4

Pairs of Random Variables

Chapter 2 and Chapter 3 analyze experiments in which an outcome is one number. This chapter and the next one analyze experiments in which an outcome is a collection of numbers. Each number is a sample value of a random variable. The probability model for such an experiment contains the properties of the individual random variables and it also contains the relationships among the random variables. Chapter 2 considers only discrete random variables and Chapter 3 considers only continuous random variables. The present chapter considers all random variables because a high proportion of the definitions and theorems apply to both discrete and continuous random variables. However, just as with individual random variables, the details of numerical calculations depend on whether random variables are discrete or continuous. Consequently we find that many formulas come in pairs. One formula, for discrete random variables, contains sums, and the other formula, for continuous random variables, contains integrals.

This chapter analyzes experiments that produce two random variables, X and Y. Chapter 5 analyzes the general case of experiments that produce n random variables, where n can be any integer. We begin with the definition of $F_{X,Y}(x, y)$, the *joint cumulative distribution function* of two random variables, a generalization of the CDF introduced in Section 2.4 and again in Section 3.1. The joint CDF is a complete probability model for any experiment that produces two random variables. However, it not very useful for analyzing practical experiments. More useful models are $P_{X,Y}(x, y)$, the *joint probability mass function* for two discrete random variables, presented in Sections 4.2 and 4.3, and $f_{X,Y}(x, y)$, the *joint probability density function* of two continuous random variables, presented in Sections 4.4 and 4.5. Sections 4.6 and 4.7 consider functions of two random variables and expectations, respectively. Sections 4.8, 4.9, and 4.10 go back to the concepts of conditional probability and independence introduced in Chapter 1. We extend the definition of independent events to define independent random variables. The subject of Section 4.11 is the special case in which X and Y are Gaussian.

Pairs of random variables appear in a wide variety of practical situations. An example of two random variables that we encounter all the time in our research is the signal (X), emitted by a radio transmitter, and the corresponding signal (Y) that eventually arrives at a receiver. In practice we observe Y, but we really want to know X. Noise and distortion prevent

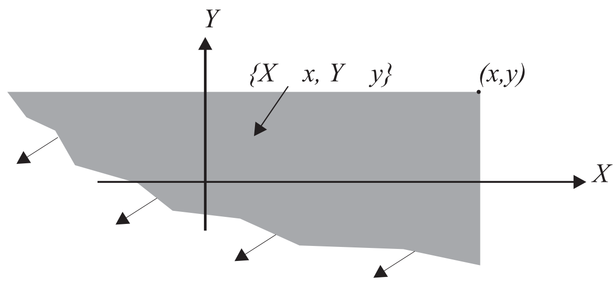


Figure 4.1 The area of the (X, Y) plane corresponding to the joint cumulative distribution function $F_{X,Y}(x, y)$.

us from observing X directly and we use the probability model $f_{X,Y}(x, y)$ to estimate X. Another example is the strength of the signal at a cellular telephone base station receiver (Y) and the distance (X) of the telephone from the base station. There are many more electrical engineering examples as well as examples throughout the physical sciences, biology, and social sciences. This chapter establishes the mathematical models for studying multiple continuous random variables.

4.1 Joint Cumulative Distribution Function

In an experiment that produces one random variable, events are points or intervals on a line. In an experiment that leads to two random variables X and Y, each outcome (x, y) is a point in a plane and events are points or areas in the plane.

Just as the CDF of one random variable, $F_X(x)$, is the probability of the interval to the left of x, the joint CDF $F_{X,Y}(x, y)$ of two random variables is the probability of the area in the plane below and to the left of (x, y). This is the infinite region that includes the shaded area in Figure 4.1 and everything below and to the left of it.

Definition 4.1 ***Joint Cumulative Distribution Function (CDF)***

*The **joint cumulative distribution function** of random variables X and Y is*

$$F_{X,Y}(x, y) = P[X \leq x, Y \leq y].$$

The joint CDF is a complete probability model. The notation is an extension of the notation convention adopted in Chapter 2. The subscripts of F, separated by a comma, are the names of the two random variables. Each name is an uppercase letter. We usually write the arguments of the function as the lowercase letters associated with the random variable names.

