1}{a} f_X\left(\frac{y-b}{a}\right) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right)$$

and the theorem is proved.

Now, armed with the multivariable concepts and techniques covered in Section 3.7, we can extend the investigation of transformations to functions defined on sets of random variables. In statistics, the most important combination of a set of random variables is often their sum, so we continue this section with the problem of finding the pdf of $X + Y$.

FINDING THE PDF OF A SUM

Theorem 3.8.3 *Suppose that X and Y are independent random variables. Let $W = X + Y$. Then*

1. *If X and Y are discrete random variables with pdfs $p_X(x)$ and $p_Y(y)$, respectively,*

$$p_W(w) = \sum_{\text{all } x} p_X(x)p_Y(w - x)$$

2. *If X and Y are continuous random variables with pdfs $f_X(x)$ and $f_Y(y)$, respectively,*

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

Proof

1. $$p_W(w) = P(W = w) = P(X + Y = w)$$
$$= P\left(\bigcup_{\text{all } x}(X = x, Y = w - x)\right) = \sum_{\text{all } x} P(X = x, Y = w - x)$$
$$= \sum_{\text{all } x} P(X = x)P(Y = w - x)$$
$$= \sum_{\text{all } x} p_X(x)p_Y(w - x)$$

   where the next-to-last equality derives from the independence of X and Y.

2. Since X and Y are continuous random variables, we can find $f_W(w)$ by differentiating the corresponding cdf, $F_W(w)$. Here, $F_W(w) = P(X + Y \leq w)$ is found by integrating $f_{X,Y}(x, y) = f_X(x) \cdot f_Y(y)$ over the shaded region R, as pictured in Figure 3.8.1.

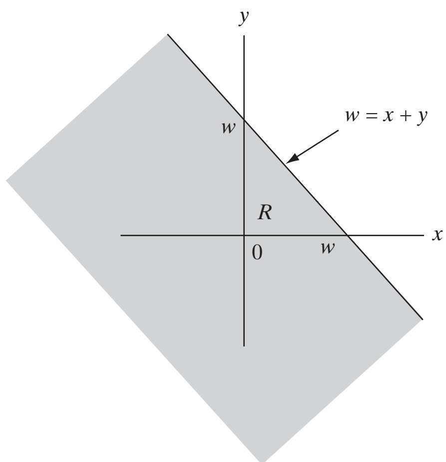


Figure 3.8.1

By inspection,

$$F_w(w) = \int_{-\infty}^{\infty}\int_{-\infty}^{w-x} f_X(x)f_Y(y)\,dy\,dx = \int_{-\infty}^{\infty} f_X(x)\left[\int_{-\infty}^{w-x} f_Y(y)\,dy\right]dx$$
$$= \int_{-\infty}^{\infty} f_X(x)F_Y(w - x)\,dx$$

Assume that the integrand in the above equation is sufficiently smooth so that differentiation and integration can be interchanged. Then we can write

$$f_W(w) = \frac{d}{dw}F_W(w) = \frac{d}{dw}\int_{-\infty}^{\infty} f_X(x)F_Y(w-x)\,dx = \int_{-\infty}^{\infty} f_X(x)\left[\frac{d}{dw}F_Y(w-x)\right]dx$$
$$= \int_{-\infty}^{\infty} f_X(x)f_Y(w-x)\,dx$$

and the theorem is proved.

Comment The integral in part (2) above is referred to as the *convolution* of the functions f_X and f_Y. Besides their frequent appearances in random variable problems, convolutions turn up in many areas of mathematics and engineering.

Example 3.8.2 Suppose that X and Y are two independent binomial random variables, each with the same success probability but defined on m and n trials, respectively. Specifically,

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

and

$$p_Y(k) = \binom{n}{k}p^k(1-p)^{n-k}, \quad k = 0, 1, \ldots, n$$

Find $p_W(w)$, where $W = X + Y$.

By Theorem 3.8.3, $p_W(w) = \sum_{\text{all } x} p_X(x)p_Y(w-x)$, but the summation over "all x" needs to be interpreted as the set of values for x and $w - x$ such that $p_X(x)$ and $p_Y(w-x)$, respectively, are both nonzero. But that will be true for all integers x from 0 to w. Therefore,

$$p_W(w) = \sum_{x=0}^{w} p_X(x)p_Y(w-x) = \sum_{x=0}^{w}\binom{m}{x}p^x(1-p)^{m-x}\binom{n}{w-x}p^{w-x}(1-p)^{n-(w-x)}$$
$$= \sum_{x=0}^{w}\binom{m}{x}\binom{n}{w-x}p^w(1-p)^{n+m-w}$$

Now, consider an urn having m red chips and n white chips. If w chips are drawn out—without replacement—the probability that exactly x red chips are in the sample is given by the hypergeometric distribution,

$$P(x \text{ reds in sample}) = \frac{\binom{m}{x}\binom{n}{w-x}}{\binom{m+n}{w}} \tag{3.8.1}$$

Summing Equation 3.8.1 from $x = 0$ to $x = w$ must equal 1 (why?), in which case

$$\sum_{x=0}^{w}\binom{m}{x}\binom{n}{w-x} = \binom{m+n}{w}$$

so

$$p_W(w) = \binom{m+n}{w}p^w(1-p)^{n+m-w}, \quad w = 0, 1, \ldots, n+m$$

Should we recognize $p_W(w)$? Definitely. Compare the structure of $p_W(w)$ to the statement of Theorem 3.2.1: The random variable W has a binomial distribution where the probability of success at any given trial is p and the total number of trials is $n + m$. ■

Comment Example 3.8.2 shows that the binomial distribution "reproduces" itself—that is, if X and Y are independent binomial random variables with the same value for p, their sum is also a binomial random variable. Not all random variables share that property. The sum of two independent uniform random variables, for example, is not a uniform random variable (see Question 3.8.5).

Example 3.8.3 Suppose a radiation monitor relies on an electronic sensor, whose lifetime X is modeled by the exponential pdf, $f_X(x) = \lambda e^{-\lambda x}, x > 0$. To improve the reliability of the monitor, the manufacturer has included an identical second sensor that is activated only in the event the first sensor malfunctions. (This is called *cold redundancy*.) Let the random variable Y denote the operating lifetime of the second sensor, in which case the lifetime of the monitor can be written as the sum $W = X + Y$. Find $f_W(w)$.

Since X and Y are both continuous random variables,

$$f_W(w) = \int_{-\infty}^{\infty} f_X(x) f_Y(w - x)\, dx \tag{3.8.2}$$

Notice that $f_X(x) > 0$ only if $x > 0$ and that $f_Y(w - x) > 0$ only if $x < w$. Therefore, the integral in Equation 3.8.2 that goes from $-\infty$ to ∞ reduces to an integral from 0 to w, and we can write

$$\begin{aligned} f_W(w) &= \int_0^w f_X(x) f_Y(w - x)\, dx = \int_0^w \lambda e^{-\lambda x} \lambda e^{-\lambda(w-x)}\, dx = \lambda^2 \int_0^w e^{-\lambda x} e^{-\lambda(w-x)}\, dx \\ &= \lambda^2 e^{-\lambda w} \int_0^w dx = \lambda^2 w e^{-\lambda w}, \quad w \geq 0 \end{aligned}$$

■

Comment By integrating $f_X(x)$ and $f_W(w)$, we can assess the improvement in the monitor's reliability afforded by the cold redundancy. Since X is an exponential random variable, $E(X) = 1/\lambda$ (recall Question 3.5.11). How different, for example, are $P(X \geq 1/\lambda)$ and $P(W \geq 1/\lambda)$? A simple calculation shows that the latter is actually *twice* the magnitude of the former:

$$P(X \geq 1/\lambda) = \int_{1/\lambda}^{\infty} \lambda e^{-\lambda x}\, dx = -e^{-u}\Big|_1^{\infty} = e^{-1} = 0.37$$

$$P(W \geq 1/\lambda) = \int_{1/\lambda}^{\infty} \lambda^2 w e^{-\lambda w}\, dw = e^{-u}(-u - 1)\Big|_1^{\infty} = 2e^{-1} = 0.74$$

FINDING THE PDFS OF QUOTIENTS AND PRODUCTS

We conclude this section by considering the pdfs for the quotient and product of two independent random variables. That is, given X and Y, we are looking for $f_W(w)$, where (1) $W = Y/X$ and (2) $W = XY$. Neither of the resulting formulas is as important as the pdf for the *sum* of two random variables, but both formulas will play key roles in several derivations in Chapter 7. Example 3.8.4 gives an introduction to the theorems in this section.

Example 3.8.4

Suppose that X and Y have a joint uniform density over the unit square:

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

Find the pdf for their product—that is, find $f_Z(z)$, where $Z = XY$.

For $0 < z < 1$, $F_Z(z)$ is the volume above the shaded region in Figure 3.8.2. Specifically,

$$F_Z(z) = P(Z \le z) = P(XY \le z) = \iint\limits_R f_{X,Y}(x,y)\, dy\, dx$$

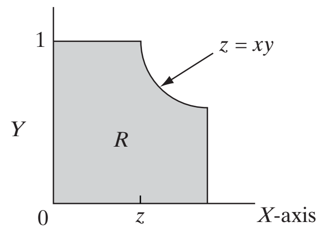


Figure 3.8.2

By inspection, we see that the double integral over R can be split up into *two* double integrals—one letting x range from 0 to z, the other having x-values extend from z to 1:

$$F_Z(z) = \int_0^z \left(\int_0^1 1\, dy \right) dx + \int_z^1 \left(\int_0^{z/x} 1\, dy \right) dx$$

But

$$\int_0^z \left(\int_0^1 1\, dy \right) dx = \int_0^z 1\, dx = z$$

and

$$\int_z^1 \left(\int_0^{z/x} 1\, dy \right) dx = \int_z^1 \left(\frac{z}{x} \right) dx = z \ln x \Big|_z^1 = -z \ln z$$

It follows that

$$F_Z(z) = \begin{cases} 0 & z \le 0 \\ z - z\ln z & 0 < z < 1 \\ 1 & z \ge 1 \end{cases}$$

in which case

$$f_Z(z) = \begin{cases} -\ln z & 0 < z < 1 \\ 0 & \text{elsewhere} \end{cases}$$

■

Theorem 3.8.4

Let X and Y be independent continuous random variables, with pdfs $f_X(x)$ and $f_Y(y)$, respectively. Assume that X is zero for at most a set of isolated points. Let $W = Y/X$. Then

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

(Continued on next page)

(Theorem 3.8.4 continued)

Proof

$$
\begin{aligned}
F_W(w) &= P(Y/X \le w) \\
&= P(Y/X \le w \quad \text{and} \quad X \ge 0) + P(Y/X \le w \quad \text{and} \quad X < 0) \\
&= P(Y \le wX \quad \text{and} \quad X \ge 0) + P(Y \ge wX \quad \text{and} \quad X < 0) \\
&= P(Y \le wX \quad \text{and} \quad X \ge 0) + 1 - P(Y \le wX \quad \text{and} \quad X < 0) \\
&= \int_0^\infty \int_{-\infty}^{wx} f_X(x) f_Y(y)\,dy\,dx + 1 - \int_{-\infty}^0 \int_{-\infty}^{wx} f_X(x) f_Y(y)\,dy\,dx
\end{aligned}
$$

Then we differentiate $F_W(w)$ to obtain

$$
\begin{aligned}
f_W(w) = \frac{d}{dw} F_W(w) &= \frac{d}{dw} \int_0^\infty \int_{-\infty}^{wx} f_X(x) f_Y(y)\,dy\,dx - \frac{d}{dw} \int_{-\infty}^0 \int_{-\infty}^{wx} f_X(x) f_Y(y)\,dy\,dx \\
&= \int_0^\infty f_X(x) \left(\frac{d}{dw} \int_{-\infty}^{wx} f_Y(y)\,dy \right) dx - \int_{-\infty}^0 f_X(x) \left(\frac{d}{dw} \int_{-\infty}^{wx} f_Y(y)\,dy \right) dx
\end{aligned}
\tag{3.8.3}
$$

(Note that we are assuming sufficient regularity of the functions to permit interchange of integration and differentiation.)

