Chapter 3

Random Variables and Distributions

3.1 Random Variables and Discrete Distributions

A random variable is a real-valued function defined on a sample space. Random variables are the main tools used for modeling unknown quantities in statistical analyses. For each random variable X and each set C of real numbers, we could calculate the probability that X takes its value in C. The collection of all of these probabilities is the distribution of X. There are two major classes of distributions and random variables: discrete (this section) and continuous (Sec. 3.2). Discrete distributions are those that assign positive probability to at most countably many different values. A discrete distribution can be characterized by its probability function (p.f.), which specifies the probability that the random variable takes each of the different possible values. A random variable with a discrete distribution will be called a discrete random variable.

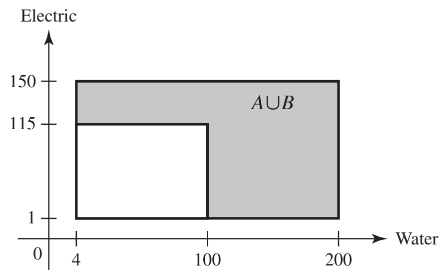
Definition of a Random Variable

Example 3.1.1 Tossing a Coin. Consider an experiment in which a fair coin is tossed 10 times. In this experiment, the sample space S can be regarded as the set of outcomes consisting of the 2^{10} different sequences of 10 heads and/or tails that are possible. We might be interested in the number of heads in the observed outcome. We can let X stand for the real-valued function defined on S that counts the number of heads in each outcome. For example, if s is the sequence HHTTTHTTTH, then $X(s) = 4$. For each possible sequence s consisting of 10 heads and/or tails, the value $X(s)$ equals the number of heads in the sequence. The possible values for the function X are 0, 1, . . . , 10. ◄

Definition 3.1.1 Random Variable. Let S be the sample space for an experiment. A real-valued function that is defined on S is called a *random variable*.

For example, in Example 3.1.1, the number X of heads in the 10 tosses is a random variable. Another random variable in that example is $Y = 10 - X$, the number of tails.

Figure 3.1 The event that at least one utility demand is high in Example 3.1.3.



Example 3.1.2 Measuring a Person's Height. Consider an experiment in which a person is selected at random from some population and her height in inches is measured. This height is a random variable. ◀

Example 3.1.3 Demands for Utilities. Consider the contractor in Example 1.5.4 on page 19 who is concerned about the demands for water and electricity in a new office complex. The sample space was pictured in Fig. 1.5 on page 12, and it consists of a collection of points of the form (x, y), where x is the demand for water and y is the demand for electricity. That is, each point $s \in S$ is a pair $s = (x, y)$. One random variable that is of interest in this problem is the demand for water. This can be expressed as $X(s) = x$ when $s = (x, y)$. The possible values of X are the numbers in the interval [4, 200]. Another interesting random variable is Y, equal to the electricity demand, which can be expressed as $Y(s) = y$ when $s = (x, y)$. The possible values of Y are the numbers in the interval [1, 150]. A third possible random variable Z is an indicator of whether or not at least one demand is high. Let A and B be the two events described in Example 1.5.4. That is, A is the event that water demand is at least 100, and B is the event that electric demand is at least 115. Define

$$Z(s) = \begin{cases} 1 & \text{if } s \in A \cup B, \\ 0 & \text{if } s \notin A \cup B. \end{cases}$$

The possible values of Z are the numbers 0 and 1. The event $A \cup B$ is indicated in Fig. 3.1. ◀

The Distribution of a Random Variable

When a probability measure has been specified on the sample space of an experiment, we can determine probabilities associated with the possible values of each random variable X. Let C be a subset of the real line such that $\{X \in C\}$ is an event, and let $\Pr(X \in C)$ denote the probability that the value of X will belong to the subset C. Then $\Pr(X \in C)$ is equal to the probability that the outcome s of the experiment will be such that $X(s) \in C$. In symbols,

$$\Pr(X \in C) = \Pr(\{s \colon X(s) \in C\}). \tag{3.1.1}$$

Definition 3.1.2 Distribution. Let X be a random variable. The *distribution* of X is the collection of all probabilities of the form $\Pr(X \in C)$ for all sets C of real numbers such that $\{X \in C\}$ is an event.

It is a straightforward consequence of the definition of the distribution of X that this distribution is itself a probability measure on the set of real numbers. The set

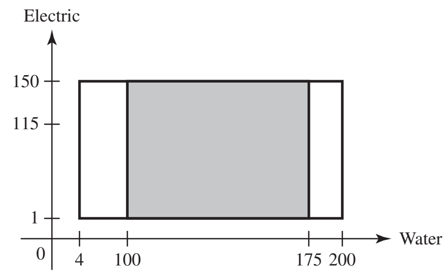


Figure 3.2 The event that water demand is between 50 and 175 in Example 3.1.5.

$\{X \in C\}$ will be an event for every set C of real numbers that most readers will be able to imagine.

Example 3.1.4 Tossing a Coin. Consider again an experiment in which a fair coin is tossed 10 times, and let X be the number of heads that are obtained. In this experiment, the possible values of X are 0, 1, 2, . . . , 10. For each x, $\Pr(X = x)$ is the sum of the probabilities of all of the outcomes in the event $\{X = x\}$. Because the coin is fair, each outcome has the same probability $1/2^{10}$, and we need only count how many outcomes s have $X(s) = x$. We know that $X(s) = x$ if and only if exactly x of the 10 tosses are H. Hence, the number of outcomes s with $X(s) = x$ is the same as the number of subsets of size x (to be the heads) that can be chosen from the 10 tosses, namely, $\binom{10}{x}$, according to Definitions 1.8.1 and 1.8.2. Hence,

$$\Pr(X = x) = \binom{10}{x}\frac{1}{2^{10}} \quad \text{for } x = 0, 1, 2, \ldots, 10. \qquad \blacktriangleleft$$

Example 3.1.5 Demands for Utilities. In Example 1.5.4, we actually calculated some features of the distributions of the three random variables X, Y, and Z defined in Example 3.1.3. For example, the event A, defined as the event that water demand is at least 100, can be expressed as $A = \{X \geq 100\}$, and $\Pr(A) = 0.5102$. This means that $\Pr(X \geq 100) = 0.5102$. The distribution of X consists of all probabilities of the form $\Pr(X \in C)$ for all sets C such that $\{X \in C\}$ is an event. These can all be calculated in a manner similar to the calculation of $\Pr(A)$ in Example 1.5.4. In particular, if C is a subinterval of the interval [4, 200], then

$$\Pr(X \in C) = \frac{(150 - 1) \times (\text{length of interval } C)}{29{,}204}. \tag{3.1.2}$$

For example, if C is the interval [50,175], then its length is 125, and $\Pr(X \in C) = 149 \times 125/29{,}204 = 0.6378$. The subset of the sample space whose probability was just calculated is drawn in Fig. 3.2. ◀

The general definition of distribution in Definition 3.1.2 is awkward, and it will be useful to find alternative ways to specify the distributions of random variables. In the remainder of this section, we shall introduce a few such alternatives.

Discrete Distributions

Definition 3.1.3 Discrete Distribution/Random Variable. We say that a random variable X has a *discrete distribution* or that X is a *discrete random variable* if X can take only a finite number k of different values $x_1, \ldots, x_k$ or, at most, an infinite sequence of different values $x_1, x_2, \ldots$.

Random variables that can take every value in an interval are said to have *continuous distributions* and are discussed in Sec. 3.2.

Definition 3.1.4 Probability Function/p.f./Support. If a random variable X has a discrete distribution, the *probability function* (abbreviated *p.f.*) of X is defined as the function f such that for every real number x,

$$f(x) = \Pr(X = x).$$

The closure of the set $\{x : f(x) > 0\}$ is called the *support of (the distribution of) X*.

Some authors refer to the probability function as the *probability mass function,* or p.m.f. We will not use that term again in this text.

Example 3.1.6 Demands for Utilities. The random variable Z in Example 3.1.3 equals 1 if at least one of the utility demands is high, and $Z = 0$ if neither demand is high. Since Z takes only two different values, it has a discrete distribution. Note that $\{s : Z(s) = 1\} = A \cup B$, where A and B are defined in Example 1.5.4. We calculated $\Pr(A \cup B) = 0.65253$ in Example 1.5.4. If Z has p.f. f, then

$$f(z) = \begin{cases} 0.65253 & \text{if } z = 1, \\ 0.34747 & \text{if } z = 0, \\ 0 & \text{otherwise.} \end{cases}$$

The support of Z is the set $\{0, 1\}$, which has only two elements. ◄

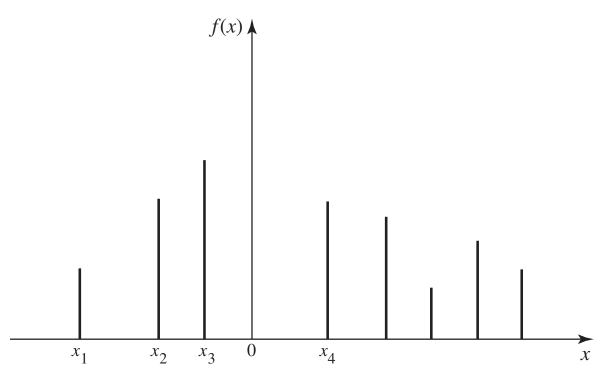
Example 3.1.7 Tossing a Coin. The random variable X in Example 3.1.4 has only 11 different possible values. Its p.f. f is given at the end of that example for the values $x = 0, \dots, 10$ that constitute the support of X; $f(x) = 0$ for all other values of x. ◄

Here are some simple facts about probability functions

Theorem 3.1.1 Let X be a discrete random variable with p.f. f. If x is not one of the possible values of X, then $f(x) = 0$. Also, if the sequence $x_1, x_2, \dots$ includes all the possible values of X, then $\sum_{i=1}^{\infty} f(x_i) = 1$. ■

A typical p.f. is sketched in Fig. 3.3, in which each vertical segment represents the value of $f(x)$ corresponding to a possible value x. The sum of the heights of the vertical segments in Fig. 3.3 must be 1.

Figure 3.3 An example of a p.f.



Theorem 3.1.2 shows that the p.f. of a discrete random variable characterizes its distribution, and it allows us to dispense with the general definition of distribution when we are discussing discrete random variables.

Theorem 3.1.2 If X has a discrete distribution, the probability of each subset C of the real line can be determined from the relation

$$\Pr(X \in C) = \sum_{x_i \in C} f(x_i). \quad ■$$

Some random variables have distributions that appear so frequently that the distributions are given names. The random variable Z in Example 3.1.6 is one such.

Definition 3.1.5 Bernoulli Distribution/Random Variable. A random variable Z that takes only two values 0 and 1 with $\Pr(Z = 1) = p$ has the *Bernoulli distribution with parameter* p. We also say that Z is a *Bernoulli random variable with parameter* p.

The Z in Example 3.1.6 has the Bernoulli distribution with parameter 0.65252. It is easy to see that the name of each Bernoulli distribution is enough to allow us to compute the p.f., which, in turn, allows us to characterize its distribution.

We conclude this section with illustrations of two additional families of discrete distributions that arise often enough to have names.

Uniform Distributions on Integers

Example 3.1.8 Daily Numbers. A popular state lottery game requires participants to select a three-digit number (leading 0s allowed). Then three balls, each with one digit, are chosen at random from well-mixed bowls. The sample space here consists of all triples (i_1, i_2, i_3) where $i_j \in \{0, \dots, 9\}$ for $j = 1, 2, 3$. If $s = (i_1, i_2, i_3)$, define $X(s) = 100i_1 + 10i_2 + i_3$. For example, $X(0, 1, 5) = 15$. It is easy to check that $\Pr(X = x) = 0.001$ for each integer $x \in \{0, 1, \dots, 999\}$. ◂

Definition 3.1.6 Uniform Distribution on Integers. Let $a \le b$ be integers. Suppose that the value of a random variable X is equally likely to be each of the integers $a, \dots, b$. Then we say that X has the *uniform distribution on the integers* $a, \dots, b$.

The X in Example 3.1.8 has the uniform distribution on the integers $0, 1, \dots, 999$. A uniform distribution on a set of k integers has probability $1/k$ on each integer. If $b > a$, there are $b - a + 1$ integers from a to b including a and b. The next result follows immediately from what we have just seen, and it illustrates how the name of the distribution characterizes the distribution.

Theorem 3.1.3 If X has the uniform distribution on the integers $a, \dots, b$, the p.f. of X is

$$f(x) = \begin{cases} \dfrac{1}{b - a + 1} & \text{for } x = a, \dots, b, \\ 0 & \text{otherwise.} \end{cases} \quad ■$$

The uniform distribution on the integers $a, \dots, b$ represents the outcome of an experiment that is often described by saying that one of the integers $a, \dots, b$ is *chosen at random*. In this context, the phrase "at random" means that each of the $b - a + 1$ integers is equally likely to be chosen. In this same sense, it is not possible to choose an integer at random from the set of *all* positive integers, because it is not possible to assign the same probability to every one of the positive integers and still make the sum of these probabilities equal to 1. In other words, a uniform distribution cannot be assigned to an infinite sequence of possible values, but such a distribution can be assigned to any finite sequence.

Note: Random Variables Can Have the Same Distribution without Being the Same Random Variable. Consider two consecutive daily number draws as in Example 3.1.8. The sample space consists of all 6-tuples $(i_1, \ldots, i_6)$, where the first three coordinates are the numbers drawn on the first day and the last three are the numbers drawn on the second day (all in the order drawn). If $s = (i_1, \ldots, i_6)$, let $X_1(s) = 100i_1 + 10i_2 + i_3$ and let $X_2(s) = 100i_4 + 10i_5 + i_6$. It is easy to see that X_1 and X_2 are different functions of s and are not the same random variable. Indeed, there is only a small probability that they will take the same value. But they have the same distribution because they assume the same values with the same probabilities. If a businessman has 1000 customers numbered $0, \ldots, 999$, and he selects one at random and records the number Y, the distribution of Y will be the same as the distribution of X_1 and of X_2, but Y is not like X_1 or X_2 in any other way.

Binomial Distributions

Example 3.1.9 Defective Parts. Consider again Example 2.2.5 from page 69. In that example, a machine produces a defective item with probability p $(0 < p < 1)$ and produces a nondefective item with probability $1 - p$. We assumed that the events that the different items were defective were mutually independent. Suppose that the experiment consists of examining n of these items. Each outcome of this experiment will consist of a list of which items are defective and which are not, in the order examined. For example, we can let 0 stand for a nondefective item and 1 stand for a defective item. Then each outcome is a string of n digits, each of which is 0 or 1. To be specific, if, say, $n = 6$, then some of the possible outcomes are

$$010010,\ 100100,\ 000011,\ 110000,\ 100001,\ 000000,\ \text{etc.} \tag{3.1.3}$$

We will let X denote the number of these items that are defective. Then the random variable X will have a discrete distribution, and the possible values of X will be $0, 1, 2, \ldots, n$. For example, the first four outcomes listed in Eq. (3.1.3) all have $X(s) = 2$. The last outcome listed has $X(s) = 0$. ◀

Example 3.1.9 is a generalization of Example 2.2.5 with n items inspected rather than just six, and rewritten in the notation of random variables. For $x = 0, 1, \ldots, n$, the probability of obtaining each particular ordered sequence of n items containing exactly x defectives and $n - x$ nondefectives is $p^x(1-p)^{n-x}$, just as it was in Example 2.2.5. Since there are $\binom{n}{x}$ different ordered sequences of this type, it follows that

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

Therefore, the p.f. of X will be as follows:

$$f(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & \text{for } x = 0, 1, \ldots, n, \\ 0 & \text{otherwise.} \end{cases} \tag{3.1.4}$$

Definition 3.1.7 Binomial Distribution/Random Variable. The discrete distribution represented by the p.f. in (3.1.4) is called the *binomial distribution with parameters n and p*. A random variable with this distribution is said to be a *binomial random variable with parameters n and p*.

The reader should be able to verify that the random variable X in Example 3.1.4, the number of heads in a sequence of 10 independent tosses of a fair coin, has the binomial distribution with parameters 10 and 1/2.

Since the name of each binomial distribution is sufficient to construct its p.f., it follows that the name is enough to identify the distribution. The name of each distribution includes the two parameters. The binomial distributions are very important in probability and statistics and will be discussed further in later chapters of this book.

A short table of values of certain binomial distributions is given at the end of this book. It can be found from this table, for example, that if X has the binomial distribution with parameters $n = 10$ and $p = 0.2$, then $\Pr(X = 5) = 0.0264$ and $\Pr(X \geq 5) = 0.0328$.

As another example, suppose that a clinical trial is being run. Suppose that the probability that a patient recovers from her symptoms during the trial is p and that the probability is $1 - p$ that the patient does not recover. Let Y denote the number of patients who recover out of n independent patients in the trial. Then the distribution of Y is also binomial with parameters n and p. Indeed, consider a general experiment that consists of observing n independent repititions (trials) with only two possible results for each trial. For convenience, call the two possible results "success" and "failure." Then the distribution of the number of trials that result in success will be binomial with parameters n and p, where p is the probability of success on each trial.

Note: Names of Distributions. In this section, we gave names to several families of distributions. The name of each distribution includes any numerical parameters that are part of the definition. For example, the random variable X in Example 3.1.4 has the binomial distribution with parameters 10 and 1/2. It is a correct statement to say that X has a binomial distribution or that X has a discrete distribution, but such statements are only partial descriptions of the distribution of X. Such statements are *not* sufficient to name the distribution of X, and hence they are not sufficient as answers to the question "What is the distribution of X?" The same considerations apply to all of the named distributions that we introduce elsewhere in the book. When attempting to specify the distribution of a random variable by giving its name, one must give the full name, including the values of any parameters. Only the full name is sufficient for determining the distribution.

Summary

A random variable is a real-valued function defined on a sample space. The distribution of a random variable X is the collection of all probabilities $\Pr(X \in C)$ for all subsets C of the real numbers such that $\{X \in C\}$ is an event. A random variable X is discrete if there are at most countably many possible values for X. In this case, the distribution of X can be characterized by the probability function (p.f.) of X, namely, $f(x) = \Pr(X = x)$ for x in the set of possible values. Some distributions are so famous that they have names. One collection of such named distributions is the collection of uniform distributions on finite sets of integers. A more famous collection is the collection of binomial distributions whose parameters are n and p, where n is a positive integer and $0 < p < 1$, having p.f. (3.1.4). The binomial distribution with parameters $n = 1$ and p is also called the Bernoulli distribution with parameter p. The names of these distributions also characterize the distributions.

Exercises

1. Suppose that a random variable X has the uniform distribution on the integers $10, \dots, 20$. Find the probability that X is even.

2. Suppose that a random variable X has a discrete distribution with the following p.f.:

$$f(x) = \begin{cases} cx & \text{for } x = 1, \dots, 5, \\ 0 & \text{otherwise.} \end{cases}$$

Determine the value of the constant c.

3. Suppose that two balanced dice are rolled, and let X denote the absolute value of the difference between the two numbers that appear. Determine and sketch the p.f. of X.

4. Suppose that a fair coin is tossed 10 times independently. Determine the p.f. of the number of heads that will be obtained.

5. Suppose that a box contains seven red balls and three blue balls. If five balls are selected at random, without replacement, determine the p.f. of the number of red balls that will be obtained.

6. Suppose that a random variable X has the binomial distribution with parameters $n = 15$ and $p = 0.5$. Find $\Pr(X < 6)$.

7. Suppose that a random variable X has the binomial distribution with parameters $n = 8$ and $p = 0.7$. Find $\Pr(X \geq 5)$ by using the table given at the end of this book. *Hint:* Use the fact that $\Pr(X \geq 5) = \Pr(Y \leq 3)$, where Y has the binomial distribution with parameters $n = 8$ and $p = 0.3$.

8. If 10 percent of the balls in a certain box are red, and if 20 balls are selected from the box at random, with replacement, what is the probability that more than three red balls will be obtained?

9. Suppose that a random variable X has a discrete distribution with the following p.f.:

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

Find the value of the constant c.

10. A civil engineer is studying a left-turn lane that is long enough to hold seven cars. Let X be the number of cars in the lane at the end of a randomly chosen red light. The engineer believes that the probability that $X = x$ is proportional to $(x + 1)(8 - x)$ for $x = 0, \dots, 7$ (the possible values of X).

- **a.** Find the p.f. of X.
- **b.** Find the probability that X will be at least 5.

11. Show that there does not exist any number c such that the following function would be a p.f.:

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

3.2 Continuous Distributions

Next, we focus on random variables that can assume every value in an interval (bounded or unbounded). If a random variable X has associated with it a function f such that the integral of f over each interval gives the probability that X is in the interval, then we call f the probability density function (p.d.f.) of X and we say that X has a continuous distribution.

The Probability Density Function

Example 3.2.1 **Demands for Utilities.** In Example 3.1.5, we determined the distribution of the demand for water, X. From Fig. 3.2, we see that the smallest possible value of X is 4 and the largest is 200. For each interval $C = [c_0, c_1] \subset [4, 200]$, Eq. (3.1.2) says that

$$\Pr(c_0 \leq X \leq c_1) = \frac{149(c_1 - c_0)}{29204} = \frac{c_1 - c_0}{196} = \int_{c_0}^{c_1} \frac{1}{196} dx.$$

So, if we define

$$f(x) = \begin{cases} \dfrac{1}{196} & \text{if } 4 \le x \le 200, \\ 0 & \text{otherwise,} \end{cases} \tag{3.2.1}$$

we have that

$$\Pr(c_0 \le X \le c_1) = \int_{c_0}^{c_1} f(x)dx. \tag{3.2.2}$$

Because we defined $f(x)$ to be 0 for x outside of the interval [4, 200], we see that Eq. (3.2.2) holds for all $c_0 \le c_1$, even if $c_0 = -\infty$ and/or $c_1 = \infty$. ◄

The water demand X in Example 3.2.1 is an example of the following.

Definition 3.2.1 Continuous Distribution/Random Variable. We say that a random variable X has a *continuous distribution* or that X is a *continuous random variable* if there exists a nonnegative function f, defined on the real line, such that for every interval of real numbers (bounded or unbounded), the probability that X takes a value in the interval is the integral of f over the interval.

For example, in the situation described in Definition 3.2.1, for each bounded closed interval $[a, b]$,

$$\Pr(a \le X \le b) = \int_{a}^{b} f(x)\,dx. \tag{3.2.3}$$

Similarly, $\Pr(X \ge a) = \int_a^\infty f(x)\,dx$ and $\Pr(X \le b) = \int_{-\infty}^{b} f(x)\,dx$.

We see that the function f characterizes the distribution of a continuous random variable in much the same way that the probability function characterizes the distribution of a discrete random variable. For this reason, the function f plays an important role, and hence we give it a name.

Definition 3.2.2 Probability Density Function/p.d.f./Support. If X has a continuous distribution, the function f described in Definition 3.2.1 is called the *probability density function* (abbreviated *p.d.f.*) of X. The closure of the set $\{x : f(x) > 0\}$ is called the *support of (the distribution of) X*.

Example 3.2.1 demonstrates that the water demand X has p.d.f. given by (3.2.1).

Every p.d.f. f must satisfy the following two requirements:

$$f(x) \ge 0, \quad \text{for all x,} \tag{3.2.4}$$

and

$$\int_{-\infty}^{\infty} f(x)\,dx = 1. \tag{3.2.5}$$

A typical p.d.f. is sketched in Fig. 3.4. In that figure, the total area under the curve must be 1, and the value of $\Pr(a \le X \le b)$ is equal to the area of the shaded region.

Note: Continuous Distributions Assign Probability 0 to Individual Values. The integral in Eq. (3.2.3) also equals $\Pr(a < X \le b)$ as well as $\Pr(a < X < b)$ and $\Pr(a \le X < b)$. Hence, it follows from the definition of continuous distributions that, if X has a continuous distribution, $\Pr(X = a) = 0$ for each number a. As we noted on page 20, the fact that $\Pr(X = a) = 0$ does not imply that $X = a$ is impossible. If it did,

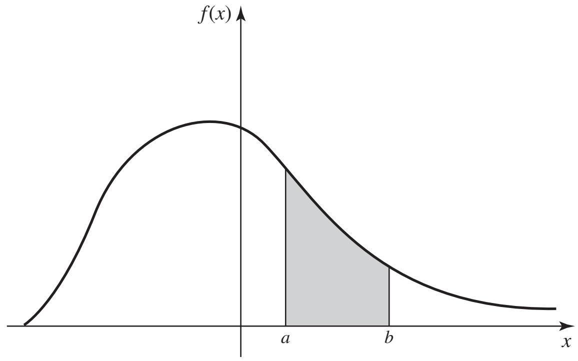


Figure 3.4 An example of a p.d.f.

all values of X would be impossible and X couldn't assume any value. What happens is that the probability in the distribution of X is spread so thinly that we can only see it on sets like nondegenerate intervals. It is much the same as the fact that lines have 0 area in two dimensions, but that does not mean that lines are not there. The two vertical lines indicated under the curve in Fig. 3.4 have 0 area, and this signifies that $\Pr(X = a) = \Pr(X = b) = 0$. However, for each $\epsilon > 0$ and each a such that $f(a) > 0$, $\Pr(a - \epsilon \le X \le a + \epsilon) \approx 2\epsilon f(a) > 0$.

Nonuniqueness of the p.d.f.

If a random variable X has a continuous distribution, then $\Pr(X = x) = 0$ for every individual value x. Because of this property, the values of each p.d.f. can be changed at a finite number of points, or even at certain infinite sequences of points, without changing the value of the integral of the p.d.f. over any subset A. In other words, the values of the p.d.f. of a random variable X can be changed arbitrarily at many points without affecting any probabilities involving X, that is, without affecting the probability distribution of X. At exactly which sets of points we can change a p.d.f. depends on subtle features of the definition of the Riemann integral. We shall not deal with this issue in this text, and we shall only contemplate changes to p.d.f.'s at finitely many points.

To the extent just described, the p.d.f. of a random variable is not unique. In many problems, however, there will be one version of the p.d.f. that is more natural than any other because for this version the p.d.f. will, wherever possible, be continuous on the real line. For example, the p.d.f. sketched in Fig. 3.4 is a continuous function over the entire real line. This p.d.f. could be changed arbitrarily at a few points without affecting the probability distribution that it represents, but these changes would introduce discontinuities into the p.d.f. without introducing any apparent advantages.

Throughout most of this book, we shall adopt the following practice: If a random variable X has a continuous distribution, we shall give only one version of the p.d.f. of X and we shall refer to that version as *the* p.d.f. of X, just as though it had been uniquely determined. It should be remembered, however, that there is some freedom in the selection of the particular version of the p.d.f. that is used to represent each continuous distribution. The most common place where such freedom will arise is in cases like Eq. (3.2.1) where the p.d.f. is required to have discontinuities. Without making the function f any less continuous, we could have defined the p.d.f. in that example so that $f(4) = f(200) = 0$ instead of $f(4) = f(200) = 1/196$. Both of these choices lead to the same calculations of all probabilities associated with X, and they are both equally valid. Because the support of a continuous distribution is the closure of the set where the p.d.f. is strictly positive, it can be shown that the support is unique. A sensible approach would then be to choose the version of the p.d.f. that was strictly positive on the support whenever possible.

The reader should note that "continuous distribution" is *not* the name of a distribution, just as "discrete distribution" is not the name of a distribution. There are many distributions that are discrete and many that are continuous. Some distributions of each type have names that we either have introduced or will introduce later.

We shall now present several examples of continuous distributions and their p.d.f.'s.

Uniform Distributions on Intervals

Example 3.2.2 Temperature Forecasts. Television weather forecasters announce high and low temperature forecasts as integer numbers of degrees. These forecasts, however, are the results of very sophisticated weather models that provide more precise forecasts that the television personalities round to the nearest integer for simplicity. Suppose that the forecaster announces a high temperature of y. If we wanted to know what temperature X the weather models actually produced, it might be safe to assume that X was equally likely to be any number in the interval from $y - 1/2$ to $y + 1/2$. ◄

The distribution of X in Example 3.2.2 is a special case of the following.

Definition 3.2.3 Uniform Distribution on an Interval. Let a and b be two given real numbers such that $a < b$. Let X be a random variable such that it is known that $a \le X \le b$ and, for every subinterval of $[a, b]$, the probability that X will belong to that subinterval is proportional to the length of that subinterval. We then say that the random variable X has the *uniform distribution on the interval* $[a, b]$.

A random variable X with the uniform distribution on the interval $[a, b]$ represents the outcome of an experiment that is often described by saying that a point is chosen *at random* from the interval $[a, b]$. In this context, the phrase "at random" means that the point is just as likely to be chosen from any particular part of the interval as from any other part of the same length.

Theorem 3.2.1 Uniform Distribution p.d.f. If X has the uniform distribution on an interval $[a, b]$, then the p.d.f. of X is

$$f(x) = \begin{cases} \dfrac{1}{b-a} & \text{for } a \le x \le b, \\ 0 & \text{otherwise.} \end{cases} \tag{3.2.6}$$

Proof X must take a value in the interval $[a, b]$. Hence, the p.d.f. $f(x)$ of X must be 0 outside of $[a, b]$. Furthermore, since any particular subinterval of $[a, b]$ having a given length is as likely to contain X as is any other subinterval having the same length, regardless of the location of the particular subinterval in $[a, b]$, it follows that $f(x)$ must be constant throughout $[a, b]$, and that interval is then the support of the distribution. Also,

$$\int_{-\infty}^{\infty} f(x)\, dx = \int_a^b f(x)\, dx = 1. \tag{3.2.7}$$

Therefore, the constant value of $f(x)$ throughout $[a, b]$ must be $1/(b-a)$, and the p.d.f. of X must be (3.2.6). ■

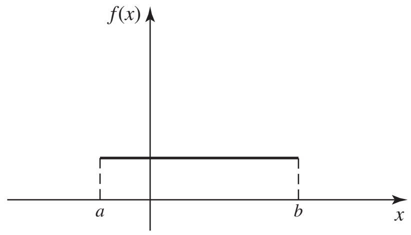


Figure 3.5 The p.d.f. for the uniform distribution on the interval $[a, b]$.

Th p.d.f. (3.2.6) is sketched in Fig. 3.5. As an example, the random variable X (demand for water) in Example 3.2.1 has the uniform distribution on the interval $[4, 200]$.

Note: Density Is Not Probability. The reader should note that the p.d.f. in (3.2.6) can be greater than 1, particularly if $b - a < 1$. Indeed, p.d.f.'s can be unbounded, as we shall see in Example 3.2.6. The p.d.f. of X, $f(x)$, itself does not equal the probability that X is near x. The integral of f over values near x gives the probability that X is near x, and the integral is never greater than 1.

It is seen from Eq. (3.2.6) that the p.d.f. representing a uniform distribution on a given interval is constant over that interval, and the constant value of the p.d.f. is the reciprocal of the length of the interval. It is not possible to define a uniform distribution over an unbounded interval, because the length of such an interval is infinite.

Consider again the uniform distribution on the interval $[a, b]$. Since the probability is 0 that one of the endpoints a or b will be chosen, it is irrelevant whether the distribution is regarded as a uniform distribution on the *closed* interval $a \le x \le b$, or as a uniform distribution on the *open* interval $a < x < b$, or as a uniform distribution on the half-open and half-closed interval $(a, b]$ in which one endpoint is included and the other endpoint is excluded.

For example, if a random variable X has the uniform distribution on the interval $[-1, 4]$, then the p.d.f. of X is

$$f(x) = \begin{cases} 1/5 & \text{for } -1 \le x \le 4, \\ 0 & \text{otherwise.} \end{cases}$$

Furthermore,

$$\Pr(0 \le X < 2) = \int_0^2 f(x)\, dx = \frac{2}{5}.$$

Notice that we defined the p.d.f. of X to be strictly positive on the closed interval $[-1, 4]$ and 0 outside of this closed interval. It would have been just as sensible to define the p.d.f. to be strictly positive on the open interval $(-1, 4)$ and 0 outside of this open interval. The probability distribution would be the same either way, including the calculation of $\Pr(0 \le X < 2)$ that we just performed. After this, when there are several equally sensible choices for how to define a p.d.f., we will simply choose one of them without making any note of the other choices.

Other Continuous Distributions

Example 3.2.3 Incompletely Specified p.d.f. Suppose that the p.d.f. of a certain random variable X has the following form:

$$f(x) = \begin{cases} cx & \text{for } 0 < x < 4, \\ 0 & \text{otherwise,} \end{cases}$$

where c is a given constant. We shall determine the value of c.

For every p.d.f., it must be true that $\int_{-\infty}^{\infty} f(x) = 1$. Therefore, in this example,

$$\int_0^4 cx\, dx = 8c = 1.$$

Hence, $c = 1/8$. ◂

Note: Calculating Normalizing Constants. The calculation in Example 3.2.3 illustrates an important point that simplifies many statistical results. The p.d.f. of X was specified without explicitly giving the value of the constant c. However, we were able to figure out what was the value of c by using the fact that the integral of a p.d.f. must be 1. It will often happen, especially in Chapter 8 where we find sampling distributions of summaries of observed data, that we can determine the p.d.f. of a random variable except for a constant factor. That constant factor must be the unique value such that the integral of the p.d.f. is 1, even if we cannot calculate it directly.

Example 3.2.4 Calculating Probabilities from a p.d.f. Suppose that the p.d.f. of X is as in Example 3.2.3, namely,

$$f(x) = \begin{cases} \dfrac{x}{8} & \text{for } 0 < x < 4, \\ 0 & \text{otherwise.} \end{cases}$$

We shall now determine the values of $\Pr(1 \le X \le 2)$ and $\Pr(X > 2)$. Apply Eq. (3.2.3) to get

$$\Pr(1 \le X \le 2) = \int_1^2 \frac{1}{8} x\, dx = \frac{3}{16}$$

and

$$\Pr(X > 2) = \int_2^4 \frac{1}{8} x\, dx = \frac{3}{4}.$$ ◂

Example 3.2.5 Unbounded Random Variables. It is often convenient and useful to represent a continuous distribution by a p.d.f. that is positive over an unbounded interval of the real line. For example, in a practical problem, the voltage X in a certain electrical system might be a random variable with a continuous distribution that can be approximately represented by the p.d.f.

$$f(x) = \begin{cases} 0 & \text{for } x \le 0, \\ \dfrac{1}{(1+x)^2} & \text{for } x > 0. \end{cases} \tag{3.2.8}$$

It can be verified that the properties (3.2.4) and (3.2.5) required of all p.d.f.'s are satisfied by $f(x)$.

Even though the voltage X may actually be bounded in the real situation, the p.d.f. (3.2.8) may provide a good approximation for the distribution of X over its full range of values. For example, suppose that it is known that the maximum possible value of X is 1000, in which case $\Pr(X > 1000) = 0$. When the p.d.f. (3.2.8) is used, we compute $\Pr(X > 1000) = 0.001$. If (3.2.8) adequately represents the variability of X over the interval (0, 1000), then it may be more convenient to use the p.d.f. (3.2.8) than a p.d.f. that is similar to (3.2.8) for $x \le 1000$, except for a new normalizing constant, and is 0 for $x > 1000$. This can be especially true if we do not know for sure that the maximum voltage is only 1000. ◂

Example 3.2.6

Unbounded p.d.f.'s. Since a value of a p.d.f. is a probability density, rather than a probability, such a value can be larger than 1. In fact, the values of the following p.d.f. are unbounded in the neighborhood of $x = 0$:

$$f(x) = \begin{cases} \frac{2}{3}x^{-1/3} & \text{for } 0 < x < 1, \\ 0 & \text{otherwise.} \end{cases} \tag{3.2.9}$$

It can be verified that even though the p.d.f. (3.2.9) is unbounded, it satisfies the properties (3.2.4) and (3.2.5) required of a p.d.f. ◂

Mixed Distributions

Most distributions that are encountered in practical problems are either discrete or continuous. We shall show, however, that it may sometimes be necessary to consider a distribution that is a mixture of a discrete distribution and a continuous distribution.

Example 3.2.7

Truncated Voltage. Suppose that in the electrical system considered in Example 3.2.5, the voltage X is to be measured by a voltmeter that will record the actual value of X if $X \le 3$ but will simply record the value 3 if $X > 3$. If we let Y denote the value recorded by the voltmeter, then the distribution of Y can be derived as follows.