The joint CDF has properties that are direct consequences of the definition. For example, we note that the event $\{X \leq x\}$ suggests that Y can have any value so long as the condition on X is met. This corresponds to the joint event $\{X \leq x, Y < \infty\}$. Therefore,

$$F_X(x) = P[X \leq x] = P[X \leq x, Y < \infty] = \lim_{y \to \infty} F_{X,Y}(x, y) = F_{X,Y}(x, \infty). \quad (4.1)$$

We obtain a similar result when we consider the event $\{Y \leq y\}$. The following theorem summarizes some basic properties of the joint CDF.

Theorem 4.1 *For any pair of random variables, X, Y,*

(a) $0 \leq F_{X,Y}(x, y) \leq 1$,

(b) $F_X(x) = F_{X,Y}(x, \infty)$,

(c) $F_Y(y) = F_{X,Y}(\infty, y)$,

(d) $F_{X,Y}(-\infty, y) = F_{X,Y}(x, -\infty) = 0$,

(e) If $x \leq x_1$ and $y \leq y_1$, then $F_{X,Y}(x, y) \leq F_{X,Y}(x_1, y_1)$,

(f) $F_{X,Y}(\infty, \infty) = 1$.

Although its definition is simple, we rarely use the joint CDF to study probability models. It is easier to work with a probability mass function when the random variables are discrete, or a probability density function if they are continuous.

Quiz 4.1 *Express the following extreme values of the joint CDF $F_{X,Y}(x, y)$ as numbers or in terms of the CDFs $F_X(x)$ and $F_Y(y)$.*

(1) $F_{X,Y}(-\infty, 2)$ *(2)* $F_{X,Y}(\infty, \infty)$

(3) $F_{X,Y}(\infty, y)$ *(4)* $F_{X,Y}(\infty, -\infty)$

4.2 Joint Probability Mass Function

Corresponding to the PMF of a single discrete random variable, we have a probability mass function of two variables.

Definition 4.2 ***Joint Probability Mass Function (PMF)***

*The **joint probability mass function** of discrete random variables X and Y is*

$$P_{X,Y}(x, y) = P[X = x, Y = y].$$

For a pair of discrete random variables, the joint PMF $P_{X,Y}(x, y)$ is a complete probability model. For any pair of real numbers, the PMF is the probability of observing these numbers. The notation is consistent with that of the joint CDF. The uppercase subscripts of P, separated by a comma, are the names of the two random variables. We usually write the arguments of the function as the lowercase letters associated with the random variable names. Corresponding to S_X, the range of a single discrete random variable, we use the notation $S_{X,Y}$ to denote the set of possible values of the pair (X, Y). That is,

$$S_{X,Y} = \left\{(x, y) | P_{X,Y}(x, y) > 0\right\}. \tag{4.2}$$

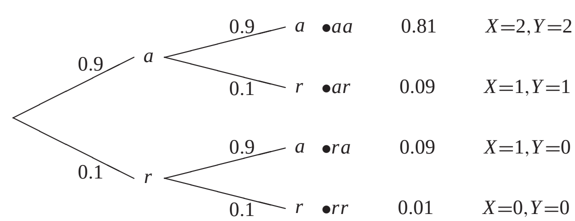
Keep in mind that $\{X = x, Y = y\}$ is an event in an experiment. That is, for this experiment, there is a set of observations that leads to both $X = x$ and $Y = y$. For any x and y, we find $P_{X,Y}(x, y)$ by summing the probabilities of all outcomes of the experiment for which $X = x$ and $Y = y$.

There are various ways to represent a joint PMF. We use three of them in the following example: a list, a matrix, and a graph.

Example 4.1 Test two integrated circuits one after the other. On each test, the possible outcomes are a (accept) and r (reject). Assume that all circuits are acceptable with probability 0.9 and that the outcomes of successive tests are independent. Count the number of acceptable circuits X and count the number of successful tests Y before you observe the first reject. (If both tests are successful, let $Y = 2$.) Draw a tree diagram for the experiment and find the joint PMF of X and Y.

....

The experiment has the following tree diagram.