To proceed, we need to differentiate the function $G(w) = \int_{-\infty}^{wx} f_Y(y)\,dy$ with respect to w. By the Fundamental Theorem of Calculus and the chain rule, we find

$$
\frac{d}{dw} G(w) = \frac{d}{dw} \int_{-\infty}^{wx} f_Y(y)\,dy = f_Y(wx) \frac{d}{dw} wx = x f_Y(wx)
$$

Putting this result into Equation 3.8.3 gives

$$
\begin{aligned}
f_W(w) &= \int_0^\infty x f_X(x) f_Y(wx)\,dx - \int_{-\infty}^0 x f_X(x) f_Y(wx)\,dx \\
&= \int_0^\infty x f_X(x) f_Y(wx)\,dx + \int_{-\infty}^0 (-x) f_X(x) f_Y(wx)\,dx \\
&= \int_0^\infty |x| f_X(x) f_Y(wx)\,dx + \int_{-\infty}^0 |x| f_X(x) f_Y(wx)\,dx \\
&= \int_{-\infty}^\infty |x| f_X(x) f_Y(wx)\,dx
\end{aligned}
$$

which completes the proof.

Example 3.8.5

Let X and Y be independent random variables with pdfs $f_X(x) = \lambda e^{-\lambda x}$, $x > 0$, and $f_Y(y) = \lambda e^{-\lambda y}$, $y > 0$, respectively. Define $W = Y/X$. Find $f_W(w)$.

Substituting into the formula given in Theorem 3.8.4, we can write

$$
\begin{aligned}
f_W(w) &= \int_0^\infty x(\lambda e^{-\lambda x})(\lambda e^{-\lambda x w})\,dx = \lambda^2 \int_0^\infty x e^{-\lambda(1+w)x}\,dx \\
&= \frac{\lambda^2}{\lambda(1+w)} \int_0^\infty x\lambda(1+w) e^{-\lambda(1+w)x}\,dx
\end{aligned}
$$

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 3.8.6 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 Y be a continuous random variable with $f_Y(y) = \frac{1}{2}(1+y)$, $-1 \leq y \leq 1$. Define the random variable W by $W = -4Y + 7$. Find $f_W(w)$. Be sure to specify those values of w for which $f_W(w) \neq 0$.

3.8.2. Let $f_Y(y) = \frac{3}{14}(1+y^2)$, $0 \leq y \leq 2$. Define the random variable W by $W = 3Y + 2$. Find $f_W(w)$. Be sure to specify the values of w for which $f_W(w) \neq 0$.

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

3.8.4. 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 $X + Y$.

3.8.5. 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 $1 \leq w \leq 2$.)

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

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

3.8.7. 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 $F_W(w)$.)

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

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

3.8.10. Suppose the velocity of a gas molecule of mass m is a random variable with pdf $f_Y(y) = ay^2e^{-by^2}$, $y \ge 0$, 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.

3.8.11. 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 $f_Y(y) = 2y, 0 \le y \le 1$

3.8.12. 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 $f_Y(y) = 2y, 0 \le y \le 1$

3.8.13. 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 (136) 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

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

Table 3.9.1

X \ Y	0	1	2	3
0	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{8}$	0
1	0	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{8}$

Table 3.9.2

z	0	1	2	3
$f_Z(z)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	0

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 that an expected value is a weighted average, we see that $E[g(X, Y)]$ is equal to 1:

$$\begin{aligned} E[g(X,Y)] = E(Z) &= \sum_{\text{all } z} z \cdot f_Z(z) \\ &= 0 \cdot \frac{1}{4} + 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{4} + 3 \cdot 0 \\ &= 1 \end{aligned}$$

The same answer is obtained by applying Theorem 3.9.1 to the joint pdf given in Figure 3.9.1:

$$\begin{aligned} E[g(X,Y)] &= 0 \cdot \frac{1}{8} + 1 \cdot \frac{1}{4} + 2 \cdot \frac{1}{8} + 3 \cdot 0 + 3 \cdot 0 + 2 \cdot \frac{1}{8} + 1 \cdot \frac{1}{4} + 0 \cdot \frac{1}{8} \\ &= 1 \end{aligned}$$

The advantage, of course, enjoyed by the latter solution is that we avoid the intermediate step of having to determine $f_Z(z)$. ■

Example 3.9.2 An electrical circuit has three resistors, R_X, R_Y, and R_Z, wired in parallel (see Figure 3.9.1). The nominal resistance of each is fifteen ohms, but their *actual* resistances, X, Y, and Z, vary between ten and twenty according to the joint pdf

$$f_{X,Y,Z}(x, y, z) = \frac{1}{675{,}000}(xy + xz + yz), \qquad \begin{array}{l} 10 \le x \le 20 \\ 10 \le y \le 20 \\ 10 \le z \le 20 \end{array}$$

What is the expected resistance for the circuit?

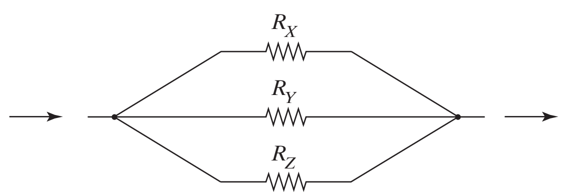


Figure 3.9.1

Let R denote the circuit's resistance. A well-known result in physics holds that

$$\frac{1}{R} = \frac{1}{X} + \frac{1}{Y} + \frac{1}{Z}$$

or, equivalently,

$$R = \frac{XYZ}{XY + XZ + YZ} = R(X, Y, Z)$$

Integrating $R(x, y, z) \cdot f_{X,Y,Z}(x, y, z)$ shows that the expected resistance is five:

$$\begin{aligned} E(R) &= \int_{10}^{20} \int_{10}^{20} \int_{10}^{20} \frac{xyz}{xy + xz + yz} \cdot \frac{1}{675{,}000}(xy + xz + yz)\,dx\,dy\,dz \\ &= \frac{1}{675{,}000} \int_{10}^{20} \int_{10}^{20} \int_{10}^{20} xyz\,dx\,dy\,dz \\ &= 5.0 \end{aligned}$$

■

Theorem 3.9.2 *Let X and Y be any two random variables (discrete or continuous, dependent or independent), and let a and b be any two constants. Then*

$$E(aX + bY) = aE(X) + bE(Y)$$

provided $E(X)$ and $E(Y)$ are both finite.

Proof Consider the continuous case (the discrete case is proved much the same way). Let $f_{X,Y}(x, y)$ be the joint pdf of X and Y, and define $g(X, Y) = aX + bY$. By Theorem 3.9.1,

$$\begin{aligned} E(aX + bY) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (ax + by) f_{X,Y}(x, y)\,dx\,dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (ax) f_{X,Y}(x, y)\,dx\,dy + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (by) f_{X,Y}(x, y)\,dx\,dy \\ &= a \int_{-\infty}^{\infty} x \left[\int_{-\infty}^{\infty} f_{X,Y}(x, y)\,dy \right] dx + b \int_{-\infty}^{\infty} y \left[\int_{-\infty}^{\infty} f_{X,Y}(x, y)\,dx \right] dy \\ &= a \int_{-\infty}^{\infty} x f_X(x)\,dx + b \int_{-\infty}^{\infty} y f_Y(y)\,dy \\ &= aE(X) + bE(Y) \end{aligned}$$

Corollary 3.9.1 *Let $W_1, W_2, \ldots, W_n$ be any random variables for which $E(|W_i|) < \infty$, $i = 1, 2, \ldots, n$, and let $a_1, a_2, \ldots, a_n$ be any set of constants. Then*

$$E(a_1W_1 + a_2W_2 + \cdots + a_nW_n) = a_1E(W_1) + a_2E(W_2) + \cdots + a_nE(W_n)$$

Example 3.9.3 Let X be a binomial random variable defined on n independent trials, each trial resulting in success with probability p. Find $E(X)$.

Note, first, that X can be thought of as a sum, $X = X_1 + X_2 + \cdots + X_n$, where X_i represents the number of successes occurring at the ith trial:

$$X_i = \begin{cases} 1 & \text{if the } i\text{th trial produces a success} \\ 0 & \text{if the } i\text{th trial produces a failure} \end{cases}$$

(Any X_i defined in this way on an individual trial is called a *Bernoulli* random variable. Every binomial random variable, then, can be thought of as the sum of n independent Bernoullis.) By assumption, $p_{X_i}(1) = p$ and $p_{X_i}(0) = 1 - p$, $i = 1, 2, \ldots, n$. Using the corollary,

$$\begin{aligned} E(X) &= E(X_1) + E(X_2) + \cdots + E(X_n) \\ &= n \cdot E(X_1) \end{aligned}$$

the last step being a consequence of the X_i's having identical distributions. But

$$E(X_1) = 1 \cdot p + 0 \cdot (1 - p) = p$$

so $E(X) = np$, which is what we found before (recall Theorem 3.5.1). ■

Comment The problem-solving implications of Theorem 3.9.2 and its corollary should not be underestimated. There are many real-world events that can be modeled as a linear combination $a_1W_1 + a_2W_2 + \cdots + a_nW_n$, where the W_i's are relatively simple random variables. Finding $E(a_1W_1 + a_2W_2 + \cdots + a_nW_n)$ *directly* may be prohibitively difficult because of the inherent complexity of the linear combination. It may very well be the case, though, that calculating the individual $E(W_i)$'s is easy. Compare, for instance, Example 3.9.3 with Theorem 3.5.1. Both derive the formula that $E(X) = np$ when X is a binomial random variable. However, the approach taken in Example 3.9.3 (i.e., using Theorem 3.9.2) is *much* easier. The next several examples further explore the technique of using linear combinations to facilitate the calculation of expected values.