First, $\Pr(Y = 3) = \Pr(X \ge 3) = 1/4$. Since the single value $Y = 3$ has probability 1/4, it follows that $\Pr(0 < Y < 3) = 3/4$. Furthermore, since $Y = X$ for $0 < X < 3$, this probability 3/4 for Y is distributed over the interval (0, 3) according to the same p.d.f. (3.2.8) as that of X over the same interval. Thus, the distribution of Y is specified by the combination of a p.d.f. over the interval (0, 3) and a positive probability at the point $Y = 3$. ◂

Summary

A continuous distribution is characterized by its probability density function (p.d.f.). A nonnegative function f is the p.d.f. of the distribution of X if, for every interval $[a, b]$, $\Pr(a \le X \le b) = \int_a^b f(x)\,dx$. Continuous random variables satisfy $\Pr(X = x) = 0$ for every value x. If the p.d.f. of a distribution is constant on an interval $[a, b]$ and is 0 off the interval, we say that the distribution is uniform on the interval $[a, b]$.

Exercises

1. Let X be a random variable with the p.d.f. specified in Example 3.2.6. Compute $\Pr(X \le 8/27)$.

2. Suppose that the p.d.f. of a random variable X is as follows:

$$f(x) = \begin{cases} \frac{4}{3}(1 - x^3) & \text{for } 0 < x < 1, \\ 0 & \text{otherwise.} \end{cases}$$

Sketch this p.d.f. and determine the values of the following probabilities: **a.** $\Pr\left(X < \frac{1}{2}\right)$ **b.** $\Pr\left(\frac{1}{4} < X < \frac{3}{4}\right)$ **c.** $\Pr\left(X > \frac{1}{3}\right)$.

3. Suppose that the p.d.f. of a random variable X is as follows:

$$f(x) = \begin{cases} \frac{1}{36}(9 - x^2) & \text{for } -3 \le x \le 3, \\ 0 & \text{otherwise.} \end{cases}$$

Sketch this p.d.f. and determine the values of the following probabilities: **a.** $\Pr(X < 0)$ **b.** $\Pr(-1 \le X \le 1)$ **c.** $\Pr(X > 2)$.

4. Suppose that the p.d.f. of a random variable X is as follows:

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

a. Find the value of the constant c and sketch the p.d.f.

b. Find the value of $\Pr(X > 3/2)$.

5. Suppose that the p.d.f. of a random variable X is as follows:

$$f(x) = \begin{cases} \frac{1}{8}x & \text{for } 0 \le x \le 4, \\ 0 & \text{otherwise.} \end{cases}$$

a. Find the value of t such that $\Pr(X \le t) = 1/4$.

b. Find the value of t such that $\Pr(X \ge t) = 1/2$.

6. Let X be a random variable for which the p.d.f. is as given in Exercise 5. After the value of X has been observed, let Y be the integer closest to X. Find the p.f. of the random variable Y.

7. Suppose that a random variable X has the uniform distribution on the interval $[-2, 8]$. Find the p.d.f. of X and the value of $\Pr(0 < X < 7)$.

8. Suppose that the p.d.f. of a random variable X is as follows:

$$f(x) = \begin{cases} ce^{-2x} & \text{for } x > 0, \\ 0 & \text{otherwise.} \end{cases}$$

a. Find the value of the constant c and sketch the p.d.f.

b. Find the value of $\Pr(1 < X < 2)$.

9. Show that there does not exist any number c such that the following function $f(x)$ would be a p.d.f.:

$$f(x) = \begin{cases} \frac{c}{1+x} & \text{for } x > 0, \\ 0 & \text{otherwise.} \end{cases}$$

10. Suppose that the p.d.f. of a random variable X is as follows:

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

a. Find the value of the constant c and sketch the p.d.f.

b. Find the value of $\Pr(X \le 1/2)$.

11. Show that there does not exist any number c such that the following function $f(x)$ would be a p.d.f.:

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

12. In Example 3.1.3 on page 94, determine the distribution of the random variable Y, the electricity demand. Also, find $\Pr(Y < 50)$.

13. An ice cream seller takes 20 gallons of ice cream in her truck each day. Let X stand for the number of gallons that she sells. The probability is 0.1 that $X = 20$. If she doesn't sell all 20 gallons, the distribution of X follows a continuous distribution with a p.d.f. of the form

$$f(x) = \begin{cases} cx & \text{for } 0 < x < 20, \\ 0 & \text{otherwise,} \end{cases}$$

where c is a constant that makes $\Pr(X < 20) = 0.9$. Find the constant c so that $\Pr(X < 20) = 0.9$ as described above.

3.3 The Cumulative Distribution Function

Although a discrete distribution is characterized by its p.f. and a continuous distribution is characterized by its p.d.f., every distribution has a common characterization through its (cumulative) distribution function (c.d.f.). The inverse of the c.d.f. is called the quantile function, and it is useful for indicating where the probability is located in a distribution.

Example 3.3.1

Voltage. Consider again the voltage X from Example 3.2.5. The distribution of X is characterized by the p.d.f. in Eq. (3.2.8). An alternative characterization that is more directly related to probabilities associated with X is obtained from the following function:

$$F(x) = \Pr(X \le x) = \int_{-\infty}^{x} f(y)dy = \begin{cases} 0 & \text{for } x \le 0, \\ \int_0^x \frac{dy}{(1+y)^2} & \text{for } x > 0, \end{cases}$$
$$= \begin{cases} 0 & \text{for } x \le 0, \\ 1 - \frac{1}{1+x} & \text{for } x > 0. \end{cases} \tag{3.3.1}$$

So, for example, $\Pr(X \le 3) = F(3) = 3/4$. ◀

Definition and Basic Properties

Definition 3.3.1 (Cumulative) Distribution Function. The *distribution function* or *cumulative distribution function* (abbreviated *c.d.f.*) F of a random variable X is the function

$$F(x) = \Pr(X \le x) \quad \text{for } -\infty < x < \infty. \tag{3.3.2}$$

It should be emphasized that the cumulative distribution function is defined as above for every random variable X, regardless of whether the distribution of X is discrete, continuous, or mixed. For the continuous random variable in Example 3.3.1, the c.d.f. was calculated in Eq. (3.3.1). Here is a discrete example:

Example 3.3.2 Bernoulli c.d.f. Let X have the Bernoulli distribution with parameter p defined in Definition 3.1.5. Then $\Pr(X = 0) = 1 - p$ and $\Pr(X = 1) = p$. Let F be the c.d.f. of X. It is easy to see that $F(x) = 0$ for $x < 0$ because $X \ge 0$ for sure. Similarly, $F(x) = 1$ for $x \ge 1$ because $X \le 1$ for sure. For $0 \le x < 1$, $\Pr(X \le x) = \Pr(X = 0) = 1 - p$ because 0 is the only possible value of X that is in the interval $(-\infty, x]$. In summary,

$$F(x) = \begin{cases} 0 & \text{for } x < 0, \\ 1 - p & \text{for } 0 \le x < 1, \\ 1 & \text{for } x \ge 1. \end{cases}$$ ◀

We shall soon see (Theorem 3.3.2) that the c.d.f. allows calculation of all interval probabilities; hence, it characterizes the distribution of a random variable. It follows from Eq. (3.3.2) that the c.d.f. of each random variable X is a function F defined on the real line. The value of F at every point x must be a number $F(x)$ in the interval $[0, 1]$ because $F(x)$ is the probability of the event $\{X \le x\}$. Furthermore, it follows from Eq. (3.3.2) that the c.d.f. of every random variable X must have the following three properties.

Property 3.3.1 Nondecreasing. *The function $F(x)$ is nondecreasing as x increases; that is, if $x_1 < x_2$, then $F(x_1) \le F(x_2)$.*

Proof If $x_1 < x_2$, then the event $\{X \le x_1\}$ is a subset of the event $\{X \le x_2\}$. Hence, $\Pr\{X \le x_1\} \le \Pr\{X \le x_2\}$ according to Theorem 1.5.4. ■

An example of a c.d.f. is sketched in Fig. 3.6. It is shown in that figure that $0 \le F(x) \le 1$ over the entire real line. Also, $F(x)$ is always nondecreasing as x increases, although $F(x)$ is constant over the interval $x_1 \le x \le x_2$ and for $x \ge x_4$.

Property 3.3.2 Limits at $\pm\infty$. $\lim_{x\to-\infty} F(x) = 0$ *and* $\lim_{x\to\infty} F(x) = 1$.

Proof As in the proof of Property 3.3.1, note that $\{X \le x_1\} \subset \{X \le x_2\}$ whenever $x_1 < x_2$. The fact that $\Pr(X \le x)$ approaches 0 as $x \to -\infty$ now follows from Exercise 13 in

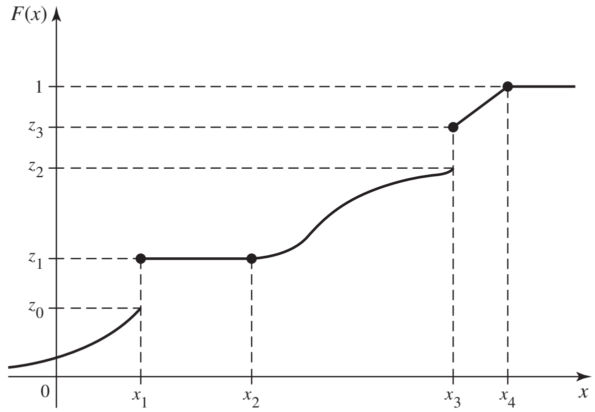


Figure 3.6 An example of a c.d.f.

Section 1.10. Similarly, the fact that $\Pr(X \le x)$ approaches 1 as $x \to \infty$ follows from Exercise 12 in Sec. 1.10. ■

The limiting values specified in Property 3.3.2 are indicated in Fig. 3.6. In this figure, the value of $F(x)$ actually becomes 1 at $x = x_4$ and then remains 1 for $x > x_4$. Hence, it may be concluded that $\Pr(X \le x_4) = 1$ and $\Pr(X > x_4) = 0$. On the other hand, according to the sketch in Fig. 3.6, the value of $F(x)$ approaches 0 as $x \to -\infty$, but does not actually become 0 at any finite point x. Therefore, for every finite value of x, no matter how small, $\Pr(X \le x) > 0$.

A c.d.f. need not be continuous. In fact, the value of $F(x)$ may jump at any finite or countable number of points. In Fig. 3.6, for instance, such jumps or points of discontinuity occur where $x = x_1$ and $x = x_3$. For each fixed value x, we shall let $F(x^-)$ denote the limit of the values of $F(y)$ as y approaches x from the left, that is, as y approaches x through values smaller than x. In symbols,

$$F(x^-) = \lim_{\substack{y \to x \\ y < x}} F(y).$$

Similarly, we shall define $F(x^+)$ as the limit of the values of $F(y)$ as y approaches x from the right. Thus,

$$F(x^+) = \lim_{\substack{y \to x \\ y > x}} F(y).$$

If the c.d.f. is continuous at a given point x, then $F(x^-) = F(x^+) = F(x)$ at that point.

Property 3.3.3 Continuity from the Right. *A c.d.f. is always continuous from the right; that is,* $F(x) = F(x^+)$ *at every point* x.

Proof Let $y_1 > y_2 > \cdots$ be a sequence of numbers that are decreasing such that $\lim_{n\to\infty} y_n = x$. Then the event $\{X \le x\}$ is the intersection of all the events $\{X \le y_n\}$ for $n = 1, 2, \ldots$. Hence, by Exercise 13 of Sec. 1.10,

$$F(x) = \Pr(X \le x) = \lim_{n\to\infty} \Pr(X \le y_n) = F(x^+). \qquad ■$$

It follows from Property 3.3.3 that at every point x at which a jump occurs,

$$F(x^+) = F(x) \text{ and } F(x^-) < F(x).$$

In Fig. 3.6 this property is illustrated by the fact that, at the points of discontinuity $x = x_1$ and $x = x_3$, the value of $F(x_1)$ is taken as z_1 and the value of $F(x_3)$ is taken as z_3.

Determining Probabilities from the Distribution Function

Example 3.3.3 Voltage. In Example 3.3.1, suppose that we want to know the probability that X lies in the interval [2, 4]. That is, we want $\Pr(2 \le X \le 4)$. The c.d.f. allows us to compute $\Pr(X \le 4)$ and $\Pr(X \le 2)$. These are related to the probability that we want as follows: Let $A = \{2 < X \le 4\}$, $B = \{X \le 2\}$, and $C = \{X \le 4\}$. Because X has a continuous distribution, $\Pr(A)$ is the same as the probability that we desire. We see that $A \cup B = C$, and it is clear that A and B are disjoint. Hence, $\Pr(A) + \Pr(B) = \Pr(C)$. It follows that

$$\Pr(A) = \Pr(C) - \Pr(B) = F(4) - F(2) = \frac{4}{5} - \frac{3}{4} = \frac{1}{20}. \quad ◀$$

The type of reasoning used in Example 3.3.3 can be extended to find the probability that an arbitrary random variable X will lie in any specified interval of the real line from the c.d.f. We shall derive this probability for four different types of intervals.

Theorem 3.3.1 For every value x,

$$\Pr(X > x) = 1 - F(x). \tag{3.3.3}$$

Proof The events $\{X > x\}$ and $\{X \le x\}$ are disjoint, and their union is the whole sample space S whose probability is 1. Hence, $\Pr(X > x) + \Pr(X \le x) = 1$. Now, Eq. (3.3.3) follows from Eq. (3.3.2). ■

Theorem 3.3.2 For all values x_1 and x_2 such that $x_1 < x_2$,

$$\Pr(x_1 < X \le x_2) = F(x_2) - F(x_1). \tag{3.3.4}$$

Proof Let $A = \{x_1 < X \le x_2\}$, $B = \{X \le x_1\}$, and $C = \{X \le x_2\}$. As in Example 3.3.3, A and B are disjoint, and their union is C, so

$$\Pr(x_1 < X \le x_2) + \Pr(X \le x_1) = \Pr(X \le x_2).$$

Subtracting $\Pr(X \le x_1)$ from both sides of this equation and applying Eq. (3.3.2) yields Eq. (3.3.4). ■

For example, if the c.d.f. of X is as sketched in Fig. 3.6, then it follows from Theorems 3.3.1 and 3.3.2 that $\Pr(X > x_2) = 1 - z_1$ and $\Pr(x_2 < X \le x_3) = z_3 - z_1$. Also, since $F(x)$ is constant over the interval $x_1 \le x \le x_2$, then $\Pr(x_1 < X \le x_2) = 0$.

It is important to distinguish carefully between the strict inequalities and the weak inequalities that appear in all of the preceding relations and also in the next theorem. If there is a jump in $F(x)$ at a given value x, then the values of $\Pr(X \le x)$ and $\Pr(X < x)$ will be different.

Theorem 3.3.3 For each value x,

$$\Pr(X < x) = F(x^-). \tag{3.3.5}$$

Proof Let $y_1 < y_2 < \cdots$ be an increasing sequence of numbers such that $\lim_{n\to\infty} y_n = x$. Then it can be shown that

$$\{X < x\} = \bigcup_{n=1}^{\infty} \{X \le y_n\}.$$

Therefore, it follows from Exercise 12 of Sec. 1.10 that

$$\begin{aligned} \Pr(X < x) &= \lim_{n\to\infty} \Pr(X \le y_n) \\ &= \lim_{n\to\infty} F(y_n) = F(x^-). \end{aligned}$$

■

For example, for the c.d.f. sketched in Fig. 3.6, $\Pr(X < x_3) = z_2$ and $\Pr(X < x_4) = 1$.

Finally, we shall show that for every value x, $\Pr(X = x)$ is equal to the amount of the jump that occurs in F at the point x. If F is continuous at the point x, that is, if there is no jump in F at x, then $\Pr(X = x) = 0$.

Theorem 3.3.4 For every value x,

$$\Pr(X = x) = F(x) - F(x^-). \tag{3.3.6}$$

Proof It is always true that $\Pr(X = x) = \Pr(X \le x) - \Pr(X < x)$. The relation (3.3.6) follows from the fact that $\Pr(X \le x) = F(x)$ at every point and from Theorem 3.3.3. ■

In Fig. 3.6, for example, $\Pr(X = x_1) = z_1 - z_0$, $\Pr(X = x_3) = z_3 - z_2$, and the probability of every other individual value of X is 0.

The c.d.f. of a Discrete Distribution

From the definition and properties of a c.d.f. $F(x)$, it follows that if $a < b$ and if $\Pr(a < X < b) = 0$, then $F(x)$ will be constant and horizontal over the interval $a < x < b$. Furthermore, as we have just seen, at every point x such that $\Pr(X = x) > 0$, the c.d.f. will jump by the amount $\Pr(X = x)$.

Suppose that X has a discrete distribution with the p.f. $f(x)$. Together, the properties of a c.d.f. imply that $F(x)$ must have the following form: $F(x)$ will have a jump of magnitude $f(x_i)$ at each possible value x_i of X, and $F(x)$ will be constant between every pair of successive jumps. The distribution of a discrete random variable X can be represented equally well by either the p.f. or the c.d.f. of X.

The c.d.f. of a Continuous Distribution

Theorem 3.3.5 Let X have a continuous distribution, and let $f(x)$ and $F(x)$ denote its p.d.f. and the c.d.f., respectively. Then F is continuous at every x,

$$F(x) = \int_{-\infty}^{x} f(t)\, dt, \tag{3.3.7}$$

and

$$\frac{dF(x)}{dx} = f(x), \tag{3.3.8}$$

at all x such that f is continuous.

Proof Since the probability of each individual point x is 0, the c.d.f. $F(x)$ will have no jumps. Hence, $F(x)$ will be a continuous function over the entire real line.

By definition, $F(x) = \Pr(X \le x)$. Since f is the p.d.f. of X, we have from the definition of p.d.f. that $\Pr(X \le x)$ is the right-hand side of Eq. (3.3.7).

It follows from Eq. (3.3.7) and the relation between integrals and derivatives (the fundamental theorem of calculus) that, for every x at which f is continuous, Eq. (3.3.8) holds. ■

Thus, the c.d.f. of a continuous random variable X can be obtained from the p.d.f. and vice versa. Eq. (3.3.7) is how we found the c.d.f. in Example 3.3.1. Notice that the derivative of the F in Example 3.3.1 is

$$F'(x) = \begin{cases} 0 & \text{for } x < 0, \\ \dfrac{1}{(1+x)^2} & \text{for } x > 0, \end{cases}$$

and F' does not exist at $x = 0$. This verifies Eq (3.3.8) for Example 3.3.1. Here, we have used the popular shorthand notation $F'(x)$ for the derivative of F at the point x.

Example 3.3.4 Calculating a p.d.f. from a c.d.f. Let the c.d.f. of a random variable be

$$F(x) = \begin{cases} 0 & \text{for } x < 0, \\ x^{2/3} & \text{for } 0 \le x \le 1, \\ 1 & \text{for } x > 1. \end{cases}$$

This function clearly satisfies the three properties required of every c.d.f., as given earlier in this section. Furthermore, since this c.d.f. is continuous over the entire real line and is differentiable at every point except $x = 0$ and $x = 1$, the distribution of X is continuous. Therefore, the p.d.f. of X can be found at every point other than $x = 0$ and $x = 1$ by the relation (3.3.8). The value of $f(x)$ at the points $x = 0$ and $x = 1$ can be assigned arbitrarily. When the derivative $F'(x)$ is calculated, it is found that $f(x)$ is as given by Eq. (3.2.9) in Example 3.2.6. Conversely, if the p.d.f. of X is given by Eq. (3.2.9), then by using Eq. (3.3.7) it is found that $F(x)$ is as given in this example. ◄

The Quantile Function

Example 3.3.5 Fair Bets. Suppose that X is the amount of rain that will fall tomorrow, and X has c.d.f. F. Suppose that we want to place an even-money bet on X as follows: If $X \le x_0$, we win one dollar and if $X > x_0$ we lose one dollar. In order to make this bet fair, we need $\Pr(X \le x_0) = \Pr(X > x_0) = 1/2$. We could search through all of the real numbers x trying to find one such that $F(x) = 1/2$, and then we would let x_0 equal the value we found. If F is a one-to-one function, then F has an inverse F^{-1} and $x_0 = F^{-1}(1/2)$. ◄

The value x_0 that we seek in Example 3.3.5 is called the 0.5 *quantile* of X or the 50th *percentile* of X because 50% of the distribution of X is at or below x_0.

Definition 3.3.2 Quantiles/Percentiles. Let X be a random variable with c.d.f. F. For each p strictly between 0 and 1, define $F^{-1}(p)$ to be the smallest value x such that $F(x) \ge p$. Then $F^{-1}(p)$ is called the p *quantile* of X or the $100p$ *percentile* of X. The function F^{-1} defined here on the open interval (0, 1) is called the *quantile function* of X.

Example 3.3.6 Standardized Test Scores. Many universities in the United States rely on standardized test scores as part of their admissions process. Thousands of people take these tests each time that they are offered. Each examinee's score is compared to the collection of scores of all examinees to see where it fits in the overall ranking. For example, if 83% of all test scores are at or below your score, your test report will say that you scored at the 83rd percentile. ◀

The notation $F^{-1}(p)$ in Definition 3.3.2 deserves some justification. Suppose first that the c.d.f. F of X is continuous and one-to-one over the whole set of possible values of X. Then the inverse F^{-1} of F exists, and for each $0 < p < 1$, there is one and only one x such that $F(x) = p$. That x is $F^{-1}(p)$. Definition 3.3.2 extends the concept of inverse function to nondecreasing functions (such as c.d.f.'s) that may be neither one-to-one nor continuous.

Quantiles of Continuous Distributions When the c.d.f. of a random variable X is continuous and one-to-one over the whole set of possible values of X, the inverse F^{-1} of F exists and equals the quantile function of X.

Example 3.3.7 Value at Risk. The manager of an investment portfolio is interested in how much money the portfolio might lose over a fixed time horizon. Let X be the change in value of the given portfolio over a period of one month. Suppose that X has the p.d.f. in Fig. 3.7. The manager computes a quantity known in the world of risk management as *Value at Risk* (denoted by VaR). To be specific, let $Y = -X$ stand for the loss incurred by the portfolio over the one month. The manager wants to have a level of confidence about how large Y might be. In this example, the manager specifies a probability level, such as 0.99 and then finds y_0, the 0.99 quantile of Y. The manager is now 99% sure that $Y \le y_0$, and y_0 is called the VaR. If X has a continuous distribution, then it is easy to see that y_0 is closely related to the 0.01 quantile of the distribution of X. The 0.01 quantile x_0 has the property that $\Pr(X < x_0) = 0.01$. But $\Pr(X < x_0) = \Pr(Y > -x_0) = 1 - \Pr(Y \le -x_0)$. Hence, $-x_0$ is a 0.99 quantile of Y. For the p.d.f. in Fig. 3.7, we see that $x_0 = -4.14$, as the shaded region indicates. Then $y_0 = 4.14$ is VaR for one month at probability level 0.99. ◀

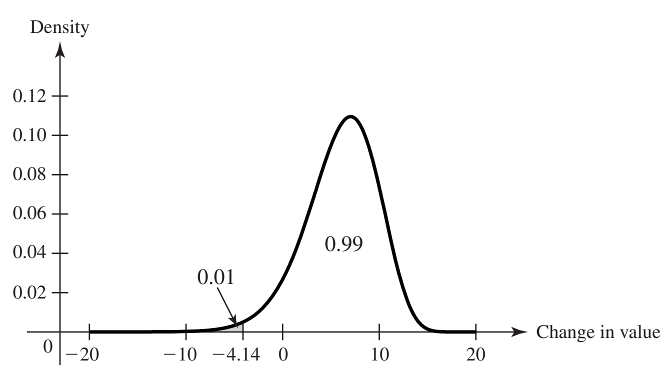


Figure 3.7 The p.d.f. of the change in value of a portfolio with lower 1% indicated.

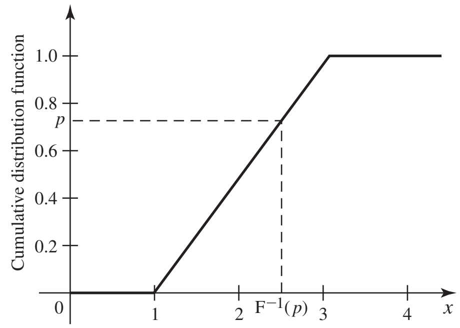


Figure 3.8 The c.d.f. of a uniform distribution indicating how to solve for a quantile.

Example 3.3.8 Uniform Distribution on an Interval. Let X have the uniform distribution on the interval $[a, b]$. The c.d.f. of X is

$$F(x) = \Pr(X \le x) = \begin{cases} 0 & \text{if } x \le a, \\ \displaystyle\int_a^x \frac{1}{b-a} du & \text{if } a < x \le b, \\ 1 & \text{if } x > b. \end{cases}$$

The integral above equals $(x-a)/(b-a)$. So, $F(x) = (x-a)/(b-a)$ for all $a < x < b$, which is a strictly increasing function over the entire interval of possible values of X. The inverse of this function is the quantile function of X, which we obtain by setting $F(x)$ equal to p and solving for x:

$$\begin{aligned} \frac{x-a}{b-a} &= p, \\ x - a &= p(b-a), \\ x &= a + p(b-a) = pb + (1-p)a. \end{aligned}$$

Figure 3.8 illustrates how the calculation of a quantile relates to the c.d.f.

The quantile function of X is $F^{-1}(p) = pb + (1-p)a$ for $0 < p < 1$. In particular, $F^{-1}(1/2) = (b+a)/2$. ◀

Note: Quantiles, Like c.d.f.'s, Depend on the Distribution Only. Any two random variables with the same distribution have the same quantile function. When we refer to a quantile of X, we mean a quantile of the distribution of X.

Quantiles of Discrete Distributions It is convenient to be able to calculate quantiles for discrete distributions as well. The quantile function of Definition 3.3.2 exists for all distributions whether discrete, continuous, or otherwise. For example, in Fig. 3.6, let $z_0 \le p \le z_1$. Then the smallest x such that $F(x) \ge p$ is x_1. For every value of $x < x_1$, we have $F(x) < z_0 \le p$ and $F(x_1) = z_1$. Notice that $F(x) = z_1$ for all x between x_1 and x_2, but since x_1 is the smallest of all those numbers, x_1 is the p quantile. Because distribution functions are continuous from the right, the smallest x such that $F(x) \ge p$ exists for all $0 < p < 1$. For $p = 1$, there is no guarantee that such an x will exist. For example, in Fig. 3.6, $F(x_4) = 1$, but in Example 3.3.1, $F(x) < 1$ for all x. For $p = 0$, there is never a smallest x such that $F(x) = 0$ because $\lim_{x\to-\infty} F(x) = 0$. That is, if $F(x_0) = 0$, then $F(x) = 0$ for all $x < x_0$. For these reasons, we never talk about the 0 or 1 quantiles.

Table 3.1 Quantile function for Example 3.3.9

p	$F^{-1}(p)$
(0, 0.1681]	0
(0.1681, 0.5283]	1
(0.5283, 0.8370]	2
(0.8370, 0.9693]	3
(0.9693, 0.9977]	4
(0.9977, 1)	5

Example 3.3.9 Quantiles of a Binomial Distribution. Let X have the binomial distribution with parameters 5 and 0.3. The binomial table in the back of the book has the p.f. f of X, which we reproduce here together with the c.d.f. F:

x	0	1	2	3	4	5
$f(x)$	0.1681	0.3602	0.3087	0.1323	0.0284	0.0024
$F(x)$	0.1681	0.5283	0.8370	0.9693	0.9977	1

(A little rounding error occurred in the p.f.) So, for example, the 0.5 quantile of this distribution is 1, which is also the 0.25 quantile and the 0.20 quantile. The entire quantile function is in Table 3.1. So, the 90th percentile is 3, which is also the 95th percentile, etc. ◀

Certain quantiles have special names.

Definition 3.3.3 Median/Quartiles. The 1/2 quantile or the 50th percentile of a distribution is called its *median*. The 1/4 quantile or 25th percentile is the *lower quartile*. The 3/4 quantile or 75th percentile is called the *upper quartile*.

Note: The Median Is Special. The median of a distribution is one of several special features that people like to use when sumarizing the distribution of a random variable. We shall discuss summaries of distributions in more detail in Chapter 4. Because the median is such a popular summary, we need to note that there are several different but similar "definitions" of median. Recall that the 1/2 quantile is the *smallest* number x such that $F(x) \geq 1/2$. For some distributions, usually discrete distributions, there will be an interval of numbers $[x_1, x_2)$ such that for all $x \in [x_1, x_2)$, $F(x) = 1/2$. In such cases, it is common to refer to all such x (including x_2) as medians of the distribution. (See Definition 4.5.1.) Another popular convention is to call $(x_1 + x_2)/2$ the median. This last is probably the most common convention. The readers should be aware that, whenever they encounter a median, it might be any one of the things that we just discussed. Fortunately, they all mean nearly the same thing, namely that the number divides the distribution in half as closely as is possible.

Example 3.3.10 Uniform Distribution on Integers. Let X have the uniform distribution on the integers 1, 2, 3, 4. (See Definition 3.1.6.) The c.d.f. of X is

$$F(x) = \begin{cases} 0 & \text{if } x < 1, \\ 1/4 & \text{if } 1 \le x < 2, \\ 1/2 & \text{if } 2 \le x < 3, \\ 3/4 & \text{if } 3 \le x < 4, \\ 1 & \text{if } x \ge 4. \end{cases}$$

The 1/2 quantile is 2, but every number in the interval [2, 3] might be called a median. The most popular choice would be 2.5. ◂

One advantage to describing a distribution by the quantile function rather than by the c.d.f. is that quantile functions are easier to display in tabular form for multiple distributions. The reason is that the domain of the quantile function is always the interval (0, 1) no matter what the possible values of X are. Quantiles are also useful for summarizing distributions in terms of where the probability is. For example, if one wishes to say where the middle half of a distribution is, one can say that it lies between the 0.25 quantile and the 0.75 quantile. In Sec. 8.5, we shall see how to use quantiles to help provide estimates of unknown quantities after observing data.

In Exercise 19, you can show how to recover the c.d.f. from the quantile function. Hence, the quantile function is an alternative way to characterize a distribution.

Summary

The c.d.f. F of a random variable X is $F(x) = \Pr(X \le x)$ for all real x. This function is continuous from the right. If we let $F(x^-)$ equal the limit of $F(y)$ as y approaches x from below, then $F(x) - F(x^-) = \Pr(X = x)$. A continuous distribution has a continuous c.d.f. and $F'(x) = f(x)$, the p.d.f. of the distribution, for all x at which F is differentiable. A discrete distribution has a c.d.f. that is constant between the possible values and jumps by $f(x)$ at each possible value x. The quantile function $F^{-1}(p)$ is equal to the smallest x such that $F(x) \ge p$ for $0 < p < 1$.

Exercises

1. Suppose that a random variable X has the Bernoulli distribution with parameter $p = 0.7$. (See Definition 3.1.5.) Sketch the c.d.f. of X.

2. Suppose that a random variable X can take only the values −2, 0, 1, and 4, and that the probabilities of these values are as follows: $\Pr(X = -2) = 0.4$, $\Pr(X = 0) = 0.1$, $\Pr(X = 1) = 0.3$, and $\Pr(X = 4) = 0.2$. Sketch the c.d.f. of X.

3. Suppose that a coin is tossed repeatedly until a head is obtained for the first time, and let X denote the number of tosses that are required. Sketch the c.d.f. of X.

4. Suppose that the c.d.f. F of a random variable X is as sketched in Fig. 3.9. Find each of the following probabilities:

a. $\Pr(X = -1)$ **b.** $\Pr(X < 0)$
c. $\Pr(X \le 0)$ **d.** $\Pr(X = 1)$
e. $\Pr(0 < X \le 3)$ **f.** $\Pr(0 < X < 3)$
g. $\Pr(0 \le X \le 3)$ **h.** $\Pr(1 < X \le 2)$
i. $\Pr(1 \le X \le 2)$ **j.** $\Pr(X > 5)$
k. $\Pr(X \ge 5)$ **l.** $\Pr(3 \le X \le 4)$

5. Suppose that the c.d.f. of a random variable X is as follows:

$$F(x) = \begin{cases} 0 & \text{for } x \le 0, \\ \frac{1}{9}x^2 & \text{for } 0 < x \le 3, \\ 1 & \text{for } x > 3. \end{cases}$$

Find and sketch the p.d.f. of X.

6. Suppose that the c.d.f. of a random variable X is as follows:

$$F(x) = \begin{cases} e^{x-3} & \text{for } x \le 3, \\ 1 & \text{for } x > 3. \end{cases}$$

Find and sketch the p.d.f. of X.

7. Suppose, as in Exercise 7 of Sec. 3.2, that a random variable X has the uniform distribution on the interval $[-2, 8]$. Find and sketch the c.d.f. of X.

8. Suppose that a point in the xy-plane is chosen at random from the interior of a circle for which the equation is $x^2 + y^2 = 1$; and suppose that the probability that the point will belong to each region inside the circle is proportional to the area of that region. Let Z denote a random variable representing the distance from the center of the circle to the point. Find and sketch the c.d.f. of Z.

9. Suppose that X has the uniform distribution on the interval $[0, 5]$ and that the random variable Y is defined by $Y = 0$ if $X \le 1$, $Y = 5$ if $X \ge 3$, and $Y = X$ otherwise. Sketch the c.d.f. of Y.

10. For the c.d.f. in Example 3.3.4, find the quantile function.

11. For the c.d.f. in Exercise 5, find the quantile function.

12. For the c.d.f. in Exercise 6, find the quantile function.

13. Suppose that a broker believes that the change in value X of a particular investment over the next two months has the uniform distribution on the interval $[-12, 24]$. Find the value at risk VaR for two months at probability level 0.95.

14. Find the quartiles and the median of the binomial distribution with parameters $n = 10$ and $p = 0.2$.

15. Suppose that X has the p.d.f.

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

Find and sketch the c.d.f. or X.

16. Find the quantile function for the distribution in Example 3.3.1.

17. Prove that the quantile function F^{-1} of a general random variable X has the following three properties that are analogous to properties of the c.d.f.:

a. F^{-1} is a nondecreasing function of p for $0 < p < 1$.

b. Let $x_0 = \lim_{\substack{p \to 0 \\ p > 0}} F^{-1}(p)$ and $x_1 = \lim_{\substack{p \to 1 \\ p < 1}} F^{-1}(p)$. Then x_0 equals the greatest lower bound on the set of numbers c such that $\Pr(X \le c) > 0$, and x_1 equals the least upper bound on the set of numbers d such that $\Pr(X \ge d) > 0$.

c. F^{-1} is continuous from the left; that is $F^{-1}(p) = F^{-1}(p^-)$ for all $0 < p < 1$.

18. Let X be a random variable with quantile function F^{-1}. Assume the following three conditions: (i) $F^{-1}(p) = c$ for all p in the interval (p_0, p_1), (ii) either $p_0 = 0$ or $F^{-1}(p_0) < c$, and (iii) either $p_1 = 1$ or $F^{-1}(p) > c$ for $p > p_1$. Prove that $\Pr(X = c) = p_1 - p_0$.

19. Let X be a random variable with c.d.f. F and quantile function F^{-1}. Let x_0 and x_1 be as defined in Exercise 17. (Note that $x_0 = -\infty$ and/or $x_1 = \infty$ are possible.) Prove that for all x in the open interval (x_0, x_1), $F(x)$ is the largest p such that $F^{-1}(p) \le x$.

20. In Exercise 13 of Sec. 3.2, draw a sketch of the c.d.f. F of X and find $F(10)$.

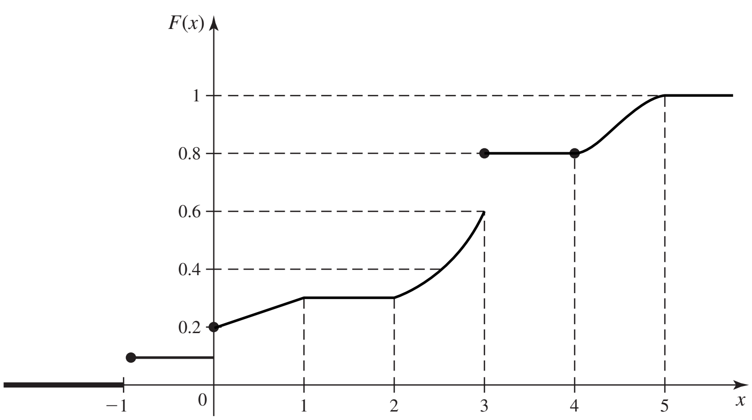


Figure 3.9 The c.d.f. for Exercise 4.

3.4 Bivariate Distributions

We generalize the concept of distribution of a random variable to the joint distribution of two random variables. In doing so, we introduce the joint p.f. for two discrete random variables, the joint p.d.f. for two continuous random variables, and the joint c.d.f. for any two random variables. We also introduce a joint hybrid of p.f. and p.d.f. for the case of one discrete random variable and one continuous random variable.