The sample space of the experiment is

$$S = \{aa, ar, ra, rr\}. \tag{4.3}$$

Observing the tree diagram, we compute

$$P[aa] = 0.81, \qquad P[ar] = P[ra] = 0.09, \qquad P[rr] = 0.01. \tag{4.4}$$

Each outcome specifies a pair of values X and Y. Let $g(s)$ be the function that transforms each outcome s in the sample space S into the pair of random variables (X, Y). Then

$$g(aa) = (2, 2), \quad g(ar) = (1, 1), \quad g(ra) = (1, 0), \quad g(rr) = (0, 0). \tag{4.5}$$

For each pair of values x, y, $P_{X,Y}(x, y)$ is the sum of the probabilities of the outcomes for which $X = x$ and $Y = y$. For example, $P_{X,Y}(1, 1) = P[ar]$. The joint PMF can be given as a set of labeled points in the x, y plane where each point is a possible value (probability > 0) of the pair (x, y), or as a simple list:

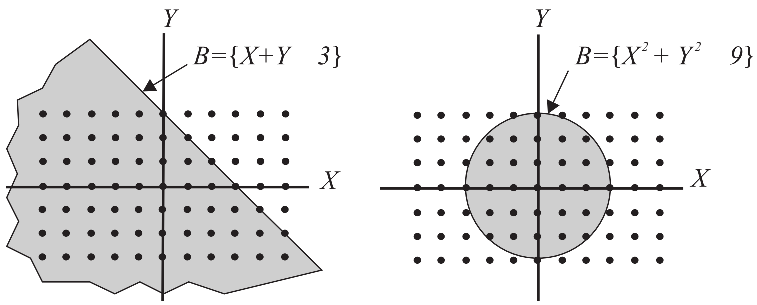
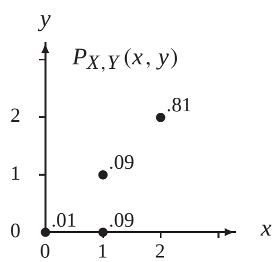


Figure 4.2 Subsets B of the (X, Y) plane. Points $(X, Y) \in S_{X,Y}$ are marked by bullets.



$$
P_{X,Y}(x, y) = \begin{cases} 0.81 & x = 2, y = 2, \\ 0.09 & x = 1, y = 1, \\ 0.09 & x = 1, y = 0, \\ 0.01 & x = 0, y = 0. \\ 0 & \text{otherwise} \end{cases} \tag{4.6}
$$

A third representation of $P_{X,Y}(x, y)$ is the matrix:

$$
\begin{array}{c|ccc} P_{X,Y}(x, y) & y = 0 & y = 1 & y = 2 \\ \hline x = 0 & 0.01 & 0 & 0 \\ x = 1 & 0.09 & 0.09 & 0 \\ x = 2 & 0 & 0 & 0.81 \end{array} \tag{4.7}
$$

Note that all of the probabilities add up to 1. This reflects the second axiom of probability (Section 1.3) that states $P[S] = 1$. Using the notation of random variables, we write this as

$$
\sum_{x \in S_X} \sum_{y \in S_Y} P_{X,Y}(x, y) = 1. \tag{4.8}
$$

As defined in Chapter 2, the range S_X is the set of all values of X with nonzero probability and similarly for S_Y. It is easy to see the role of the first axiom of probability in the PMF: $P_{X,Y}(x, y) \geq 0$ for all pairs x, y. The third axiom, which has to do with the union of disjoint events, takes us to another important property of the joint PMF.

We represent an event B as a region in the X, Y plane. Figure 4.2 shows two examples of events. We would like to find the probability that the pair of random variables (X, Y) is in the set B. When $(X, Y) \in B$, we say the event B occurs. Moreover, we write $P[B]$ as a shorthand for $P[(X, Y) \in B]$. The next theorem says that we can find $P[B]$ by adding the probabilities of all points (x, y) with nonzero probability that are in B.

Theorem 4.2 *For discrete random variables X and Y and any set B in the X, Y plane, the probability of the event $\{(X, Y) \in B\}$ is*

$$P[B] = \sum_{(x,y)\in B} P_{X,Y}(x, y).$$

The following example uses Theorem 4.2.

Example 4.2 Continuing Example 4.1, find the probability of the event B that X, the number of acceptable circuits, equals Y, the number of tests before observing the first failure.

Mathematically, B is the event $\{X = Y\}$. The elements of B with nonzero probability are

$$B \cap S_{X,Y} = \{(0, 0), (1, 1), (2, 2)\}. \tag{4.9}$$

Therefore,

$$P[B] = P_{X,Y}(0, 0) + P_{X,Y}(1, 1) + P_{X,Y}(2, 2) \tag{4.10}$$

$$= 0.01 + 0.09 + 0.81 = 0.91. \tag{4.11}$$

If we view x, y as the outcome of an experiment, then Theorem 4.2 simply says that to find the probability of an event, we sum over all the outcomes in that event. In essence, Theorem 4.2 is a restatement of Theorem 1.5 in terms of random variables X and Y and joint PMF $P_{X,Y}(x, y)$.