Example 3.9.4

A disgruntled secretary is upset about having to stuff envelopes. Handed a box of n letters and n envelopes, she vents her frustration by putting the letters into the envelopes *at random*. How many people, on the average, will receive their correct mail?

If X denotes the number of envelopes properly stuffed, what we want is $E(X)$. However, applying Definition 3.5.1 here would prove formidable because of the difficulty in getting a workable expression for $p_X(k)$. By using the corollary to Theorem 3.9.2, though, we can solve the problem quite easily.

Let X_i denote a random variable equal to the number of correct letters put into the ith envelope, $i = 1, 2, \ldots, n$. Then X_i equals 0 or 1, and

$$p_{X_i}(k) = P(X_i = k) = \begin{cases} \dfrac{1}{n} & \text{for } k = 1 \\ \dfrac{n-1}{n} & \text{for } k = 0 \end{cases}$$

But $X = X_1 + X_2 + \cdots + X_n$ and $E(X) = E(X_1) + E(X_2) + \cdots + E(X_n)$. Furthermore, each of the X_i's has the same expected value, $1/n$:

$$E(X_i) = \sum_{k=0}^{1} k \cdot P(X_i = k) = 0 \cdot \frac{n-1}{n} + 1 \cdot \frac{1}{n} = \frac{1}{n}$$

It follows that

$$\begin{aligned} E(X) &= \sum_{i=1}^{n} E(X_i) = n \cdot \left(\frac{1}{n}\right) \\ &= 1 \end{aligned}$$

showing that, *regardless of n,* the expected number of properly stuffed envelopes is one. (Are the X_i's independent? Does it matter?) ■

Example 3.9.5 Ten fair dice are rolled. Calculate the expected value of the sum of the faces showing.

If the random variable X denotes the sum of the faces showing on the ten dice, then

$$X = X_1 + X_2 + \cdots + X_{10}$$

where X_i is the number showing on the ith die, $i = 1, 2, \ldots, 10$. By assumption, $p_{X_i}(k) = \frac{1}{6}$ for $k = 1, 2, 3, 4, 5, 6$, so $E(X_i) = \sum_{k=1}^{6} k \cdot \frac{1}{6} = \frac{1}{6} \sum_{k=1}^{6} k = \frac{1}{6} \cdot \frac{6(7)}{2} = 3.5$. By the corollary to Theorem 3.9.2,

$$\begin{aligned} E(X) &= E(X_1) + E(X_2) + \cdots + E(X_{10}) \\ &= 10(3.5) \\ &= 35 \end{aligned}$$

Notice that $E(X)$ can also be deduced here by appealing to the notion that expected values are centers of gravity. It should be clear from our work with combinatorics that $P(X = 10) = P(X = 60)$, $P(X = 11) = P(X = 59)$, $P(X = 12) = P(X = 58)$, and so on. In other words, the probability function $p_X(k)$ is symmetric, which implies that its center of gravity is the midpoint of the range of its X-values. It must be the case, then, that $E(X)$ equals $\frac{10+60}{2}$ or 35. ■

Example 3.9.6 The honor count in a (thirteen-card) bridge hand can vary from zero to thirty-seven according to the formula:

$$\begin{aligned} \text{honor count} = {} & 4 \cdot (\text{number of aces}) + 3 \cdot (\text{number of kings}) + 2 \cdot (\text{number of queens}) \\ & + 1 \cdot (\text{number of jacks}) \end{aligned}$$

What is the expected honor count of North's hand?

The solution here is a bit unusual in that we use the corollary to Theorem 3.9.2 *backward*. If X_i, $i = 1, 2, 3, 4$, denotes the honor count for players North, South, East, and West, respectively, and if X denotes the analogous sum for the entire deck, we can write

$$X = X_1 + X_2 + X_3 + X_4$$

But

$$X = E(X) = 4 \cdot 4 + 3 \cdot 4 + 2 \cdot 4 + 1 \cdot 4 = 40$$

By symmetry, $E(X_i) = E(X_j)$, $i \neq j$, so it follows that $40 = 4 \cdot E(X_1)$, which implies that *ten* is the expected honor count of North's hand. (Try doing this problem directly, without making use of the fact that the deck's honor count is forty.) ■

EXPECTED VALUES OF PRODUCTS: A SPECIAL CASE

We know from Theorem 3.9.1 that for any two random variables X and Y,

$$E(XY) = \begin{cases} \displaystyle\sum_{\text{all } x} \sum_{\text{all } y} xy\, p_{X,Y}(x, y) & \text{if } X \text{ and } Y \text{ are discrete} \\ \displaystyle\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy\, f_{X,Y}(x, y)\, dx\, dy & \text{if } X \text{ and } Y \text{ are continuous} \end{cases}$$

If, however, X and Y are independent, there is an easier way to calculate $E(XY)$.

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 four times the number of bull's-eyes made with the first gun plus six 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 $E(X_i) = \mu \neq 0$, $i = 1, 2, \ldots, n$. 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 3.8.8.)

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*

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

Proof For convenience, denote $E(X)$ by μ_X and $E(Y)$ by μ_Y. Then $E(aX + bY) = a\mu_X + b\mu_Y$ and

$$\begin{aligned}\mathrm{Var}(aX + bY) &= E[(aX + bY)^2] - (a\mu_X + b\mu_Y)^2 \\ &= E(a^2X^2 + b^2Y^2 + 2abXY) - (a^2\mu_X^2 + b^2\mu_Y^2 + 2ab\mu_X\mu_Y) \\ &= [E(a^2X^2) - a^2\mu_X^2] + [E(b^2Y^2) - b^2\mu_Y^2] + [2abE(XY) - 2ab\mu_X\mu_Y] \\ &= a^2[E(X^2) - \mu_X^2] + b^2[E(Y^2) - \mu_Y^2] + 2ab[E(XY) - \mu_X\mu_Y] \\ &= a^2\,\mathrm{Var}(X) + b^2\,\mathrm{Var}(Y) + 2ab\mathrm{Cov}(X, Y)\end{aligned}$$

Example 3.9.8 For the joint pdf $f_{X,Y}(x, y) = x + y, 0 \le x \le 1, 0 \le y \le 1$, find the variance of $X + Y$.

Since X and Y are not independent,

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X, Y)$$

The pdf is symmetric in X and Y, so $\mathrm{Var}(X) = \mathrm{Var}(Y)$, and we can write $\mathrm{Var}(X + Y) = 2[\mathrm{Var}(X) + \mathrm{Cov}(X, Y)]$.

To calculate $\mathrm{Var}(X)$, the marginal pdf of X is needed. But

$$f_X(x) = \int_0^1 (x + y)dy = x + \frac{1}{2}$$

$$\mu_X = \int_0^1 x(x + \frac{1}{2})dx = \int_0^1 (x^2 + \frac{x}{2})dx = \frac{7}{12}$$

$$E(X^2) = \int_0^1 x^2(x + \frac{1}{2})dx = \int_0^1 (x^3 + \frac{x^2}{2})dx = \frac{5}{12}$$

$$\mathrm{Var}(X) = E(X^2) - \mu_X^2 = \frac{5}{12} - \left(\frac{7}{12}\right)^2 = \frac{11}{144}$$

Then

$$E(XY) = \int_0^1 \int_0^1 xy(x + y)dydx = \int_0^1 \left(\frac{x^2}{2} + \frac{x}{3}\right) dx = \left.\frac{x^3}{6} + \frac{x^2}{6}\right|_0^1 = \frac{1}{3}$$

so, putting all of the pieces together,

$$\mathrm{Cov}(X, Y) = 1/3 - (7/12)(7/12) = -1/144$$

and, finally, $\mathrm{Var}(X + Y) = 2[11/144 + (-1/144)] = 5/36$ ■

The two corollaries that follow are straightforward extensions of Theorem 3.9.5 to n variables. The details of the proof will be left as an exercise.

Corollary *Suppose that $W_1, W_2, \ldots, W_n$ are random variables with finite variances. Then*

$$\mathrm{Var}\left(\sum_{i=1}^{a} a_i W_i\right) = \sum_{i=1}^{n} a_i^2 \mathrm{Var}(W_i) + 2\sum_{i<j} a_i a_j \mathrm{Cov}(W_i, W_j)$$

Corollary *Suppose that $W_1, W_2, \ldots, W_n$ are independent random variables with finite variances. Then*

$$\mathrm{Var}(W_1 + W_2 + \cdots + W_n) = \mathrm{Var}(W_1) + \mathrm{Var}(W_2) + \cdots + \mathrm{Var}(W_n)$$

More discussion of the covariance and its role in measuring the relationship between random variables occurs in Section 11.4.

Example 3.9.9 The binomial random variable, being a sum of n independent Bernoullis, is an obvious candidate for the corollary to Theorem 3.9.5 on the sum of independent random variables. Let X_i denote the number of successes occurring on the ith trial. Then

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

and

$$X = X_1 + X_2 + \cdots + X_n = \text{total number of successes in } n \text{ trials}$$

Find Var(X).

Note that

$$E(X_i) = 1 \cdot p + 0 \cdot (1-p) = p$$

and

$$E\left(X_i^2\right) = (1)^2 \cdot p + (0)^2 \cdot (1-p) = p$$

so

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

It follows, then, that the *variance of a binomial random variable is $np(1-p)$*:

$$\text{Var}(X) = \sum_{i=1}^{n} \text{Var}(X_i) = np(1-p)$$

■

Example 3.9.10 Recall the hypergeometric model—an urn contains N chips, r red and w white ($r + w = N$); a random sample of size n is selected without replacement and the random variable X is defined to be the number of red chips in the sample. As in the previous example, write X as a sum of simple random variables.

$$X_i = \begin{cases} 1 & \text{if the } i\text{th chip drawn is red} \\ 0 & \text{otherwise} \end{cases}$$

Then $X = X_1 + X_2 + \cdots + X_n$. Clearly,

$$E(X_i) = 1 \cdot \frac{r}{N} + 0 \cdot \frac{w}{N} = \frac{r}{N}$$

and $E(X) = n\left(\frac{r}{N}\right) = np$, where $p = \frac{r}{N}$.