Example 3.4.1 **Demands for Utilities.** In Example 3.1.5, we found the distribution of the random variable X that represented the demand for water. But there is another random variable, Y, the demand for electricity, that is also of interest. When discussing two random variables at once, it is often convenient to put them together into an ordered pair, (X, Y). As early as Example 1.5.4 on page 19, we actually calculated some probabilities associated with the pair (X, Y). In that example, we defined two events A and B that we now can express as $A = \{X \geq 115\}$ and $B = \{Y \geq 110\}$. In Example 1.5.4, we computed $\Pr(A \cap B)$ and $\Pr(A \cup B)$. We can express $A \cap B$ and $A \cup B$ as events involving the pair (X, Y). For example, define the set of ordered pairs $C = \{(x, y) : x \geq 115 \text{ and } y \geq 110\}$ so that that the event $\{(X, Y) \in C)\} = A \cap B$. That is, the event that the pair of random variables lies in the set C is the same as the intersection of the two events A and B. In Example 1.5.4, we computed $\Pr(A \cap B) = 0.1198$. So, we can now assert that $\Pr((X, Y) \in C) = 0.1198$. ◄

Definition 3.4.1 **Joint/Bivariate Distribution.** Let X and Y be random variables. The *joint distribution* or *bivariate distribution* of X and Y is the collection of all probabilities of the form $\Pr[(X, Y) \in C]$ for all sets C of pairs of real numbers such that $\{(X, Y) \in C\}$ is an event.

It is a straightforward consequence of the definition of the joint distribution of X and Y that this joint distribution is itself a probability measure on the set of ordered pairs of real numbers. The set $\{(X, Y) \in C\}$ will be an event for every set C of pairs of real numbers that most readers will be able to imagine.

In this section and the next two sections, we shall discuss convenient ways to characterize and do computations with bivariate distributions. In Sec. 3.7, these considerations will be extended to the joint distribution of an arbitrary finite number of random variables.

Discrete Joint Distributions

Example 3.4.2 **Theater Patrons.** Suppose that a sample of 10 people is selected at random from a theater with 200 patrons. One random variable of interest might be the number X of people in the sample who are over 60 years of age, and another random variable might be the number Y of people in the sample who live more than 25 miles from the theater. For each ordered pair (x, y) with $x = 0, \dots, 10$ and $y = 0, \dots, 10$, we might wish to compute $\Pr((X, Y) = (x, y))$, the probability that there are x people in the sample who are over 60 years of age and there are y people in the sample who live more than 25 miles away. ◄

Definition 3.4.2 **Discrete Joint Distribution.** Let X and Y be random variables, and consider the ordered pair (X, Y). If there are only finitely or at most countably many different possible values (x, y) for the pair (X, Y), then we say that X and Y have a *discrete joint distribution*.

The two random variables in Example 3.4.2 have a discrete joint distribution.

Theorem 3.4.1 Suppose that two random variables X and Y each have a discrete distribution. Then X and Y have a discrete joint distribution.

Proof If both X and Y have only finitely many possible values, then there will be only a finite number of different possible values (x, y) for the pair (X, Y). On the other hand, if either X or Y or both can take a countably infinite number of possible values, then there will also be a countably infinite number of possible values for the pair (X, Y). In all of these cases, the pair (X, Y) has a discrete joint distribution. ■

When we define continuous joint distribution shortly, we shall see that the obvious analog of Theorem 3.4.1 is not true.

Definition 3.4.3 Joint Probability Function, p.f. The *joint probability function*, or the *joint p.f.*, of X and Y is defined as the function f such that for every point (x, y) in the xy-plane,

$$f(x, y) = \Pr(X = x \text{ and } Y = y).$$

The following result is easy to prove because there are at most countably many pairs (x, y) that must account for all of the probability a discrete joint distribution.

Theorem 3.4.2 Let X and Y have a discrete joint distribution. If (x, y) is not one of the possible values of the pair (X, Y), then $f(x, y) = 0$. Also,

$$\sum_{\text{All } (x,y)} f(x, y) = 1.$$

Finally, for each set C of ordered pairs,

$$\Pr[(X, Y) \in C] = \sum_{(x,y)\in C} f(x, y).$$ ■

Example 3.4.3 Specifying a Discrete Joint Distribution by a Table of Probabilities. In a certain suburban area, each household reported the number of cars and the number of television sets that they owned. Let X stand for the number of cars owned by a randomly selected household in this area. Let Y stand for the number of television sets owned by that same randomly selected household. In this case, X takes only the values 1, 2, and 3; Y takes only the values 1, 2, 3, and 4; and the joint p.f. f of X and Y is as specified in Table 3.2.

Table 3.2 Joint p.f. $f(x, y)$ for Example 3.4.3

	y			
x	1	2	3	4
1	0.1	0	0.1	0
2	0.3	0	0.1	0.2
3	0	0.2	0	0

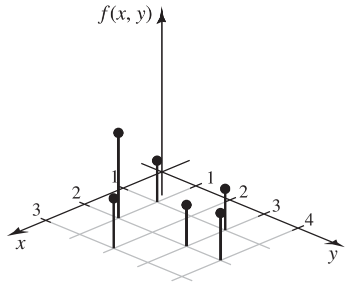


Figure 3.10 The joint p.f. of X and Y in Example 3.4.3.

This joint p.f. is sketched in Fig. 3.10. We shall determine the probability that the randomly selected household owns at least two of both cars and televisions. In symbols, this is $\Pr(X \geq 2 \text{ and } Y \geq 2)$.

By summing $f(x, y)$ over all values of $x \geq 2$ and $y \geq 2$, we obtain the value

$$\begin{aligned}\Pr(X \geq 2 \text{ and } Y \geq 2) &= f(2, 2) + f(2, 3) + f(2, 4) + f(3, 2) \\ &\quad + f(3, 3) + f(3, 4) \\ &= 0.5.\end{aligned}$$

Next, we shall determine the probability that the randomly selected household owns exactly one car, namely $\Pr(X = 1)$. By summing the probabilities in the first row of the table, we obtain the value

$$\Pr(X = 1) = \sum_{y=1}^{4} f(1, y) = 0.2. \qquad ◀$$

Continuous Joint Distributions

Example 3.4.4 **Demands for Utilities.** Consider again the joint distribution of X and Y in Example 3.4.1. When we first calculated probabilities for these two random variables back in Example 1.5.4 on page 19 (even before we named them or called them random variables), we assumed that the probability of each subset of the sample space was proportional to the area of the subset. Since the area of the sample space is 29,204, the probability that the pair (X, Y) lies in a region C is the area of C divided by 29,204. We can also write this relation as

$$\Pr((X, Y) \in C\} = \int_C \int \frac{1}{29{,}204}\, dx\, dy, \qquad (3.4.1)$$

assuming that the integral exists. ◀

If one looks carefully at Eq. (3.4.1), one will notice the similarity to Eqs. (3.2.2) and (3.2.1). We formalize this connection as follows.

Definition 3.4.4 **Continuous Joint Distribution/Joint p.d.f./Support.** Two random variables X and Y have a *continuous joint distribution* if there exists a nonnegative function f defined over the entire xy-plane such that for every subset C of the plane,

$$\Pr[(X, Y) \in C] = \int_C \int f(x, y)\, dx\, dy,$$

if the integral exists. The function f is called the *joint probability density function* (abbreviated *joint p.d.f.*) of X and Y. The closure of the set $\{(x, y) : f(x, y) > 0\}$ is called the *support of (the distribution of)* (X, Y).

Example 3.4.5

Demands for Utilities. In Example 3.4.4, it is clear from Eq. (3.4.1) that the joint p.d.f. of X and Y is the function

$$f(x, y) = \begin{cases} \dfrac{1}{29{,}204} & \text{for } 4 \le x \le 200 \text{ and } 1 \le y \le 150, \\ 0 & \text{otherwise.} \end{cases} \tag{3.4.2}$$

◀

It is clear from Definition 3.4.4 that the joint p.d.f. of two random variables characterizes their joint distribution. The following result is also straightforward.

Theorem 3.4.3

A joint p.d.f. must satisfy the following two conditions:

$$f(x, y) \ge 0 \quad \text{for } -\infty < x < \infty \text{ and } -\infty < y < \infty,$$

and

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\, dx\, dy = 1.$$

■

Any function that satisfies the two displayed formulas in Theorem 3.4.3 is the joint p.d.f. for some probability distribution.

An example of the graph of a joint p.d.f. is presented in Fig. 3.11.
The total volume beneath the surface $z = f(x, y)$ and above the xy-plane must be 1. The probability that the pair (X, Y) will belong to the rectangle C is equal to the volume of the solid figure with base A shown in Fig. 3.11. The top of this solid figure is formed by the surface $z = f(x, y)$.

In Sec. 3.5, we will show that if X and Y have a continuous joint distribution, then X and Y each have a continuous distribution when considered separately. This seems reasonable intutively. However, the converse of this statement is not true, and the following result helps to show why.

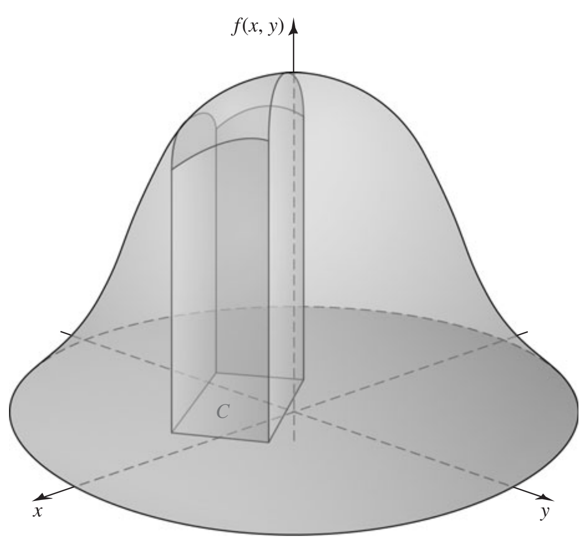


Figure 3.11 An example of a joint p.d.f.

Theorem 3.4.4 For every continuous joint distribution on the xy-plane, the following two statements hold:

i. Every individual point, and every infinite sequence of points, in the xy-plane has probability 0.
ii. Let f be a continuous function of one real variable defined on a (possibly unbounded) interval (a, b). The sets $\{(x, y) : y = f(x), a < x < b\}$ and $\{(x, y) : x = f(y), a < y < b\}$ have probability 0.

Proof According to Definition 3.4.4, the probability that a continuous joint distribution assigns to a specified region of the xy-plane can be found by integrating the joint p.d.f. $f(x, y)$ over that region, if the integral exists. If the region is a single point, the integral will be 0. By Axiom 3 of probability, the probability for any countable collection of points must also be 0. The integral of a function of two variables over the graph of a continuous function in the xy-plane is also 0. ■

Example 3.4.6 Not a Continuous Joint Distribution. It follows from (ii) of Theorem 3.4.4 that the probability that (X, Y) will lie on each specified straight line in the plane is 0. If X has a continuous distribution and if $Y = X$, then both X and Y have continuous distributions, but the probability is 1 that (X, Y) lies on the straight line $y = x$. Hence, X and Y cannot have a continuous joint distribution. ◄

Example 3.4.7 Calculating a Normalizing Constant. Suppose that the joint p.d.f. of X and Y is specified as follows:

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

We shall determine the value of the constant c.

The support S of (X, Y) is sketched in Fig. 3.12. Since $f(x, y) = 0$ outside S, it follows that

$$\begin{aligned} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)\,dx\,dy &= \int_S\int f(x, y)\,dx\,dy \\ &= \int_{-1}^{1}\int_{x^2}^{1} cx^2y\,dy\,dx = \frac{4}{21}c. \end{aligned} \tag{3.4.3}$$

Since the value of this integral must be 1, the value of c must be 21/4.

The limits of integration on the last integral in (3.4.3) were determined as follows. We have our choice of whether to integrate x or y as the inner integral, and we chose y. So, we must find, for each x, the interval of y values over which to integrate. From Fig. 3.12, we see that, for each x, y runs from the curve where $y = x^2$ to the line where $y = 1$. The interval of x values for the outer integral is from -1 to 1 according to Fig. 3.12. If we had chosen to integrate x on the inside, then for each y, we see that x runs from $-\sqrt{y}$ to $\sqrt{y}$, while y runs from 0 to 1. The final answer would have been the same. ◄

Example 3.4.8 Calculating Probabilities from a Joint p.d.f. For the joint distribution in Example 3.4.7, we shall now determine the value of $\Pr(X \ge Y)$.

The subset S_0 of S where $x \ge y$ is sketched in Fig. 3.13. Hence,

$$\Pr(X \ge Y) = \int_{S_0}\int f(x, y)\,dx\,dy = \int_0^1\int_{x^2}^{x} \frac{21}{4}x^2y\,dy\,dx = \frac{3}{20}.$$ ◄

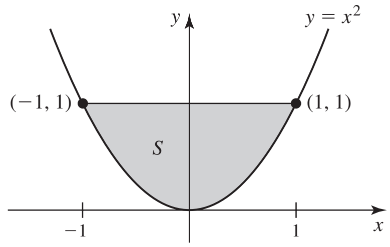


Figure 3.12 The support S of (X, Y) in Example 3.4.8.

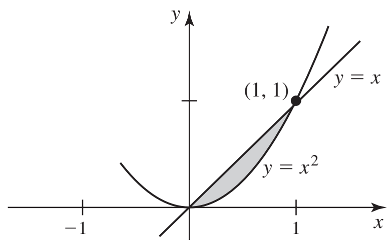


Figure 3.13 The subset S_0 of the support S where $x \geq y$ in Example 3.4.8.

Example 3.4.9

Determining a Joint p.d.f. by Geometric Methods. Suppose that a point (X, Y) is selected at random from inside the circle $x^2 + y^2 \leq 9$. We shall determine the joint p.d.f. of X and Y.

The support of (X, Y) is the set S of points on and inside the circle $x^2 + y^2 \leq 9$. The statement that the point (X, Y) is selected at random from inside the circle is interpreted to mean that the joint p.d.f. of X and Y is constant over S and is 0 outside S. Thus,

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

We must have

$$\int_S \int f(x, y)\, dx\, dy = c \times (\text{area of } S) = 1.$$

Since the area of the circle S is 9π, the value of the constant c must be $1/(9\pi)$. ◀

Mixed Bivariate Distributions

Example 3.4.10

A Clinical Trial. Consider a clinical trial (such as the one described in Example 2.1.12) in which each patient with depression receives a treatment and is followed to see whether they have a relapse into depression. Let X be the indicator of whether or not the first patient is a "success" (no relapse). That is $X = 1$ if the patient does not relapse and $X = 0$ if the patient relapses. Also, let P be the proportion of patients who have no replapse among all patients who might receive the treatment. It is clear that X must have a discrete distribution, but it might be sensible to think of P as a continuous random variable taking its value anywhere in the interval [0, 1]. Even though X and P can have neither a joint discrete distribution nor a joint continuous distribution, we can still be interested in the joint distribution of X and P. ◀

Prior to Example 3.4.10, we have discussed bivariate distributions that were either discrete or continuous. Occasionally, one must consider a mixed bivariate distribution in which one of the random variables is discrete and the other is continuous. We shall use a function $f(x, y)$ to characterize such a joint distribution in much the same way that we use a joint p.f. to characterize a discrete joint distribution or a joint p.d.f. to characterize a continuous joint distribution.

Definition 3.4.5 Joint p.f./p.d.f. Let X and Y be random variables such that X is discrete and Y is continuous. Suppose that there is a function $f(x, y)$ defined on the xy-plane such that, for every pair A and B of subsets of the real numbers,

$$\Pr(X \in A \text{ and } Y \in B) = \int_B \sum_{x \in A} f(x, y)dy, \tag{3.4.4}$$

if the integral exists. Then the function f is called the *joint p.f./p.d.f.* of X and Y.

Clearly, Definition 3.4.5 can be modified in an obvious way if Y is discrete and X is continuous. Every joint p.f./p.d.f. must satisfy two conditions. If X is the discrete random variable with possible values $x_1, x_2, \ldots$ and Y is the continuous random variable, then $f(x, y) \geq 0$ for all x, y and

$$\int_{-\infty}^{\infty} \sum_{i=1}^{\infty} f(x_i, y)dy = 1. \tag{3.4.5}$$

Because f is nonnegative, the sum and integral in Eqs. (3.4.4) and (3.4.5) can be done in whichever order is more convenient.

Note: Probabilities of More General Sets. For a general set C of pairs of real numbers, we can compute $\Pr((X, Y) \in C)$ using the joint p.f./p.d.f. of X and Y. For each x, let $C_x = \{y : (x, y) \in C\}$. Then

$$\Pr((X, Y) \in C) = \sum_{\text{All } x} \int_{C_x} f(x, y)dy,$$

if all of the integrals exist. Alternatively, for each y, define $C^y = \{x : (x, y) \in C\}$, and then

$$\Pr((X, Y) \in C) = \int_{-\infty}^{\infty} \left[\sum_{x \in C^y} f(x, y) \right] dy,$$

if the integral exists.

Example 3.4.11 A Joint p.f./p.d.f. Suppose that the joint p.f./p.d.f. of X and Y is

$$f(x, y) = \frac{xy^{x-1}}{3}, \quad \text{for } x = 1, 2, 3 \text{ and } 0 < y < 1.$$

We should check to make sure that this function satisfies (3.4.5). It is easier to integrate over the y values first, so we compute

$$\sum_{x=1}^{3} \int_0^1 \frac{xy^{x-1}}{3} dy = \sum_{x=1}^{3} \frac{1}{3} = 1.$$

Suppose that we wish to compute the probability that $Y \geq 1/2$ and $X \geq 2$. That is, we want $\Pr(X \in A \text{ and } Y \in B)$ with $A = [2, \infty)$ and $B = [1/2, \infty)$. So, we apply Eq. (3.4.4) to get the probability

$$\sum_{x=2}^{3}\int_{1/2}^{1}\frac{xy^{x-1}}{3}dy=\sum_{x=2}^{3}\left(\frac{1-(1/2)^x}{3}\right)=0.5417.$$

For illustration, we shall compute the sum and integral in the other order also. For each $y \in [1/2, 1)$, $\sum_{x=2}^{3} f(x, y) = 2y/3 + y^2$. For $y \ge 1/2$, the sum is 0. So, the probability is

$$\int_{1/2}^{1}\left[\frac{2}{3}y+y^2\right]dy=\frac{1}{3}\left[1-\left(\frac{1}{2}\right)^2\right]+\frac{1}{3}\left[1-\left(\frac{1}{2}\right)^3\right]=0.5417. \blacktriangleleft$$

Example 3.4.12

A Clinical Trial. A possible joint p.f./p.d.f. for X and P in Example 3.4.10 is

$$f(x, p) = p^x(1-p)^{1-x}, \quad \text{for } x = 0, 1 \text{ and } 0 < p < 1.$$

Here, X is discrete and P is continuous. The function f is nonnegative, and the reader should be able to demonstrate that it satisfies (3.4.5). Suppose that we wish to compute $\Pr(X \le 0 \text{ and } P \le 1/2)$. This can be computed as

$$\int_0^{1/2}(1-p)dp=-\frac{1}{2}[(1-1/2)^2-(1-0)^2]=\frac{3}{8}.$$

Suppose that we also wish to compute $\Pr(X = 1)$. This time, we apply Eq. (3.4.4) with $A = \{1\}$ and $B = (0, 1)$. In this case,

$$\Pr(X=1)=\int_0^1 p\,dp=\frac{1}{2}. \blacktriangleleft$$

A more complicated type of joint distribution can also arise in a practical problem.

Example 3.4.13

A Complicated Joint Distribution. Suppose that X and Y are the times at which two specific components in an electronic system fail. There might be a certain probability p $(0 < p < 1)$ that the two components will fail at the same time and a certain probability $1 - p$ that they will fail at different times. Furthermore, if they fail at the same time, then their common failure time might be distributed according to a certain p.d.f. $f(x)$; if they fail at different times, then these times might be distributed according to a certain joint p.d.f. $g(x, y)$.

The joint distribution of X and Y in this example is not continuous, because there is positive probability p that (X, Y) will lie on the line $x = y$. Nor does the joint distribution have a joint p.f./p.d.f. or any other simple function to describe it. There are ways to deal with such joint distributions, but we shall not discuss them in this text. ◀

Bivariate Cumulative Distribution Functions

The first calculation in Example 3.4.12, namely, $\Pr(X \le 0 \text{ and } Y \le 1/2)$, is a generalization of the calculation of a c.d.f. to a bivariate distribution. We formalize the generalization as follows.

Definition 3.4.6

Joint (Cumulative) Distribution Function/c.d.f. The *joint distribution function* or *joint cumulative distribution function* (*joint c.d.f.*) of two random variables X and Y is

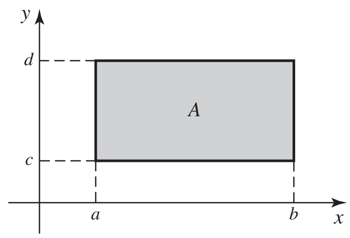


Figure 3.14 The probability of a rectangle.

defined as the function F such that for all values of x and y ($-\infty < x < \infty$ and $-\infty < y < \infty$),

$$F(x, y) = \Pr(X \le x \text{ and } Y \le y).$$

It is clear from Definition 3.4.6 that $F(x, y)$ is monotone increasing in x for each fixed y and is monotone increasing in y for each fixed x.

If the joint c.d.f. of two arbitrary random variables X and Y is F, then the probability that the pair (X, Y) will lie in a specified rectangle in the xy-plane can be found from F as follows: For given numbers $a < b$ and $c < d$,

$$\begin{aligned}
&\Pr(a < X \le b \text{ and } c < Y \le d) \\
&\quad = \Pr(a < X \le b \text{ and } Y \le d) - \Pr(a < X \le b \text{ and } Y \le c) \\
&\quad = [\Pr(X \le b \text{ and } Y \le d) - \Pr(X \le a \text{ and } Y \le d)] \\
&\qquad - [\Pr(X \le b \text{ and } Y \le c) - \Pr(X \le a \text{ and } Y \le c)] \\
&\quad = F(b, d) - F(a, d) - F(b, c) + F(a, c).
\end{aligned} \tag{3.4.6}$$

Hence, the probability of the rectangle C sketched in Fig. 3.14 is given by the combination of values of F just derived. It should be noted that two sides of the rectangle are included in the set C and the other two sides are excluded. Thus, if there are points or line segments on the boundary of C that have positive probability, it is important to distinguish between the weak inequalities and the strict inequalities in Eq. (3.4.6).

Theorem 3.4.5 Let X and Y have a joint c.d.f. F. The c.d.f. F_1 of just the single random variable X can be derived from the joint c.d.f. F as $F_1(x) = \lim_{y\to\infty} F(x, y)$. Similarly, the c.d.f. F_2 of Y equals $F_2(y) = \lim_{x\to\infty} F(x, y)$, for $0 < y < \infty$.

Proof We prove the claim about F_1 as the claim about F_2 is similar. Let $-\infty < x < \infty$. Define

$$\begin{aligned}
B_0 &= \{X \le x \text{ and } Y \le 0\}, \\
B_n &= \{X \le x \text{ and } n - 1 < Y \le n\}, \quad \text{for } n = 1, 2, \ldots, \\
A_m &= \bigcup_{n=0}^{m} B_n, \quad \text{for } m = 1, 2, \ldots .
\end{aligned}$$

Then $\{X \le x\} = \bigcup_{n=-0}^{\infty} B_n$, and $A_m = \{X \le x \text{ and } Y \le m\}$ for $m = 1, 2, \ldots$. It follows that $\Pr(A_m) = F(x, m)$ for each m. Also,

$$
\begin{aligned}
F_1(x) = \Pr(X \le x) &= \Pr\left(\bigcup_{n=1}^{\infty} B_n\right) \\
&= \sum_{n=0}^{\infty} \Pr(B_n) = \lim_{m\to\infty} \Pr(A_m) \\
&= \lim_{m\to\infty} F(x, m) = \lim_{y\to\infty} F(x, y),
\end{aligned}
$$

where the third equality follows from countable additivity and the fact that the B_n events are disjoint, and the last equality follows from the fact that $F(x, y)$ is monotone increasing in y for each fixed x. ■

Other relationships involving the univariate distribution of X, the univariate distribution of Y, and their joint bivariate distribution will be presented in the next section.

Finally, if X and Y have a continuous joint distribution with joint p.d.f. f, then the joint c.d.f. at (x, y) is

$$F(x, y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f(r, s)\, dr\, ds.$$

Here, the symbols r and s are used simply as dummy variables of integration. The joint p.d.f. can be derived from the joint c.d.f. by using the relations

$$f(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y} = \frac{\partial^2 F(x, y)}{\partial y \partial x}$$

at every point (x, y) at which these second-order derivatives exist.

Example 3.4.14 **Determining a Joint p.d.f. from a Joint c.d.f.** Suppose that X and Y are random variables that take values only in the intervals $0 \le X \le 2$ and $0 \le Y \le 2$. Suppose also that the joint c.d.f. of X and Y, for $0 \le x \le 2$ and $0 \le y \le 2$, is as follows:

$$F(x, y) = \frac{1}{16} xy(x + y). \tag{3.4.7}$$

We shall first determine the c.d.f. F_1 of just the random variable X and then determine the joint p.d.f. f of X and Y.

The value of $F(x, y)$ at any point (x, y) in the xy-plane that does not represent a pair of possible values of X and Y can be calculated from (3.4.7) and the fact that $F(x, y) = \Pr(X \le x \text{ and } Y \le y)$. Thus, if either $x < 0$ or $y < 0$, then $F(x, y) = 0$. If both $x > 2$ and $y > 2$, then $F(x, y) = 1$. If $0 \le x \le 2$ and $y > 2$, then $F(x, y) = F(x, 2)$, and it follows from Eq. (3.4.7) that

$$F(x, y) = \frac{1}{8} x(x + 2).$$

Similarly, if $0 \le y \le 2$ and $x > 2$, then

$$F(x, y) = \frac{1}{8} y(y + 2).$$

The function $F(x, y)$ has now been specified for every point in the xy-plane.

By letting $y \to \infty$, we find that the c.d.f. of just the random variable X is

$$
F_1(x) = \begin{cases} 0 & \text{for } x < 0, \\ \frac{1}{8} x(x + 2) & \text{for } 0 \le x \le 2, \\ 1 & \text{for } x > 2. \end{cases}
$$

Furthermore, for $0 < x < 2$ and $0 < y < 2$,

$$\frac{\partial^2 F(x, y)}{\partial x \partial y} = \frac{1}{8}(x + y).$$

Also, if $x < 0$, $y < 0$, $x > 2$, or $y > 2$, then

$$\frac{\partial^2 F(x, y)}{\partial x \partial y} = 0.$$

Hence, the joint p.d.f. of X and Y is

$$f(x, y) = \begin{cases} \frac{1}{8}(x + y) & \text{for } 0 < x < 2 \text{ and } 0 < y < 2, \\ 0 & \text{otherwise.} \end{cases}$$ ◀

Example 3.4.15 **Demands for Utilities.** We can compute the joint c.d.f. for water and electric demand in Example 3.4.4 by using the joint p.d.f. that was given in Eq. (3.4.2). If either $x \le 4$ or $y \le 1$, then $F(x, y) = 0$ because either $X \le x$ or $Y \le y$ would be impossible. Similarly, if both $x \ge 200$ and $y \ge 150$, $F(x, y) = 1$ because both $X \le x$ and $Y \le y$ would be sure events. For other values of x and y, we compute

$$F(x, y) = \begin{cases} \displaystyle\int_4^x \int_1^y \frac{1}{29{,}204} dy dx = \frac{xy}{29{,}204} & \text{for } 4 \le x \le 200,\ 1 \le y \le 150, \\ \displaystyle\int_4^x \int_1^{150} \frac{1}{29{,}204} dy dx = \frac{x}{196} & \text{for } 4 \le x \le 200,\ y > 150, \\ \displaystyle\int_4^{200} \int_1^y \frac{1}{29{,}204} dy dx = \frac{y}{149} & \text{for } x > 200,\ 1 \le y \le 150. \end{cases}$$

The reason that we need three cases in the formula for $F(x, y)$ is that the joint p.d.f. in Eq. (3.4.2) drops to 0 when x crosses above 200 or when y crosses above 150; hence, we never want to integrate 1/29,204 beyond $x = 200$ or beyond $y = 150$. If one takes the limit as $y \to \infty$ of $F(x, y)$ (for fixed $4 \le x \le 200$), one gets the second case in the formula above, which then is the c.d.f. of X, $F_1(x)$. Similarly, if one takes the $\lim_{x\to\infty} F(x, y)$ (for fixed $1 \le y \le 150$), one gets the third case in the formula, which then is the c.d.f. of Y, $F_2(y)$. ◀

Summary

The joint c.d.f. of two random variables X and Y is $F(x, y) = \Pr(X \le x \text{ and } Y \le y)$. The joint p.d.f. of two continuous random variables is a nonnegative function f such that the probability of the pair (X, Y) being in a set C is the integral of $f(x, y)$ over the set C, if the integral exists. The joint p.d.f. is also the second mixed partial derivative of the joint c.d.f. with respect to both variables. The joint p.f. of two discrete random variables is a nonnegative function f such that the probability of the pair (X, Y) being in a set C is the sum of $f(x, y)$ over all points in C. A joint p.f. can be strictly positive at countably many pairs (x, y) at most. The joint p.f./p.d.f. of a discrete random variable X and a continuous random variable Y is a nonnegative function f such that the probability of the pair (X, Y) being in a set C is obtained by summing $f(x, y)$ over all x such that $(x, y) \in C$ for each y and then integrating the resulting function of y.

Exercises

1. Suppose that the joint p.d.f. of a pair of random variables (X, Y) is constant on the rectangle where $0 \le x \le 2$ and $0 \le y \le 1$, and suppose that the p.d.f. is 0 off of this rectangle.

a. Find the constant value of the p.d.f. on the rectangle.

b. Find $\Pr(X \ge Y)$.

2. Suppose that in an electric display sign there are three light bulbs in the first row and four light bulbs in the second row. Let X denote the number of bulbs in the first row that will be burned out at a specified time t, and let Y denote the number of bulbs in the second row that will be burned out at the same time t. Suppose that the joint p.f. of X and Y is as specified in the following table:

	Y				
X	0	1	2	3	4
0	0.08	0.07	0.06	0.01	0.01
1	0.06	0.10	0.12	0.05	0.02
2	0.05	0.06	0.09	0.04	0.03
3	0.02	0.03	0.03	0.03	0.04

Determine each of the following probabilities:

a. $\Pr(X = 2)$ **b.** $\Pr(Y \ge 2)$
c. $\Pr(X \le 2 \text{ and } Y \le 2)$ **d.** $\Pr(X = Y)$
e. $\Pr(X > Y)$

3. Suppose that X and Y have a discrete joint distribution for which the joint p.f. is defined as follows:

$$f(x, y) = \begin{cases} c|x + y| & \text{for } x = -2, -1, 0, 1, 2 \text{ and} \\ & \quad y = -2, -1, 0, 1, 2, \\ 0 & \text{otherwise.} \end{cases}$$

Determine **(a)** the value of the constant c; **(b)** $\Pr(X = 0 \text{ and } Y = -2)$; **(c)** $\Pr(X = 1)$; **(d)** $\Pr(|X - Y| \le 1)$.

4. Suppose that X and Y have a continuous joint distribution for which the joint p.d.f. is defined as follows:

$$f(x, y) = \begin{cases} cy^2 & \text{for } 0 \le x \le 2 \text{ and } 0 \le y \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

Determine **(a)** the value of the constant c; **(b)** $\Pr(X + Y > 2)$; **(c)** $\Pr(Y < 1/2)$; **(d)** $\Pr(X \le 1)$; **(e)** $\Pr(X = 3Y)$.

5. Suppose that the joint p.d.f. of two random variables X and Y is as follows:

$$f(x, y) = \begin{cases} c(x^2 + y) & \text{for } 0 \le y \le 1 - x^2, \\ 0 & \text{otherwise.} \end{cases}$$

Determine **(a)** the value of the constant c; **(b)** $\Pr(0 \le X \le 1/2)$; **(c)** $\Pr(Y \le X + 1)$; **(d)** $\Pr(Y = X^2)$.

6. Suppose that a point (X, Y) is chosen at random from the region S in the xy-plane containing all points (x, y) such that $x \ge 0$, $y \ge 0$, and $4y + x \le 4$.

a. Determine the joint p.d.f. of X and Y.

b. Suppose that S_0 is a subset of the region S having area α and determine $\Pr[(X, Y) \in S_0]$.

7. Suppose that a point (X, Y) is to be chosen from the square S in the xy-plane containing all points (x, y) such that $0 \le x \le 1$ and $0 \le y \le 1$. Suppose that the probability that the chosen point will be the corner (0, 0) is 0.1, the probability that it will be the corner (1, 0) is 0.2, the probability that it will be the corner (0, 1) is 0.4, and the probability that it will be the corner (1, 1) is 0.1. Suppose also that if the chosen point is not one of the four corners of the square, then it will be an interior point of the square and will be chosen according to a constant p.d.f. over the interior of the square. Determine **(a)** $\Pr(X \le 1/4)$ and **(b)** $\Pr(X + Y \le 1)$.

8. Suppose that X and Y are random variables such that (X, Y) must belong to the rectangle in the xy-plane containing all points (x, y) for which $0 \le x \le 3$ and $0 \le y \le 4$. Suppose also that the joint c.d.f. of X and Y at every point (x, y) in this rectangle is specified as follows:

$$F(x, y) = \frac{1}{156} xy(x^2 + y).$$

Determine **(a)** $\Pr(1 \le X \le 2 \text{ and } 1 \le Y \le 2)$; **(b)** $\Pr(2 \le X \le 4 \text{ and } 2 \le Y \le 4)$; **(c)** the c.d.f. of Y; **(d)** the joint p.d.f. of X and Y; **(e)** $\Pr(Y \le X)$.

9. In Example 3.4.5, compute the probability that water demand X is greater than electric demand Y.

10. Let Y be the rate (calls per hour) at which calls arrive at a switchboard. Let X be the number of calls during a two-hour period. A popular choice of joint p.f./p.d.f. for (X, Y) in this example would be one like

$$f(x, y) = \begin{cases} \frac{(2y)^x}{x!} e^{-3y} & \text{if } y > 0 \text{ and } x = 0, 1, \ldots, \\ 0 & \text{otherwise.} \end{cases}$$

a. Verify that f is a joint p.f./p.d.f. *Hint:* First, sum over the x values using the well-known formula for the power series expansion of e^{2y}.

b. Find $\Pr(X = 0)$.

11. Consider the clinical trial of depression drugs in Example 2.1.4. Suppose that a patient is selected at random from the 150 patients in that study and we record Y, an

Table 3.3 Proportions in clinical depression study for Exercise 11

	Treatment group (Y)			
Response (X)	Imipramine (1)	Lithium (2)	Combination (3)	Placebo (4)
Relapse (0)	0.120	0.087	0.146	0.160
No relapse (1)	0.147	0.166	0.107	0.067

indicator of the treatment group for that patient, and X, an indicator of whether or not the patient relapsed. Table 3.3 contains the joint p.f. of X and Y.

a. Calculate the probability that a patient selected at random from this study used Lithium (either alone or in combination with Imipramine) and did not relapse.

b. Calculate the probability that the patient had a relapse (without regard to the treatment group).

3.5 Marginal Distributions

Earlier in this chapter, we introduced distributions for random variables, and in Sec. 3.4 we discussed a generalization to joint distributions of two random variables simultaneously. Often, we start with a joint distribution of two random variables and we then want to find the distribution of just one of them. The distribution of one random variable X computed from a joint distribution is also called the marginal distribution of X. Each random variable will have a marginal c.d.f. as well as a marginal p.d.f. or p.f. We also introduce the concept of independent random variables, which is a natural generalization of independent events.

Deriving a Marginal p.f. or a Marginal p.d.f.

We have seen in Theorem 3.4.5 that if the joint c.d.f. F of two random variables X and Y is known, then the c.d.f. F_1 of the random variable X can be derived from F. We saw an example of this derivation in Example 3.4.15. If X has a continuous distribution, we can also derive the p.d.f. of X from the joint distribution.

Example 3.5.1 **Demands for Utilities.** Look carefully at the formula for $F(x, y)$ in Example 3.4.15, specifically the last two branches that we identified as $F_1(x)$ and $F_2(y)$, the c.d.f.'s of the two individual random variables X and Y. It is apparent from those two formulas and Theorem 3.3.5 that the p.d.f. of X alone is

$$f_1(x) = \begin{cases} \dfrac{1}{196} & \text{for } 4 \le x \le 200, \\ 0 & \text{otherwise,} \end{cases}$$

which matches what we already found in Example 3.2.1. Similarly, the p.d.f. of Y alone is

$$f_2(y) = \begin{cases} \dfrac{1}{149} & \text{for } 1 \le y \le 150, \\ 0 & \text{otherwise.} \end{cases}$$ ◀

The ideas employed in Example 3.5.1 lead to the following definition.