Since $X_i^2 = X_i$, $E(X_i^2) = E(X_i) = \frac{r}{N}$ and

$$\text{Var}(X_i) = E(X_i^2) - [E(X_i)]^2 = \frac{r}{N} - \left(\frac{r}{N}\right)^2 = p(1-p)$$

Also, for any $j \neq k$,

$$\begin{aligned} \text{Cov}(X_j, X_k) &= E(X_j X_k) - E(X_j)E(X_k) \\ &= 1 \cdot P(X_j X_k = 1) - \left(\frac{r}{N}\right)^2 \\ &= \frac{r}{N} \cdot \frac{r-1}{N-1} - \frac{r^2}{N^2} = -\frac{r}{N} \cdot \frac{N-r}{N} \cdot \frac{1}{N-1} \end{aligned}$$

From the first corollary to Theorem 3.9.5, then,

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

■

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 $\text{Var}(\overline{W}) = \sigma^2/n$. 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

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

■

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 $\text{Cov}(X, Y)$.

3.9.14. Show that

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

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 $\text{Cov}(X, Y) = 0$.

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 $\text{Cov}(X, Y) = 0$ 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 $\text{Var}(X + Y)$. (*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 $\text{Var}(X + Y)$. (*Hint:* See Question 3.9.3.)

3.9.19. Suppose that $f_{X,Y}(x, y) = \frac{3}{2}(x^2 + y^2)$, $0 \le x \le 1$, $0 \le y \le 1$. Find $\text{Var}(X + Y)$.

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 $\text{Var}(W)$, where $W = 4X + 6Y$.

3.9.21. A *Poisson random variable* has pdf $p_X(k) = e^{-\lambda}\frac{\lambda^k}{k!}$, $k = 0, 1, 2, \ldots$ and $\lambda > 0$ (see Section 4.2). Also, $E(X) = \lambda$. Suppose the Poisson random variable U is the number of calls for technical assistance received by a computer company during the firm's nine normal workday hours, with the average number of calls per hour equal 7.0. Also suppose 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. Assume the average number of calls per hour is four 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 pdf with $\mu = 80$ and $\sigma = 15$. The pdf describing an *ordered* observation, though, is *not* the same as the pdf describing a *random* observation. Intuitively, that makes sense. If a single observation is drawn from a normal distribution with $\mu = 80$ and $\sigma = 15$, it would not be surprising if that observation were to take on a value near eighty. On the other hand, if a random sample of $n = 100$ observations is drawn from that same distribution, we would not expect the smallest observation—that is, Y_{min}—to be anywhere near eighty. Common sense tells us that that smallest observation is likely to be much smaller than eighty, just as the largest observation, Y_{max}, is likely to be much larger than eighty.

It follows, then, that before we can do any probability calculations—or any applications whatsoever—involving order statistics, we need to know the pdf of Y_i' for $i = 1, 2, \ldots, n$. We begin by investigating the pdfs of the "extreme" order statistics, $f_{Y_{\text{max}}}(y)$ and $f_{Y_{\text{min}}}(y)$. These are the simplest to work with. At the end of the section we return to the more general problems of finding (1) the pdf of Y_i' for any i and (2) the joint pdf of Y_i' and Y_j', where $i < j$.

Theorem 3.10.1 *Suppose that $Y_1, Y_2, \ldots, Y_n$ is a random sample of continuous random variables, each having pdf $f_Y(y)$ and cdf $F_Y(y)$. Then*

a. *The pdf of the largest order statistic is*

$$f_{Y_{\text{max}}}(y) = f_{Y_n'}(y) = n[F_Y(y)]^{n-1} f_Y(y)$$

b. *The pdf of the smallest order statistic is*

$$f_{Y_{\text{min}}}(y) = f_{Y_1'}(y) = n[1 - F_Y(y)]^{n-1} f_Y(y)$$

Proof Finding the pdfs of Y_{max} and Y_{min} is accomplished by using the now-familiar technique of differentiating a random variable's cdf. Consider, for example, the case of the largest order statistic, Y_n':

$$\begin{aligned} F_{Y_n'}(y) = F_{Y_{\text{max}}}(y) &= P(Y_{\text{max}} \le y) \\ &= P(Y_1 \le y, Y_2 \le y, \cdots, Y_n \le y) \\ &= P(Y_1 \le y) \cdot P(Y_2 \le y) \cdots P(Y_n \le y) \qquad \text{(why?)} \\ &= [F_Y(y)]^n \end{aligned}$$

Therefore,

$$f_{Y_n'}(y) = d/dy[[F_Y(y)]^n] = n[F_Y(y)]^{n-1} f_Y(y)$$

Similarly, for the smallest order statistic ($i = 1$),

$$\begin{aligned} F_{Y_1'}(y) = F_{Y_{\text{min}}}(y) &= P(Y_{\text{min}} \le y) \\ &= 1 - P(Y_{\text{min}} > y) = 1 - P(Y_1 > y) \cdot P(Y_2 > y) \cdots P(Y_n > y) \\ &= 1 - [1 - F_Y(y)]^n \end{aligned}$$

Therefore,

$$f_{Y_1'}(y) = d/dy[1 - [1 - F_Y(y)]^n] = n[1 - F_Y(y)]^{n-1} f_Y(y)$$

Example 3.10.2 Suppose a random sample of $n = 3$ observations, Y_1, Y_2, and Y_3, is taken from the exponential pdf, $f_Y(y) = e^{-y}$, $y \ge 0$. Compare $f_{Y_1}(y)$ with $f_{Y_1'}(y)$. Intuitively, which will be larger, $P(Y_1 < 1)$ or $P(Y_1' < 1)$?

The pdf for Y_1, of course, is just the pdf of the distribution being sampled, that is,

$$f_{Y_1}(y) = f_Y(y) = e^{-y}, \quad y \geq 0$$

To find the pdf for Y_1' requires that we apply the formula given in the proof of Theorem 3.10.1 for $f_{Y_{\min}}(y)$. Note, first of all, that

$$F_Y(y) = \int_0^y e^{-t}dt = -e^{-t}\Big|_0^y = 1 - e^{-y}$$

Then, since $n = 3$ (and $i = 1$), we can write

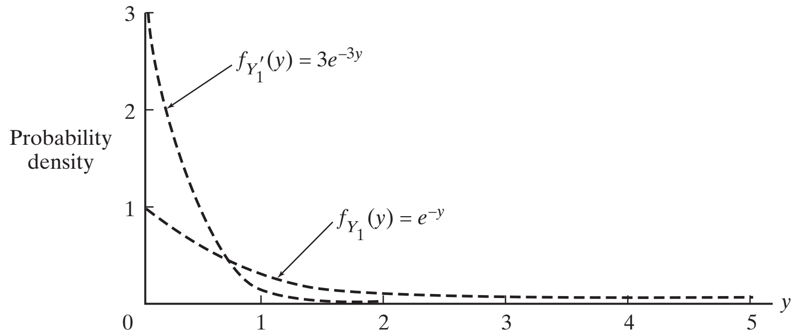
$$\begin{aligned} f_{Y_1'}(y) &= 3[1 - (1 - e^{-y})]^2 e^{-y} \\ &= 3e^{-3y}, \quad y \geq 0 \end{aligned}$$

Figure 3.10.1 shows the two pdfs plotted on the same set of axes. Compared to $f_{Y_1}(y)$, the pdf for Y_1' has more of its area located above the smaller values of y (where Y_1' is more likely to lie). For example, the probability that the smallest observation (out of three) is less than 1 is 95%, while the probability that a random observation is less than 1 is only 63%:

$$\begin{aligned} P(Y_1' < 1) &= \int_0^1 3e^{-3y}\,dy = \int_0^3 e^{-u}\,du = -e^{-u}\Big|_0^3 = 1 - e^{-3} \\ &= 0.95 \end{aligned}$$

$$\begin{aligned} P(Y_1 < 1) &= \int_0^1 e^{-y}\,dy = -e^{-y}\Big|_0^1 = 1 - e^{-1} \\ &= 0.63 \end{aligned}$$

Figure 3.10.1



■

Example 3.10.3

Suppose a random sample of size 10 is drawn from a continuous pdf $f_Y(y)$. What is the probability that the largest observation, Y_{10}', is less than the pdf's median, m?

Using the formula for $f_{Y_{10}'}(y) = f_{Y_{\max}}(y)$ given in the proof of Theorem 3.10.1, it is certainly true that

$$P(Y_{10}' < m) = \int_{-\infty}^{m} 10 f_Y(y)[F_Y(y)]^9 dy \tag{3.10.1}$$

but the problem does not specify $f_Y(y)$, so Equation 3.10.1 is of no help.

Fortunately, a much simpler solution is available, even if $f_Y(y)$ were specified: The event "$Y'_{10} < m$" is equivalent to the event "$Y_1 < m \cap Y_2 < m \cap \cdots \cap Y_{10} < m$." Therefore,

$$P(Y'_{10} < m) = P(Y_1 < m, Y_2 < m, \ldots, Y_{10} < m) \tag{3.10.2}$$

But the ten observations here are independent, so the intersection probability implicit on the right-hand side of Equation 3.10.2 factors into a product of ten terms. Moreover, each of those terms equals $\frac{1}{2}$ (by definition of the median), so

$$\begin{aligned} P(Y'_{10} < m) &= P(Y_1 < m) \cdot P(Y_2 < m) \cdots P(Y_{10} < m) \\ &= \left(\tfrac{1}{2}\right)^{10} \\ &= 0.00098 \end{aligned}$$

■

Example 3.10.4 To find order statistics for discrete pdfs, the probability arguments of the type used in the proof of Theorem 3.10.1 can be be employed. The example of finding the pdf of $X_{\min}$ for the discrete density function $p_X(k)$, $k = 0, 1, 2, \ldots$ suffices to demonstrate this point.

Given a random sample $X_1, X_2, \ldots, X_n$ from $p_X(k)$, choose an arbitrary nonnegative integer m. Recall that the cdf in this case is given by $F_X(m) = \sum_{k=0}^{m} p_k$.

Consider the events

$$\begin{aligned} A &= (m \le X_1 \cap m \le X_2 \cap \cdots \cap m \le X_n) \text{ and} \\ B &= (m+1 \le X_1 \cap m+1 \le X_2 \cap \cdots \cap m+1 \le X_n) \end{aligned}$$

Then $p_{X_{\min}}(m) = P(A \cap B^C) = P(A) - P(A \cap B) = P(A) - P(B)$, where $A \cap B = B$, since $B \subset A$.

Now $P(A) = P(m \le X_1) \cdot P(m \le X_2) \cdot \ldots \cdot P(m \le X_n) = [1 - F_X(m-1)]^n$ by the independence of the X_i. Similarly $P(B) = [1 - F_X(m)]^n$, so

$$p_{Y_{\min}}(m) = [1 - F_X(m-1)]^n - [1 - F_X(m)]^n$$

■

A GENERAL FORMULA FOR $f_{Y'_i}(y)$

Having discussed two special cases of order statistics, $Y_{\min}$ and $Y_{\max}$, we now turn to the more general problem of finding the pdf for the ith order statistic, where i can be any integer from 1 through n.

Theorem 3.10.2 *Let $Y_1, Y_2, \ldots, Y_n$ be a random sample of continuous random variables drawn from a distribution having pdf $f_Y(y)$ and cdf $F_Y(y)$. The pdf of the ith order statistic is given by*

$$f_{Y'_i}(y) = \frac{n!}{(i-1)!(n-i)!}[F_Y(y)]^{i-1}[1 - F_Y(y)]^{n-i} f_Y(y)$$

for $1 \le i \le n$.

Proof We will give a heuristic argument that draws on the similarity between the statement of Theorem 3.10.2 and the binomial distribution. For a formal induction proof verifying the expression given for $f_{Y'_i}(y)$, see (105).

Recall the derivation of the binomial probability function, $p_X(k) = P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$, where X is the number of successes in n independent

(Continued on next page)

(Theorem 3.10.2 continued)

trials, and p is the probability that any given trial ends in success. Central to that derivation was the recognition that the event "$X = k$" is actually a union of all the different (mutually exclusive) sequences having exactly k successes and $n - k$ failures. Because the trials are independent, the probability of any such sequence is $p^k(1-p)^{n-k}$ and the number of such sequences (by Theorem 2.6.2) is $n!/[k!(n-k)!]$ (or $\binom{n}{k}$), so the probability that $X = k$ is the product $\binom{n}{k}p^k(1-p)^{n-k}$.

Here we are looking for the pdf of the ith order statistic at some point y, that is, $f_{Y_i'}(y)$. As was the case with the binomial, that pdf will reduce to a combinatorial term times the probability associated with an intersection of independent events. The only fundamental difference is that Y_i' is a continuous random variable, whereas the binomial X is discrete, which means that what we find here will be a probability *density* function.

By Theorem 2.6.2, there are $n!/[(i-1)!1!(n-i)!]$ ways that n observations can be parceled into three groups such that the ith largest is at the point y (see Figure 3.10.2). Moreover, the likelihood associated with any particular set of points having the configuration pictured in Figure 3.10.2 will be the probability that $i-1$ (independent) observations are all less than y, $n-i$ observations are greater than y, and one observation is at y. The probability density associated with those constraints for a given set of points would be $[F_Y(y)]^{i-1}[1-F_Y(y)]^{n-i}f_Y(y)$. The probability density, then, that the ith order statistic is located at the point y is the product

$$f_{Y_i'}(y) = \frac{n!}{(i-1)!(n-i)!}[F_Y(y)]^{i-1}[1-F_Y(y)]^{n-i}f_Y(y)$$

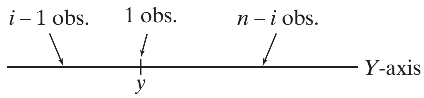


Figure 3.10.2

Example 3.10.5

Suppose that many years of observation have confirmed that the annual maximum flood tide Y (in feet) for a certain river can be modeled by the pdf

$$f_Y(y) = \frac{1}{20}, \quad 20 < y < 40$$

(*Note:* It is unlikely that flood tides would be described by anything as simple as a uniform pdf. We are making that choice here solely to facilitate the mathematics.) The Army Corps of Engineers is planning to build a levee along a certain portion of the river, and they want to make it high enough so that there is only a 30% chance that the second worst flood in the next thirty-three years will overflow the embankment. How high should the levee be? (We assume that there will be only one potential flood per year.)

Let h be the desired height. If $Y_1, Y_2, \ldots, Y_{33}$ denote the flood tides for the next $n = 33$ years, what we require of h is that

$$P(Y'_{32} > h) = 0.30$$

As a starting point, notice that for $20 < y < 40$,

$$F_Y(y) = \int_{20}^{y} \frac{1}{20}\,dy = \frac{y}{20} - 1$$

Therefore,

$$f_{Y'_{32}}(y) = \frac{33!}{31!1!}\left(\frac{y}{20}-1\right)^{31}\left(2-\frac{y}{20}\right)^{1}\cdot\frac{1}{20}$$

and h is the solution of the integral equation

$$\int_h^{40}(33)(32)\left(\frac{y}{20}-1\right)^{31}\left(2-\frac{y}{20}\right)^{1}\cdot\frac{dy}{20} = 0.30 \tag{3.10.3}$$

If we make the substitution

$$u = \frac{y}{20} - 1$$

Equation 3.10.3 simplifies to

$$\begin{aligned} P(Y'_{32} > h) &= 33(32)\int_{(h/20)-1}^{1} u^{31}(1-u)\,du \\ &= 1 - 33\left(\frac{h}{20}-1\right)^{32} + 32\left(\frac{h}{20}-1\right)^{33} \end{aligned} \tag{3.10.4}$$

Setting the right-hand side of Equation 3.10.4 equal to 0.30 and solving for h by trial and error gives

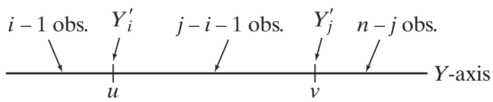
$$h = 39.33 \text{ feet}$$

■

JOINT PDFS OF ORDER STATISTICS

Finding the joint pdf of two or more order statistics is easily accomplished by generalizing the argument that derived from Figure 3.10.2. Suppose, for example, that each of n observations in a random sample has pdf $f_Y(y)$ and cdf $F_Y(y)$. The joint pdf for order statistics Y'_i and Y'_j at points u and v, where $i < j$ and $u < v$, can be deduced from Figure 3.10.3, which shows how the n points must be distributed if the ith and jth order statistics are to be located at points u and v, respectively.

Figure 3.10.3



By Theorem 2.6.2, the number of ways to divide a set of n observations into groups of sizes $i-1$, 1, $j-i-1$, 1, and $n-j$ is the quotient

$$\frac{n!}{(i-1)!1!(j-i-1)!1!(n-j)!}$$

Also, given the independence of the n observations, the probability that $i-1$ are less than u is $[F_Y(u)]^{i-1}$, the probability that $j-i-1$ are between u and v is $[F_Y(v)-F_Y(u)]^{j-i-1}$, and the probability that $n-j$ are greater than v is $[1-F_Y(v)]^{n-j}$. Multiplying, then, by the pdfs describing the likelihoods that Y'_i and Y'_j would be at points u and v, respectively, gives the joint pdf of the two order statistics:

$$\begin{aligned} f_{Y'_i,Y'_j}(u,v) = {} & \frac{n!}{(i-1)!(j-i-1)!(n-j)!}[F_Y(u)]^{i-1}[F_Y(v)-F_Y(u)]^{j-i-1}\cdot \\ & [1-F_Y(v)]^{n-j}\, f_Y(u)\, f_Y(v) \end{aligned} \tag{3.10.5}$$

for $i < j$ and $u < v$.

Example 3.10.6 Let Y_1, Y_2, and Y_3 be a random sample of size $n = 3$ from the uniform pdf defined over the unit interval, $f_Y(y) = 1, 0 \le y \le 1$. By definition, the *range*, R, of a sample is the difference between the largest and smallest order statistics, in this case

$$R = \text{range} = Y_{\max} - Y_{\min} = Y_3' - Y_1'$$

Find $f_R(r)$, the pdf for the range.

We will begin by finding the joint pdf of Y_1' and Y_3'. Then $f_{Y_1',Y_3'}(u, v)$ is integrated over the region $Y_3' - Y_1' \le r$ to find the cdf, $F_R(r) = P(R \le r)$. The final step is to differentiate the cdf and make use of the fact that $f_R(r) = F_R'(r)$.

If $f_Y(y) = 1, 0 \le y \le 1$, it follows that

$$F_Y(y) = P(Y \le y) = \begin{cases} 0, & y < 0 \\ y, & 0 \le y \le 1 \\ 1, & y > 1 \end{cases}$$

Applying Equation 3.10.5, then, with $n = 3$, $i = 1$, and $j = 3$, gives the joint pdf of Y_1' and Y_3'. Specifically,

$$\begin{aligned} f_{Y_1',Y_3'}(u, v) &= \frac{3!}{0!1!0!} u^0 (v-u)^1 (1-v)^0 \cdot 1 \cdot 1 \\ &= 6(v-u), \quad 0 \le u < v \le 1 \end{aligned}$$

Moreover, we can write the cdf for R in terms of Y_1' and Y_3':

$$F_R(r) = P(R \le r) = P(Y_3' - Y_1' \le r) = P(Y_3' \le Y_1' + r)$$

Figure 3.10.4 shows the region in the $Y_1'Y_3'$-plane corresponding to the event that $R \le r$. Integrating the joint pdf of Y_1' and Y_3' over the shaded region gives

$$F_R(r) = P(R \le r) = \int_0^{1-r} \int_u^{u+r} 6(v-u)\, dv\, du + \int_{1-r}^1 \int_u^1 6(v-u)\, dv\, du$$

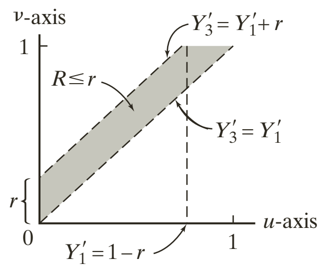


Figure 3.10.4

The first double integral equals $3r^2 - 3r^3$; the second equals r^3. Therefore,

$$F_R(r) = 3r^2 - 3r^3 + r^3 = 3r^2 - 2r^3$$

which implies that

$$f_R(r) = F_R'(r) = 6r - 6r^2, \quad 0 \le r \le 1$$

■

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 $f_Y(y) = 3y^2, 0 \leq y \leq 1$. 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 $f_Y(y) = 2y, 0 \leq y \leq 1$. 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 $f_y(y) = e^{-y}, y \geq 0$. 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 $f_Y(y) = 2y, 0 \leq y \leq 1$. On the same set of axes, graph the pdfs for Y_2, Y_1', and Y_5'.

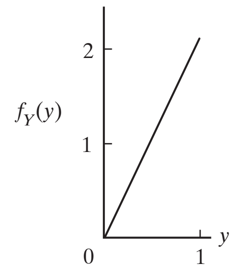
3.10.9. Suppose that n observations are taken at random from the pdf

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

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, for instance, can be considered a pair of random variables. While it is easy to measure girth, it can be difficult to determine height; thus it might be of interest to a lumberman to know the probabilities of a ponderosa pine's attaining certain heights given a known value for its girth. Or consider the plight of a school board member agonizing over which way to vote on a proposed budget increase. Her task would be that much easier if she knew the conditional probability that x additional tax dollars would stimulate an average increase of y points among twelfth graders taking a standardized proficiency exam.