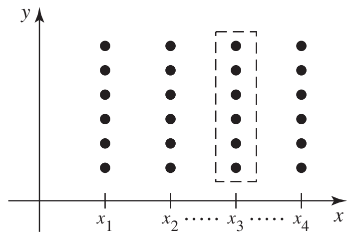


Figure 3.15 Computing $f_1(x)$ from the joint p.f.

Definition 3.5.1 Marginal c.d.f./p.f./p.d.f. Suppose that X and Y have a joint distribution. The c.d.f. of X derived by Theorem 3.4.5 is called the *marginal c.d.f.* of X. Similarly, the p.f. or p.d.f. of X associated with the marginal c.d.f. of X is called the *marginal p.f.* or *marginal p.d.f.* of X.

To obtain a specific formula for the marginal p.f. or marginal p.d.f., we start with a discrete joint distribution.

Theorem 3.5.1 If X and Y have a discrete joint distribution for which the joint p.f. is f, then the marginal p.f. f_1 of X is

$$f_1(x) = \sum_{\text{All } y} f(x, y). \tag{3.5.1}$$

Similarly, the marginal p.f. f_2 of Y is $f_2(y) = \sum_{\text{All } x} f(x, y)$.

Proof We prove the result for f_1, as the proof for f_2 is similar. We illustrate the proof in Fig. 3.15. In that figure, the set of points in the dashed box is the set of pairs with first coordinate x. The event $\{X = x\}$ can be expressed as the union of the events represented by the pairs in the dashed box, namely, $B_y = \{X = x \text{ and } Y = y\}$ for all possible y. The B_y events are disjoint and $\Pr(B_y) = f(x, y)$. Since $\Pr(X = x) = \sum_{\text{All } y} \Pr(B_y)$, Eq. (3.5.1) holds. ■

Example 3.5.2 Deriving a Marginal p.f. from a Table of Probabilities. Suppose that X and Y are the random variables in Example 3.4.3 on page 119. These are respectively the numbers of cars and televisions owned by a radomly selected household in a certain suburban area. Table 3.2 on page 119 gives their joint p.f., and we repeat that table in Table 3.4 together with row and column totals added to the margins.

The marginal p.f. f_1 of X can be read from the row totals of Table 3.4. The numbers were obtained by summing the values in each row of this table from the four columns in the central part of the table (those labeled $y = 1, 2, 3, 4$). In this way, it is found that $f_1(1) = 0.2$, $f_1(2) = 0.6$, $f_1(3) = 0.2$, and $f_1(x) = 0$ for all other values of x. This marginal p.f. gives the probabilities that a randomly selected household owns 1, 2, or 3 cars. Similarly, the marginal p.f. f_2 of Y, the probabilities that a household owns 1, 2, 3, or 4 televisions, can be read from the column totals. These numbers were obtained by adding the numbers in each of the columns from the three rows in the central part of the table (those labeled $x = 1, 2, 3$.) ◀

The name *marginal distribution* derives from the fact that the marginal distributions are the totals that appear in the margins of tables like Table 3.4.

If X and Y have a continuous joint distribution for which the joint p.d.f. is f, then the marginal p.d.f. f_1 of X is again determined in the manner shown in Eq. (3.5.1), but

Table 3.4 Joint p.f. $f(x, y)$ with marginal p.f.'s for Example 3.5.2

x	y: 1	2	3	4	Total
1	0.1	0	0.1	0	0.2
2	0.3	0	0.1	0.2	0.6
3	0	0.2	0	0	0.2
Total	0.4	0.2	0.2	0.2	1.0

the sum over all possible values of Y is now replaced by the integral over all possible values of Y.

Theorem 3.5.2 If X and Y have a continuous joint distribution with joint p.d.f. f, then the marginal p.d.f. f_1 of X is

$$f_1(x) = \int_{-\infty}^{\infty} f(x, y)\, dy \quad \text{for } -\infty < x < \infty. \tag{3.5.2}$$

Similarly, the marginal p.d.f. f_2 of Y is

$$f_2(y) = \int_{-\infty}^{\infty} f(x, y)\, dx \quad \text{for } -\infty < y < \infty. \tag{3.5.3}$$

Proof We prove (3.5.2) as the proof of (3.5.3) is similar. For each x, $\Pr(X \le x)$ can be written as $\Pr((X, Y) \in C)$, where $C = \{(r, s)) : r \le x\}$. We can compute this probability directly from the joint p.d.f. of X and Y as

$$\begin{aligned} \Pr((X, Y) \in C) &= \int_{-\infty}^{x} \int_{-\infty}^{\infty} f(r, s) ds dr \\ &= \int_{-\infty}^{x} \left[\int_{-\infty}^{\infty} f(r, s) ds \right] dr \end{aligned} \tag{3.5.4}$$

The inner integral in the last expression of Eq. (3.5.4) is a function of r and it can easily be recognized as $f_1(r)$, where f_1 is defined in Eq. (3.5.2). It follows that $\Pr(X \le x) = \int_{-\infty}^{x} f_1(r) dr$, so f_1 is the marginal p.d.f. of X. ■

Example 3.5.3 Deriving a Marginal p.d.f. Suppose that the joint p.d.f. of X and Y is as specified in Example 3.4.8, namely,

$$f(x, y) = \begin{cases} \frac{21}{4} x^2 y & \text{for } x^2 \le y \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

The set S of points (x, y) for which $f(x, y) > 0$ is sketched in Fig. 3.16. We shall determine first the marginal p.d.f. f_1 of X and then the marginal p.d.f. f_2 of Y.

It can be seen from Fig. 3.16 that X cannot take any value outside the interval $[-1, 1]$. Therefore, $f_1(x) = 0$ for $x < -1$ or $x > 1$. Furthermore, for $-1 \le x \le 1$, it is seen from Fig. 3.16 that $f(x, y) = 0$ unless $x^2 \le y \le 1$. Therefore, for $-1 \le x \le 1$,

$$f_1(x) = \int_{-\infty}^{\infty} f(x, y)\, dy = \int_{x^2}^{1} \left(\frac{21}{4}\right) x^2 y\, dy = \left(\frac{21}{8}\right) x^2 (1 - x^4).$$

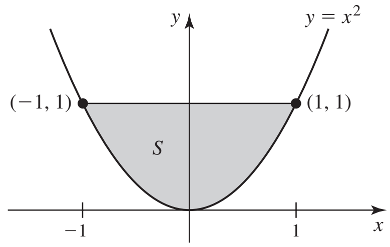


Figure 3.16 The set S where $f(x, y) > 0$ in Example 3.5.3.

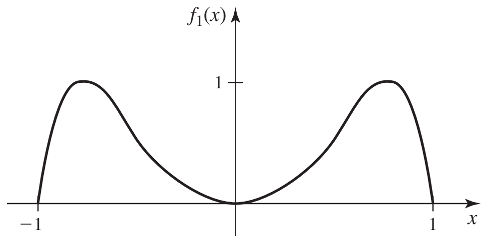


Figure 3.17 The marginal p.d.f. of X in Example 3.5.3.

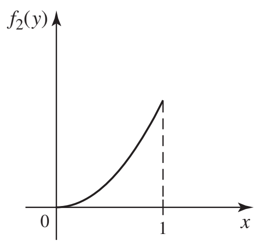


Figure 3.18 The marginal p.d.f. of Y in Example 3.5.3.

This marginal p.d.f. of X is sketched in Fig. 3.17.

Next, it can be seen from Fig. 3.16 that Y cannot take any value outside the interval [0, 1]. Therefore, $f_2(y) = 0$ for $y < 0$ or $y > 1$. Furthermore, for $0 \le y \le 1$, it is seen from Fig. 3.12 that $f(x, y) = 0$ unless $-\sqrt{y} \le x \le \sqrt{y}$. Therefore, for $0 \le y \le 1$,

$$f_2(y) = \int_{-\infty}^{\infty} f(x, y)\,dx = \int_{-\sqrt{y}}^{\sqrt{y}} \left(\frac{21}{4}\right) x^2 y\,dx = \left(\frac{7}{2}\right) y^{5/2}.$$

This marginal p.d.f. of Y is sketched in Fig. 3.18. ◂

If X has a discrete distribution and Y has a continuous distribution, we can derive the marginal p.f. of X and the marginal p.d.f. of Y from the joint p.f./p.d.f. in the same ways that we derived a marginal p.f. or a marginal p.d.f. from a joint p.f. or a joint p.d.f. The following result can be proven by combining the techniques used in the proofs of Theorems 3.5.1 and 3.5.2.

Theorem 3.5.3 Let f be the joint p.f./p.d.f. of X and Y, with X discrete and Y continuous. Then the marginal p.f. of X is

$$f_1(x) = \Pr(X = x) = \int_{-\infty}^{\infty} f(x, y)\,dy, \quad \text{for all } x,$$

and the marginal p.d.f. of Y is

$$f_2(y) = \sum_x f(x, y), \quad \text{for } -\infty < y < \infty. \qquad \blacksquare$$

Example 3.5.4

Determining a Marginal p.f. and Marginal p.d.f. from a Joint p.f./p.d.f. Suppose that the joint p.f./p.d.f. of X and Y is as in Example 3.4.11 on page 124. The marginal p.f. of X is obtained by integrating

$$f_1(x) = \int_0^1 \frac{xy^{x-1}}{3} dy = \frac{1}{3},$$

for $x = 1, 2, 3$. The marginal p.d.f. of Y is obtained by summing

$$f_2(y) = \frac{1}{3} + \frac{2y}{3} + y^2, \quad \text{for } 0 < y < 1. \qquad \blacktriangleleft$$

Although the marginal distributions of X and Y can be derived from their joint distribution, it is not possible to reconstruct the joint distribution of X and Y from their marginal distributions without additional information. For instance, the marginal p.d.f.'s sketched in Figs. 3.17 and 3.18 reveal no information about the relationship between X and Y. In fact, by definition, the marginal distribution of X specifies probabilities for X without regard for the values of any other random variables. This property of a marginal p.d.f. can be further illustrated by another example.

Example 3.5.5

Marginal and Joint Distributions. Suppose that a penny and a nickel are each tossed n times so that every pair of sequences of tosses (n tosses in each sequence) is equally likely to occur. Consider the following two definitions of X and Y: (i) X is the number of heads obtained with the penny, and Y is the number of heads obtained with the nickel. (ii) Both X and Y are the number of heads obtained with the penny, so the random variables X and Y are actually identical.

In case (i), the marginal distribution of X and the marginal distribution of Y will be identical binomial distributions. The same pair of marginal distributions of X and Y will also be obtained in case (ii). However, the joint distribution of X and Y will not be the same in the two cases. In case (i), X and Y can take different values. Their joint p.f. is

$$f(x, y) = \begin{cases} \binom{n}{x}\binom{n}{y}\left(\frac{1}{2}\right)^{x+y} & \text{for } x = 0, 1 \ldots, n, \ y = 0, 1, \ldots, n, \\ 0 & \text{otherwise.} \end{cases}$$

In case (ii), X and Y must take the same value, and their joint p.f. is

$$f(x, y) = \begin{cases} \binom{n}{x}\left(\frac{1}{2}\right)^{x} & \text{for } x = y = 0, 1 \ldots, n, \\ 0 & \text{otherwise.} \end{cases} \qquad \blacktriangleleft$$

Independent Random Variables

Example 3.5.6

Demands for Utilities. In Examples 3.4.15 and 3.5.1, we found the marginal c.d.f.'s of water and electric demand were, respectively,

$$F_1(x) = \begin{cases} 0 & \text{for } x < 4, \\ \dfrac{x}{196} & \text{for } 4 \le x \le 200, \\ 1 & \text{for } x > 200, \end{cases} \qquad F_2(y) = \begin{cases} 0 & \text{for } y < 1, \\ \dfrac{y}{149} & \text{for } 1 \le y \le 150, \\ 1 & \text{for } y > 150. \end{cases}$$

The product of these two functions is precisely the same as the joint c.d.f. of X and Y given in Example 3.5.1. One consequence of this fact is that, for every x and y, $\Pr(X \le x,\ \text{and } Y \le y) = \Pr(X \le x)\Pr(Y \le y)$. This equation makes X and Y an example of the next definition. ◀

Definition 3.5.2 Independent Random Variables. It is said that two random variables X and Y are *independent* if, for every two sets A and B of real numbers such that $\{X \in A\}$ and $\{Y \in B\}$ are events,

$$\Pr(X \in A \text{ and } Y \in B) = \Pr(X \in A)\Pr(Y \in B). \tag{3.5.5}$$

In other words, let E be any event the occurrence or nonoccurrence of which depends only on the value of X (such as $E = \{X \in A\}$), and let D be any event the occurrence or nonoccurrence of which depends only on the value of Y (such as $D = \{Y \in B\}$). Then X and Y are independent random variables if and only if E and D are independent events for all such events E and D.

If X and Y are independent, then for all real numbers x and y, it must be true that

$$\Pr(X \le x \text{ and } Y \le y) = \Pr(X \le x)\Pr(Y \le y). \tag{3.5.6}$$

Moreover, since all probabilities for X and Y of the type appearing in Eq. (3.5.5) can be derived from probabilities of the type appearing in Eq. (3.5.6), it can be shown that if Eq. (3.5.6) is satisfied for all values of x and y, then X and Y must be independent. The proof of this statement is beyond the scope of this book and is omitted, but we summarize it as the following theorem.

Theorem 3.5.4 Let the joint c.d.f. of X and Y be F, let the marginal c.d.f. of X be F_1, and let the marginal c.d.f. of Y be F_2. Then X and Y are independent if and only if, for all real numbers x and y, $F(x, y) = F_1(x)F_2(y)$. ■

For example, the demands for water and electricity in Example 3.5.6 are independent. If one returns to Example 3.5.1, one also sees that the product of the marginal p.d.f.'s of water and electric demand equals their joint p.d.f. given in Eq. (3.4.2). This relation is characteristic of independent random variables whether discrete or continuous.

Theorem 3.5.5 Suppose that X and Y are random variables that have a joint p.f., p.d.f., or p.f./p.d.f. f. Then X and Y will be independent if and only if f can be represented in the following form for $-\infty < x < \infty$ and $-\infty < y < \infty$:

$$f(x, y) = h_1(x)h_2(y), \tag{3.5.7}$$

where h_1 is a nonnegative function of x alone and h_2 is a nonnegative function of y alone.

Proof We shall give the proof only for the case in which X is discrete and Y is continuous. The other cases are similar. For the "if" part, assume that Eq. (3.5.7) holds. Write

$$f_1(x) = \int_{-\infty}^{\infty} h_1(x)h_2(y)dy = c_1 h_1(x),$$

where $c_1 = \int_{-\infty}^{\infty} h_2(y)dy$ must be finite and strictly positive, otherwise f_1 wouldn't be a p.f. So, $h_1(x) = f_1(x)/c_1$. Similarly,

$$f_2(y) = \sum_x h_1(x)h_2(y) = h_2(y)\sum_x \frac{1}{c_1} f_1(x) = \frac{1}{c_1} h_2(y).$$

So, $h_2(y) = c_1 f_2(y)$. Since $f(x, y) = h_1(x)h_2(y)$, it follows that

$$f(x, y) = \frac{f_1(x)}{c_1} c_1 f_2(y) = f_1(x) f_2(y). \tag{3.5.8}$$

Now let A and B be sets of real numbers. Assuming the integrals exist, we can write

$$\begin{aligned}
\Pr(X \in A \text{ and } Y \in B) &= \sum_{x \in A} \int_B f(x, y) dy \\
&= \int_B \sum_{x \in A} f_1(x) f_2(y) dy, \\
&= \sum_{x \in A} f_1(x) \int_B f_2(y) dy,
\end{aligned}$$

where the first equality is from Definition 3.4.5, the second is from Eq. (3.5.8), and the third is straightforward rearrangement. We now see that X and Y are independent according to Definition 3.5.2.

For the "only if" part, assume that X and Y are independent. Let A and B be sets of real numbers. Let f_1 be the marginal p.d.f. of X, and let f_2 be the marginal p.f. of Y. Then

$$\begin{aligned}
\Pr(X \in A \text{ and } Y \in B) &= \sum_{x \in A} f_1(x) \int_B f_2(y) dy \\
&= \int_B \sum_{x \in A} f_1(x) f_2(y) dy,
\end{aligned}$$

(if the integral exists) where the first equality follows from Definition 3.5.2 and the second is a straightforward rearrangement. We now see that $f_1(x) f_2(y)$ satisfies the conditions needed to be $f(x, y)$ as stated in Definition 3.4.5. ∎

A simple corollary follows from Theorem 3.5.5.

Corollary 3.5.1 Two random variables X and Y are independent if and only if the following factorization is satisfied for all real numbers x and y:

$$f(x, y) = f_1(x) f_2(y). \tag{3.5.9}$$

∎

As stated in Sec. 3.2 (see page 102), in a continuous distribution the values of a p.d.f. can be changed arbitrarily at any countable set of points. Therefore, for such a distribution it would be more precise to state that the random variables X and Y are independent if and only if it is possible to choose versions of f, f_1, and f_2 such that Eq. (3.5.9) is satisfied for $-\infty < x < \infty$ and $-\infty < y < \infty$.

The Meaning of Independence We have given a mathematical definition of independent random variables in Definition 3.5.2, but we have not yet given any interpretation of the concept of independent random variables. Because of the close connection between independent events and independent random variables, the interpretation of independent random variables should be closely related to the interpretation of independent events. We model two events as independent if learning that one of them occurs does not change the probability that the other one occurs. It is easiest to extend this idea to discrete random variables. Suppose that X and Y

Table 3.5 Joint p.f. $f(x, y)$ for Example 3.5.7

x	y = 1	2	3	4	5	6	Total
0	1/24	1/24	1/24	1/24	1/24	1/24	1/4
1	1/12	1/12	1/12	1/12	1/12	1/12	1/2
2	1/24	1/24	1/24	1/24	1/24	1/24	1/4
Total	1/6	1/6	1/6	1/6	1/6	1/6	1.000

have a discrete joint distribution. If, for each y, learning that $Y = y$ does not change any of the probabilities of the events $\{X = x\}$, we would like to say that X and Y are independent. From Corollary 3.5.1 and the definition of marginal p.f., we see that indeed X and Y are independent if and only if, for each y and x such that $\Pr(Y = y) > 0$, $\Pr(X = x|Y = y) = \Pr(X = x)$, that is, learning the value of Y doesn't change any of the probabilities associated with X. When we formally define conditional distributions in Sec. 3.6, we shall see that this interpretation of independent discrete random variables extends to all bivariate distributions. In summary, if we are trying to decide whether or not to model two random variables X and Y as independent, we should think about whether we would change the distribution of X after we learned the value of Y or vice versa.

Example 3.5.7 Games of Chance. A carnival game consists of rolling a fair die, tossing a fair coin two times, and recording both outcomes. Let Y stand for the number on the die, and let X stand for the number of heads in the two tosses. It seems reasonable to believe that all of the events determined by the roll of the die are independent of all of the events determined by the flips of the coin. Hence, we can assume that X and Y are independent random variables. The marginal distribution of Y is the uniform distribution on the integers $1, \dots, 6$, while the distribution of X is the binomial distribution with parameters 2 and 1/2. The marginal p.f.'s and the joint p.f. of X and Y are given in Table 3.5, where the joint p.f. was constructed using Eq. (3.5.9). The Total column gives the marginal p.f. f_1 of X, and the Total row gives the marginal p.f. f_2 of Y. ◄

Example 3.5.8 Determining Whether Random Variables Are Independent in a Clinical Trial. Return to the clinical trial of depression drugs in Exercise 11 of Sec. 3.4 (on page 129). In that trial, a patient is selected at random from the 150 patients in the study and we record Y, an indicator of the treatment group for that patient, and X, an indicator of whether or not the patient relapsed. Table 3.6 repeats the joint p.f. of X and Y along with the marginal distributions in the margins. We shall determine whether or not X and Y are independent.

In Eq. (3.5.9), $f(x, y)$ is the probability in the xth row and the yth column of the table, $f_1(x)$ is the number in the Total column in the xth row, and $f_2(y)$ is the number in the Total row in the yth column. It is seen in the table that $f(1, 2) = 0.087$, while $f_1(1) = 0.513$, and $f_2(1) = 0.253$. Hence, $f(1, 2) \neq f_1(1)f_2(1) = 0.129$. It follows that X and Y are not independent. ◄

It should be noted from Examples 3.5.7 and 3.5.8 that X and Y will be independent if and only if the rows of the table specifying their joint p.f. are proportional to

Table 3.6 Proportions marginals in Example 3.5.8

	Treatment group (Y)				
Response (X)	Imipramine (1)	Lithium (2)	Combination (3)	Placebo (4)	Total
Relapse (0)	0.120	0.087	0.146	0.160	0.513
No relapse (1)	0.147	0.166	0.107	0.067	0.487
Total	0.267	0.253	0.253	0.227	1.0

one another, or equivalently, if and only if the columns of the table are proportional to one another.

Example 3.5.9 Calculating a Probability Involving Independent Random Variables. Suppose that two measurements X and Y are made of the rainfall at a certain location on May 1 in two consecutive years. It might be reasonable, given knowledge of the history of rainfall on May 1, to treat the random variables X and Y as independent. Suppose that the p.d.f. g of each measurement is as follows:

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

We shall determine the value of $\Pr(X + Y \le 1)$.

Since X and Y are independent and each has the p.d.f. g, it follows from Eq. (3.5.9) that for all values of x and y the joint p.d.f. $f(x, y)$ of X and Y will be specified by the relation $f(x, y) = g(x)g(y)$. Hence,

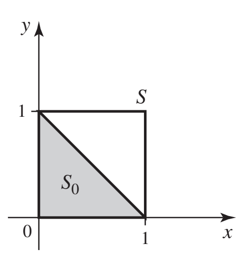
$$f(x, y) = \begin{cases} 4xy & \text{for } 0 \le x \le 1 \text{ and } 0 \le y \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

The set S in the xy-plane, where $f(x, y) > 0$, and the subset S_0, where $x + y \le 1$, are sketched in Fig. 3.19. Thus,

$$\Pr(X + Y \le 1) = \int_{S_0}\int f(x, y)\, dx\, dy = \int_0^1 \int_0^{1-x} 4xy\, dy\, dx = \frac{1}{6}.$$

As a final note, if the two measurements X and Y had been made on the same day at nearby locations, then it might not make as much sense to treat them as independent, since we would expect them to be more similar to each other than to historical rainfalls. For example, if we first learn that X is small compared to historical rainfall on the date in question, we might then expect Y to be smaller than the historical distribution would suggest. ◂

Figure 3.19 The subset S_0 where $x + y \le 1$ in Example 3.5.9.



Theorem 3.5.5 says that X and Y are independent if and only if, for all values of x and y, f can be factored into the product of an arbitrary nonnegative function of x and an arbitrary nonnegative function of y. However, it should be emphasized that, just as in Eq. (3.5.9), the factorization in Eq. (3.5.7) must be satisfied for all values of x and y ($-\infty < x < \infty$ and $-\infty < y < \infty$).

Example 3.5.10 Dependent Random Variables. Suppose that the joint p.d.f. of X and Y has the following form:

$$f(x, y) = \begin{cases} kx^2y^2 & \text{for } x^2 + y^2 \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

We shall show that X and Y are not independent.

It is evident that at each point inside the circle $x^2 + y^2 \le 1$, $f(x, y)$ can be factored as in Eq. (3.5.7). However, this same factorization cannot also be satisfied at every point outside this circle. For example, $f(0.9, 0.9) = 0$, but neither $f_1(0.9) = 0$ nor $f_2(0.9) = 0$. (In Exercise 13, you can verify this feature of f_1 and f_2.)

The important feature of this example is that the values of X and Y are constrained to lie inside a circle. The joint p.d.f. of X and Y is positive inside the circle and zero outside the circle. Under these conditions, X and Y cannot be independent, because for every given value y of Y, the possible values of X will depend on y. For example, if $Y = 0$, then X can have any value such that $X^2 \le 1$; if $Y = 1/2$, then X must have a value such that $X^2 \le 3/4$. ◀

Example 3.5.10 shows that one must be careful when trying to apply Theorem 3.5.5. The situation that arose in that example will occur whenever $\{(x, y) : f(x, y) > 0\}$ has boundaries that are curved or not parallel to the coordinate axes. There is one important special case in which it is easy to check the conditions of Theorem 3.5.5. The proof is left as an exercise.

Theorem 3.5.6 Let X and Y have a continuous joint distribution. Suppose that $\{(x, y) : f(x, y) > 0\}$ is a rectangular region R (possibly unbounded) with sides (if any) parallel to the coordinate axes. Then X and Y are independent if and only if Eq. (3.5.7) holds for all $(x, y) \in R$. ■

Example 3.5.11 Verifying the Factorization of a Joint p.d.f. Suppose that the joint p.d.f. f of X and Y is as follows:

$$f(x, y) = \begin{cases} ke^{-(x+2y)} & \text{for } x \ge 0 \text{ and } y \ge 0, \\ 0 & \text{otherwise,} \end{cases}$$

where k is some constant. We shall first determine whether X and Y are independent and then determine their marginal p.d.f.'s.

In this example, $f(x, y) = 0$ outside of an unbounded rectangular region R whose sides are the lines $x = 0$ and $y = 0$. Furthermore, at each point inside R, $f(x, y)$ can be factored as in Eq. (3.5.7) by letting $h_1(x) = ke^{-x}$ and $h_2(y) = e^{-2y}$. Therefore, X and Y are independent.

It follows that in this case, except for constant factors, $h_1(x)$ for $x \ge 0$ and $h_2(y)$ for $y \ge 0$ must be the marginal p.d.f.'s of X and Y. By choosing constants that make $h_1(x)$ and $h_2(y)$ integrate to unity, we can conclude that the marginal p.d.f.'s f_1 and f_2 of X and Y must be as follows:

$$f_1(x) = \begin{cases} e^{-x} & \text{for } x \ge 0, \\ 0 & \text{otherwise,} \end{cases}$$

and

$$f_2(y) = \begin{cases} 2e^{-2y} & \text{for } y \ge 0, \\ 0 & \text{otherwise.} \end{cases}$$

If we multiply $f_1(x)$ times $f_2(y)$ and compare the product to $f(x, y)$, we see that $k = 2$. ◀

Note: Separate Functions of Independent Random Variables Are Independent. If X and Y are independent, then $h(X)$ and $g(Y)$ are independent no matter what the functions h and g are. This is true because for every t, the event $\{h(X) \le t\}$ can always be written as $\{X \in A\}$, where $A = \{x : h(x) \le t\}$. Similarly, $\{g(Y) \le u\}$ can be written as $\{Y \in B\}$, so Eq. (3.5.6) for $h(X)$ and $g(Y)$ follows from Eq. (3.5.5) for X and Y.

Summary

Let $f(x, y)$ be a joint p.f., joint p.d.f., or joint p.f./p.d.f. of two random variables X and Y. The marginal p.f. or p.d.f. of X is denoted by $f_1(x)$, and the marginal p.f. or p.d.f. of Y is denoted by $f_2(y)$. To obtain $f_1(x)$, compute $\sum_y f(x, y)$ if Y is discrete or $\int_{-\infty}^{\infty} f(x, y)\,dy$ if Y is continuous. Similarly, to obtain $f_2(y)$, compute $\sum_x f(x, y)$ if X is discrete or $\int_{-\infty}^{\infty} f(x, y)\,dx$ if X is continuous. The random variables X and Y are independent if and only if $f(x, y) = f_1(x) f_2(y)$ for *all* x and y. This is true regardless of whether X and/or Y is continuous or discrete. A sufficient condition for two continuous random variables to be independent is that $R = \{(x, y) : f(x, y) > 0\}$ be rectangular with sides parallel to the coordinate axes and that $f(x, y)$ factors into separate functions of x of y in R.

Exercises

1. Suppose that X and Y have a continuous joint distribution for which the joint p.d.f. is

$$f(x, y) = \begin{cases} k & \text{for } a \le x \le b \text{ and } c \le y \le d, \\ 0 & \text{otherwise,} \end{cases}$$

where $a < b$, $c < d$, and $k > 0$. Find the marginal distributions of X and Y.

2. Suppose that X and Y have a discrete joint distribution for which the joint p.f. is defined as follows:

$$f(x, y) = \begin{cases} \frac{1}{30}(x + y) & \text{for } x = 0, 1, 2 \text{ and } y = 0, 1, 2, 3, \\ 0 & \text{otherwise.} \end{cases}$$

a. Determine the marginal p.f.'s of X and Y.

b. Are X and Y independent?

3. Suppose that X and Y have a continuous joint distribution for which the joint p.d.f. is defined as follows:

$$f(x, y) = \begin{cases} \frac{3}{2}y^2 & \text{for } 0 \le x \le 2 \text{ and } 0 \le y \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

a. Determine the marginal p.d.f.'s of X and Y.

b. Are X and Y independent?

c. Are the event $\{X < 1\}$ and the event $\{Y \ge 1/2\}$ independent?

4. Suppose that the joint p.d.f. of X and Y is as follows:

$$f(x, y) = \begin{cases} \frac{15}{4}x^2 & \text{for } 0 \le y \le 1 - x^2, \\ 0 & \text{otherwise.} \end{cases}$$

a. Determine the marginal p.d.f.'s of X and Y.

b. Are X and Y independent?

5. A certain drugstore has three public telephone booths. For $i = 0, 1, 2, 3$, let p_i denote the probability that exactly i telephone booths will be occupied on any Monday evening at 8:00 P.M.; and suppose that $p_0 = 0.1$, $p_1 = 0.2$, $p_2 = 0.4$, and $p_3 = 0.3$. Let X and Y denote the number of booths that will be occupied at 8:00 P.M. on two independent Monday evenings. Determine: **(a)** the joint p.f. of X and Y; **(b)** $\Pr(X = Y)$; **(c)** $\Pr(X > Y)$.

6. Suppose that in a certain drug the concentration of a particular chemical is a random variable with a continuous distribution for which the p.d.f. g is as follows:

$$g(x) = \begin{cases} \frac{3}{8}x^2 & \text{for } 0 \le x \le 2, \\ 0 & \text{otherwise.} \end{cases}$$

Suppose that the concentrations X and Y of the chemical in two separate batches of the drug are independent random variables for each of which the p.d.f. is g. Determine **(a)** the joint p.d.f. of X and Y; **(b)** $\Pr(X = Y)$; **(c)** $\Pr(X > Y)$; **(d)** $\Pr(X + Y \le 1)$.

7. Suppose that the joint p.d.f. of X and Y is as follows:

$$f(x, y) = \begin{cases} 2xe^{-y} & \text{for } 0 \le x \le 1 \text{ and } 0 < y < \infty, \\ 0 & \text{otherwise.} \end{cases}$$

Are X and Y independent?

8. Suppose that the joint p.d.f. of X and Y is as follows:

$$f(x, y) = \begin{cases} 24xy & \text{for } x \ge 0, y \ge 0, \text{ and } x + y \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

Are X and Y independent?

9. Suppose that a point (X, Y) is chosen at random from the rectangle S defined as follows:

$$S = \{(x, y) : 0 \le x \le 2 \text{ and } 1 \le y \le 4\}.$$

a. Determine the joint p.d.f. of X and Y, the marginal p.d.f. of X, and the marginal p.d.f. of Y.

b. Are X and Y independent?

10. Suppose that a point (X, Y) is chosen at random from the circle S defined as follows:

$$S = \{(x, y) : x^2 + y^2 \le 1\}.$$

a. Determine the joint p.d.f. of X and Y, the marginal p.d.f. of X, and the marginal p.d.f. of Y.

b. Are X and Y independent?

11. Suppose that two persons make an appointment to meet between 5 P.M. and 6 P.M. at a certain location, and they agree that neither person will wait more than 10 minutes for the other person. If they arrive independently at random times between 5 P.M. and 6 P.M., what is the probability that they will meet?

12. Prove Theorem 3.5.6.

13. In Example 3.5.10, verify that X and Y have the same marginal p.d.f.'s and that

$$f_1(x) = \begin{cases} 2kx^2(1 - x^2)^{3/2}/3 & \text{if } -1 \le x \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

14. For the joint p.d.f. in Example 3.4.7, determine whether or not X and Y are independent.

15. A painting process consists of two stages. In the first stage, the paint is applied, and in the second stage, a protective coat is added. Let X be the time spent on the first stage, and let Y be the time spent on the second stage. The first stage involves an inspection. If the paint fails the inspection, one must wait three minutes and apply the paint again. After a second application, there is no further inspection. The joint p.d.f. of X and Y is

$$f(x, y) = \begin{cases} \frac{1}{3} & \text{if } 1 < x < 3 \text{ and } 0 < y < 1, \\ \frac{1}{6} & \text{if } 6 < x < 8 \text{ and } 0 < y < 1, \\ 0 & \text{otherwise.} \end{cases}$$

a. Sketch the region where $f(x, y) > 0$. Note that it is not exactly a rectangle.

b. Find the marginal p.d.f.'s of X and Y.

c. Show that X and Y are independent.

This problem does not contradict Theorem 3.5.6. In that theorem the conditions, including that the set where $f(x, y) > 0$ be rectangular, are sufficient but not necessary.

3.6 Conditional Distributions

We generalize the concept of conditional probability to conditional distributions. Recall that distributions are just collections of probabilities of events determined by random variables. Conditional distributions will be the probabilities of events determined by some random variables conditional on events determined by other random variables. The idea is that there will typically be many random variables of interest in an applied problem. After we observe some of those random variables, we want to be able to adjust the probabilities associated with the ones that have not yet been observed. The conditional distribution of one random variable X given another Y will be the distribution that we would use for X after we learn the value of Y.

Table 3.7 Joint p.f. for Example 3.6.1

	Brand Y					
Stolen X	1	2	3	4	5	Total
0	0.129	0.298	0.161	0.280	0.108	0.976
1	0.010	0.010	0.001	0.002	0.001	0.024
Total	0.139	0.308	0.162	0.282	0.109	1.000

Discrete Conditional Distributions

Example 3.6.1

Auto Insurance. Insurance companies keep track of how likely various cars are to be stolen. Suppose that a company in a particular area computes the joint distribution of car brands and the indicator of whether the car will be stolen during a particular year that appears in Table 3.7.

We let $X = 1$ mean that a car is stolen, and we let $X = 0$ mean that the car is not stolen. We let Y take one of the values from 1 to 5 to indicate the brand of car as indicated in Table 3.7. If a customer applies for insurance for a particular brand of car, the company needs to compute the distribution of the random variable X as part of its premium determination. The insurance company might adjust their premium according to a risk factor such as likelihood of being stolen. Although, overall, the probability that a car will be stolen is 0.024, if we assume that we know the brand of car, the probability might change quite a bit. This section introduces the formal concepts for addressing this type of problem. ◂

Suppose that X and Y are two random variables having a discrete joint distribution for which the joint p.f. is f. As before, we shall let f_1 and f_2 denote the marginal p.f.'s of X and Y, respectively. After we observe that $Y = y$, the probability that the random variable X will take a particular value x is specified by the following conditional probability:

$$\Pr(X = x|Y = y) = \frac{\Pr(X = x \text{ and } Y = y)}{\Pr(Y = y)} = \frac{f(x, y)}{f_2(y)}. \tag{3.6.1}$$

In other words, if it is known that $Y = y$, then the probability that $X = x$ will be updated to the value in Eq. (3.6.1). Next, we consider the entire distribution of X after learning that $Y = y$.

Definition 3.6.1

Conditional Distribution/p.f. Let X and Y have a discrete joint distribution with joint p.f. f. Let f_2 denote the marginal p.f. of Y. For each y such that $f_2(y) > 0$, define

$$g_1(x|y) = \frac{f(x, y)}{f_2(y)}. \tag{3.6.2}$$

Then g_1 is called the *conditional p.f. of X given Y*. The discrete distribution whose p.f. is $g_1(\cdot|y)$ is called the *conditional distribution of X given that $Y = y$*.

Table 3.8 Conditional p.f. of Y given X for Example 3.6.3

	Brand Y				
Stolen X	1	2	3	4	5
0	0.928	0.968	0.994	0.993	0.991
1	0.072	0.032	0.006	0.007	0.009

We should verify that $g_1(x|y)$ is actually a p.f. as a function of x for each y. Let y be such that $f_2(y) > 0$. Then $g_1(x|y) \geq 0$ for all x and

$$\sum_x g_1(x|y) = \frac{1}{f_2(y)} \sum_x f(x, y) = \frac{1}{f_2(y)} f_2(y) = 1.$$

Notice that we do not bother to define $g_1(x|y)$ for those y such that $f_2(y) = 0$.