FINDING CONDITIONAL PDFS FOR DISCRETE RANDOM VARIABLES

In the case of discrete random variables, a conditional pdf can be treated in the same way as a conditional probability. Note the similarity between Definitions 3.11.1 and 2.4.1.

Definition 3.11.1
Let X and Y be discrete random variables. The *conditional probability density function of Y given x*—that is, the probability that Y takes on the value y given that X is equal to x—is denoted $p_{Y|x}(y)$ and given by

$$p_{Y|x}(y) = P(Y = y \mid X = x) = \frac{p_{X,Y}(x, y)}{p_X(x)}$$

for $p_X(x) \neq 0$.

Example 3.11.1 A fair coin is tossed five times. Let the random variable Y denote the total number of heads that occur, and let X denote the number of heads occurring on the last two tosses. Find the conditional pdf $p_{Y|x}(y)$ for all x and y.

Clearly, there will be three different conditional pdfs, one for each possible value of X ($x = 0, x = 1$, and $x = 2$). Moreover, for each value of x there will be four possible values of Y, based on whether the first three tosses yield zero, one, two, or three heads.

For example, suppose no heads occur on the last two tosses. Then $X = 0$, and

$$\begin{aligned} p_{Y|0}(y) &= P(Y = y \mid X = 0) = P(y \text{ heads occur on first three tosses}) \\ &= \binom{3}{y}\left(\frac{1}{2}\right)^y\left(1-\frac{1}{2}\right)^{3-y} \\ &= \binom{3}{y}\left(\frac{1}{2}\right)^3, \quad y = 0, 1, 2, 3 \end{aligned}$$

Now, suppose that $X = 1$. The corresponding conditional pdf in that case becomes

$$p_{Y|x}(y) = P(Y = y \mid X = 1)$$

Notice that $Y = 1$ if zero heads occur in the first three tosses, $Y = 2$ if one head occurs in the first three trials, and so on. Therefore,

$$\begin{aligned} p_{Y|1}(y) &= \binom{3}{y-1}\left(\frac{1}{2}\right)^{y-1}\left(1-\frac{1}{2}\right)^{3-(y-1)} \\ &= \binom{3}{y-1}\left(\frac{1}{2}\right)^3, \quad y = 1, 2, 3, 4 \end{aligned}$$

Similarly,

$$p_{Y|2}(y) = P(Y = y \mid X = 2) = \binom{3}{y-2}\left(\frac{1}{2}\right)^3, \quad y = 2, 3, 4, 5$$

Figure 3.11.1 shows the three conditional pdfs. Each has the same shape, but the possible values of Y are different for each value of X.

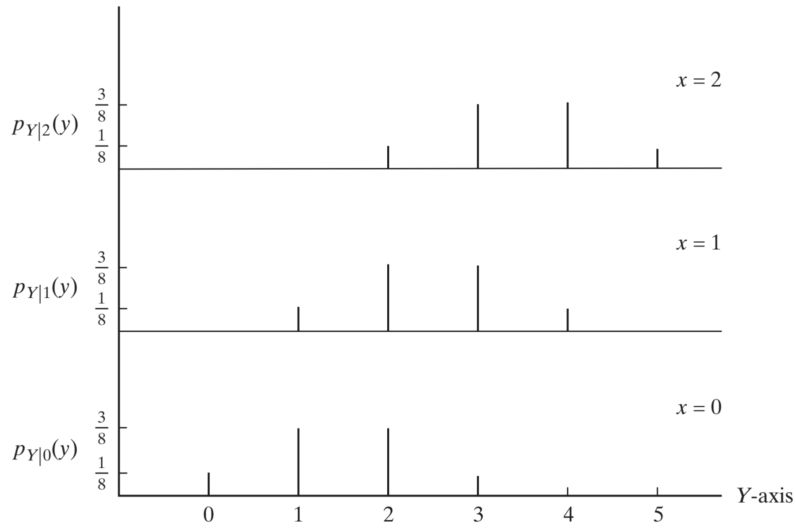


Figure 3.11.1 ■

Example 3.11.2

Assume that the probabilistic behavior of a pair of discrete random variables X and Y is described by the joint pdf

$$p_{X,Y}(x, y) = xy^2/39$$

defined over the four points (1, 2), (1, 3), (2, 2), and (2, 3). Find the conditional probability that $X = 1$ given that $Y = 2$.

By definition,

$$\begin{aligned} p_{X|2}(1) &= P(X = 1 \text{ given that } Y = 2) \\ &= \frac{p_{X,Y}(1, 2)}{p_Y(2)} \\ &= \frac{1 \cdot 2^2/39}{1 \cdot 2^2/39 + 2 \cdot 2^2/39} \\ &= 1/3 \end{aligned}$$

■

Example 3.11.3

Suppose that X and Y are two independent binomial random variables, each defined on n trials and each having the same success probability p. Let $Z = X + Y$. Show that the conditional pdf $p_{X|z}(x)$ is a hypergeometric distribution.

We know from Example 3.8.2 that Z has a binomial distribution with parameters $2n$ and p. That is,

$$p_Z(z) = P(Z = z) = \binom{2n}{z} p^z (1-p)^{2n-z}, \quad z = 0, 1, \ldots, 2n.$$

By Definition 3.11.1,

$$
\begin{aligned}
p_{X|z}(x) = P(X = x|Z = z) &= \frac{p_{X,Z}(x, z)}{p_Z(z)} \\
&= \frac{P(X = x \text{ and } Z = z)}{P(Z = z)} \\
&= \frac{P(X = x \text{ and } Y = z - x)}{P(Z = z)} \\
&= \frac{P(X = x) \cdot P(Y = z - x)}{P(Z = z)} \quad \text{(because } X \text{ and } Y \text{ are independent)} \\
&= \frac{\binom{n}{x} p^x (1-p)^{n-x} \cdot \binom{n}{z-x} p^{z-x} (1-p)^{n-(z-x)}}{\binom{2n}{z} p^z (1-p)^{2n-z}} \\
&= \frac{\binom{n}{x}\binom{n}{z-x}}{\binom{2n}{z}}
\end{aligned}
$$

which we recognize as being the hypergeometric distribution.

Comment The notion of a conditional pdf generalizes easily to situations involving more than two discrete random variables. For example, if X, Y, and Z have the joint pdf $p_{X,Y,Z}(x, y, z)$, the *joint conditional pdf* of, say, X and Y given that $Z = z$ is the ratio

$$p_{X,Y|z}(x, y) = \frac{p_{X,Y,Z}(x, y, z)}{p_Z(z)}$$

■

Example 3.11.4 Suppose that random variables X, Y, and Z have the joint pdf

$$p_{X,Y,Z}(x, y, z) = xy/9z$$

for points (1, 1, 1), (2, 1, 2), (1, 2, 2), (2, 2, 2), and (2, 2, 1). Find $p_{X,Y|z}(x, y)$ for all values of z.

To begin, we see from the points for which $p_{X,Y,Z}(x, y, z)$ is defined that Z has two possible values, 1 and 2. Suppose $z = 1$. Then

$$p_{X,Y|1}(x, y) = \frac{p_{X,Y,Z}(x, y, 1)}{p_Z(1)}$$

But

$$
\begin{aligned}
p_Z(1) = P(Z = 1) &= P[(1, 1, 1) \cup (2, 2, 1)] \\
&= 1 \cdot \frac{1}{9} \cdot 1 + 2 \cdot \frac{2}{9} \cdot 1 \\
&= \frac{5}{9}
\end{aligned}
$$

Therefore,

$$p_{X,Y|1}(x, y) = \frac{xy/9}{\frac{5}{9}} = xy/5 \quad \text{for} \quad (x, y) = (1, 1) \quad \text{and} \quad (2, 2)$$

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 $p_{Y|x}(x)$.

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 Poisson 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 3.8.3.)

3.11.10. Suppose Compositor A is preparing a manuscript to be published. Assume that she makes X errors per page, where X has a Poisson pdf, with $\lambda = 2$ (see Question 3.9.21). A second compositor, B, is also working on the book. He makes Y errors on a page, where Y is Poisson with $\lambda = 3$. 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 five hundred twenty errors. Write a formula for the exact probability that fewer than half of the errors are due to Compositor A.

FINDING CONDITIONAL PDFS FOR CONTINUOUS RANDOM VARIABLES

If the variables X and Y are continuous, we can still appeal to the quotient $f_{X,Y}(x, y)/f_X(x)$ as the definition of $f_{Y|x}(y)$ and argue its propriety by analogy. A more satisfying approach, though, is to arrive at the same conclusion by taking the limit of Y's "conditional" *cdf*.

If X is continuous, a direct evaluation of $F_{Y|x}(y) = P(Y \leq y|X = x)$, via Definition 2.4.1, is impossible, since the denominator would be zero. Alternatively, we can think of $P(Y \leq y|X = x)$ as a limit:

$$P(Y \leq y|X = x) = \lim_{h \to 0} P(Y \leq y|x \leq X \leq x + h)$$

$$= \lim_{h \to 0} \frac{\int_x^{x+h} \int_{-\infty}^{y} f_{X,Y}(t, u)\, du\, dt}{\int_x^{x+h} f_X(t)\, dt}$$

Evaluating the quotient of the limits gives $\frac{0}{0}$, so l'Hôpital's rule is indicated:

$$P(Y \leq y|X = x) = \lim_{h \to 0} \frac{\frac{d}{dh} \int_x^{x+h} \int_{-\infty}^{y} f_{X,Y}(t, u)\, du\, dt}{\frac{d}{dh} \int_x^{x+h} f_X(t)\, dt} \tag{3.11.1}$$

By the Fundamental Theorem of Calculus,

$$\frac{d}{dh} \int_x^{x+h} g(t)\, dt = g(x + h)$$

which simplifies Equation 3.11.1 to

$$P(Y \leq y|X = x) = \lim_{h \to 0} \frac{\int_{-\infty}^{y} f_{X,Y}[(x + h), u]\, du}{f_X(x + h)}$$

$$= \frac{\int_{-\infty}^{y} \lim_{h \to 0} f_{X,Y}(x + h, u)\, du}{\lim_{h \to 0} f_X(x + h)} = \int_{-\infty}^{y} \frac{f_{X,Y}(x, u)}{f_X(x)}\, du$$

provided that the limit operation and the integration can be interchanged [see (9) for a discussion of when such an interchange is valid]. It follows from this last expression that $f_{X,Y}(x, y)/f_X(x)$ behaves as a conditional probability density function should, and we are justified in extending Definition 3.11.1 to the continuous case.