Similarly, if x is a given value of X such that $f_1(x) = \Pr(X = x) > 0$, and if $g_2(y|x)$ is the *conditional p.f. of Y given that $X = x$*, then

$$g_2(y|x) = \frac{f(x, y)}{f_1(x)}. \tag{3.6.3}$$

For each x such that $f_1(x) > 0$, the function $g_2(y|x)$ will be a p.f. as a function of y.

Example 3.6.2 Calculating a Conditional p.f. from a Joint p.f. Suppose that the joint p.f. of X and Y is as specified in Table 3.4 in Example 3.5.2. We shall determine the conditional p.f. of Y given that $X = 2$.

The marginal p.f. of X appears in the Total column of Table 3.4, so $f_1(2) = \Pr(X = 2) = 0.6$. Therefore, the conditional probability $g_2(y|2)$ that Y will take a particular value y is

$$g_2(y|2) = \frac{f(2, y)}{0.6}.$$

It should be noted that for all possible values of y, the conditional probabilities $g_2(y|2)$ must be proportional to the joint probabilities $f(2, y)$. In this example, each value of $f(2, y)$ is simply divided by the constant $f_1(2) = 0.6$ in order that the sum of the results will be equal to 1. Thus,

$$g_2(1|2) = 1/2, \quad g_2(2|2) = 0, \quad g_2(3|2) = 1/6, \quad g_2(4|2) = 1/3.$$ ◀

Example 3.6.3 Auto Insurance. Consider again the probabilities of car brands and cars being stolen in Example 3.6.1. The conditional distribution of X (being stolen) given Y (brand) is given in Table 3.8. It appears that Brand 1 is much more likely to be stolen than other cars in this area, with Brand 1 also having a significant chance of being stolen. ◀

Continuous Conditional Distributions

Example 3.6.4 Processing Times. A manufacturing process consists of two stages. The first stage takes Y minutes, and the whole process takes X minutes (which includes the first

Y minutes). Suppose that X and Y have a joint continuous distribution with joint p.d.f.

$$f(x, y) = \begin{cases} e^{-x} & \text{for } 0 \le y \le x < \infty, \\ 0 & \text{otherwise.} \end{cases}$$

After we learn how much time Y that the first stage takes, we want to update our distribution for the total time X. In other words, we would like to be able to compute a conditional distribution for X given $Y = y$. We cannot argue the same way as we did with discrete joint distributions, because $\{Y = y\}$ is an event with probability 0 for all y. ◀

To facilitate the solutions of problems such as the one posed in Example 3.6.4, the concept of conditional probability will be extended by considering the definition of the conditional p.f. of X given in Eq. (3.6.2) and the analogy between a p.f. and a p.d.f.

Definition 3.6.2 **Conditional p.d.f.** Let X and Y have a continuous joint distribution with joint p.d.f. f and respective marginals f_1 and f_2. Let y be a value such that $f_2(y) > 0$. Then the *conditional p.d.f.* g_1 *of* X *given that* $Y = y$ is defined as follows:

$$g_1(x|y) = \frac{f(x, y)}{f_2(y)} \quad \text{for } -\infty < x < \infty. \tag{3.6.4}$$

For values of y such that $f_2(y) = 0$, we are free to define $g_1(x|y)$ however we wish, so long as $g_1(x|y)$ is a p.d.f. as a function of x.

It should be noted that Eq. (3.6.2) and Eq. (3.6.4) are identical. However, Eq. (3.6.2) was *derived* as the conditional probability that $X = x$ given that $Y = y$, whereas Eq. (3.6.4) was *defined* to be the value of the conditional p.d.f. of X given that $Y = y$. In fact, we should verify that $g_1(x|y)$ as defined above really is a p.d.f.

Theorem 3.6.1 For each y, $g_1(x|y)$ defined in Definition 3.6.2 is a p.d.f. as a function of x.

Proof If $f_2(y) = 0$, then g_1 is defined to be any p.d.f. we wish, and hence it is a p.d.f. If $f_2(y) > 0$, g_1 is defined by Eq. (3.6.4). For each such y, it is clear that $g_1(x|y) \ge 0$ for all x. Also, if $f_2(y) > 0$, then

$$\int_{-\infty}^{\infty} g_1(x|y)\, dx = \frac{\int_{-\infty}^{\infty} f(x, y)\, dx}{f_2(y)} = \frac{f_2(y)}{f_2(y)} = 1,$$

by using the formula for $f_2(y)$ in Eq. (3.5.3). ■

Example 3.6.5 **Processing Times.** In Example 3.6.4, Y is the time that the first stage of a process takes, while X is the total time of the two stages. We want to calculate the conditional p.d.f. of X given Y. We can calculate the marginal p.d.f. of Y as follows: For each y, the possible values of X are all $x \ge y$, so for each $y > 0$,

$$f_2(y) = \int_{y}^{\infty} e^{-x} dx = e^{-y},$$

and $f_2(y) = 0$ for $y < 0$. For each $y \ge 0$, the conditional p.d.f. of X given $Y = y$ is then

$$g_1(x|y) = \frac{f(x, y)}{f_2(y)} = \frac{e^{-x}}{e^{-y}} = e^{y-x}, \text{ for } x \ge y,$$

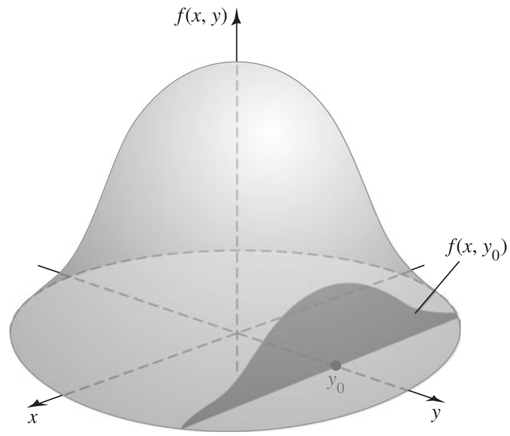


Figure 3.20 The conditional p.d.f. $g_1(x|y_0)$ is proportional to $f(x, y_0)$.

and $g_1(x|y) = 0$ for $x < y$. So, for example, if we observe $Y = 4$ and we want the conditional probability that $X \geq 9$, we compute

$$\Pr(X \geq 9|Y = 4) = \int_9^\infty e^{4-x} dx = e^{-5} = 0.0067. \quad ◂$$

Definition 3.6.2 has an interpretation that can be understood by considering Fig. 3.20. The joint p.d.f. f defines a surface over the xy-plane for which the height $f(x, y)$ at each point (x, y) represents the relative likelihood of that point. For instance, if it is known that $Y = y_0$, then the point (x, y) must lie on the line $y = y_0$ in the xy-plane, and the relative likelihood of any point (x, y_0) on this line is $f(x, y_0)$. Hence, the conditional p.d.f. $g_1(x|y_0)$ of X should be proportional to $f(x, y_0)$. In other words, $g_1(x|y_0)$ is essentially the same as $f(x, y_0)$, but it includes a constant factor $1/[f_2(y_0)]$, which is required to make the conditional p.d.f. integrate to unity over all values of x.

Similarly, for each value of x such that $f_1(x) > 0$, the *conditional p.d.f. of Y given that $X = x$* is defined as follows:

$$g_2(y|x) = \frac{f(x, y)}{f_1(x)} \quad \text{for } -\infty < y < \infty. \tag{3.6.5}$$

This equation is identical to Eq. (3.6.3), which was derived for discrete distributions. If $f_1(x) = 0$, then $g_2(y|x)$ is arbitrary so long as it is a p.d.f. as a function of y.

Example 3.6.6 **Calculating a Conditional p.d.f. from a Joint p.d.f.** Suppose that the joint p.d.f. of X and Y is as specified in Example 3.4.8 on page 122. We shall first determine the conditional p.d.f. of Y given that $X = x$ and then determine some probabilities for Y given the specific value $X = 1/2$.

The set S for which $f(x, y) > 0$ was sketched in Fig. 3.12 on page 123. Furthermore, the marginal p.d.f. f_1 was derived in Example 3.5.3 on page 132 and sketched in Fig. 3.17 on page 133. It can be seen from Fig. 3.17 that $f_1(x) > 0$ for $-1 < x < 1$ but not for $x = 0$. Therefore, for each given value of x such that $-1 < x < 0$ or $0 < x < 1$, the conditional p.d.f. $g_2(y|x)$ of Y will be as follows:

$$g_2(y|x) = \begin{cases} \dfrac{2y}{1 - x^4} & \text{for } x^2 \leq y \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

In particular, if it is known that $X = 1/2$, then $\Pr\left(Y \geq \frac{1}{4} \middle| X = \frac{1}{2}\right) = 1$ and

$$\Pr\left(Y \geq \frac{3}{4} \middle| X = \frac{1}{2}\right) = \int_{3/4}^{1} g_2\left(y \middle| \frac{1}{2}\right) dy = \frac{7}{15}. \quad ◀$$

Note: A Conditional p.d.f. Is Not the Result of Conditioning on a Set of Probability Zero. The conditional p.d.f. $g_1(x|y)$ of X given $Y = y$ is the p.d.f. we would use for X if we were to learn that $Y = y$. This sounds as if we were conditioning on the event $\{Y = y\}$, which has zero probability if Y has a continuous distribution. Actually, for the cases we shall see in this text, the value of $g_1(x|y)$ is a limit:

$$g_1(x|y) = \lim_{\epsilon \to 0} \frac{\partial}{\partial x} \Pr(X \leq x | y - \epsilon < Y \leq y + \epsilon). \tag{3.6.6}$$

The conditioning event $\{y - \epsilon \leq Y \leq y + \epsilon\}$ in Eq. (3.6.6) has positive probability if the marginal p.d.f. of Y is positive at y. The mathematics required to make this rigorous is beyond the scope of this text. (See Exercise 11 in this section and Exercises 25 and 26 in Sec. 3.11 for results that we can prove.) Another way to think about conditioning on a continuous random variable is to notice that the conditional p.d.f.'s that we compute are typically continuous as a function of the conditioning variable. This means that conditioning on $Y = y$ or on $Y = y + \epsilon$ for small ϵ will produce nearly the same conditional distribution for X. So it does not matter much if we use $Y = y$ as a surogate for Y close to y. Nevertheless, it is important to keep in mind that the conditional p.d.f. of X given $Y = y$ is better thought of as the conditional p.d.f. of X given that Y is very close to y. This wording is awkward, so we shall not use it, but we must remember the distinction between the conditional p.d.f. and conditioning on an event with probability 0. Despite this distinction, it is still legitimate to treat Y as the constant y when dealing with the conditional distribution of X given $Y = y$.

For mixed joint distributions, we continue to use Eqs. (3.6.2) and (3.6.3) to define conditional p.f.'s and p.d.f.'s.

Definition 3.6.3 Conditional p.f. or p.d.f. from Mixed Distribution. Let X be discrete and let Y be continuous with joint p.f./p.d.f. f. Then the *conditional p.f. of X given $Y = y$* is defined by Eq. (3.6.2), and the *conditional p.d.f. of Y given $X = x$* is defined by Eq. (3.6.3).

Construction of the Joint Distribution

Example 3.6.7 Defective Parts. Suppose that a certain machine produces defective and nondefective parts, but we do not know what proportion of defectives we would find among all parts that could be produced by this machine. Let P stand for the unknown proportion of defective parts among all possible parts produced by the machine. If we were to learn that $P = p$, we might be willing to say that the parts were independent of each other and each had probability p of being defective. In other words, if we condition on $P = p$, then we have the situation described in Example 3.1.9. As in that example, suppose that we examine n parts and let X stand for the number of defectives among the n examined parts. The distribution of X, assuming that we know $P = p$, is the binomial distribution with parameters n and p. That is, we can let the binomial p.f. (3.1.4) be the conditional p.f. of X given $P = p$, namely,

$$g_1(x|p) = \binom{n}{x} p^x (1 - p)^{n-x}, \text{ for } x = 0, \ldots, n.$$

We might also believe that P has a continuous distribution with p.d.f. such as $f_2(p) = 1$ for $0 \le p \le 1$. (This means that P has the uniform distribution on the interval [0, 1].) We know that the conditional p.f. g_1 of X given $P = p$ satisfies

$$g_1(x|p) = \frac{f(x, p)}{f_2(p)},$$

where f is the joint p.f./p.d.f. of X and P. If we multiply both sides of this equation by $f_2(p)$, it follows that the joint p.f./p.d.f. of X and P is

$$f(x, p) = g_1(x|p) f_2(p) = \binom{n}{x} p^x (1-p)^{n-x}, \quad \text{for } x = 0, \ldots, n, \text{ and } 0 \le p \le 1.$$

◀

The construction in Example 3.6.7 is available in general, as we explain next.

Generalizing the Multiplication Rule for Conditional Probabilities A special case of Theorem 2.1.2, the multiplication rule for conditional probabilities, says that if A and B are two events, then $\Pr(A \cap B) = \Pr(A)\Pr(B|A)$. The following theorem, whose proof is immediate from Eqs. (3.6.4) and (3.6.5), generalizes Theorem 2.1.2 to the case of two random variables.

Theorem 3.6.2 Multiplication Rule for Distributions. Let X and Y be random variables such that X has p.f. or p.d.f. $f_1(x)$ and Y has p.f. or p.d.f. $f_2(y)$. Also, assume that the conditional p.f. or p.d.f. of X given $Y = y$ is $g_1(x|y)$ while the conditional p.f. or p.d.f. of Y given $X = x$ is $g_2(y|x)$. Then for each y such that $f_2(y) > 0$ and each x,

$$f(x, y) = g_1(x|y) f_2(y), \tag{3.6.7}$$

where f is the joint p.f., p.d.f., or p.f./p.d.f. of X and Y. Similarly, for each x such that $f_1(x) > 0$ and each y,

$$f(x, y) = f_1(x) g_2(y|x). \tag{3.6.8}$$

■

In Theorem 3.6.2, if $f_2(y_0) = 0$ for some value y_0, then it can be assumed without loss of generality that $f(x, y_0) = 0$ for all values of x. In this case, both sides of Eq. (3.6.7) will be 0, and the fact that $g_1(x|y_0)$ is not uniquely defined becomes irrelevant. Hence, Eq. (3.6.7) will be satisfied for *all* values of x and y. A similar statement applies to Eq. (3.6.8).

Example 3.6.8 Waiting in a Queue. Let X be the amount of time that a person has to wait for service in a queue. The faster the server works in the queue, the shorter should be the waiting time. Let Y stand for the rate at which the server works, which we will take to be unknown. A common choice of conditional distribution for X given $Y = y$ has conditional p.d.f. for each $y > 0$:

$$g_1(x|y) = \begin{cases} ye^{-xy} & \text{for } x \ge 0, \\ 0 & \text{otherwise.} \end{cases}$$

We shall assume that Y has a continuous distribution with p.d.f. $f_2(y) = e^{-y}$ for $y > 0$. Now we can construct the joint p.d.f. of X and Y using Theorem 3.6.2:

$$f(x, y) = g_1(x|y) f_2(y) = \begin{cases} ye^{-y(x+1)} & \text{for } x \ge 0, y > 0, \\ 0 & \text{otherwise.} \end{cases}$$

◀

Example 3.6.9

Defective Parts. Let X be the number of defective parts in a sample of size n, and let P be the proportion of defectives among all parts, as in Example 3.6.7. The joint p.f./p.d.f of X and $P = p$ was calculated there as

$$f(x, p) = g_1(x|p)f_2(p) = \binom{n}{x} p^x(1-p)^{n-x}, \quad \text{for } x = 0, \dots, n \text{ and } 0 \le p \le 1.$$

We could now compute the conditional p.d.f. of P given $X = x$ by first finding the marginal p.f. of X:

$$f_1(x) = \int_0^1 \binom{n}{x} p^x(1-p)^{n-x} dp, \tag{3.6.9}$$

The conditional p.d.f. of P given $X = x$ is then

$$g_2(p|x) = \frac{f(x, p)}{f_1(x)} = \frac{p^x(1-p)^{n-x}}{\int_0^1 q^x(1-q)^{n-x} dq}, \quad \text{for } 0 < p < 1. \tag{3.6.10}$$

The integral in the denominator of Eq. (3.6.10) can be tedious to calculate, but it can be found. For example, if $n = 2$ and $x = 1$, we get

$$\int_0^1 q(1-q)dq = \frac{1}{2} - \frac{1}{3} = \frac{1}{6}.$$

In this case, $g_2(p|1) = 6p(1-p)$ for $0 \le p \le 1$. ◀

Bayes' Theorem and the Law of Total Probability for Random Variables The calculation done in Eq. (3.6.9) is an example of the generalization of the law of total probability to random variables. Also, the calculation in Eq. (3.6.10) is an example of the generalization of Bayes' theorem to random variables. The proofs of these results are straightforward and not given here.

Theorem 3.6.3

Law of Total Probability for Random Variables. If $f_2(y)$ is the marginal p.f. or p.d.f. of a random variable Y and $g_1(x|y)$ is the conditional p.f. or p.d.f. of X given $Y = y$, then the marginal p.f. or p.d.f. of X is

$$f_1(x) = \sum_y g_1(x|y)f_2(y), \tag{3.6.11}$$

if Y is discrete. If Y is continuous, the marginal p.f. or p.d.f. of X is

$$f_1(x) = \int_{-\infty}^{\infty} g_1(x|y)f_2(y)\, dy. \tag{3.6.12}$$

■

There are versions of Eqs. (3.6.11) and (3.6.12) with x and y switched and the subscripts 1 and 2 switched. These versions would be used if the joint distribution of X and Y were constructed from the conditional distribution of Y given X and the marginal distribution of X.

Theorem 3.6.4

Bayes' Theorem for Random Variables. If $f_2(y)$ is the marginal p.f. or p.d.f. of a random variable Y and $g_1(x|y)$ is the conditional p.f. or p.d.f. of X given $Y = y$, then the conditional p.f. or p.d.f. of Y given $X = x$ is

$$g_2(y|x) = \frac{g_1(x|y)f_2(y)}{f_1(x)}, \tag{3.6.13}$$

where $f_1(x)$ is obtained from Eq. (3.6.11) or (3.6.12). Similarly, the conditional p.f. or p.d.f. of X given $Y = y$ is

$$g_1(x|y) = \frac{g_2(y|x)f_1(x)}{f_2(y)}, \tag{3.6.14}$$

where $f_2(y)$ is obtained from Eq. (3.6.11) or (3.6.12) with x and y switched and with the subscripts 1 and 2 switched. ■

Example 3.6.10 Choosing Points from Uniform Distributions. Suppose that a point X is chosen from the uniform distribution on the interval [0, 1], and that after the value $X = x$ has been observed $(0 < x < 1)$, a point Y is then chosen from the uniform distribution on the interval $[x, 1]$. We shall derive the marginal p.d.f. of Y.

Since X has a uniform distribution, the marginal p.d.f. of X is as follows:

$$f_1(x) = \begin{cases} 1 & \text{for } 0 < x < 1, \\ 0 & \text{otherwise.} \end{cases}$$

Similarly, for each value $X = x$ $(0 < x < 1)$, the conditional distribution of Y is the uniform distribution on the interval $[x, 1]$. Since the length of this interval is $1 - x$, the conditional p.d.f. of Y given that $X = x$ will be

$$g_2(y|x) = \begin{cases} \dfrac{1}{1-x} & \text{for } x < y < 1, \\ 0 & \text{otherwise.} \end{cases}$$

It follows from Eq. (3.6.8) that the joint p.d.f. of X and Y will be

$$f(x, y) = \begin{cases} \dfrac{1}{1-x} & \text{for } 0 < x < y < 1, \\ 0 & \text{otherwise.} \end{cases} \tag{3.6.15}$$

Thus, for $0 < y < 1$, the value of the marginal p.d.f. $f_2(y)$ of Y will be

$$f_2(y) = \int_{-\infty}^{\infty} f(x, y)\, dx = \int_0^y \frac{1}{1-x}\, dx = -\log(1-y). \tag{3.6.16}$$

Furthermore, since Y cannot be outside the interval $0 < y < 1$, then $f_2(y) = 0$ for $y \le 0$ or $y \ge 1$. This marginal p.d.f. f_2 is sketched in Fig. 3.21. It is interesting to note that in this example the function f_2 is unbounded.

We can also find the conditional p.d.f. of X given $Y = y$ by applying Bayes' theorem (3.6.14). The product of $g_2(y|x)$ and $f_1(x)$ was already calculated in Eq. (3.6.15).

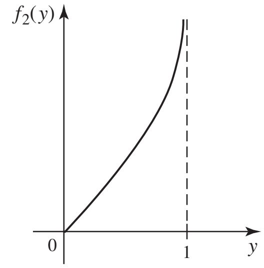


Figure 3.21 The marginal p.d.f. of Y in Example 3.6.10.

The ratio of this product to $f_2(y)$ from Eq. (3.6.16) is

$$g_1(x|y) = \begin{cases} \dfrac{-1}{(1-x)\log(1-y)} & \text{for } 0 < x < y, \\ 0 & \text{otherwise.} \end{cases}$$ ◀

Theorem 3.6.5 Independent Random Variables. Suppose that X and Y are two random variables having a joint p.f., p.d.f., or p.f./p.d.f. f. Then X and Y are independent if and only if for every value of y such that $f_2(y) > 0$ and every value of x,

$$g_1(x|y) = f_1(x). \tag{3.6.17}$$

Proof Theorem 3.5.4 says that X and Y are independent if and only if $f(x, y)$ can be factored in the following form for $-\infty < x < \infty$ and $-\infty < y < \infty$:

$$f(x, y) = f_1(x)f_2(y),$$

which holds if and only if, for all x and all y such that $f_2(y) > 0$,

$$f_1(x) = \frac{f(x, y)}{f_2(y)}. \tag{3.6.18}$$

But the right side of Eq. (3.6.18) is the formula for $g_1(x|y)$. Hence, X and Y are independent if and only if Eq. (3.6.17) holds for all x and all y such that $f_2(y) > 0$. ■

Theorem 3.6.5 says that X and Y are independent if and only if the conditional p.f. or p.d.f. of X given $Y = y$ is the same as the marginal p.f. or p.d.f. of X for all y such that $f_2(y) > 0$. Because $g_1(x|y)$ is arbitrary when $f_2(y) = 0$, we cannot expect Eq. (3.6.17) to hold in that case.

Similarly, it follows from Eq. (3.6.8) that X and Y are independent if and only if

$$g_2(y|x) = f_2(y), \tag{3.6.19}$$

for every value of x such that $f_1(x) > 0$. Theorem 3.6.5 and Eq. (3.6.19) give the mathematical justification for the meaning of independence that we presented on page 136.

Note: Conditional Distributions Behave Just Like Distributions. As we noted on page 59, conditional probabilities behave just like probabilities. Since distributions are just collections of probabilities, it follows that conditional distributions behave just like distributions. For example, to compute the conditional probability that a discrete random variable X is in some interval $[a, b]$ given $Y = y$, we must add $g_1(x|y)$ for all values of x in the interval. Also, theorems that we have proven or shall prove about distributions will have versions conditional on additional random variables. We shall postpone examples of such theorems until Sec. 3.7 because they rely on joint distributions of more than two random variables.

Summary

The conditional distribution of one random variable X given an observed value y of another random variable Y is the distribution we would use for X if we were to learn that $Y = y$. When dealing with the conditional distribution of X given $Y = y$, it is safe to behave as if Y were the constant y. If X and Y have joint p.f., p.d.f., or p.f./p.d.f. $f(x, y)$, then the conditional p.f. or p.d.f. of X given $Y = y$ is $g_1(x|y) =$

$f(x, y)/f_2(y)$, where f_2 is the marginal p.f. or p.d.f. of Y. When it is convenient to specify a conditional distribution directly, the joint distribution can be constructed from the conditional together with the other marginal. For example,

$$f(x, y) = g_1(x|y)f_2(y) = f_1(x)g_2(y|x).$$

In this case, we have versions of the law of total probability and Bayes' theorem for random variables that allow us to calculate the other marginal and conditional.

Two random variables X and Y are independent if and only if the conditional p.f. or p.d.f. of X given $Y = y$ is the same as the marginal p.f. or p.d.f. of X for all y such that $f_2(y) > 0$. Equivalently, X and Y are independent if and only if the conditional p.f. of p.d.f. of Y given $X = x$ is the same as the marginal p.f. or p.d.f. of Y for all x such that $f_1(x) > 0$.

Exercises

1. Suppose that two random variables X and Y have the joint p.d.f. in Example 3.5.10 on page 139. Compute the conditional p.d.f. of X given $Y = y$ for each y.

2. Each student in a certain high school was classified according to her year in school (freshman, sophomore, junior, or senior) and according to the number of times that she had visited a certain museum (never, once, or more than once). The proportions of students in the various classifications are given in the following table:

	Never	Once	More than once
Freshmen	0.08	0.10	0.04
Sophomores	0.04	0.10	0.04
Juniors	0.04	0.20	0.09
Seniors	0.02	0.15	0.10

a. If a student selected at random from the high school is a junior, what is the probability that she has never visited the museum?

b. If a student selected at random from the high school has visited the museum three times, what is the probability that she is a senior?

3. Suppose that a point (X, Y) is chosen at random from the disk S defined as follows:

$$S = \{(x, y) : (x - 1)^2 + (y + 2)^2 \le 9\}.$$

Determine **(a)** the conditional p.d.f. of Y for every given value of X, and **(b)** $\Pr(Y > 0|X = 2)$.

4. Suppose that the joint p.d.f. of two random variables X and Y is as follows:

$$f(x, y) = \begin{cases} c(x + y^2) & \text{for } 0 \le x \le 1 \text{ and } 0 \le y \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

Determine **(a)** the conditional p.d.f. of X for every given value of Y, and **(b)** $\Pr(X < \frac{1}{2}|Y = \frac{1}{2})$.

5. Suppose that the joint p.d.f. of two points X and Y chosen by the process described in Example 3.6.10 is as given by Eq. (3.6.15). Determine **(a)** the conditional p.d.f. of X for every given value of Y, and **(b)** $\Pr\left(X > \frac{1}{2}\middle|Y = \frac{3}{4}\right)$.

6. Suppose that the joint p.d.f. of two random variables X and Y is as follows:

$$f(x, y) = \begin{cases} c \sin x & \text{for } 0 \le x \le \pi/2 \text{ and } 0 \le y \le 3, \\ 0 & \text{otherwise.} \end{cases}$$

Determine **(a)** the conditional p.d.f. of Y for every given value of X, and **(b)** $\Pr(1 < Y < 2|X = 0.73)$.

7. Suppose that the joint p.d.f. of two random variables X and Y is as follows:

$$f(x, y) = \begin{cases} \frac{3}{16}(4 - 2x - y) & \text{for } x > 0,\ y > 0, \\ & \quad \text{and } 2x + y < 4, \\ 0 & \text{otherwise.} \end{cases}$$

Determine **(a)** the conditional p.d.f. of Y for every given value of X, and **(b)** $\Pr(Y \ge 2|X = 0.5)$.

8. Suppose that a person's score X on a mathematics aptitude test is a number between 0 and 1, and that his score Y on a music aptitude test is also a number between 0 and 1. Suppose further that in the population of all college students in the United States, the scores X and Y are distributed according to the following joint p.d.f.:

$$f(x, y) = \begin{cases} \frac{2}{5}(2x + 3y) & \text{for } 0 \le x \le 1 \text{ and } 0 \le y \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

a. What proportion of college students obtain a score greater than 0.8 on the mathematics test?

b. If a student's score on the music test is 0.3, what is the probability that his score on the mathematics test will be greater than 0.8?

c. If a student's score on the mathematics test is 0.3, what is the probability that his score on the music test will be greater than 0.8?

9. Suppose that either of two instruments might be used for making a certain measurement. Instrument 1 yields a measurement whose p.d.f. h_1 is

$$h_1(x) = \begin{cases} 2x & \text{for } 0 < x < 1, \\ 0 & \text{otherwise.} \end{cases}$$

Instrument 2 yields a measurement whose p.d.f. h_2 is

$$h_2(x) = \begin{cases} 3x^2 & \text{for } 0 < x < 1, \\ 0 & \text{otherwise.} \end{cases}$$

Suppose that one of the two instruments is chosen at random and a measurement X is made with it.

a. Determine the marginal p.d.f. of X.

b. If the value of the measurement is $X = 1/4$, what is the probability that instrument 1 was used?

10. In a large collection of coins, the probability X that a head will be obtained when a coin is tossed varies from one coin to another, and the distribution of X in the collection is specified by the following p.d.f.:

$$f_1(x) = \begin{cases} 6x(1-x) & \text{for } 0 < x < 1, \\ 0 & \text{otherwise.} \end{cases}$$

Suppose that a coin is selected at random from the collection and tossed once, and that a head is obtained. Determine the conditional p.d.f. of X for this coin.

11. The definition of the conditional p.d.f. of X given $Y = y$ is arbitrary if $f_2(y) = 0$. The reason that this causes no serious problem is that it is highly unlikely that we will observe Y close to a value y_0 such that $f_2(y_0) = 0$. To be more precise, let $f_2(y_0) = 0$, and let $A_0 = [y_0 - \epsilon, y_0 + \epsilon]$. Also, let y_1 be such that $f_2(y_1) > 0$, and let $A_1 = [y_1 - \epsilon, y_1 + \epsilon]$. Assume that f_2 is continuous at both y_0 and y_1. Show that

$$\lim_{\epsilon \to 0} \frac{\Pr(Y \in A_0)}{\Pr(Y \in A_1)} = 0.$$

That is, the probability that Y is close to y_0 is much smaller than the probability that Y is close to y_1.

12. Let Y be the rate (calls per hour) at which calls arrive at a switchboard. Let X be the number of calls during a two-hour period. Suppose that the marginal p.d.f. of Y is

$$f_2(y) = \begin{cases} e^{-y} & \text{if } y > 0, \\ 0 & \text{otherwise,} \end{cases}$$

and that the conditional p.f. of X given $Y = y$ is

$$g_1(x|y) = \begin{cases} \dfrac{(2y)^x}{x!} e^{-2y} & \text{if } x = 0, 1, \dots, \\ 0 & \text{otherwise.} \end{cases}$$

a. Find the marginal p.f. of X. (You may use the formula $\int_0^\infty y^k e^{-y}\, dy = k!$.)

b. Find the conditional p.d.f. $g_2(y|0)$ of Y given $X = 0$.

c. Find the conditional p.d.f. $g_2(y|1)$ of Y given $X = 1$.

d. For what values of y is $g_2(y|1) > g_2(y|0)$? Does this agree with the intuition that the more calls you see, the higher you should think the rate is?

13. Start with the joint distribution of treatment group and response in Table 3.6 on page 138. For each treatment group, compute the conditional distribution of response given the treatment group. Do they appear to be very similar or quite different?

3.7 Multivariate Distributions

In this section, we shall extend the results that were developed in Sections 3.4, 3.5, and 3.6 for two random variables X and Y to an arbitrary finite number n of random variables $X_1, \dots, X_n$. In general, the joint distribution of more than two random variables is called a multivariate distribution. *The theory of statistical inference (the subject of the part of this book beginning with Chapter 7) relies on mathematical models for observable data in which each observation is a random variable. For this reason, multivariate distributions arise naturally in the mathematical models for data. The most commonly used model will be one in which the individual data random variables are conditionally independent given one or two other random variables.*

Joint Distributions

Example 3.7.1 A Clinical Trial. Suppose that m patients with a certain medical condition are given a treatment, and each patient either recovers from the condition or fails to recover. For each $i = 1, \ldots, m$, we can let $X_i = 1$ if patient i recovers and $X_i = 0$ if not. We might also believe that there is a random variable P having a continuous distribution taking values between 0 and 1 such that, if we knew that $P = p$, we would say that the m patients recover or fail to recover independently of each other each with probability p of recovery. We now have named $n = m + 1$ random variables in which we are interested. ◀

The situation described in Example 3.7.1 requires us to construct a joint distribution for n random variables. We shall now provide definitions and examples of the important concepts needed to discuss multivariate distributions.

Definition 3.7.1 Joint Distribution Function/c.d.f. The *joint c.d.f.* of n random variables $X_1, \ldots, X_n$ is the function F whose value at every point $(x_1, \ldots, x_n)$ in n-dimensional space R^n is specified by the relation

$$F(x_1, \ldots, x_n) = \Pr(X_1 \le x_1, X_2 \le x_2, \ldots, X_n \le x_n). \tag{3.7.1}$$

Every multivariate c.d.f. satisfies properties similar to those given earlier for univariate and bivariate c.d.f.'s.

Example 3.7.2 Failure Times. Suppose that a machine has three parts, and part i will fail at time X_i for $i = 1, 2, 3$. The following function might be the joint c.d.f. of X_1, X_2, and X_3:

$$F(x_1, x_2, x_3) = \begin{cases} (1 - e^{-x_1})(1 - e^{-2x_2})(1 - e^{-3x_3}) & \text{for } x_1, x_2, x_3 \ge 0, \\ 0 & \text{otherwise.} \end{cases}$$ ◀

Vector Notation In the study of the joint distribution of n random variables $X_1, \ldots, X_n$, it is often convenient to use the vector notation $\boldsymbol{X} = (X_1, \ldots, X_n)$ and to refer to $\boldsymbol{X}$ as a *random vector*. Instead of speaking of the joint distribution of the random variables $X_1, \ldots, X_n$ with a joint c.d.f. $F(x_1, \ldots, x_n)$, we can simply speak of the distribution of the random vector $\boldsymbol{X}$ with c.d.f. $F(\boldsymbol{x})$. When this vector notation is used, it must be kept in mind that if $\boldsymbol{X}$ is an n-dimensional random vector, then its c.d.f. is defined as a function on n-dimensional space R^n. At each point $\boldsymbol{x} = (x_1, \ldots, x_n) \in R^n$, the value of $F(\boldsymbol{x})$ is specified by Eq. (3.7.1).

Definition 3.7.2 Joint Discrete Distribution/p.f. It is said that n random variables $X_1, \ldots, X_n$ have a *discrete joint distribution* if the random vector $(X_1, \ldots, X_n)$ can have only a finite number or an infinite sequence of different possible values $(x_1, \ldots, x_n)$ in R^n. The *joint p.f.* of $X_1, \ldots, X_n$ is then defined as the function f such that for every point $(x_1, \ldots, x_n) \in R^n$,

$$f(x_1, \ldots, x_n) = \Pr(X_1 = x_1, \ldots, X_n = x_n).$$

In vector notation, Definition 3.7.2 says that the random vector $\boldsymbol{X}$ has a discrete distribution and that its p.f. is specified at every point $\boldsymbol{x} \in R^n$ by the relation

$$f(\boldsymbol{x}) = \Pr(\boldsymbol{X} = \boldsymbol{x}).$$

The following result is a simple generalization of Theorem 3.4.2.

Theorem 3.7.1 If $\boldsymbol{X}$ has a joint discrete distribution with joint p.f. f, then for every subset $C \subset R^n$,

$$\Pr(\boldsymbol{X} \in C) = \sum_{x \in C} f(\boldsymbol{x}). \qquad \blacksquare$$

It is easy to show that, if each of $X_1, \ldots, X_n$ has a discrete distribution, then $\boldsymbol{X} = (X_1, \ldots, X_n)$ has a discrete joint distribution.

Example 3.7.3 A Clinical Trial. Consider the m patients in Example 3.7.1. Suppose for now that $P = p$ is known so that we don't treat it as a random variable. The joint p.f. of $\boldsymbol{X} = (X_1, \ldots, X_m)$ is

$$f(\boldsymbol{x}) = p^{x_1+\cdots+x_m}(1-p)^{m-x_1-\cdots-x_m},$$

for all $x_i \in \{0, 1\}$ and 0 otherwise. ◀

Definition 3.7.3 Continuous Distribution/p.d.f. It is said that n random variables $X_1, \ldots, X_n$ have a continuous joint distribution if there is a nonnegative function f defined on R^n such that for every subset $C \subset R^n$,

$$\Pr[(X_1, \ldots, X_n) \in C] = \int \underset{C}{\cdots} \int f(x_1, \ldots x_n)\, dx_1 \cdots dx_n, \tag{3.7.2}$$

if the integral exists. The function f is called the *joint p.d.f.* of $X_1, \ldots, X_n$.

In vector notation, $f(\boldsymbol{x})$ denotes the p.d.f. of the random vector $\boldsymbol{X}$ and Eq. (3.7.2) could be rewritten more simply in the form

$$\Pr(\boldsymbol{X} \in C) = \int \underset{C}{\cdots} \int f(\boldsymbol{x})\, d\boldsymbol{x}.$$

Theorem 3.7.2 If the joint distribution of $X_1, \ldots, X_n$ is continuous, then the joint p.d.f. f can be derived from the joint c.d.f. F by using the relation

$$f(x_1, \ldots, x_n) = \frac{\partial^n F(x_1, \ldots, x_n)}{\partial x_1 \cdots \partial x_n}$$

at all points $(x_1, \ldots, x_n)$ at which the derivative in this relation exists. ■

Example 3.7.4 Failure Times. We can find the joint p.d.f. for the three random variables in Example 3.7.2 by applying Theorem 3.7.2. The third-order mixed partial is easily calculated to be

$$f(x_1, x_2, x_3) = \begin{cases} 6e^{-x_1-2x_2-3x_3} & \text{for } x_1, x_2, x_3 > 0, \\ 0 & \text{otherwise.} \end{cases}$$ ◀

It is important to note that, even if each of $X_1, \ldots, X_n$ has a continuous distribution, the vector $\boldsymbol{X} = (X_1, \ldots, X_n)$ might not have a continuous joint distribution. See Exercise 9 in this section.

Example 3.7.5 Service Times in a Queue. A queue is a system in which customers line up for service and receive their service according to some algorithm. A simple model is the single-server queue, in which all customers wait for a single server to serve everyone ahead of them in the line and then they get served. Suppose that n customers arrive at a single-server queue for service. Let X_i be the time that the server spends serving customer i for $i = 1, \ldots, n$. We might use a joint distribution for $\boldsymbol{X} = (X_1, \ldots, X_n)$ with joint p.d.f. of the form

$$f(\boldsymbol{x}) = \begin{cases} \dfrac{c}{\left(2 + \sum_{i=1}^{n} x_i\right)^{n+1}} & \text{for all } x_i > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{3.7.3}$$

We shall now find the value of c such that the function in Eq. (3.7.3) is a joint p.d.f. We can do this by integrating over each variable $x_1, \ldots, x_n$ in succession (starting with x_n). The first integral is

$$\int_0^\infty \frac{c}{(2 + x_1 + \cdots + x_n)^{n+1}}\, dx_n = \frac{c/n}{(2 + x_1 + \cdots + x_{n-1})^n}. \tag{3.7.4}$$

The right-hand side of Eq. (3.7.4) is in the same form as the original p.d.f. except that n has been reduced to $n-1$ and c has been divided by n. It follows that when we integrate over the variable x_i (for $i = n-1, n-2, \ldots, 1$), the result will be in the same form with n reduced to $i-1$ and c divided by $n(n-1)\cdots i$. The result of integrating all coordinates except x_1 is then

$$\frac{c/n!}{(2 + x_1)^2}, \quad \text{for } x_1 > 0.$$

Integrating x_1 out of this yields $c/[2(n!)]$, which must equal 1, so $c = 2(n!)$. ◀

Mixed Distributions

Example 3.7.6 Arrivals at a Queue. In Example 3.7.5, we introduced the single-server queue and discussed service times. Some features that influence the performance of a queue are the rate at which customers arrive and the rate at which customers are served. Let Z stand for the rate at which customers are served, and let Y stand for the rate at which customers arrive at the queue. Finally, let W stand for the number of customers that arrive during one day. Then W is discrete while Y and Z could be continuous random variables. A possible joint p.f./p.d.f. for these three random variables is

$$f(y, z, w) = \begin{cases} 6e^{-3z-10y}(8y)^w/w! & \text{for } z, y > 0 \text{ and } w = 0, 1, \ldots, \\ 0 & \text{otherwise.} \end{cases}$$

We can verify this claim shortly. ◀

Definition 3.7.4 Joint p.f./p.d.f. Let $X_1, \ldots, X_n$ be random variables, some of which have a continuous joint distribution and some of which have discrete distributions; their joint distribution would then be represented by a function f that we call the *joint p.f./p.d.f*. The function has the property that the probability that $\boldsymbol{X}$ lies in a subset $C \subset R^n$ is calculated by summing $f(\boldsymbol{x})$ over the values of the coordinates of $\boldsymbol{x}$ that correspond to the discrete random variables and integrating over those coordinates that correspond to the continuous random variables for all points $\boldsymbol{x} \in C$.

Example 3.7.7 Arrivals at a Queue. We shall now verify that the proposed p.f./p.d.f. in Example 3.7.6 actually sums and integrates to 1 over all values of (y, z, w). We must sum over w and integrate over y and z. We have our choice of in what order to do them. It is not difficult to see that we can factor f as $f(y, z, w) = h_2(z)h_{13}(y, w)$, where

$$h_2(z) = \begin{cases} 6e^{-3z} & \text{for } z > 0, \\ 0 & \text{otherwise,} \end{cases}$$

$$h_{13}(y, w) = \begin{cases} e^{-10y}(8y)^w/w! & \text{for } y > 0 \text{ and } w = 0, 1, \ldots, \\ 0 & \text{otherwise.} \end{cases}$$

So we can integrate z out first to get

$$\int_{-\infty}^{\infty} f(y, z, w)dz = h_{13}(y, w) \int_0^{\infty} 6e^{-3z}dz = 2h_{13}(y, w).$$

Integrating y out of $h_{13}(y, w)$ is possible, but not pleasant. Instead, notice that $(8y)^w/w!$ is the wth term in the Taylor expansion of e^{8y}. Hence,

$$\sum_{w=0}^{\infty} 2h_{13}(y, w) = 2e^{-10y} \sum_{w=0}^{\infty} \frac{(8y)^w}{w!} = 2e^{-10y}e^{8y} = 2e^{-2y},$$

for $y > 0$ and 0 otherwise. Finally, integrating over y yields 1. ◀

Example 3.7.8 A Clinical Trial. In Example 3.7.1, one of the random variables P has a continuous distribution, and the others $X_1, \ldots, X_m$ have discrete distributions. A possible joint p.f./p.d.f. for $(X_1, \ldots, X_m, P)$ is

$$f(\boldsymbol{x}, p) = \begin{cases} p^{x_1+\cdots+x_m}(1-p)^{m-x_1-\cdots-x_m} & \text{for all } x_i \in \{0, 1\} \text{ and } 0 \le p \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

We can find probabilities based on this function. Suppose, for example, that we want the probability that there is exactly one success among the first two patients, that is, $\Pr(X_1 + X_2 = 1)$. We must integrate $f(\boldsymbol{x}, p)$ over p and sum over all values of $\boldsymbol{x}$ that have $x_1 + x_2 = 1$. For purposes of illustration, suppose that $m = 4$. First, factor out $p^{x_1+x_2}(1-p)^{2-x_1-x_2} = p(1-p)$, which yields

$$f(\boldsymbol{x}, p) = [p(1-p)]p^{x_3+x_4}(1-p)^{2-x_3-x_4},$$

for $x_3, x_4 \in \{0, 1\}$, $0 < p < 1$, and $x_1 + x_2 = 1$. Summing over x_3 yields

$$[p(1-p)]\left(p^{x_4}(1-p)^{1-x_4}(1-p) + pp^{x_4}(1-p)^{1-x_4}\right) = [p(1-p)]p^{x_4}(1-p)^{1-x_4}.$$

Summing this over x_4 gives $p(1-p)$. Next, integrate over p to get $\int_0^1 p(1-p)dp = 1/6$. Finally, note that there are two (x_1, x_2) vectors, $(1, 0)$ and $(0, 1)$, that have $x_1 + x_2 = 1$, so $\Pr(X_1 + X_2 = 1) = (1/6) + (1/6) = 1/3$. ◀

Marginal Distributions

Deriving a Marginal p.d.f. If the joint distribution of n random variables $X_1, \ldots, X_n$ is known, then the marginal distribution of each single random variable X_i can be derived from this joint distribution. For example, if the joint p.d.f. of $X_1, \ldots, X_n$ is f, then the marginal p.d.f. f_1 of X_1 is specified at every value x_1 by the relation

$$f_1(x_1) = \underbrace{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty}}_{n-1} f(x_1, \ldots, x_n)\, dx_2 \cdots dx_n.$$

More generally, the marginal joint p.d.f. of any k of the n random variables $X_1, \ldots, X_n$ can be found by integrating the joint p.d.f. over all possible values of the other $n - k$ variables. For example, if f is the joint p.d.f. of four random variables X_1, X_2, X_3, and X_4, then the marginal bivariate p.d.f. f_{24} of X_2 and X_4 is specified at each point (x_2, x_4) by the relation

$$f_{24}(x_2, x_4) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x_1, x_2, x_3, x_4)\, dx_1\, dx_3.$$

Example 3.7.9 Service Times in a Queue. Suppose that $n = 5$ in Example 3.7.5 and that we want the marginal bivariate p.d.f. of (X_1, X_4). We must integrate Eq. (3.7.3) over x_2, x_3, and x_5. Since the joint p.d.f. is symmetric with respect to permutations of the coordinates of $\boldsymbol{x}$, we shall just integrate over the last three variables and then change the names of the remaining variables to x_1 and x_4. We already saw how to do this in Example 3.7.5. The result is

$$f_{12}(x_1, x_2) = \begin{cases} \dfrac{4}{(2 + x_1 + x_2)^3} & \text{for } x_1, x_2 > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{3.7.5}$$

Then f_{14} is just like (3.7.5) with all the 2 subscripts changed to 4. The univariate marginal p.d.f. of each X_i is

$$f_i(x_i) = \begin{cases} \dfrac{2}{(2 + x_i)^2} & \text{for } x_i > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{3.7.6}$$

So, for example, if we want to know how likely it is that a customer will have to wait longer than three time units, we can calculate $\Pr(X_i > 3)$ by integrating the function in Eq. (3.7.6) from 3 to ∞. The result is 0.4. ◀

If n random variables $X_1, \ldots, X_n$ have a discrete joint distribution, then the marginal joint p.f. of each subset of the n variables can be obtained from relations similar to those for continuous distributions. In the new relations, the integrals are replaced by sums.

Deriving a Marginal c.d.f. Consider now a joint distribution for which the joint c.d.f. of $X_1, \ldots, X_n$ is F. The marginal c.d.f. F_1 of X_1 can be obtained from the following relation:

$$\begin{aligned} F_1(x_1) &= \Pr(X_1 \le x_1) = \Pr(X_1 \le x_1, X_2 < \infty, \ldots, X_n < \infty) \\ &= \lim_{x_2, \ldots, x_n \to \infty} F(x_1, x_2, \ldots, x_n). \end{aligned}$$

Example 3.7.10 Failure Times. We can find the marginal c.d.f. of X_1 from the joint c.d.f. in Example 3.7.2 by letting x_2 and x_3 go to ∞. The limit is $F_1(x_1) = 1 - e^{-x_1}$ for $x_1 \ge 0$ and 0 otherwise. ◀

More generally, the marginal joint c.d.f. of any k of the n random variables $X_1, \ldots, X_n$ can be found by computing the limiting value of the n-dimensional c.d.f. F as $x_j \to \infty$ for each of the other $n - k$ variables x_j. For example, if F is the joint c.d.f. of four random variables X_1, X_2, X_3, and X_4, then the marginal bivariate c.d.f. F_{24} of X_2 and X_4 is specified at every point (x_2, x_4) by the relation

$$F_{24}(x_2, x_4) = \lim_{x_1, x_3 \to \infty} F(x_1, x_2, x_3, x_4).$$

Example 3.7.11 Failure Times. We can find the marginal bivariate c.d.f. of X_1 and X_3 from the joint c.d.f. in Example 3.7.2 by letting x_2 go to ∞. The limit is

$$F_{13}(x_1, x_3) = \begin{cases} (1 - e^{-x_1})(1 - e^{-3x_3}) & \text{for } x_1, x_3 \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$ ◀

Independent Random Variables

Definition 3.7.5 Independent Random Variables. It is said that n random variables $X_1, \ldots, X_n$ are *independent* if, for every n sets $A_1, A_2, \ldots, A_n$ of real numbers,

$$\begin{aligned} &\Pr(X_1 \in A_1, X_2 \in A_2, \ldots, X_n \in A_n) \\ &\quad = \Pr(X_1 \in A_1) \Pr(X_2 \in A_2) \cdots \Pr(X_n \in A_n). \end{aligned}$$

If $X_1, \ldots, X_n$ are independent, it follows easily that the random variables in every nonempty subset of $X_1, \ldots, X_n$ are also independent. (See Exercise 11.)

There is a generalization of Theorem 3.5.4.

Theorem 3.7.3 Let F denote the joint c.d.f. of $X_1, \ldots, X_n$, and let F_i denote the marginal univariate c.d.f. of X_i for $i = 1, \ldots, n$. The variables $X_1, \ldots, X_n$ are independent if and only if, for all points $(x_1, x_2, \ldots, x_n) \in R^n$,

$$F(x_1, x_2, \ldots, x_n) = F_1(x_1)F_2(x_2) \cdots F_n(x_n).$$ ■

Theorem 3.7.3 says that $X_1, \ldots, X_n$ are independent if and only if their joint c.d.f. is the product of their n individual marginal c.d.f.'s. It is easy to check that the three random variables in Example 3.7.2 are independent using Theorem 3.7.3.

There is also a generalization of Corollary 3.5.1.

Theorem 3.7.4 If $X_1, \ldots, X_n$ have a continuous, discrete, or mixed joint distribution for which the joint p.d.f., joint p.f., or joint p.f./p.d.f. is f, and if f_i is the marginal univariate p.d.f. or p.f. of X_i $(i = 1, \ldots, n)$, then $X_1, \ldots, X_n$ are independent if and only if the following relation is satisfied at all points $(x_1, x_2, \ldots, x_n) \in R^n$:

$$f(x_1, x_2, \ldots, x_n) = f_1(x_1)f_2(x_2) \cdots f_n(x_n). \tag{3.7.7}$$ ■

Example 3.7.12 Service Times in a Queue. In Example 3.7.9, we can multiply together the two univariate marginal p.d.f.'s of X_1 and X_2 calculated using Eq. (3.7.6) and see that the product does *not* equal the bivariate marginal p.d.f. of (X_1, X_2) in Eq. (3.7.5). So X_1 and X_2 are not independent. ◀

Definition 3.7.6 Random Samples/i.i.d./Sample Size. Consider a given probability distribution on the real line that can be represented by either a p.f. or a p.d.f. f. It is said that n random variables $X_1, \ldots, X_n$ form a *random sample* from this distribution if these random variables are independent and the marginal p.f. or p.d.f. of each of them is f. Such random variables are also said to be *independent and identically distributed*, abbreviated *i.i.d.* We refer to the number n of random variables as the *sample size*.

Definition 3.7.6 says that $X_1, \ldots, X_n$ form a random sample from the distribution represented by f if their joint p.f. or p.d.f. g is specified as follows at all points $(x_1, x_2, \ldots, x_n) \in R^n$:

$$g(x_1, \ldots, x_n) = f(x_1)f(x_2) \cdots f(x_n).$$

Clearly, an i.i.d. sample cannot have a mixed joint distribution.

Example 3.7.13 Lifetimes of Light Bulbs. Suppose that the lifetime of each light bulb produced in a certain factory is distributed according to the following p.d.f.:

$$f(x) = \begin{cases} xe^{-x} & \text{for } x > 0, \\ 0 & \text{otherwise.} \end{cases}$$

We shall determine the joint p.d.f. of the lifetimes of a random sample of n light bulbs drawn from the factory's production.

The lifetimes $X_1, \ldots, X_n$ of the selected bulbs will form a random sample from the p.d.f. f. For typographical simplicity, we shall use the notation $\exp(v)$ to denote the exponential e^v when the expression for v is complicated. Then the joint p.d.f. g of $X_1, \ldots, X_n$ will be as follows: If $x_i > 0$ for $i = 1, \ldots, n$,

$$\begin{aligned} g(x_1, \ldots, x_n) &= \prod_{i=1}^{n} f(x_i) \\ &= \left(\prod_{i=1}^{n} x_i \right) \exp\left(-\sum_{i=1}^{n} x_i \right). \end{aligned}$$

Otherwise, $g(x_1, \ldots, x_n) = 0$.

Every probability involving the n lifetimes $X_1, \ldots, X_n$ can in principle be determined by integrating this joint p.d.f. over the appropriate subset of R^n. For example, if C is the subset of points $(x_1, \ldots, x_n)$ such that $x_i > 0$ for $i = 1, \ldots, n$ and $\sum_{i=1}^{n} x_i < a$, where a is a given positive number, then

$$\Pr\left(\sum_{i=1}^{n} X_i < a \right) = \int \underset{C}{\cdots} \int \left(\prod_{i=1}^{n} x_i \right) \exp\left(-\sum_{i=1}^{n} x_i \right) dx_1 \cdots dx_n. \qquad ◀$$

The evaluation of the integral given at the end of Example 3.7.13 may require a considerable amount of time without the aid of tables or a computer. Certain other probabilities, however, can be evaluated easily from the basic properties of continuous distributions and random samples. For example, suppose that for the conditions of Example 3.7.13 it is desired to find $\Pr(X_1 < X_2 < \cdots < X_n)$. Since the random variables $X_1, \ldots, X_n$ have a continuous joint distribution, the probability that at least two of these random variables will have the same value is 0. In fact, the probability is 0 that the vector $(X_1, \ldots, X_n)$ will belong to each specific subset of R^n for which the n-dimensional volume is 0. Furthermore, since $X_1, \ldots, X_n$ are independent and identically distributed, each of these variables is equally likely to be the smallest of the n lifetimes, and each is equally likely to be the largest. More generally, if the lifetimes $X_1, \ldots, X_n$ are arranged in order from the smallest to the largest, each particular ordering of $X_1, \ldots, X_n$ is as likely to be obtained as any other ordering. Since there are $n!$ different possible orderings, the probability that the particular ordering $X_1 < X_2 < \cdots < X_n$ will be obtained is $1/n!$. Hence,

$$\Pr(X_1 < X_2 < \cdots < X_n) = \frac{1}{n!}.$$

Conditional Distributions

Suppose that n random variables $X_1, \ldots, X_n$ have a continuous joint distribution for which the joint p.d.f. is f and that f_0 denotes the marginal joint p.d.f. of the $k < n$ random variables $X_1, \ldots, X_k$. Then for all values of $x_1, \ldots, x_k$ such that $f_0(x_1, \ldots, x_k) > 0$, the conditional p.d.f. of $(X_{k+1}, \ldots, X_n)$ given that $X_1 = x_1, \ldots, X_k = x_k$ is defined as follows:

$$g_{k+1\ldots n}(x_{k+1}, \ldots, x_n|x_1, \ldots, x_k) = \frac{f(x_1, x_2, \ldots, x_n)}{f_0(x_1, \ldots, x_k)}.$$

The definition above generalizes to arbitrary joint distributions as follows.

Definition 3.7.7 Conditional p.f., p.d.f., or p.f./p.d.f. Suppose that the random vector $\boldsymbol{X} = (X_1, \ldots, X_n)$ is divided into two subvectors $\boldsymbol{Y}$ and $\boldsymbol{Z}$, where $\boldsymbol{Y}$ is a k-dimensional random vector comprising k of the n random variables in $\boldsymbol{X}$, and $\boldsymbol{Z}$ is an $(n-k)$-dimensional random vector comprising the other $n-k$ random variables in $\boldsymbol{X}$. Suppose also that the n-dimensional joint p.f., p.d.f., or p.f./p.d.f. of $(\boldsymbol{Y}, \boldsymbol{Z})$ is f and that the marginal $(n-k)$-dimensional p.f., p.d.f., or p.f./p.d.f. of $\boldsymbol{Z}$ is f_2. Then for every given point $\boldsymbol{z} \in R^{n-k}$ such that $f_2(\boldsymbol{z}) > 0$, the conditional k-dimensional p.f., p.d.f., or p.f./p.d.f. g_1 of $\boldsymbol{Y}$ given $\boldsymbol{Z} = \boldsymbol{z}$ is defined as follows:

$$g_1(\boldsymbol{y}|\boldsymbol{z}) = \frac{f(\boldsymbol{y}, \boldsymbol{z})}{f_2(\boldsymbol{z})} \qquad \text{for } \boldsymbol{y} \in R^k. \tag{3.7.8}$$

Eq. (3.7.8) can be rewritten as

$$f(\boldsymbol{y}, \boldsymbol{z}) = g_1(\boldsymbol{y}|\boldsymbol{z}) f_2(\boldsymbol{z}), \tag{3.7.9}$$

which allows construction of the joint distribution from a conditional distribution and a marginal distribution. As in the bivariate case, it is safe to assume that $f(\boldsymbol{y}, \boldsymbol{z}) = 0$ whenever $f_2(\boldsymbol{z}) = 0$. Then Eq. (3.7.9) holds for all $\boldsymbol{y}$ and $\boldsymbol{z}$ even though $g_1(\boldsymbol{y}|\boldsymbol{z})$ is not uniquely defined.

Example 3.7.14 Service Times in a Queue. In Example 3.7.9, we calculated the marginal bivariate distribution of two service times $\boldsymbol{Z} = (X_1, X_2)$. We can now find the conditional three-dimensional p.d.f. of $\boldsymbol{Y} = (X_3, X_4, X_5)$ given $\boldsymbol{Z} = (x_1, x_2)$ for every pair (x_1, x_2) such that $x_1, x_2 > 0$:

$$\begin{aligned} g_1(x_3, x_4, x_5|x_1, x_2) &= \frac{f(x_1, \ldots, x_5)}{f_{12}(x_1, x_2)} \\ &= \left(\frac{240}{(2 + x_1 + \cdots + x_5)^6}\right)\left(\frac{4}{(2 + x_1 + x_2)^3}\right)^{-1} \\ &= \frac{60(2 + x_1 + x_2)^3}{(2 + x_1 + \cdots + x_5)^6}, \end{aligned} \tag{3.7.10}$$

for $x_3, x_4, x_5 > 0$, and 0 otherwise. The joint p.d.f. in (3.7.10) looks like a bunch of symbols, but it can be quite useful. Suppose that we observe $X_1 = 4$ and $X_2 = 6$. Then

$$g_1(x_3, x_4, x_5|4.6) = \begin{cases} \dfrac{103{,}680}{(12 + x_3 + x_4 + x_5)^6} & \text{for } x_3, x_4, x_5 > 0, \\ 0 & \text{otherwise.} \end{cases}$$

We can now calculate the conditional probability that $X_3 > 3$ given $X_1 = 4$, $X_2 = 6$:

$$\begin{aligned}
\Pr(X_3 > 3 | X_1 = 4,\ X_2 = 6) &= \int_3^\infty \int_0^\infty \int_0^\infty \frac{10{,}360}{(12 + x_3 + x_4 + x_5)^6} dx_5 dx_4 dx_3 \\
&= \int_3^\infty \int_0^\infty \frac{20{,}736}{(12 + x_3 + x_4)^5} dx_4 dx_3 \\
&= \int_3^\infty \frac{5184}{(12 + x_3)^4} dx_3 \\
&= \frac{1728}{15^3} = 0.512.
\end{aligned}$$

Compare this to the calculation of $\Pr(X_3 > 3) = 0.4$ at the end of Example 3.7.9. After learning that the first two service times are a bit longer than three time units, we revise the probability that $X_3 > 3$ upward to reflect what we learned from the first two observations. If the first two service times had been small, the conditional probability that $X_3 > 3$ would have been smaller than 0.4. For example, $\Pr(X_3 > 3 | X_1 = 1,\ X_2 = 1.5) = 0.216$. ◄

Example 3.7.15 **Determining a Marginal Bivariate p.d.f.** Suppose that Z is a random variable for which the p.d.f. f_0 is as follows:

$$f_0(z) = \begin{cases} 2e^{-2z} & \text{for } z > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{3.7.11}$$

Suppose, furthermore, that for every given value $Z = z > 0$ two other random variables X_1 and X_2 are independent and identically distributed and the conditional p.d.f. of each of these variables is as follows:

$$g(x|z) = \begin{cases} ze^{-zx} & \text{for } x > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{3.7.12}$$

We shall determine the marginal joint p.d.f. of (X_1, X_2).

Since X_1 and X_2 are i.i.d. for each given value of Z, their conditional joint p.d.f. when $Z = z > 0$ is

$$g_{12}(x_1, x_2|z) = \begin{cases} z^2 e^{-z(x_1+x_2)} & \text{for } x_1, x_2 > 0, \\ 0 & \text{otherwise.} \end{cases}$$

The joint p.d.f. f of (Z, X_1, X_2) will be positive only at those points (z, x_1, x_2) such that $x_1, x_2, z > 0$. It now follows that, at every such point,

$$f(z, x_1, x_2) = f_0(z) g_{12}(x_1, x_2|z) = 2z^2 e^{-z(2+x_1+x_2)}.$$

For $x_1 > 0$ and $x_2 > 0$, the marginal joint p.d.f. $f_{12}(x_1, x_2)$ of X_1 and X_2 can be determined either using integration by parts or some special results that will arise in Sec. 5.7:

$$f_{12}(x_1, x_2) = \int_0^\infty f(z, x_1, x_2)\, dz = \frac{4}{(2 + x_1 + x_2)^3},$$

for $x_1, x_2 > 0$. The reader will note that this p.d.f. is the same as the marginal bivariate p.d.f. of (X_1, X_2) found in Eq. (3.7.5).

From this marginal bivariate p.d.f., we can evaluate probabilities involving X_1 and X_2, such as $\Pr(X_1 + X_2 < 4)$. We have

$$\Pr(X_1 + X_2 < 4) = \int_0^4 \int_0^{4-x_2} \frac{4}{(2 + x_1 + x_2)^3}\, dx_1\, dx_2 = \frac{4}{9}.$$ ◄

Example 3.7.16 Service Times in a Queue. We can think of the random variable Z in Example 3.7.15 as the rate at which customers are served in the queue of Example 3.7.5. With this interpretation, it is useful to find the conditional distribution of the rate Z after we observe some of the service times such as X_1 and X_2.

For every value of z, the conditional p.d.f. of Z given $X_1 = x_1$ and $X_2 = x_2$ is

$$g_0(z|x_1, x_2) = \frac{f(z, x_1, x_2)}{f_{12}(x_1, x_2)} = \begin{cases} \frac{1}{2}(2 + x_1 + x_2)^3 z^2 e^{-z(2+x_1+x_2)} & \text{for } z > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{3.7.13}$$

Finally, we shall evaluate $\Pr(Z \le 1|X_1 = 1,\ X_2 = 4)$. We have

$$\Pr(Z \le 1|X_1 = 1,\ X_2 = 4) = \int_0^1 g_0(z|1, 4)\, dz = \int_0^1 171.5 z^2 e^{-7z}\, dz = 0.9704.$$ ◄

Law of Total Probability and Bayes' Theorem Example 3.7.15 contains an example of the multivariate version of the law of total probability, while Example 3.7.16 contains an example of the multivariate version of Bayes' theorem. The proofs of the general versions are straightforward consequences of Definition 3.7.7.

Theorem 3.7.5 Multivariate Law of Total Probability and Bayes' Theorem. Assume the conditions and notation given in Definition 3.7.7. If $\boldsymbol{Z}$ has a continuous joint distribution, the marginal p.d.f. of $\boldsymbol{Y}$ is

$$f_1(\boldsymbol{y}) = \underbrace{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty}}_{n-k} g_1(\boldsymbol{y}|\boldsymbol{z}) f_2(\boldsymbol{z})\, d\boldsymbol{z}, \tag{3.7.14}$$

and the conditional p.d.f. of $\boldsymbol{Z}$ given $\boldsymbol{Y} = \boldsymbol{y}$ is

$$g_2(\boldsymbol{z}|\boldsymbol{y}) = \frac{g_1(\boldsymbol{y}|\boldsymbol{z}) f_2(\boldsymbol{z})}{f_1(\boldsymbol{y})}. \tag{3.7.15}$$

If $\boldsymbol{Z}$ has a discrete joint distribution, then the multiple integral in (3.7.14) must be replaced by a multiple summation. If $\boldsymbol{Z}$ has a mixed joint distribution, the multiple integral must be replaced by integration over those coordinates with continuous distributions and summation over those coordinates with discrete distributions. ■

Conditionally Independent Random Variables In Examples 3.7.15 and 3.7.16, $\boldsymbol{Z}$ is the single random variable Z and $\boldsymbol{Y} = (X_1, X_2)$. These examples also illustrate the use of conditionally independent random variables. That is, X_1 and X_2 are conditionally independent given $Z = z$ for all $z > 0$. In Example 3.7.16, we said that Z was the rate at which customers were served. When this rate is unknown, it is a major source of uncertainty. Partitioning the sample space by the values of the rate Z and then conditioning on each value of Z removes a major source of uncertainty for part of the calculation.

In general, conditional independence for random variables is similar to conditional independence for events.

Definition 3.7.8 Conditionally Independent Random Variables. Let $\boldsymbol{Z}$ be a random vector with joint p.f., p.d.f., or p.f./p.d.f. $f_0(\boldsymbol{z})$. Several random variables $X_1, \ldots, X_n$ are *conditionally independent given* $\boldsymbol{Z}$ if, for all $\boldsymbol{z}$ such that $f_0(\boldsymbol{z}) > 0$, we have

$$g(\boldsymbol{x}|\boldsymbol{z}) = \prod_{i=1}^{n} g_i(x_i|\boldsymbol{z}),$$

where $g(\boldsymbol{x}|\boldsymbol{z})$ stands for the conditional multivariate p.f., p.d.f., or p.f./p.d.f. of $\boldsymbol{X}$ given $\boldsymbol{Z} = \boldsymbol{z}$ and $g_i(x_i|\boldsymbol{z})$ stands for the conditional univariate p.f. or p.d.f. of X_i given $\boldsymbol{Z} = \boldsymbol{z}$.

In Example 3.7.15, $g_i(x_i|z) = ze^{-zx_i}$ for $x_i > 0$ and $i = 1, 2$.

Example 3.7.17 A Clinical Trial. In Example 3.7.8, the joint p.f./p.d.f. given there was constructed by assuming that $X_1, \ldots, X_m$ were conditionally independent given $P = p$ each with the same conditional p.f., $g_i(x_i|p) = p^{x_i}(1-p)^{1-x_i}$ for $x_i \in \{0, 1\}$ and that P had the uniform distribution on the interval [0, 1]. These assumptions produce, in the notation of Definition 3.7.8,

$$g(\boldsymbol{x}|p) = \begin{cases} p^{x_1+\cdots+x_m}(1-p)^{40-x_1-\cdots-x_m} & \text{for all } x_i \in \{0, 1\} \text{ and } 0 \le p \le 1, \\ 0 & \text{otherwise,} \end{cases}$$

for $0 \le p \le 1$. Combining this with the marginal p.d.f. of P, $f_2(p) = 1$ for $0 \le p \le 1$ and 0 otherwise, we get the joint p.f./p.d.f. given in Example 3.7.8. ◀

Conditional Versions of Past and Future Theorems We mentioned earlier that conditional distributions behave just like distributions. Hence, all theorems that we have proven and will prove in the future have conditional versions. For example, the law of total probability in Eq. (3.7.14) has the following version conditional on another random vector $\boldsymbol{W} = \boldsymbol{w}$:

$$f_1(\boldsymbol{y}|\boldsymbol{w}) = \underbrace{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty}}_{n-k} g_1(\boldsymbol{y}|\boldsymbol{z}, \boldsymbol{w}) f_2(\boldsymbol{z}|\boldsymbol{w})\, d\boldsymbol{z}, \tag{3.7.16}$$

where $f_1(\boldsymbol{y}|\boldsymbol{w})$ stands for the conditional p.d.f., p.f., or p.f./p.d.f. of $\boldsymbol{Y}$ given $\boldsymbol{W} = \boldsymbol{w}$, $g_1(\boldsymbol{y}|\boldsymbol{z}, \boldsymbol{w})$ stands for the conditional p.d.f., p.f., or p.f./p.d.f. of $\boldsymbol{Y}$ given $(\boldsymbol{Z}, \boldsymbol{W}) = (\boldsymbol{z}, \boldsymbol{w})$, and $f_2(\boldsymbol{z}|\boldsymbol{w})$ stands for the conditional p.d.f. of $\boldsymbol{Z}$ given $\boldsymbol{W} = \boldsymbol{w}$. Using the same notation, the conditional version of Bayes' theorem is

$$g_2(\boldsymbol{z}|\boldsymbol{y}, \boldsymbol{w}) = \frac{g_1(\boldsymbol{y}|\boldsymbol{z}, \boldsymbol{w}) f_2(\boldsymbol{z}|\boldsymbol{w})}{f_1(\boldsymbol{y}|\boldsymbol{w})}. \tag{3.7.17}$$

Example 3.7.18 Conditioning on Random Variables in Sequence. In Example 3.7.15, we found the conditional p.d.f. of Z given $(X_1, X_2) = (x_1, x_2)$. Suppose now that there are three more observations available, X_3, X_4, and X_5, and suppose that all of $X_1, \ldots, X_5$ are conditionally i.i.d. given $Z = z$ with p.d.f. $g(x|z)$. We shall use the conditional version of Bayes' theorem to compute the conditional p.d.f. of Z given $(X_1, \ldots, X_5) = (x_1, \ldots, x_5)$. First, we shall find the conditional p.d.f. $g_{345}(x_3, x_4, x_5|x_1, x_2, z)$ of $\boldsymbol{Y} = (X_3, X_4, X_5)$ given $Z = z$ and $\boldsymbol{W} = (X_1, X_2) = (x_1, x_2)$. We shall use the notation for p.d.f.'s in the discussion immediately preceding this example. Since $X_1, \ldots, X_5$ are conditionally i.i.d. given Z, we have that $g_1(\boldsymbol{y}|z, \boldsymbol{w})$ does not depend on $\boldsymbol{w}$. In fact,

$$g_1(\boldsymbol{y}|z, \boldsymbol{w}) = g(x_3|z)g(x_4|z)g(x_5|z) = z^3 e^{-z(x_3+x_4+x_5)},$$

for $x_3, x_4, x_5 > 0$. We also need the conditional p.d.f. of Z given $\boldsymbol{W} = \boldsymbol{w}$, which was calculated in Eq. (3.7.13), and we now denote it

$$f_2(z|\boldsymbol{w}) = \frac{1}{2}(2 + x_1 + x_2)^3 z^2 e^{-z(2+x_1+x_2)}.$$

Finally, we need the conditional p.d.f. of the last three observations given the first two. This was calculated in Example 3.7.14, and we now denote it

$$f_1(\boldsymbol{y}|\boldsymbol{w}) = \frac{60(2 + x_1 + x_2)^3}{(2 + x_1 + \cdots + x_5)^6}.$$

Now combine these using Bayes' theorem (3.7.17) to obtain

$$\begin{aligned} g_2(\boldsymbol{z}|\,\boldsymbol{y}, \boldsymbol{w}) &= \frac{z^3 e^{-z(x_3+x_4+x_5)}\frac{1}{2}(2 + x_1 + x_2)^3 z^2 e^{-z(2+x_1+x_2)}}{\dfrac{60(2 + x_1 + x_2)^3}{(2 + x_1 + \cdots + x_5)^6}} \\ &= \frac{1}{120}(2 + x_1 + \cdots + x_5)^6 z^5 e^{-z(2+x_1+\cdots+x_5)}, \end{aligned}$$

for $z > 0$. ◀

Note: Simple Rule for Creating Conditional Versions of Results. If you ever wish to determine the conditional version given $\boldsymbol{W} = \boldsymbol{w}$ of a result that you have proven, here is a simple method. Just add "conditional on $\boldsymbol{W} = \boldsymbol{w}$" to every probabilistic statement in the result. This includes all probabilities, c.d.f.'s, quantiles, names of distributions, p.d.f.'s, p.f.'s, and so on. It also includes all future probabilistic concepts that we introduce in later chapters (such as expected values and variances in Chapter 4).

Note: Independence is a Special Case of Conditional Independence. Let $X_1, \ldots, X_n$ be independent random variables, and let W be a constant random variable. That is, there is a constant c such that $\Pr(W = c) = 1$. Then $X_1, \ldots, X_n$ are also conditionally independent given $W = c$. The proof is straightforward and is left to the reader (Exercise 15). This result is not particularly interesting in its own right. Its value is the following: If we prove a result for conditionally independent random variables or conditionally i.i.d. random variables, then the same result will hold for independent random variables or i.i.d. random variables as the case may be.

Histograms

Example 3.7.19

Rate of Service. In Examples 3.7.5 and 3.7.6, we considered customers arriving at a queue and being served. Let Z stand for the rate at which customers were served, and we let $X_1, X_2, \ldots$ stand for the times that the successive customers requrired for service. Assume that $X_1, X_2, \ldots$ are conditionally i.i.d. given $Z = z$ with p.d.f.

$$g(x|z) = \begin{cases} ze^{-zx} & \text{for } x > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{3.7.18}$$

This is the same as (3.7.12) from Example 3.7.15. In that example, we modeled Z as a random variable with p.d.f. $f_0(z) = 2\exp(-2z)$ for $z > 0$. In this example, we shall assume that $X_1, \ldots, X_n$ will be observed for some large value n, and we want to think about what these observations tell us about Z. To be specific, suppose that we observe $n = 100$ service times. The first 10 times are listed here:

1.39, 0.61, 2.47, 3.35, 2.56, 3.60, 0.32, 1.43, 0.51, 0.94.