Example 3.11.5

Let X and Y be continuous random variables with joint pdf

$$f_{X,Y}(x, y) = \begin{cases} \left(\frac{1}{8}\right)(6 - x - y), & 0 \leq x \leq 2, \quad 2 \leq y \leq 4 \\ 0, & \text{elsewhere} \end{cases}$$

Find (a) $f_X(x)$, (b) $f_{Y|x}(y)$, and (c) $P(2 < Y < 3|x = 1)$.

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 $0 \le x \le 1$ and $0 \le y \le 1$. 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) $P(\frac{1}{4} \le Y \le \frac{3}{4}|X = \frac{1}{2})$.
(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 $0 \le x_i \le 1$, $i = 1, 2, \ldots, 5$. 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) $P\left(Y > 1 | X > \frac{1}{2}\right)$.

3.11.21. For continuous random variables X and Y, prove that $E(Y) = E_X[E(Y|x)]$.

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 $\sum_{\text{all } k} k^r p_X(k)$ 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

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

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

Example 3.12.1

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

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

[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

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

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

The t in $M_X(t)$ can be any number in a neighborhood of zero, as long as $M_X(t) < \infty$. Here, $M_X(t)$ is an infinite sum of the terms $[(1-p)e^t]^k$, and that sum will be finite only if $(1-p)e^t < 1$, or, equivalently, if $t < \ln[1/(1-p)]$. It will be assumed, then, in what follows that $0 < t < \ln[1/(1-p)]$.

Recall that

$$\sum_{k=0}^{\infty} r^k = \frac{1}{1-r}$$

provided $0 < r < 1$. This formula can be used on Equation 3.12.1, where $r = (1-p)e^t$ and $0 < t < \ln\left[\frac{1}{(1-p)}\right]$. Specifically,

$$M_X(t) = \frac{p}{1-p}\left[\sum_{k=0}^{\infty}[(1-p)e^t]^k - [(1-p)e^t]^0\right]$$

$$= \frac{p}{1-p}\left[\frac{1}{1-(1-p)e^t} - 1\right]$$

$$= \frac{pe^t}{1-(1-p)e^t}$$

■

Example 3.12.2

Suppose that X is a binomial random variable with pdf

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

Find $M_X(t)$.

By Definition 3.12.1,

$$M_X(t) = E(e^{tX}) = \sum_{k=0}^{n} e^{tk}\binom{n}{k}p^k(1-p)^{n-k}$$

$$= \sum_{k=0}^{n}\binom{n}{k}(pe^t)^k(1-p)^{n-k} \tag{3.12.2}$$

To get a closed-form expression for $M_X(t)$—that is, to evaluate the sum indicated in Equation 3.12.2—requires a (hopefully) familiar formula from algebra: According to *Newton's binomial expansion,*

$$(x+y)^n = \sum_{k=0}^{n}\binom{n}{k}x^k y^{n-k} \tag{3.12.3}$$

for any x and y. Suppose we let $x = pe^t$ and $y = 1-p$. It follows from Equations 3.12.2 and 3.12.3, then, that

$$M_X(t) = (1-p+pe^t)^n$$

Notice in this case that $M_X(t)$ is defined for all values of t. ■

Example 3.12.3 Suppose that Y has an exponential pdf, where $f_Y(y) = \lambda e^{-\lambda y}$, $y > 0$. Find $M_Y(t)$.

Since the exponential pdf describes a continuous random variable, $M_Y(t)$ is an integral:

$$\begin{aligned} M_Y(t) = E(e^{tY}) &= \int_0^\infty e^{ty} \cdot \lambda e^{-\lambda y}\, dy \\ &= \int_0^\infty \lambda e^{-(\lambda - t)y}\, dy \end{aligned}$$

After making the substitution $u = (\lambda - t)y$, we can write

$$\begin{aligned} M_Y(t) &= \int_{u=0}^\infty \lambda e^{-u} \frac{du}{\lambda - t} \\ &= \frac{\lambda}{\lambda - t}\left[-e^{-u}\Big|_{u=0}^{\infty}\right] \\ &= \frac{\lambda}{\lambda - t}\left[1 - \lim_{u\to\infty} e^{-u}\right] = \frac{\lambda}{\lambda - t} \end{aligned}$$

Here, $M_Y(t)$ is finite and nonzero only when $u = (\lambda - t)y > 0$, which implies that t must be less than λ. For $t \geq \lambda$, $M_Y(t)$ fails to exist. ■

Example 3.12.4 The normal (or bell-shaped) curve was introduced in Example 3.4.3. Its pdf is the rather cumbersome function

$$f_Y(y) = (1/\sqrt{2\pi}\sigma) \exp\left[-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right], \quad -\infty < y < \infty$$

where $\mu = E(Y)$ and $\sigma^2 = \mathrm{Var}(Y)$. Derive the moment-generating function for this most important of all probability models.

Since Y is a continuous random variable,

$$\begin{aligned} M_Y(t) = E(e^{tY}) &= (1/\sqrt{2\pi}\sigma) \int_{-\infty}^{\infty} \exp(ty) \exp\left[-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right] dy \\ &= (1/\sqrt{2\pi}\sigma) \int_{-\infty}^{\infty} \exp\left[-\frac{y^2 - 2\mu y - 2\sigma^2 t y + \mu^2}{2\sigma^2}\right] dy \end{aligned} \tag{3.12.4}$$

Evaluating the integral in Equation 3.12.4 is best accomplished by completing the square of the numerator of the exponent (which means that the square of half the coefficient of y is added and subtracted). That is, we can write

$$\begin{aligned} &y^2 - (2\mu + 2\sigma^2 t)y + (\mu + \sigma^2 t)^2 - (\mu + \sigma^2 t)^2 + \mu^2 \\ &\qquad = [y - (\mu + \sigma^2 t)]^2 - \sigma^4 t^2 + 2\mu t \sigma^2 \end{aligned} \tag{3.12.5}$$

The last two terms on the right-hand side of Equation 3.12.5, though, do not involve y, so they can be factored out of the integral, and Equation 3.12.4 reduces to

$$M_Y(t) = \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right)(1/\sqrt{2\pi}\sigma) \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2}\left[\frac{y - (\mu + t\sigma^2)}{\sigma}\right]^2\right] dy$$

But, together, the latter two factors equal 1 (why?), implying that the moment-generating function for a normally distributed random variable is given by

$$M_Y(t) = e^{\mu t + \sigma^2 t^2/2}$$

■

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. The random variable X has a *Poisson distribution* $p_X(k) = e^{-\lambda}\lambda^k/k!$, $k = 0, 1, 2, \ldots$. 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 $r = 1$,

$$M_Y^{(1)}(0) = \frac{d}{dt}\int_{-\infty}^{\infty} e^{ty} f_Y(y)\,dy\bigg|_{t=0} = \int_{-\infty}^{\infty} \frac{d}{dt}e^{ty} f_Y(y)\,dy\bigg|_{t=0}$$

$$= \int_{-\infty}^{\infty} ye^{ty} f_Y(y)\,dy\bigg|_{t=0} = \int_{-\infty}^{\infty} ye^{0\cdot y} f_Y(y)\,dy$$

$$= \int_{-\infty}^{\infty} y f_Y(y)\,dy = E(Y)$$

(Continued on next page)

(Theorem 3.12.1 continued)

For $r = 2$,

$$M_Y^{(2)}(0) = \frac{d^2}{dt^2}\int_{-\infty}^{\infty} e^{ty} f_Y(y)\,dy\bigg|_{t=0} = \int_{-\infty}^{\infty} \frac{d^2}{dt^2} e^{ty} f_Y(y)\,dy\bigg|_{t=0}$$

$$= \int_{-\infty}^{\infty} y^2 e^{ty} f_Y(y)\,dy\bigg|_{t=0} = \int_{-\infty}^{\infty} y^2 e^{0\cdot y} f_Y(y)\,dy$$

$$= \int_{-\infty}^{\infty} y^2 f_Y(y)\,dy = E(Y^2)$$

Example 3.12.5

For a geometric random variable X with pdf

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

we saw in Example 3.12.1 that

$$M_X(t) = pe^t[1-(1-p)e^t]^{-1}$$

Find the expected value of X by differentiating its moment-generating function.

Using the product rule, we can write the first derivative of $M_X(t)$ as

$$M_X^{(1)}(t) = pe^t(-1)[1-(1-p)e^t]^{-2}(-1)(1-p)e^t + [1-(1-p)e^t]^{-1}pe^t$$

$$= \frac{p(1-p)e^{2t}}{[1-(1-p)e^t]^2} + \frac{pe^t}{1-(1-p)e^t}$$

Setting $t = 0$ shows that $E(X) = \frac{1}{p}$:

$$M_X^{(1)}(0) = E(X) = \frac{p(1-p)e^{2\cdot 0}}{[1-(1-p)e^0]^2} + \frac{pe^0}{1-(1-p)e^0}$$

$$= \frac{p(1-p)}{p^2} + \frac{p}{p}$$

$$= \frac{1}{p}$$

■

Example 3.12.6

Find the expected value of an exponential random variable with pdf

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

Use the fact that

$$M_Y(t) = \lambda(\lambda - t)^{-1}$$

(as shown in Example 3.12.3).

Differentiating $M_Y(t)$ gives

$$M_Y^{(1)}(t) = \lambda(-1)(\lambda - t)^{-2}(-1)$$

$$= \frac{\lambda}{(\lambda - t)^2}$$

Set $t = 0$. Then

$$M_Y^{(1)}(0) = \frac{\lambda}{(\lambda - 0)^2}$$

implying that

$$E(Y) = \frac{1}{\lambda}$$

■

Example 3.12.7 Find an expression for $E(X^k)$ if the moment-generating function for X is given by

$$M_X(t) = (1 - p_1 - p_2) + p_1 e^t + p_2 e^{2t}$$

The only way to deduce a formula for an arbitrary moment such as $E(X^k)$ is to calculate the first couple moments and look for a pattern that can be generalized. Here,

$$M_X^{(1)}(t) = p_1 e^t + 2 p_2 e^{2t}$$

so

$$E(X) = M_X^{(1)}(0) = p_1 e^0 + 2 p_2 e^{2 \cdot 0} = p_1 + 2p_2$$

Taking the second derivative, we see that

$$M_X^{(2)}(t) = p_1 e^t + 2^2 p_2 e^{2t}$$

implying that

$$E(X^2) = M_X^{(2)}(0) = p_1 e^0 + 2^2 p_2 e^{2 \cdot 0} = p_1 + 2^2 p_2$$

Clearly, each successive differentiation will leave the p_1 term unaffected but will multiply the p_2 term by 2. Therefore,

$$E(X^k) = M_X^{(k)}(0) = p_1 + 2^k p_2$$ ■

USING MOMENT-GENERATING FUNCTIONS TO FIND VARIANCES

In addition to providing a useful technique for calculating $E(W^r)$, moment-generating functions can also find variances, because

$$\text{Var}(W) = E(W^2) - [E(W)]^2 \tag{3.12.6}$$

for any random variable W (recall Theorem 3.6.1). Other useful "descriptors" of pdfs can also be reduced to combinations of moments. The *skewness* of a distribution, for example, is a function of $E[(W - \mu)^3]$, where $\mu = E(W)$. But

$$E[(W - \mu)^3] = E(W^3) - 3E(W^2)E(W) + 2[E(W)]^3$$

In many cases, finding $E[(W - \mu)^2]$ or $E[(W - \mu)^3]$ could be quite difficult if moment-generating functions were not available.