The smallest and largest observed service times from the entire sample are 0.004 and 9.60, respectively. It would be nice to have a graphical display of the entire sample of $n = 100$ service times without having to list them separately. ◀

The histogram, defined below, is a graphical display of a collection of numbers. It is particularly useful for displaying the observed values of a collection of random variables that have been modeled as conditionally i.i.d.

Definition 3.7.9 Histogram. Let $x_1, \dots, x_n$ be a collection of numbers that all lie between two values $a < b$. That is, $a \le x_i \le b$ for all $i = 1, \dots, n$. Choose some integer $k \ge 1$ and divide the interval $[a, b]$ into k equal-length subintervals of length $(b - a)/k$. For each subinterval, count how many of the numbers $x_1, \dots, x_n$ are in the subinterval. Let c_i be the count for subinterval i for $i = 1, \dots, k$. Choose a number $r > 0$. (Typically, $r = 1$ or $r = n$ or $r = n(b - a)/k$.) Draw a two-dimensional graph with the horizonal axis running from a to b. For each subinterval $i = 1, \dots, k$ draw a rectangular bar of width $(b - a)/k$ and height equal to c_i/r over the midpoint of the ith interval. Such a graph is called a *histogram*.

The choice of the number r in the definition of histogram depends on what one wishes to be displayed on the vertical axis. The shape of the histogram is identical regardless of what value one chooses for r. With $r = 1$, the height of each bar is the raw count for each subinterval, and counts are displayed on the vertical axis. With $r = n$, the height of each bar is the proportion of the set of numbers in each subinterval, and the vertical axis displays proportions. With $r = n(b - a)/k$, the area of each bar is the proportion of the set of numbers in each subinterval.

Example 3.7.20 Rate of Service. The $n = 100$ observed service times in Example 3.7.19 all lie between 0 and 10. It is convenient, in this example, to draw a histogram with horizontal axis running from 0 to 10 and divided into 10 subintervals of length 1 each. Other choices are possible, but this one will do for illustration. Figure 3.22 contains the histogram of the 100 observed service times with $r = 100$. One sees that the numbers of observed service times in the subintervals decrease as the center of the subinterval increses. This matches the behavior of the conditional p.d.f. $g(x|z)$ of the service times as a function of x for fixed z. ◀

Histograms are useful as more than just graphical displays of large sets of numbers. After we see the law of large numbers (Theorem 6.2.4), we can show that the

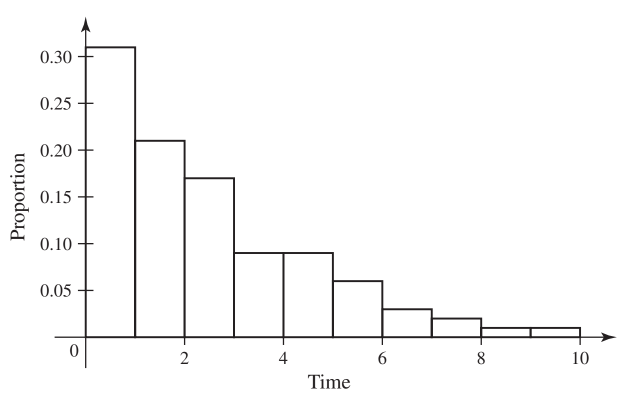


Figure 3.22 Histogram of service times for Example 3.7.20 with $a = 0$, $b = 10$, $k = 10$, and $r = 100$.

histogram of a large (conditionally) i.i.d. sample of continuous random variables is an approximation to the (conditional) p.d.f. of the random variables in the sample, so long as one uses the third choice of r, namely, $r = n(b - a)/k$.

Note: More General Histograms. Sometimes it is convenient to divide the range of the numbers to be plotted in a histogram into unequal-length subintervals. In such a case, one would typically let the height of each bar be c_i/r_i, where c_i is the raw count and r_i is proportional to the length of the ith subinterval. In this way, the area of each bar is still proportional to the count or proportion in each subinterval.

Summary

A finite collection of random variables is called a random vector. We have defined joint distributions for arbitrary random vectors. Every random vector has a joint c.d.f. Continuous random vectors have a joint p.d.f. Discrete random vectors have a joint p.f. Mixed distribution random vectors have a joint p.f./p.d.f. The coordinates of an n-dimensional random vector $\boldsymbol{X}$ are independent if the joint p.f., p.d.f., or p.f./p.d.f. $f(\boldsymbol{x})$ factors into $\prod_{i=1}^{n} f_i(x_i)$.

We can compute marginal distributions of subvectors of a random vector, and we can compute the conditional distribution of one subvector given the rest of the vector. We can construct a joint distribution for a random vector by piecing together a marginal distribution for part of the vector and a conditional distribution for the rest given the first part. There are versions of Bayes' theorem and the law of total probability for random vectors.

An n-dimensional random vector $\boldsymbol{X}$ has coordinates that are conditionally independent given $\boldsymbol{Z}$ if the conditional p.f., p.d.f., or p.f./p.d.f. $g(\boldsymbol{x}|\boldsymbol{z})$ of $\boldsymbol{X}$ given $\boldsymbol{Z} = \boldsymbol{z}$ factors into $\prod_{i=1}^{n} g_i(x_i|\boldsymbol{z})$. There are versions of Bayes' theorem, the law of total probability, and all future theorems about random variables and random vectors conditional on an arbitrary additional random vector.

Exercises

1. Suppose that three random variables X_1, X_2, and X_3 have a continuous joint distribution with the following joint p.d.f.: $f(x_1, x_2, x_3) =$

$$\begin{cases} c(x_1 + 2x_2 + 3x_3) & \text{for } 0 \le x_i \le 1 \ (i = 1, 2, 3), \\ 0 & \text{otherwise.} \end{cases}$$

Determine **(a)** the value of the constant c;
(b) the marginal joint p.d.f. of X_1 and X_3; and
(c) $\Pr\left(X_3 < \frac{1}{2} \middle| X_1 = \frac{1}{4},\ X_2 = \frac{3}{4}\right)$.

2. Suppose that three random variables X_1, X_2, and X_3 have a mixed joint distribution with p.f./p.d.f.:

$$f(x_1, x_2, x_3) = \begin{cases} c x_1^{1+x_2+x_3}(1 - x_1)^{3-x_2-x_3} & \text{if } 0 < x_1 < 1 \\ & \quad \text{and } x_2, x_3 \in \{0, 1\}, \\ 0 & \text{otherwise.} \end{cases}$$

(Notice that X_1 has a continuous distribution and X_2 and X_3 have discrete distributions.) Determine **(a)** the value of the constant c; **(b)** the marginal joint p.f. of X_2 and X_3; and **(c)** the conditional p.d.f. of X_1 given $X_2 = 1$ and $X_3 = 1$.

3. Suppose that three random variables X_1, X_2, and X_3 have a continuous joint distribution with the following joint p.d.f.: $f(x_1, x_2, x_3) =$

$$\begin{cases} c e^{-(x_1+2x_2+3x_3)} & \text{for } x_i > 0 \ (i = 1, 2, 3), \\ 0 & \text{otherwise.} \end{cases}$$

Determine **(a)** the value of the constant c; **(b)** the marginal joint p.d.f. of X_1 and X_3; and **(c)** $\Pr(X_1 < 1 | X_2 = 2,\ X_3 = 1)$.

4. Suppose that a point (X_1, X_2, X_3) is chosen at random, that is, in accordance with the uniform p.d.f., from the following set S:

$$S = \{(x_1, x_2, x_3) \colon 0 \le x_i \le 1 \text{ for } i = 1, 2, 3\}.$$

Determine:

a. $\Pr\left[\left(X_1 - \frac{1}{2}\right)^2 + \left(X_2 - \frac{1}{2}\right)^2 + \left(X_3 - \frac{1}{2}\right)^2 \le \frac{1}{4}\right]$

b. $\Pr(X_1^2 + X_2^2 + X_3^2 \le 1)$

5. Suppose that an electronic system contains n components that function independently of each other and that the probability that component i will function properly is p_i $(i = 1, \ldots, n)$. It is said that the components are connected *in series* if a necessary and sufficient condition for the system to function properly is that all n components function properly. It is said that the components are connected *in parallel* if a necessary and sufficient condition for the system to function properly is that at least one of the n components functions properly. The probability that the system will function properly is called the *reliability* of the system. Determine the reliability of the system, **(a)** assuming that the components are connected in series, and **(b)** assuming that the components are connected in parallel.

6. Suppose that the n random variables $X_1 \ldots, X_n$ form a random sample from a discrete distribution for which the p.f. is f. Determine the value of $\Pr(X_1 = X_2 = \cdots = X_n)$.

7. Suppose that the n random variables $X_1, \ldots, X_n$ form a random sample from a continuous distribution for which the p.d.f. is f. Determine the probability that at least k of these n random variables will lie in a specified interval $a \le x \le b$.

8. Suppose that the p.d.f. of a random variable X is as follows:

$$f(x) = \begin{cases} \frac{1}{n!}x^n e^{-x} & \text{for } x > 0 \\ 0 & \text{otherwise.} \end{cases}$$

Suppose also that for any given value $X = x$ $(x > 0)$, the n random variables $Y_1, \ldots, Y_n$ are i.i.d. and the conditional p.d.f. g of each of them is as follows:

$$g(y|x) = \begin{cases} \frac{1}{x} & \text{for } 0 < y < x, \\ 0 & \text{otherwise.} \end{cases}$$

Determine **(a)** the marginal joint p.d.f. of $Y_1, \ldots, Y_n$ and **(b)** the conditional p.d.f. of X for any given values of $Y_1, \ldots, Y_n$.

9. Let X be a random variable with a continuous distribution. Let $X_1 = X_2 = X$.

a. Prove that both X_1 and X_2 have a continuous distribution.

b. Prove that $\boldsymbol{X} = (X_1, X_2)$ does not have a continuous joint distribution.

10. Return to the situation described in Example 3.7.18. Let $\boldsymbol{X} = (X_1, \ldots, X_5)$ and compute the conditional p.d.f. of Z given $\boldsymbol{X} = \boldsymbol{x}$ directly in one step, as if all of $\boldsymbol{X}$ were observed at the same time.

11. Suppose that $X_1, \ldots, X_n$ are independent. Let $k < n$ and let $i_1, \ldots, i_k$ be distinct integers between 1 and n. Prove that $X_{i_1}, \ldots, X_{i_k}$ are independent.

12. Let $\boldsymbol{X}$ be a random vector that is split into three parts, $\boldsymbol{X} = (\boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{W})$. Suppose that $\boldsymbol{X}$ has a continuous joint distribution with p.d.f. $f(\boldsymbol{y}, \boldsymbol{z}, \boldsymbol{w})$. Let $g_1(\boldsymbol{y}, \boldsymbol{z}|\boldsymbol{w})$ be the conditional p.d.f. of $(\boldsymbol{Y}, \boldsymbol{Z})$ given $\boldsymbol{W} = \boldsymbol{w}$, and let $g_2(\boldsymbol{y}|\boldsymbol{w})$ be the conditional p.d.f. of $\boldsymbol{Y}$ given $\boldsymbol{W} = \boldsymbol{w}$. Prove that $g_2(\boldsymbol{y}|\boldsymbol{w}) = \int g_1(\boldsymbol{y}, \boldsymbol{z}|\boldsymbol{w})\, d\boldsymbol{z}$.

13. Let X_1, X_2, X_3 be conditionally independent given $Z = z$ for all z with the conditional p.d.f. $g(x|z)$ in Eq. (3.7.12). Also, let the marginal p.d.f. of Z be f_0 in Eq. (3.7.11). Prove that the conditional p.d.f. of X_3 given $(X_1, X_2) = (x_1, x_2)$ is $\int_0^\infty g(x_3|z) g_0(z|x_1, x_2)\, dz$, where g_0 is defined in Eq. (3.7.13). (You can prove this even if you cannot compute the integral in closed form.)

14. Consider the situation described in Example 3.7.14. Suppose that $X_1 = 5$ and $X_2 = 7$ are observed.

a. Compute the conditional p.d.f. of X_3 given $(X_1, X_2) = (5, 7)$. (You may use the result stated in Exercise 12.)

b. Find the conditional probability that $X_3 > 3$ given $(X_1, X_2) = (5, 7)$ and compare it to the value of $\Pr(X_3 > 3)$ found in Example 3.7.9. Can you suggest a reason why the conditional probability should be higher than the marginal probability?

15. Let $X_1, \ldots, X_n$ be independent random variables, and let W be a random variable such that $\Pr(W = c) = 1$ for some constant c. Prove that $X_1, \ldots, X_n$ are conditionally independent given $W = c$.

3.8 Functions of a Random Variable

Often we find that after we compute the distribution of a random variable X, we really want the distribution of some function of X. For example, if X is the rate at which customers are served in a queue, then $1/X$ is the average waiting time. If we have the distribution of X, we should be able to determine the distribution of $1/X$ or of any other function of X. How to do that is the subject of this section.

Random Variable with a Discrete Distribution

Example 3.8.1 Distance from the Middle. Let X have the uniform distribution on the integers $1, 2, \ldots, 9$. Suppose that we are interested in how far X is from the middle of the distribution, namely, 5. We could define $Y = |X - 5|$ and compute probabilities such as $\Pr(Y = 1) = \Pr(X \in \{4, 6\}) = 2/9$. ◄

Example 3.8.1 illustrates the general procedure for finding the distribution of a function of a discrete random variable. The general result is straightforward.

Theorem 3.8.1 Function of a Discrete Random Variable. Let X have a discrete distribution with p.f. f, and let $Y = r(X)$ for some function of r defined on the set of possible values of X. For each possible value y of Y, the p.f. g of Y is

$$g(y) = \Pr(Y = y) = \Pr[r(X) = y] = \sum_{x: r(x)=y} f(x). \qquad \blacksquare$$

Example 3.8.2 Distance from the Middle. The possible values of Y in Example 3.8.1 are 0, 1, 2, 3, and 4. We see that $Y = 0$ if and only if $X = 5$, so $g(0) = f(5) = 1/9$. For all other values of Y, there are two values of X that give that value of Y. For example, $\{Y = 4\} = \{X = 1\} \cup \{X = 9\}$. So, $g(y) = 2/9$ for $y = 1, 2, 3, 4$. ◄

Random Variable with a Continuous Distribution

If a random variable X has a continuous distribution, then the procedure for deriving the probability distribution of a function of X differs from that given for a discrete distribution. One way to proceed is by direct calculation as in Example 3.8.3.

Example 3.8.3 Average Waiting Time. Let Z be the rate at which customers are served in a queue, and suppose that Z has a continuous c.d.f. F. The average waiting time is $Y = 1/Z$. If we want to find the c.d.f. G of Y, we can write

$$G(y) = \Pr(Y \le y) = \Pr\left(\frac{1}{Z} \le y\right) = \Pr\left(Z \ge \frac{1}{y}\right) = \Pr\left(Z > \frac{1}{y}\right) = 1 - F\left(\frac{1}{y}\right),$$

where the fourth equality follows from the fact that Z has a continuous distribution so that $\Pr(Z = 1/y) = 0$. ◄

In general, suppose that the p.d.f. of X is f and that another random variable is defined as $Y = r(X)$. For each real number y, the c.d.f. $G(y)$ of Y can be derived as follows:

$$\begin{aligned} G(y) &= \Pr(Y \le y) = \Pr[r(X) \le y] \\ &= \int_{\{x: r(x) \le y\}} f(x)\, dx. \end{aligned}$$

If the random variable Y also has a continuous distribution, its p.d.f. g can be obtained from the relation

$$g(y) = \frac{dG(y)}{dy}.$$

This relation is satisfied at every point y at which G is differentiable.

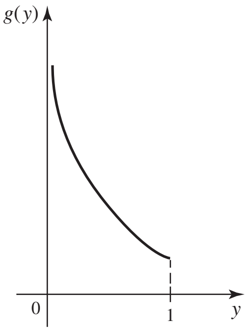


Figure 3.23 The p.d.f. of $Y = X^2$ in Example 3.8.4.

Example 3.8.4

Deriving the p.d.f. of X^2 when X Has a Uniform Distribution. Suppose that X has the uniform distribution on the interval $[-1, 1]$, so

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

We shall determine the p.d.f. of the random variable $Y = X^2$.

Since $Y = X^2$, then Y must belong to the interval $0 \le Y \le 1$. Thus, for each value of Y such that $0 \le y \le 1$, the c.d.f. $G(y)$ of Y is

$$\begin{aligned} G(y) &= \Pr(Y \le y) = \Pr(X^2 \le y) \\ &= \Pr(-y^{1/2} \le X \le y^{1/2}) \\ &= \int_{-y^{1/2}}^{y^{1/2}} f(x)\, dx = y^{1/2}. \end{aligned}$$

For $0 < y < 1$, it follows that the p.d.f. $g(y)$ of Y is

$$g(y) = \frac{dG(y)}{dy} = \frac{1}{2y^{1/2}}.$$

This p.d.f. of Y is sketched in Fig. 3.23. It should be noted that although Y is simply the square of a random variable with a uniform distribution, the p.d.f. of Y is unbounded in the neighborhood of $y = 0$. ◀

Linear functions are very useful transformations, and the p.d.f. of a linear function of a continuous random variable is easy to derive. The proof of the following result is left to the reader in Exercise 5.

Theorem 3.8.2

Linear Function. Suppose that X is a random variable for which the p.d.f. is f and that $Y = aX + b$ ($a \ne 0$). Then the p.d.f. of Y is

$$g(y) = \frac{1}{|a|} f\left(\frac{y - b}{a}\right) \quad \text{for } -\infty < y < \infty, \tag{3.8.1}$$

and 0 otherwise. ■

The Probability Integral Transformation

Example 3.8.5

Let X be a continuous random variable with p.d.f. $f(x) = \exp(-x)$ for $x > 0$ and 0 otherwise. The c.d.f. of X is $F(x) = 1 - \exp(-x)$ for $x > 0$ and 0 otherwise. If we let

F be the function r in the earlier results of this section, we can find the distribution of $Y = F(X)$. The c.d.f. or Y is, for $0 < y < 1$,

$$\begin{aligned} G(y) &= \Pr(Y \le y) = \Pr(1 - \exp(-X) \le y) = \Pr(X \le -\log(1-y)) \\ &= F(-\log(1-y)) = 1 - \exp(-[-\log(1-y)]) = y, \end{aligned}$$

which is the c.d.f. of the uniform distribution on the interval [0, 1]. It follows that Y has the uniform distribution on the interval [0, 1]. ◀

The result in Example 3.8.5 is quite general.

Theorem 3.8.3 Probability Integral Transformation. Let X have a continuous c.d.f. F, and let $Y = F(X)$. (This transformation from X to Y is called the *probability integral transformation*.) The distribution of Y is the uniform distribution on the interval [0, 1].

Proof First, because F is the c.d.f. of a random variable, then $0 \le F(x) \le 1$ for $-\infty < x < \infty$. Therefore, $\Pr(Y < 0) = \Pr(Y > 1) = 0$. Since F is continuous, the set of x such that $F(x) = y$ is a nonempty closed and bounded interval $[x_0, x_1]$ for each y in the interval (0, 1). Let $F^{-1}(y)$ denote the lower endpoint x_0 of this interval, which was called the y quantile of F in Definition 3.3.2. In this way, $Y \le y$ if and only if $X \le x_1$. Let G denote the c.d.f. of Y. Then

$$G(y) = \Pr(Y \le y) = \Pr(X \le x_1) = F(x_1) = y.$$

Hence, $G(y) = y$ for $0 < y < 1$. Because this function is the c.d.f. of the uniform distribution on the interval [0, 1], this uniform distribution is the distribution of Y. ■

Because $\Pr(X = F^{-1}(Y)) = 1$ in the proof of Theorem 3.8.3, we have the following corollary.

Corollary 3.8.1 Let Y have the uniform distribution on the interval [0, 1], and let F be a continuous c.d.f. with quantile function F^{-1}. Then $X = F^{-1}(Y)$ has c.d.f. F. ■

Theorem 3.8.3 and its corollary give us a method for transforming an arbitrary continuous random variable X into another random variable Z with any desired continuous distribution. To be specific, let X have a continuous c.d.f. F, and let G be another continuous c.d.f. Then $Y = F(X)$ has the uniform distribution on the interval [0, 1] according to Theorem 3.8.3, and $Z = G^{-1}(Y)$ has the c.d.f. G according to Corollary 3.8.1. Combining these, we see that $Z = G^{-1}[F(X)]$ has c.d.f. G.

Simulation

Pseudo-Random Numbers Most computer packages that do statistical analyses also produce what are called *pseudo-random numbers*. These numbers appear to have some of the properties that a random sample would have, even though they are generated by deterministic algorithms. The most fundamental of these programs are the ones that generate pseudo-random numbers that appear to have the uniform distribution on the interval [0, 1]. We shall refer to such functions as *uniform pseudo-random number generators*. The important features that a uniform pseudo-random number generator must have are the following.

The numbers that it produces need to be spread somewhat uniformly over the interval [0, 1], and they need to appear to be observed values of independent random variables. This last feature is very complicated to word precisely. An example of a sequence that does *not* appear to be observations of independent random variables would be one that was perfectly evenly spaced. Another example would be one with the following behavior: Suppose that we look at the sequence $X_1, X_2, \ldots$ one at a time, and every time we find an $X_i > 0.5$, we write down the next number X_{i+1}. If the subsequence of numbers that we write down is not spread approximately uniformly over the interval [0, 1], then the original sequence does not look like observations of independent random variables with the uniform distribution on the interval [0, 1]. The reason is that the conditional distribution of X_{i+1} given that $X_i > 0.5$ is supposed to be uniform over the interval [0, 1], according to independence.

Generating Pseudo-Random Numbers Having a Specified Distribution A uniform pseudo-random number generator can be used to generate values of a random variable Y having any specified continuous c.d.f. G. If a random variable X has the uniform distribution on the interval [0, 1] and if the quantile function G^{-1} is defined as before, then it follows from Corollary 3.8.1 that the c.d.f. of the random variable $Y = G^{-1}(X)$ will be G. Hence, if a value of X is produced by a uniform pseudo-random number generator, then the corresponding value of Y will have the desired property. If n independent values $X_1, \ldots, X_n$ are produced by the generator, then the corresponding values $Y_1, \ldots, Y_n$ will appear to form a random sample of size n from the distribution with the c.d.f. G.

Example 3.8.6 Generating Independent Values from a Specified p.d.f. Suppose that a uniform pseudo-random number generator is to be used to generate three independent values from the distribution for which the p.d.f. g is as follows:

$$g(y) = \begin{cases} \frac{1}{2}(2 - y) & \text{for } 0 < y < 2, \\ 0 & \text{otherwise.} \end{cases}$$

For $0 < y < 2$, the c.d.f. G of the given distribution is

$$G(y) = y - \frac{y^2}{4}.$$

Also, for $0 < x < 1$, the inverse function $y = G^{-1}(x)$ can be found by solving the equation $x = G(y)$ for y. The result is

$$y = G^{-1}(x) = 2[1 - (1 - x)^{1/2}]. \tag{3.8.2}$$

The next step is to generate three uniform pseudo-random numbers x_1, x_2, and x_3 using the generator. Suppose that the three generated values are

$$x_1 = 0.4125, \quad x_2 = 0.0894, \quad x_3 = 0.8302.$$

When these values of x_1, x_2, and x_3 are substituted successively into Eq. (3.8.2), the values of y that are obtained are $y_1 = 0.47$, $y_2 = 0.09$, and $y_3 = 1.18$. These are then treated as the observed values of three independent random variables with the distribution for which the p.d.f. is g. ◄

If G is a general c.d.f., there is a method similar to Corollary 3.8.1 that can be used to transform a uniform random variable into a random variable with c.d.f. G. See Exercise 12 in this section. There are other computer methods for generating values from certain specified distributions that are faster and more accurate than using the quantile function. These topics are discussed in the books by Kennedy and

Gentle (1980) and Rubinstein (1981). Chapter 12 of this text contains techniques and examples that show how simulation can be used to solve statistical problems.

General Function In general, if X has a continuous distribution and if $Y = r(X)$, then it is not necessarily true that Y will also have a continuous distribution. For example, suppose that $r(x) = c$, where c is a constant, for all values of x in some interval $a \le x \le b$, and that $\Pr(a \le X \le b) > 0$. Then $\Pr(Y = c) > 0$. Since the distribution of Y assigns positive probability to the value c, this distribution cannot be continuous. In order to derive the distribution of Y in a case like this, the c.d.f. of Y must be derived by applying methods like those described above. For certain functions r, however, the distribution of Y will be continuous; and it will then be possible to derive the p.d.f. of Y directly without first deriving its c.d.f. We shall develop this case in detail at the end of this section.

Direct Derivation of the p.d.f. When *r* is One-to-One and Differentiable

Example 3.8.7 Average Waiting Time. Consider Example 3.8.3 again. The p.d.f. g of Y can be computed from $G(y) = 1 - F(1/y)$ because F and $1/y$ both have derivatives at enough places. We apply the chain rule for differentiation to obtain

$$g(y) = \frac{dG(y)}{dy} = -\left.\frac{dF(x)}{dx}\right|_{x=1/y}\left(-\frac{1}{y^2}\right) = f\left(\frac{1}{y}\right)\frac{1}{y^2},$$

except at $y = 0$ and at those values of y such that $F(x)$ is not differentiable at $x = 1/y$. ◄

Differentiable One-To-One Functions The method used in Example 3.8.7 generalizes to very arbitrary differentiable one-to-one functions. Before stating the general result, we should recall some properties of differentiable one-to-one functions from calculus. Let r be a differentiable one-to-one function on the open interval (a, b). Then r is either strictly increasing or strictly decreasing. Because r is also continuous, it will map the interval (a, b) to another open interval (α, β), called the *image of* (a, b) *under* r. That is, for each $x \in (a, b)$, $r(x) \in (\alpha, \beta)$, and for each $y \in (\alpha, \beta)$ there is $x \in (a, b)$ such that $y = r(x)$ and this y is unique because r is one-to-one. So the inverse s of r will exist on the interval (α, β), meaning that for $x \in (a, b)$ and $y \in (\alpha, \beta)$ we have $r(x) = y$ if and only if $s(y) = x$. The derivative of s will exist (possibly infinite), and it is related to the derivative of r by

$$\frac{ds(y)}{dy} = \left(\left.\frac{dr(x)}{dx}\right|_{x=s(y)}\right)^{-1}.$$

Theorem 3.8.4 Let X be a random variable for which the p.d.f. is f and for which $\Pr(a < X < b) = 1$. (Here, a and/or b can be either finite or infinite.) Let $Y = r(X)$, and suppose that $r(x)$ is differentiable and one-to-one for $a < x < b$. Let (α, β) be the image of the interval (a, b) under the function r. Let $s(y)$ be the inverse function of $r(x)$ for $\alpha < y < \beta$. Then the p.d.f. g of Y is

$$g(y) = \begin{cases} f[s(y)]\left|\dfrac{ds(y)}{dy}\right| & \text{for } \alpha < y < \beta, \\ 0 & \text{otherwise.} \end{cases} \tag{3.8.3}$$

Proof If r is increasing, then s is increasing, and for each $y \in (\alpha, \beta)$,

$$G(y) = \Pr(Y \leq y) = \Pr[r(X) \leq y] = \Pr[X \leq s(y)] = F[s(y)].$$

It follows that G is differentiable at all y where both s is differentiable and where $F(x)$ is differentiable at $x = s(y)$. Using the chain rule for differentiation, it follows that the p.d.f. $g(y)$ for $\alpha < y < \beta$ will be

$$g(y) = \frac{dG(y)}{dy} = \frac{dF[s(y)]}{dy} = f[s(y)]\frac{ds(y)}{dy}. \tag{3.8.4}$$

Because s is increasing, $ds(y)/dy$ is positive; hence, it equals $|ds(y)/dy|$ and Eq. (3.8.4) implies Eq. (3.8.3). Similarly, if r is decreasing, then s is decreasing, and for each $y \in (\alpha, \beta)$,

$$G(y) = \Pr[r(X) \leq y] = \Pr[X \geq s(y)] = 1 - F[s(y)].$$

Using the chain rule again, we differentiate G to get the p.d.f. of Y

$$g(y) = \frac{dG(y)}{dy} = -f[s(y)]\frac{ds(y)}{dy}. \tag{3.8.5}$$

Since s is strictly decreasing, $ds(y)/dy$ is negative so that $-ds(y)/dy$ equals $|ds(y)/dy|$. It follows that Eq. (3.8.5) implies Eq. (3.8.3). ■

Example 3.8.8 Microbial Growth. A popular model for populations of microscopic organisms in large environments is exponential growth. At time 0, suppose that v organisms are introduced into a large tank of water, and let X be the rate of growth. After time t, we would predict a population size of ve^{Xt}. Assume that X is unknown but has a continuous distribution with p.d.f.

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

We are interested in the distribution of $Y = ve^{Xt}$ for known values of v and t. For concreteness, let $v = 10$ and $t = 5$, so that $r(x) = 10e^{5x}$.

In this example, $\Pr(0 < X < 1) = 1$ and r is a continuous and strictly increasing function of x for $0 < x < 1$. As x varies over the interval $(0, 1)$, it is found that $y = r(x)$ varies over the interval $(10, 10e^5)$. Furthermore, for $10 < y < 10e^5$, the inverse function is $s(y) = \log(y/10)/5$. Hence, for $10 < y < 10e^5$,

$$\frac{ds(y)}{dy} = \frac{1}{5y}.$$

It follows from Eq. (3.8.3) that $g(y)$ will be

$$g(y) = \begin{cases} \dfrac{3(1 - \log(y/10)/5)^2}{5y} & \text{for } 10 < y < 10e^5, \\ 0 & \text{otherwise.} \end{cases}$$

◄

Summary

We learned several methods for determining the distribution of a function of a random variable. For a random variable X with a continuous distribution having p.d.f. f, if r is strictly increasing or strictly decreasing with differentiable inverse s (i.e., $s(r(x)) = x$ and s is differentiable), then the p.d.f. of $Y = r(X)$ is $g(y) =$

$f(s(y))|ds(y)/dy|$. A special transformation allows us to transform a random variable X with the uniform distribution on the interval [0, 1] into a random variable Y with an arbitrary continuous c.d.f. G by $Y = G^{-1}(X)$. This method can be used in conjunction with a uniform pseudo-random number generator to generate random variables with arbitrary continuous distributions.

Exercises

1. Suppose that the p.d.f. of a random variable X is as follows:

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

Also, suppose that $Y = 1 - X^2$. Determine the p.d.f. of Y.

2. Suppose that a random variable X can have each of the seven values $-3, -2, -1, 0, 1, 2, 3$ with equal probability. Determine the p.f. of $Y = X^2 - X$.

3. Suppose that the p.d.f. of a random variable X is as follows:

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

Also, suppose that $Y = X(2 - X)$. Determine the c.d.f. and the p.d.f. of Y.

4. Suppose that the p.d.f. of X is as given in Exercise 3. Determine the p.d.f. of $Y = 4 - X^3$.

5. Prove Theorem 3.8.2. (*Hint:* Either apply Theorem 3.8.4 or first compute the c.d.f. seperately for $a > 0$ and $a < 0$.)

6. Suppose that the p.d.f. of X is as given in Exercise 3. Determine the p.d.f. of $Y = 3X + 2$.

7. Suppose that a random variable X has the uniform distribution on the interval [0, 1]. Determine the p.d.f. of **(a)** X^2, **(b)** $-X^3$, and **(c)** $X^{1/2}$.

8. Suppose that the p.d.f. of X is as follows:

$$f(x) = \begin{cases} e^{-x} & \text{for } x > 0, \\ 0 & \text{for } x \le 0. \end{cases}$$

Determine the p.d.f. of $Y = X^{1/2}$.

9. Suppose that X has the uniform distribution on the interval [0, 1]. Construct a random variable $Y = r(X)$ for which the p.d.f. will be

$$g(y) = \begin{cases} \frac{3}{8}y^2 & \text{for } 0 < y < 2, \\ 0 & \text{otherwise.} \end{cases}$$

10. Let X be a random variable for which the p.d.f f is as given in Exercise 3. Construct a random variable $Y = r(X)$ for which the p.d.f. g is as given in Exercise 9.

11. Explain how to use a uniform pseudo-random number generator to generate four independent values from a distribution for which the p.d.f. is

$$g(y) = \begin{cases} \frac{1}{2}(2y + 1) & \text{for } 0 < y < 1, \\ 0 & \text{otherwise.} \end{cases}$$

12. Let F be an arbitrary c.d.f. (not necessarily discrete, not necessarily continuous, not necessarily either). Let F^{-1} be the quantile function from Definition 3.3.2. Let X have the uniform distribution on the interval [0, 1]. Define $Y = F^{-1}(X)$. Prove that the c.d.f. of Y is F. *Hint:* Compute $\Pr(Y \le y)$ in two cases. First, do the case in which y is the unique value of x such that $F(x) = F(y)$. Second, do the case in which there is an entire interval of x values such that $F(x) = F(y)$.

13. Let Z be the rate at which customers are served in a queue. Assume that Z has the p.d.f.

$$f(z) = \begin{cases} 2e^{-2z} & \text{for } z > 0, \\ 0 & \text{otherwise.} \end{cases}$$

Find the p.d.f. of the average waiting time $T = 1/Z$.

14. Let X have the uniform distribution on the interval $[a, b]$, and let $c > 0$. Prove that $cX + d$ has the uniform distribution on the interval $[ca + d, cb + d]$.

15. Most of the calculation in Example 3.8.4 is quite general. Suppose that X has a continuous distribution with p.d.f. f. Let $Y = X^2$, and show that the p.d.f. of Y is

$$g(y) = \frac{1}{2y^{1/2}}[f(y^{1/2}) + f(-y^{1/2})].$$

16. In Example 3.8.4, the p.d.f. of $Y = X^2$ is much larger for values of y near 0 than for values of y near 1 despite the fact that the p.d.f. of X is flat. Give an intuitive reason why this occurs in this example.

17. An insurance agent sells a policy which has a \$100 deductible and a \$5000 cap. This means that when the policy holder files a claim, the policy holder must pay the first

\$100. After the first \$100, the insurance company pays the rest of the claim up to a maximum payment of \$5000. Any excess must be paid by the policy holder. Suppose that the dollar amount X of a claim has a continuous distribution with p.d.f. $f(x) = 1/(1 + x)^2$ for $x > 0$ and 0 otherwise. Let Y be the amount that the insurance company has to pay on the claim.

a. Write Y as a function of X, i.e., $Y = r(X)$.

b. Find the c.d.f. of Y.

c. Explain why Y has neither a continuous nor a discrete distribution.

3.9 Functions of Two or More Random Variables

When we observe data consisting of the values of several random variables, we need to summarize the observed values in order to be able to focus on the information in the data. Summarizing consists of constructing one or a few functions of the random variables that capture the bulk of the information. In this section, we describe the techniques needed to determine the distribution of a function of two or more random variables.

Random Variables with a Discrete Joint Distribution

Example 3.9.1 Bull Market. Three different investment firms are trying to advertise their mutual funds by showing how many perform better than a recognized standard. Each company has 10 funds, so there are 30 in total. Suppose that the first 10 funds belong to the first firm, the next 10 to the second firm, and the last 10 to the third firm. Let $X_i = 1$ if fund i performs better than the standard and $X_i = 0$ otherwise, for $i = 1, \dots, 30$. Then, we are interested in the three functions

$$Y_1 = X_1 + \cdots + X_{10},$$
$$Y_2 = X_{11} + \cdots + X_{20},$$
$$Y_3 = X_{21} + \cdots + X_{30}.$$

We would like to be able to determine the joint distribution of Y_1, Y_2, and Y_3 from the joint distribution of $X_1, \dots, X_{30}$. ◀

The general method for solving problems like those of Example 3.9.1 is a straightforward extension of Theorem 3.8.1.

Theorem 3.9.1 Functions of Discrete Random Variables. Suppose that n random variables $X_1, \dots, X_n$ have a discrete joint distribution for which the joint p.f. is f, and that m functions $Y_1, \dots, Y_m$ of these n random variables are defined as follows:

$$\begin{aligned} Y_1 &= r_1(X_1, \dots, X_n), \\ Y_2 &= r_2(X_1, \dots, X_n), \\ &\vdots \\ Y_m &= r_m(X_1, \dots, X_n). \end{aligned}$$

For given values $y_1, \ldots, y_m$ of the m random variables $Y_1, \ldots, Y_m$, let A denote the set of all points $(x_1, \ldots, x_n)$ such that

$$\begin{aligned} r_1(x_1, \ldots, x_n) &= y_1, \\ r_2(x_1, \ldots, x_n) &= y_2, \\ &\vdots \\ r_m(x_1, \ldots, x_n) &= y_m. \end{aligned}$$

Then the value of the joint p.f. g of $Y_1, \ldots, Y_m$ is specified at the point $(y_1, \ldots, y_m)$ by the relation

$$g(y_1, \ldots, y_m) = \sum_{(x_1, \ldots, x_n) \in A} f(x_1, \ldots, x_n). \qquad \blacksquare$$

Example 3.9.2 Bull Market. Recall the situation in Example 3.9.1. Suppose that we want the joint p.f. g of (Y_1, Y_2, Y_3) at the point $(3, 5, 8)$. That is, we want $g(3, 5, 8) = \Pr(Y_1 = 3, Y_2 = 5, Y_3 = 8)$. The set A as defined in Theorem 3.9.1 is

$$A = \{(x_1, \ldots, x_{30}) : x_1 + \cdots + x_{10} = 3, x_{11} + \cdots + x_{20} = 5, x_{21} + \cdots + x_{30} = 8\}.$$

Two of the points in the set A are

$$(1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0),$$
$$(1, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 1, 1, 0, 0, 1, 0, 1, 0, 1, 1, 0, 1, 1, 1, 0, 1, 1, 1, 1).$$

A counting argument like those developed in Sec. 1.8 can be used to discover that there are

$$\binom{10}{3}\binom{10}{5}\binom{10}{8} = 1{,}360{,}800$$

points in A. Unless the joint distribution of $X_1, \ldots, X_{30}$ has some simple structure, it will be extremely tedious to compute $g(3, 5, 8)$ as well as most other values of g. For example, if all of the 2^{30} possible values of the vector $(X_1, \ldots, X_{30})$ are equally likely, then

$$g(3, 5, 8) = \frac{1{,}360{,}800}{2^{30}} = 1.27 \times 10^{-3}. \qquad \blacktriangleleft$$

The next result gives an important example of a function of discrete random variables.

Theorem 3.9.2 Binomial and Bernoulli Distributions. Assume that $X_1, \ldots, X_n$ are i.i.d. random variables having the Bernoulli distribution with parameter p. Let $Y = X_1 + \cdots + X_n$. Then Y has the binomial distribution with parameters n and p.

Proof It is clear that $Y = y$ if and only if exactly y of $X_1, \ldots, X_n$ equal 1 and the other $n - y$ equal 0. There are $\binom{n}{y}$ distinct possible values for the vector $(X_1, \ldots, X_n)$ that have y ones and $n - y$ zeros. Each such vector has probability $p^y(1-p)^{n-y}$ of being observed; hence the probability that $Y = y$ is the sum of the probabilities of those vectors, namely, $\binom{n}{y}p^y(1-p)^{n-y}$ for $y = 0, \ldots, n$. From Definition 3.1.7, we see that Y has the binomial distribution with parameters n and p. $\blacksquare$

Example 3.9.3 Sampling Parts. Suppose that two machines are producing parts. For $i = 1, 2$, the probability is p_i that machine i will produce a defective part, and we shall assume that all parts from both machines are independent. Assume that the first n_1 parts are produced by machine 1 and that the last n_2 parts are produced by machine 2, with $n = n_1 + n_2$ being the total number of parts sampled. Let $X_i = 1$ if the ith part is defective and $X_i = 0$ otherwise for $i = 1, \dots, n$. Define $Y_1 = X_1 + \cdots + X_{n_1}$ and $Y_2 = X_{n_1+1} + \cdots + X_n$. These are the total numbers of defective parts produced by each machine. The assumptions stated in the problem allow us to conclude that Y_1 and Y_2 are independent according to the note about separate functions of independent random variables on page 140. Furthermore, Theorem 3.9.2 says that Y_j has the binomial distribution with parameters n_j and p_j for $j = 1, 2$. These two marginal distributions, together with the fact that Y_1 and Y_2 are independent, give the entire joint distribution. So, for example, if g is the joint p.f. of Y_1 and Y_2, we can compute

$$g(y_1, y_2) = \binom{n_1}{y_1} p_1^{y_1}(1 - p_1)^{n_1-y_1} \binom{n_2}{y_2} p_2^{y_2}(1 - p)^{n_2-y_2},$$

for $y_1 = 0, \dots, n_1$ and $y_2 = 0, \dots, n_2$, while $g(y_1, y_2) = 0$ otherwise. There is no need to find a set A as in Example 3.9.2, because of the simplifying structure of the joint distribution of $X_1, \dots, X_n$. ◄

Random Variables with a Continuous Joint Distribution

Example 3.9.4

Total Service Time. Suppose that the first two customers in a queue plan to leave together. Let X_i be the time it takes to serve customer i for $i = 1, 2$. Suppose also that X_1 and X_2 are independent random variables with common distribution having p.d.f. $f(x) = 2e^{-2x}$ for $x > 0$ and 0 otherwise. Since the customers will leave together, they are interested in the total time it takes to serve both of them, namely, $Y = X_1 + X_2$. We can now find the p.d.f. of Y.

For each y, let

$$A_y = \{(x_1, x_2) : x_1 + x_2 \le y\}.$$

Then $Y \le y$ if and only if $(X_1, X_2) \in A_y$. The set A_y is pictured in Fig. 3.24. If we let $G(y)$ denote the c.d.f. of Y, then, for $y > 0$,

$$\begin{aligned} G(y) = \Pr((X_1, X_2) \in A_y) &= \int_0^y \int_0^{y-x_2} 4e^{-2x_1-2x_2} dx_1 dx_2 \\ &= \int_0^y 2e^{-2x_2}\left[1 - e^{-2(y-x_2)}\right] dx_2 = \int_0^y \left[2e^{-2x_2} - 2e^{-2y}\right] dx_2 \\ &= 1 - e^{-2y} - 2ye^{-2y}. \end{aligned}$$

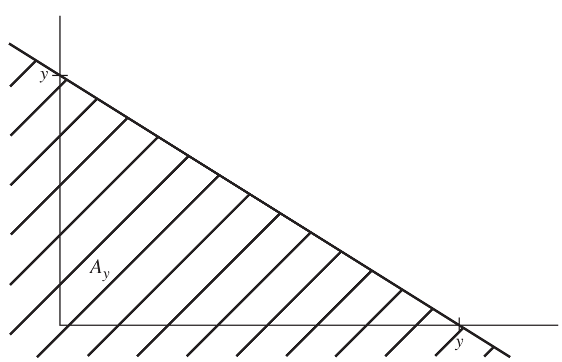


Figure 3.24 The set A_y in Example 3.9.4 and in the proof of Theorem 3.9.4.

Taking the derivative of $G(y)$ with respect to y, we get the p.d.f.

$$g(y) = \frac{d}{dy}\left[1 - e^{-2y} - ye^{-2y}\right] = 4ye^{-2y},$$

for $y > 0$ and 0 otherwise. ◂

The transformation in Example 3.9.4 is an example of a brute-force method that is always available for finding the distribution of a function of several random variables, however, it might be difficult to apply in individual cases.

Theorem 3.9.3 Brute-Force Distribution of a Function. Suppose that the joint p.d.f. of $\boldsymbol{X} = (X_1, \ldots, X_n)$ is $f(\boldsymbol{x})$ and that $Y = r(\boldsymbol{X})$. For each real number y, define $A_y = \{\boldsymbol{x} : r(\boldsymbol{x}) \le y\}$. Then the c.d.f. $G(y)$ of Y is

$$G(y) = \int \underset{A_y}{\cdots} \int f(\boldsymbol{x})\, d\boldsymbol{x}. \tag{3.9.1}$$

Proof From the definition of c.d.f.,

$$G(y) = \Pr(Y \le y) = \Pr[r(\boldsymbol{X}) \le y] = \Pr(\boldsymbol{X} \in A_y),$$

which equals the right side of Eq. (3.9.1) by Definition 3.7.3. ∎

If the distribution of Y also is continuous, then the p.d.f. of Y can be found by differentiating the c.d.f. $G(y)$.

A popular special case of Theorem 3.9.3 is the following.

Theorem 3.9.4 Linear Function of Two Random Variables. Let X_1 and X_2 have joint p.d.f. $f(x_1, x_2)$, and let $Y = a_1X_1 + a_2X_2 + b$ with $a_1 \ne 0$. Then Y has a continuous distribution whose p.d.f. is

$$g(y) = \int_{-\infty}^{\infty} f\left(\frac{y - b - a_2x_2}{a_1}, x_2\right) \frac{1}{|a_1|} dx_2. \tag{3.9.2}$$

Proof First, we shall find the c.d.f. G of Y whose derivative we will see is the function g in Eq. (3.9.2). For each y, let $A_y = \{(x_1, x_2) : a_1x_1 + a_2x_2 + b \le y\}$. The set A_y has the same general form as the set in Fig. 3.24. We shall write the integral over the set A_y with x_2 in the outer integral and x_1 in the inner integral. Assume that $a_1 > 0$. The other case is similar. According to Theorem 3.9.3,

$$G(y) = \int_{A_y}\int f(x_1, x_2)dx_1dx_2 = \int_{-\infty}^{\infty}\int_{-\infty}^{(y-b-a_2x_2)/a_1} f(x_1, x_2)dx_1dx_2. \tag{3.9.3}$$

For the inner integral, perform the change of variable $z = a_1x_1 + a_2x_2 + b$ whose inverse is $x_1 = (z - b - a_2x_2)/a_1$, so that $dx_1 = dz/a_1$. The inner integral, after this change of variable, becomes

$$\int_{-\infty}^{y} f\left(\frac{z - b - a_2x_2}{a_1}, x_2\right) \frac{1}{a_1} dz.$$

We can now substitute this expression for the inner integral into Eq. (3.9.3):

$$\begin{aligned} G(y) &= \int_{-\infty}^{\infty}\int_{-\infty}^{y} f\left(\frac{z - b - a_2x_2}{a_1}, x_2\right) \frac{1}{a_1} dz dx_2 \\ &= \int_{-\infty}^{y}\int_{-\infty}^{\infty} f\left(\frac{z - b - a_2x_2}{a_1}, x_2\right) \frac{1}{a_1} dx_2 dz. \end{aligned} \tag{3.9.4}$$

Let $g(z)$ denote the inner integral on the far right side of Eq. (3.9.4). Then we have $G(y) = \int_{-\infty}^{y} g(z)dz$, whose derivative is $g(y)$, the function in Eq. (3.9.2). ■

The special case of Theorem 3.9.4 in which X_1 and X_2 are independent, $a_1 = a_2 = 1$, and $b = 0$ is called *convolution*.

Definition 3.9.1 Convolution. Let X_1 and X_2 be independent continuous random variables and let $Y = X_1 + X_2$. The distribution of Y is called the *convolution* of the distributions of X_1 and X_2. The p.d.f. of Y is sometimes called the convolution of the p.d.f.'s of X_1 and X_2.

If we let the p.d.f. of X_i be f_i for $i = 1, 2$ in Definition 3.9.1, then Theorem 3.9.4 (with $a_1 = a_2 = 1$ and $b = 0$) says that the p.d.f. of $Y = X_1 + X_2$ is

$$g(y) = \int_{-\infty}^{\infty} f_1(y - z) f_2(z)dz. \tag{3.9.5}$$

Equivalently, by switching the names of X_1 and X_2, we obtain the alternative form for the convolution:

$$g(y) = \int_{-\infty}^{\infty} f_1(z) f_2(y - z)\, dz. \tag{3.9.6}$$

The p.d.f. found in Example 3.9.4 is the special case of (3.9.5) with $f_1(x) = f_2(x) = 2e^{-2x}$ for $x > 0$ and 0 otherwise.

Example 3.9.5 An Investment Portfolio. Suppose that an investor wants to purchase both stocks and bonds. Let X_1 be the value of the stocks at the end of one year, and let X_2 be the value of the bonds at the end of one year. Suppose that X_1 and X_2 are independent. Let X_1 have the uniform distribution on the interval [1000, 4000], and let X_2 have the uniform distribution on the interval [800, 1200]. The sum $Y = X_1 + X_2$ is the value at the end of the year of the portfolio consisting of both the stocks and the bonds. We shall find the p.d.f. of Y. The function $f_1(z) f_2(y - z)$ in Eq. (3.9.6) is

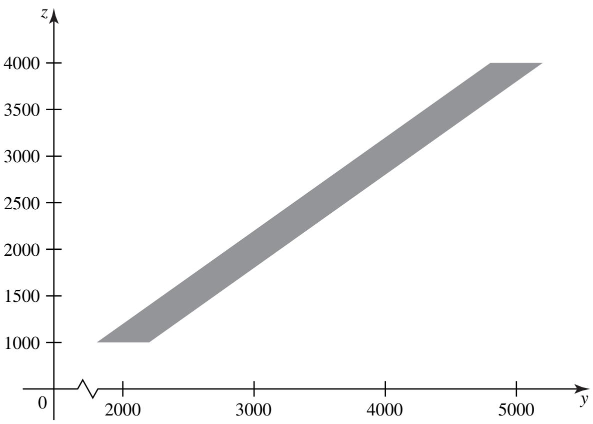
$$f_1(z) f_2(y - z) = \begin{cases} 8.333 \times 10^{-7} & \text{for } 1000 \le z \le 4000 \\ & \quad \text{and } 800 \le y - z \le 1200, \\ 0 & \text{otherwise.} \end{cases} \tag{3.9.7}$$

We need to integrate the function in Eq. (3.9.7) over z for each value of y to get the marginal p.d.f. of Y. It is helpful to look at a graph of the set of (y, z) pairs for which the function in Eq. (3.9.7) is positive. Figure 3.25 shows the region shaded. For $1800 < y \le 2200$, we must integrate z from 1000 to $y - 800$. For $2200 < y \le 4800$, we must integrate z from $y - 1200$ to $y - 800$. For $4800 < y < 5200$, we must integrate z from $y - 1200$ to 4000. Since the function in Eq. (3.9.7) is constant when it is positive, the integral equals the constant times the length of the interval of z values. So, the p.d.f. of Y is

$$g(y) = \begin{cases} 8.333 \times 10^{-7}(y - 1800) & \text{for } 1800 < y \le 2200, \\ 3.333 \times 10^{-4} & \text{for } 2200 < y \le 4800, \\ 8.333 \times 10^{-7}(5200 - y) & \text{for } 4800 < y < 5200, \\ 0 & \text{otherwise.} \end{cases}$$ ◄

As another example of the brute-force method, we consider the largest and smallest observations in a random sample. These functions give an idea of how spread out the sample is. For example, meteorologists often report record high and low

Figure 3.25 The region where the function in Eq. (3.9.7) is positive.



temperatures for specific days as well as record high and low rainfalls for months and years.

Example 3.9.6 Maximum and Minimum of a Random Sample. Suppose that $X_1, \ldots, X_n$ form a random sample of size n from a distribution for which the p.d.f. is f and the c.d.f. is F. The largest value Y_n and the smallest value Y_1 in the random sample are defined as follows:

$$\begin{aligned} Y_n &= \max\{X_1, \ldots, X_n\}, \\ Y_1 &= \min\{X_1, \ldots, X_n\}. \end{aligned} \tag{3.9.8}$$

Consider Y_n first. Let G_n stand for its c.d.f., and let g_n be its p.d.f. For every given value of y $(-\infty < y < \infty)$,

$$\begin{aligned} G_n(y) &= \Pr(Y_n \le y) = \Pr(X_1 \le y, X_2 \le y, \ldots, X_n \le y) \\ &= \Pr(X_1 \le y)\Pr(X_2 \le y) \cdots \Pr(X_n \le y) \\ &= F(y)F(y) \cdots F(y) = [F(y)]^n, \end{aligned}$$

where the third equality follows from the fact that the X_i are independent and the fourth follows from the fact that all of the X_i have the same c.d.f. F. Thus, $G_n(y) = [F(y)]^n$.

Now, g_n can be determined by differentiating the c.d.f. G_n. The result is

$$g_n(y) = n[F(y)]^{n-1} f(y) \quad \text{for } -\infty < y < \infty.$$

Next, consider Y_1 with c.d.f. G_1 and p.d.f. g_1. For every given value of y $(-\infty < y < \infty)$,

$$\begin{aligned} G_1(y) &= \Pr(Y_1 \le y) = 1 - \Pr(Y_1 > y) \\ &= 1 - \Pr(X_1 > y, X_2 > y, \ldots, X_n > y) \\ &= 1 - \Pr(X_1 > y)\Pr(X_2 > y) \cdots \Pr(X_n > y) \\ &= 1 - [1 - F(y)][1 - F(y)] \cdots [1 - F(y)] \\ &= 1 - [1 - F(y)]^n. \end{aligned}$$

Thus, $G_1(y) = 1 - [1 - F(y)]^n$.

Then g_1 can be determined by differentiating the c.d.f. G_1. The result is

$$g_1(y) = n[1 - F(y)]^{n-1} f(y) \quad \text{for } -\infty < y < \infty.$$

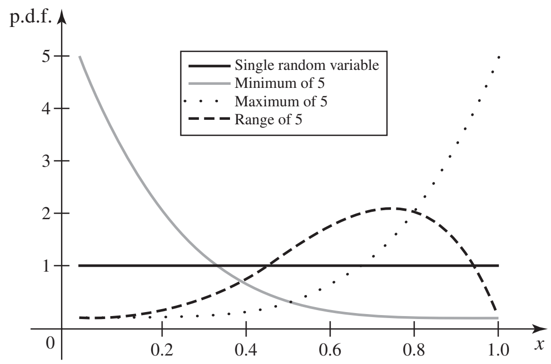


Figure 3.26 The p.d.f. of the uniform distribution on the interval [0, 1] together with the p.d.f.'s of the minimum and maximum of samples of size $n = 5$. The p.d.f. of the range of a sample of size $n = 5$ (see Example 3.9.7) is also included.

Figure 3.26 shows the p.d.f. of the uniform distribution on the interval [0, 1] together with the p.d.f.'s of Y_1 and Y_n for the case $n = 5$. It also shows the p.d.f. of $Y_5 - Y_1$, which will be derived in Example 3.9.7. Notice that the p.d.f. of Y_1 is highest near 0 and lowest near 1, while the opposite is true of the p.d.f. of Y_n, as one would expect.

Finally, we shall determine the joint distribution of Y_1 and Y_n. For every pair of values (y_1, y_n) such that $-\infty < y_1 < y_n < \infty$, the event $\{Y_1 \le y_1\} \cap \{Y_n \le y_n\}$ is the same as $\{Y_n \le y_n\} \cap \{Y_1 > y_1\}^c$. If G denotes the bivariate joint c.d.f. of Y_1 and Y_n, then

$$\begin{aligned} G(y_1, y_n) &= \Pr(Y_1 \le y_1 \text{ and } Y_n \le y_n) \\ &= \Pr(Y_n \le y_n) - \Pr(Y_n \le y_n \text{ and } Y_1 > y_1) \\ &= \Pr(Y_n \le y_n) \\ &\quad - \Pr(y_1 < X_1 \le y_n, y_1 < X_2 \le y_n, \ldots, y_1 < X_n \le y_n) \\ &= G_n(y_n) - \prod_{i=1}^{n} \Pr(y_1 < X_i \le y_n) \\ &= [F(y_n)]^n - [F(y_n) - F(y_1)]^n. \end{aligned}$$

The bivariate joint p.d.f. g of Y_1 and Y_n can be found from the relation

$$g(y_1, y_n) = \frac{\partial^2 G(y_1, y_n)}{\partial y_1 \partial y_n}.$$

Thus, for $-\infty < y_1 < y_n < \infty$,

$$g(y_1, y_n) = n(n-1)[F(y_n) - F(y_1)]^{n-2} f(y_1) f(y_n). \tag{3.9.9}$$

Also, for all other values of y_1 and y_n, $g(y_1, y_n) = 0$. ◄

A popular way to describe how spread out is a random sample is to use the distance from the minimum to the maximum, which is called the *range* of the random sample. We can combine the result from the end of Example 3.9.6 with Theorem 3.9.4 to find the p.d.f. of the range.

Example 3.9.7 **The Distribution of the Range of a Random Sample.** Consider the same situation as in Example 3.9.6. The random variable $W = Y_n - Y_1$ is called the *range* of the sample. The joint p.d.f. $g(y_1, y_n)$ of Y_1 and Y_n was presented in Eq. (3.9.9). We can now apply Theorem 3.9.4 with $a_1 = -1$, $a_2 = 1$, and $b = 0$ to get the p.d.f. h of W:

$$h(w) = \int_{-\infty}^{\infty} g(y_n - w, y_n)dy_n = \int_{-\infty}^{\infty} g(z, z + w)dz, \qquad (3.9.10)$$

where, for the last equality, we have made the change of variable $z = y_n - w$. ◀

Here is a special case in which the integral of Eq. 3.9.10 can be computed in closed form.

Example 3.9.8 **The Range of a Random Sample from a Uniform Distribution.** Suppose that the n random variables $X_1, \ldots, X_n$ form a random sample from the uniform distribution on the interval [0, 1]. We shall determine the p.d.f. of the range of the sample.

In this example,

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

Also, $F(x) = x$ for $0 < x < 1$. We can write $g(y_1, y_n)$ from Eq. (3.9.9) in this case as

$$g(y_1, y_n) = \begin{cases} n(n-1)(y_n - y_1)^{n-2} & \text{for } 0 < y_1 < y_n < 1, \\ 0 & \text{otherwise.} \end{cases}$$

Therefore, in Eq. (3.9.10), $g(z, z + w) = 0$ unless $0 < w < 1$ and $0 < z < 1 - w$. For values of w and z satisfying these conditions, $g(z, w + z) = n(n-1)w^{n-2}$. The p.d.f. in Eq. (3.9.10) is then, for $0 < w < 1$,

$$h(w) = \int_0^{1-w} n(n-1)w^{n-2}\, dz = n(n-1)w^{n-2}(1 - w).$$

Otherwise, $h(w) = 0$. This p.d.f. is shown in Fig. 3.26 for the case $n = 5$. ◀

Direct Transformation of a Multivariate p.d.f.

Next, we state without proof a generalization of Theorem 3.8.4 to the case of several random variables. The proof of Theorem 3.9.5 is based on the theory of differentiable one-to-one transformations in advanced calculus.

Theorem 3.9.5 **Multivariate Transformation.** Let $X_1, \ldots, X_n$ have a continuous joint distribution for which the joint p.d.f. is f. Assume that there is a subset S of R^n such that $\Pr[(X_1, \ldots, X_n) \in S] = 1$. Define n new random variables $Y_1, \ldots, Y_n$ as follows:

$$\begin{aligned} Y_1 &= r_1(X_1, \ldots, X_n), \\ Y_2 &= r_2(X_1, \ldots, X_n), \\ &\vdots \\ Y_n &= r_n(X_1, \ldots, X_n), \end{aligned} \qquad (3.9.11)$$

where we assume that the n functions $r_1, \ldots, r_n$ define a one-to-one differentiable transformation of S onto a subset T of R^n. Let the inverse of this transformation be given as follows:

$$\begin{aligned} x_1 &= s_1(y_1, \ldots, y_n), \\ x_2 &= s_2(y_1, \ldots, y_n), \\ &\vdots \\ x_n &= s_n(y_1, \ldots, y_n). \end{aligned} \qquad (3.9.12)$$

Then the joint p.d.f. g of $Y_1, \ldots, Y_n$ is

$$g(y_1, \ldots, y_n) = \begin{cases} f(s_1, \ldots, s_n)|J| & \text{for } (y_1, \ldots, y_n) \in T, \\ 0 & \text{otherwise,} \end{cases} \tag{3.9.13}$$

where J is the determinant

$$J = \det \begin{bmatrix} \dfrac{\partial s_1}{\partial y_1} & \cdots & \dfrac{\partial s_1}{\partial y_n} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial s_n}{\partial y_1} & \cdots & \dfrac{\partial s_n}{\partial y_n} \end{bmatrix}$$

and $|J|$ denotes the absolute value of the determinant J. ■

Thus, the joint p.d.f. $g(y_1, \ldots, y_n)$ is obtained by starting with the joint p.d.f. $f(x_1, \ldots, x_n)$, replacing each value x_i by its expression $s_i(y_1, \ldots, y_n)$ in terms of $y_1, \ldots, y_n$, and then multiplying the result by $|J|$. This determinant J is called the *Jacobian* of the transformation specified by the equations in (3.9.12).

Note: The Jacobian Is a Generalization of the Derivative of the Inverse. Eqs. (3.8.3) and (3.9.13) are very similar. The former gives the p.d.f. of a single function of a single random variable. Indeed, if $n = 1$ in (3.9.13), $J = ds_1(y_1)/dy_1$ and Eq. (3.9.13) becomes the same as (3.8.3). The Jacobian merely generalizes the derivative of the inverse of a single function of one variable to n functions of n variables.

Example 3.9.9

The Joint p.d.f. of the Quotient and the Product of Two Random Variables. Suppose that two random variables X_1 and X_2 have a continuous joint distribution for which the joint p.d.f. is as follows:

$$f(x_1, x_2) = \begin{cases} 4x_1x_2 & \text{for } 0 < x_1 < 1 \text{ and } 0 < x_2 < 1, \\ 0 & \text{otherwise.} \end{cases}$$

We shall determine the joint p.d.f. of two new random variables Y_1 and Y_2, which are defined by the relations

$$Y_1 = \frac{X_1}{X_2} \text{ and } Y_2 = X_1X_2.$$

In the notation of Theorem 3.9.5, we would say that $Y_1 = r_1(X_1, X_2)$ and $Y_2 = r_2(X_1, X_2)$, where

$$r_1(x_1, x_2) = \frac{x_1}{x_2} \text{ and } r_2(x_1, x_2) = x_1x_2. \tag{3.9.14}$$

The inverse of the transformation in Eq. (3.9.14) is found by solving the equations $y_1 = r_1(x_1, x_2)$ and $y_2 = r_2(x_1, x_2)$ for x_1 and x_2 in terms of y_1 and y_2. The result is

$$\begin{aligned} x_1 &= s_1(y_1, y_2) = (y_1y_2)^{1/2}, \\ x_2 &= s_2(y_1, y_2) = \left(\frac{y_2}{y_1}\right)^{1/2}. \end{aligned} \tag{3.9.15}$$

Let S denote the set of points (x_1, x_2) such that $0 < x_1 < 1$ and $0 < x_2 < 1$, so that $\Pr[(X_1, X_2) \in S] = 1$. Let T be the set of (y_1, y_2) pairs such that $(y_1, y_2) \in T$ if and only if $(s_1(y_1, y_2), s_2(y_1, y_2)) \in S$. Then $\Pr[(Y_1, Y_2) \in T] = 1$. The transformation defined by the equations in (3.9.14) or, equivalently, by the equations in (3.9.15) specifies a one-to-one relation between the points in S and the points in T.

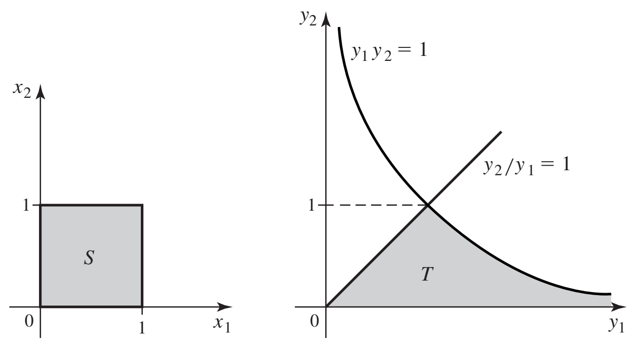


Figure 3.27 The sets S and T in Example 3.9.9.

We shall now show how to find the set T. We know that $(x_1, x_2) \in S$ if and only if the following inequalities hold:

$$x_1 > 0, \quad x_1 < 1, \quad x_2 > 0, \text{ and } x_2 < 1. \tag{3.9.16}$$

We can substitute the formulas for x_1 and x_2 in terms of y_1 and y_2 from Eq. (3.9.15) into the inequalities in (3.9.16) to obtain

$$(y_1y_2)^{1/2} > 0, \quad (y_1y_2)^{1/2} < 1, \quad \left(\frac{y_2}{y_1}\right)^{1/2} > 0,$$
$$\text{and } \left(\frac{y_2}{y_1}\right)^{1/2} < 1. \tag{3.9.17}$$

The first inequality transforms to ($y_1 > 0$ and $y_2 > 0$) or ($y_1 < 0$ and $y_2 < 0$). However, since $y_1 = x_1/x_2$, we cannot have $y_1 < 0$, so we get only $y_1 > 0$ and $y_2 > 0$. The third inequality in (3.9.17) transforms to the same thing. The second inequality in (3.9.17) becomes $y_2 < 1/y_1$. The fourth inequality becomes $y_2 < y_1$. The region T where (y_1, y_2) satisfy these new inequalities is shown in the right panel of Fig. 3.27 with the set S in the left panel.

For the functions in (3.9.15),

$$\frac{\partial s_1}{\partial y_1} = \frac{1}{2}\left(\frac{y_2}{y_1}\right)^{1/2}, \quad \frac{\partial s_1}{\partial y_2} = \frac{1}{2}\left(\frac{y_1}{y_2}\right)^{1/2},$$
$$\frac{\partial s_2}{\partial y_1} = -\frac{1}{2}\left(\frac{y_2}{y_1^3}\right)^{1/2}, \quad \frac{\partial s_2}{\partial y_2} = \frac{1}{2}\left(\frac{1}{y_1y_2}\right)^{1/2}.$$

Hence,

$$J = \det\begin{bmatrix} \frac{1}{2}\left(\frac{y_2}{y_1}\right)^{1/2} & \frac{1}{2}\left(\frac{y_1}{y_2}\right)^{1/2} \\ -\frac{1}{2}\left(\frac{y_2}{y_1^3}\right)^{1/2} & \frac{1}{2}\left(\frac{1}{y_1y_2}\right)^{1/2} \end{bmatrix} = \frac{1}{2y_1}.$$

Since $y_1 > 0$ throughout the set T, $|J| = 1/(2y_1)$.

The joint p.d.f. $g(y_1, y_2)$ can now be obtained directly from Eq. (3.9.13) in the following way: In the expression for $f(x_1, x_2)$, replace x_1 with $(y_1y_2)^{1/2}$, replace x_2

with $(y_2/y_1)^{1/2}$, and multiply the result by $|J| = 1/(2y_1)$. Therefore,

$$g(y_1, y_2) = \begin{cases} 2\left(\frac{y_2}{y_1}\right) & \text{for } (y_1, y_2) \in T, \\ 0 & \text{otherwise.} \end{cases}$$ ◂

Example 3.9.10

Service Time in a Queue. Let X be the time that the server in a single-server queue will spend on a particular customer, and let Y be the rate at which the server can operate. A popular model for the conditional distribution of X given $Y = y$ is to say that the conditional p.d.f. of X given $Y = y$ is

$$g_1(x|y) = \begin{cases} ye^{-xy} & \text{for } x > 0, \\ 0 & \text{otherwise.} \end{cases}$$

Let Y have the p.d.f. $f_2(y)$. The joint p.d.f. of (X, Y) is then $g_1(x|y)f_2(y)$. Because $1/Y$ can be interpreted as the average service time, $Z = XY$ measures how quickly, compared to average, that the customer is served. For example, $Z = 1$ corresponds to an average service time, while $Z > 1$ means that this customer took longer than average, and $Z < 1$ means that this customer was served more quickly than the average customer. If we want the distribution of Z, we could compute the joint p.d.f. of (Z, Y) directly using the methods just illustrated. We could then integrate the joint p.d.f. over y to obtain the marginal p.d.f. of Z. However, it is simpler to transform the conditional distribution of X given $Y = y$ into the conditional distribution of Z given $Y = y$, since conditioning on $Y = y$ allows us to treat Y as the constant y. Because $X = Z/Y$, the inverse transformation is $x = s(z)$, where $s(z) = z/y$. The derivative of this is $1/y$, and the conditional p.d.f. of Z given $Y = y$ is

$$h_1(z|y) = \frac{1}{y} g_1\left(\frac{z}{y}\middle| y\right).$$

Because Y is a rate, $Y \geq 0$ and $X = Z/Y > 0$ if and only if $Z > 0$. So,

$$h_1(z|y) = \begin{cases} e^{-z} & \text{for } z > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{3.9.18}$$

Notice that h_1 does not depend on y, so Z is independent of Y and h_1 is the marginal p.d.f. of Z. The reader can verify all of this in Exercise 17. ◂

Note: Removing Dependence. The formula $Z = XY$ in Example 3.9.10 makes it look as if Z should depend on Y. In reality, however, multiplying X by Y removes the dependence that X already has on Y and makes the result independent of Y. This type of transformation that removes the dependence of one random variable on another is a very powerful technique for finding marginal distributions of transformations of random variables.

In Example 3.9.10, we mentioned that there was another, more straightforward but more tedious, way to compute the distribution of Z. That method, which is useful in many settings, is to transform (X, Y) into (Z, W) for some uninteresting random variable W and then integrate w out of the joint p.d.f. All that matters in the choice of W is that the transformation be one-to-one with differentiable inverse and that the calculations are feasible. Here is a specific example.

Example 3.9.11

One Function of Two Variables. In Example 3.9.9, suppose that we were interested only in the quotient $Y_1 = X_1/X_2$ rather than both the quotient and the product $Y_2 = X_1X_2$. Since we already have the joint p.d.f. of (Y_1, Y_2), we will merely integrate y_2 out rather than start from scratch. For each value of $y_1 > 0$, we need to look at the set T in Fig. 3.27 and find the interval of y_2 values to integrate over. For $0 < y_1 < 1$, we integrate over $0 < y_2 < y_1$. For $y_1 > 1$, we integrate over $0 < y_2 < 1/y_1$. (For $y_1 = 1$ both intervals are the same.) So, the marginal p.d.f. of Y_1 is

$$g_1(y_1) = \begin{cases} \int_0^{y_1} 2\left(\frac{y_2}{y_1}\right) dy_2 & \text{for } 0 < y_1 < 1, \\ \int_0^{1/y_1} 2\left(\frac{y_2}{y_1}\right) dy_2 & \text{for } y_1 > 1, \end{cases}$$

$$= \begin{cases} y_1 & \text{for } 0 < y_1 < 1, \\ \frac{1}{y_1^3} & \text{for } y_1 > 1. \end{cases}$$

There are other transformations that would have made the calculation of g_1 simpler if that had been all we wanted. See Exercise 21 for an example. ◀

Theorem 3.9.6 Linear Transformations. Let $\boldsymbol{X} = (X_1, \ldots, X_n)$ have a continuous joint distribution for which the joint p.d.f. is f. Define $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$ by

$$\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X}, \tag{3.9.19}$$

where $\boldsymbol{A}$ is a nonsingular $n \times n$ matrix. Then $\boldsymbol{Y}$ has a continuous joint distribution with p.d.f.

$$g(\boldsymbol{y}) = \frac{1}{|\det \boldsymbol{A}|} f(\boldsymbol{A}^{-1}\boldsymbol{y}) \quad \text{for } \boldsymbol{y} \in R^n, \tag{3.9.20}$$

where $\boldsymbol{A}^{-1}$ is the inverse of $\boldsymbol{A}$.

Proof Each Y_i is a linear combination of $X_1, \ldots, X_n$. Because $\boldsymbol{A}$ is nonsingular, the transformation in Eq. (3.9.19) is a one-to-one transformation of the entire space R^n onto itself. At every point $\boldsymbol{y} \in R^n$, the inverse transformation can be represented by the equation

$$\boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{y}. \tag{3.9.21}$$

The Jacobian J of the transformation that is defined by Eq. (3.9.21) is simply $J = \det \boldsymbol{A}^{-1}$. Also, it is known from the theory of determinants that

$$\det \boldsymbol{A}^{-1} = \frac{1}{\det \boldsymbol{A}}.$$

Therefore, at every point $\boldsymbol{y} \in R^n$, the joint p.d.f. $g(\boldsymbol{y})$ can be evaluated in the following way, according to Theorem 3.9.5: First, for $i = 1, \ldots, n$, the component x_i in $f(x_1, \ldots, x_n)$ is replaced with the ith component of the vector $\boldsymbol{A}^{-1}\boldsymbol{y}$. Then, the result is divided by $|\det \boldsymbol{A}|$. This produces Eq. (3.9.20). ■

Summary

We extended the construction of the distribution of a function of a random variable to the case of several functions of several random variables. If one only wants the distribution of one function r_1 of n random variables, the usual way to find this is to first find $n-1$ additional functions $r_2, \ldots, r_n$ so that the n functions together compose a one-to-one transformation. Then find the joint p.d.f. of the n functions and finally find the marginal p.d.f. of the first function by integrating out the extra $n-1$ variables. The method is illustrated for the cases of the sum and the range of several random variables.