Example 3.12.8 We know from Example 3.12.2 that if X is a binomial random variable with parameters n and p, then

$$M_X(t) = (1 - p + pe^t)^n$$

Use $M_X(t)$ to find the variance of X.

The first two derivatives of $M_X(t)$ are

$$M_X^{(1)}(t) = n(1 - p + pe^t)^{n-1} \cdot pe^t$$

and

$$M_X^{(2)}(t) = pe^t \cdot n(n-1)(1-p+pe^t)^{n-2} \cdot pe^t + n(1-p+pe^t)^{n-1} \cdot pe^t$$

Setting $t = 0$ gives

$$M_X^{(1)}(0) = np = E(X)$$

and

$$M_X^{(2)}(0) = n(n-1)p^2 + np = E(X^2)$$

From Equation 3.12.6, then,

$$\begin{aligned}\text{Var}(X) &= n(n-1)p^2 + np - (np)^2 \\ &= np(1-p)\end{aligned}$$

(the same answer we found in Example 3.9.9). ■

Example 3.12.9 It can be shown (see Question 3.12.7) that the moment-generating function for a Poisson random variable is given by

$$M_X(t) = e^{-\lambda+\lambda e^t}$$

Use $M_X(t)$ to find $E(X)$ and $\text{Var}(X)$.

Taking the first derivative of $M_X(t)$ gives

$$M_X^{(1)}(t) = e^{-\lambda+\lambda e^t} \cdot \lambda e^t$$

so

$$\begin{aligned}E(X) = M_X^{(1)}(0) &= e^{-\lambda+\lambda e^0} \cdot \lambda e^0 \\ &= \lambda\end{aligned}$$

Applying the product rule to $M_X^{(1)}(t)$ yields the second derivative,

$$M_X^{(2)}(t) = e^{-\lambda+\lambda e^t} \cdot \lambda e^t + \lambda e^t e^{-\lambda+\lambda e^t} \cdot \lambda e^t$$

For $t = 0$,

$$\begin{aligned}M_X^{(2)}(0) = E(X^2) &= e^{-\lambda+\lambda e^0} \cdot \lambda e^0 + \lambda e^0 \cdot e^{-\lambda+\lambda e^0} \cdot \lambda e^0 \\ &= \lambda + \lambda^2\end{aligned}$$

The variance of a Poisson random variable, then, proves to be the same as its mean:

$$\begin{aligned}\text{Var}(X) &= E(X^2) - [E(X)]^2 \\ &= M_X^{(2)}(0) - \left[M_X^{(1)}(0)\right]^2 \\ &= \lambda^2 + \lambda - \lambda^2 \\ &= \lambda\end{aligned}$$

■

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 Example 3.4.1.

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 (103).

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,

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

and

$$p_{X_2}(k) = P(X_2 = k) = \frac{e^{-\lambda_2}\lambda_2^k}{k!}, \quad k = 0, 1, 2, \ldots$$

Let $X = X_1 + X_2$. What is the pdf for X?

According to Example 3.12.9, the moment-generating functions for X_1 and X_2 are

$$M_{X_1}(t) = e^{-\lambda_1+\lambda_1 e^t}$$

and

$$M_{X_2}(t) = e^{-\lambda_2+\lambda_2 e^t}$$

Moreover, if $X = X_1 + X_2$, then by part (b) of Theorem 3.12.3,

$$\begin{aligned} M_X(t) &= M_{X_1}(t) \cdot M_{X_2}(t) \\ &= e^{-\lambda_1+\lambda_1 e^t} \cdot e^{-\lambda_2+\lambda_2 e^t} \\ &= e^{-(\lambda_1+\lambda_2)+(\lambda_1+\lambda_2)e^t} \end{aligned} \tag{3.12.7}$$

But, by inspection, Equation 3.12.7 is the moment-generating function that a Poisson random variable with $\lambda = \lambda_1 + \lambda_2$ would have. It follows, then, by Theorem 3.12.2 that

$$p_X(k) = \frac{e^{-(\lambda_1+\lambda_2)}(\lambda_1 + \lambda_2)^k}{k!}, \quad k = 0, 1, 2, \ldots$$

Comment The Poisson random variable reproduces itself in the sense that the sum of independent Poissons is also a Poisson. A similar property holds for independent normal random variables (see Question 3.12.19) and, under certain conditions, for independent binomial random variables (recall Example 3.8.2). ■

Example 3.12.11 We saw in Example 3.12.4 that a normal random variable, Y, with mean μ and variance σ^2 has pdf

$$f_Y(y) = (1/\sqrt{2\pi}\sigma) \exp\left[-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right], \quad -\infty < y < \infty$$

and mgf

$$M_Y(t) = e^{\mu t + \sigma^2 t^2/2}$$

By definition, a *standard normal random variable* is a normal random variable for which $\mu = 0$ and $\sigma = 1$. Denoted Z, the pdf and mgf for a standard normal random variable are $f_Z(z) = (1/\sqrt{2\pi})e^{-z^2/2}$, $-\infty < z < \infty$, and $M_Z(t) = e^{t^2/2}$, respectively. Show that the ratio

$$\frac{Y-\mu}{\sigma}$$

is a standard normal random variable, Z.

Write $\frac{Y-\mu}{\sigma}$ as $\frac{1}{\sigma}Y - \frac{\mu}{\sigma}$. By part (a) of Theorem 3.12.3,

$$\begin{aligned} M_{(Y-\mu)/\sigma}(t) &= e^{-\mu t/\sigma} M_Y\left(\frac{t}{\sigma}\right) \\ &= e^{-\mu t/\sigma} e^{[\mu t/\sigma + \sigma^2 (t/\sigma)^2/2]} \\ &= e^{t^2/2} \end{aligned}$$

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 $f_Y(y) = \lambda^2 y e^{-\lambda y}$, $0 \le y$.

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 $P(X \le 2)$ 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 $P(W \le 1)$. (*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.

In the lingo of applied statistics, there are actually two fundamentally different types of "means"—*population means* and *sample means*. The term "population mean" is a synonym for what mathematical statisticians would call an expected value—that is, a population mean (μ) is a weighted average of the possible values associated with a theoretical probability model, either $p_X(k)$ or $f_Y(y)$, depending on whether the underlying random variable is discrete or continuous. A *sample mean* is the arithmetic average of a set of measurements. If, for example, n observations—$y_1, y_2, \ldots, y_n$—are taken on a continuous random variable Y, the sample mean is denoted $\bar{y}$, where

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

Conceptually, sample means are *estimates* of population means, where the "quality" of the estimation is a function of (1) the sample size and (2) the standard deviation (σ) associated with the individual measurements. Intuitively, as the sample size gets larger and/or the standard deviation gets smaller, the approximation will tend to get better.

Interpreting means (either $\bar{y}$ or μ) is not always easy. To be sure, what they imply *in principle* is clear enough—both $\bar{y}$ and μ are measuring the centers of their respective distributions. Still, many a wrong conclusion can be traced directly to researchers misunderstanding the value of a mean. Why? Because the distributions that $\bar{y}$ and/or μ are *actually* representing may be dramatically different from the distributions we *think* they are representing.

An interesting case in point arises in connection with SAT scores. Each fall the average SATs earned by students in each of the fifty states and the District of Columbia are released by the Educational Testing Service (ETS). At the state and

Table 3.13.1

State	Average SAT Score	State	Average SAT Score
North Dakota	1816	Arizona	1547
Illinois	1802	Oregon	1544
Iowa	1794	Virginia	1530
South Dakota	1792	New Jersey	1526
Minnesota	1786	Connecticut	1525
Michigan	1784	West Virginia	1522
Wisconsin	1782	Washington	1519
Missouri	1771	California	1504
Wyoming	1762	Alaska	1485
Kansas	1753	North Carolina	1483
Kentucky	1746	Pennsylvania	1481
Nebraska	1745	Rhode Island	1480
Colorado	1735	Indiana	1474
Mississippi	1714	Maryland	1468
Tennessee	1714	New York	1468
Arkansas	1698	Hawaii	1460
Oklahoma	1697	Nevada	1458
Utah	1690	Florida	1448
Louisiana	1667	Georgia	1445
Ohio	1652	South Carolina	1443
Montana	1637	Texas	1432
Alabama	1617	Maine	1387
New Mexico	1617	Idaho	1364
New Hampshire	1566	Delaware	1359
Massachusetts	1556	DC	1309
Vermont	1554		

federal level, these SAT averages are often used as indicators of educational success or failure. Does it make sense, though, to use SAT averages to characterize the quality of a state's education system? Absolutely not! Averages of this sort refer to very different distributions from state to state. Any attempt to interpret them at face value will necessarily be misleading.

One such state-by-state SAT comparison that appeared in 2014 is reproduced in Table 3.13.1. Notice that North Dakota's entry is 1816, which is the highest average listed. Does it follow that North Dakota's educational system is among the best in the nation? Probably not. So, why did their students do so well on the SAT?

The answer to that question lies in the academic profiles of the students who take the SAT. North Dakota primarily uses the ACT for its college admission test. Only 2% of college-bound students took the SAT, approximately one hundred sixty seniors. The SAT is primarily used by private schools, where admissions tend to be more competitive. As a result, the students in North Dakota who take the SAT are not representative of the entire population of students in that state. A disproportionate number are exceptionally strong academically, those being the students who feel that they have the ability to be competitive at Ivy League–type schools. The number 1816, then, is the average of *something* (in this case, an elite subset of all students), but it does not correspond to the center of the SAT distribution for *all* students.

The moral here is that analyzing data effectively requires that we look beyond the obvious. What we have learned in Chapter 3 about random variables and probability distributions and expected values will be helpful only if we take the time to learn about the context and the idiosyncrasies of the phenomenon being studied. To do otherwise is likely to lead to conclusions that are, at best, superficial and, at worst, incorrect.