Quiz 4.2 *The joint PMF $P_{Q,G}(q, g)$ for random variables Q and G is given in the following table:*

$P_{Q,G}(q, g)$	$g = 0$	$g = 1$	$g = 2$	$g = 3$
$q = 0$	0.06	0.18	0.24	0.12
$q = 1$	0.04	0.12	0.16	0.08

(4.12)

Calculate the following probabilities:

(1) $P[Q = 0]$ *(2)* $P[Q = G]$

(3) $P[G > 1]$ *(4)* $P[G > Q]$

4.3 Marginal PMF

In an experiment that produces two random variables X and Y, it is always possible to consider one of the random variables, Y, and ignore the other one, X. In this case, we can use the methods of Chapter 2 to analyze the experiment and derive $P_Y(y)$, which contains the probability model for the random variable of interest. On the other hand, if we have already analyzed the experiment to derive the joint PMF $P_{X,Y}(x, y)$, it would be convenient to derive $P_Y(y)$ from $P_{X,Y}(x, y)$ without reexamining the details of the experiment.

To do so, we view x, y as the outcome of an experiment and observe that $P_{X,Y}(x, y)$ is the probability of an outcome. Moreover, $\{Y = y\}$ is an event, so that $P_Y(y) = P[Y = y]$ is the probability of an event. Theorem 4.2 relates the probability of an event to the joint

PMF. It implies that we can find $P_Y(y)$ by summing $P_{X,Y}(x, y)$ over all points in $S_{X,Y}$ with the property $Y = y$. In the sum, y is a constant, and each term corresponds to a value of $x \in S_X$. Similarly, we can find $P_X(x)$ by summing $P_{X,Y}(x, y)$ over all points X, Y such that $X = x$. We state this mathematically in the next theorem.

Theorem 4.3 *For discrete random variables X and Y with joint PMF $P_{X,Y}(x, y)$,*

$$P_X(x) = \sum_{y \in S_Y} P_{X,Y}(x, y), \qquad P_Y(y) = \sum_{x \in S_X} P_{X,Y}(x, y).$$

Theorem 4.3 shows us how to obtain the probability model (PMF) of X, and the probability model of Y given a probability model (joint PMF) of X and Y. When a random variable X is part of an experiment that produces two random variables, we sometimes refer to its PMF as a *marginal probability mass function*. This terminology comes from the matrix representation of the joint PMF. By adding rows and columns and writing the results in the margins, we obtain the marginal PMFs of X and Y. We illustrate this by reference to the experiment in Example 4.1.

Example 4.3 In Example 4.1, we found the joint PMF of X and Y to be

$P_{X,Y}(x, y)$	$y = 0$	$y = 1$	$y = 2$
$x = 0$	0.01	0	0
$x = 1$	0.09	0.09	0
$x = 2$	0	0	0.81

(4.13)

Find the marginal PMFs for the random variables X and Y.

. . . .

To find $P_X(x)$, we note that both X and Y have range $\{0, 1, 2\}$. Theorem 4.3 gives

$$P_X(0) = \sum_{y=0}^{2} P_{X,Y}(0, y) = 0.01 \qquad P_X(1) = \sum_{y=0}^{2} P_{X,Y}(1, y) = 0.18 \tag{4.14}$$

$$P_X(2) = \sum_{y=0}^{2} P_{X,Y}(2, y) = 0.81 \qquad P_X(x) = 0 \quad x \neq 0, 1, 2 \tag{4.15}$$

For the PMF of Y, we obtain

$$P_Y(0) = \sum_{x=0}^{2} P_{X,Y}(x, 0) = 0.10 \qquad P_Y(1) = \sum_{x=0}^{2} P_{X,Y}(x, 1) = 0.09 \tag{4.16}$$

$$P_Y(2) = \sum_{x=0}^{2} P_{X,Y}(x, 2) = 0.81 \qquad P_Y(y) = 0 \quad y \neq 0, 1, 2 \tag{4.17}$$

Referring to the matrix representation of $P_{X,Y}(x, y)$ in Example 4.1, we observe that each value of $P_X(x)$ is the result of adding all the entries in one row of the matrix. Each value of $P_Y(y)$ is a column sum. We display $P_X(x)$ and $P_Y(y)$ by rewriting the matrix in Example 4.1 and placing the row sums and column sums in the margins.

$P_{X,Y}(x,y)$	$y=0$	$y=1$	$y=2$	$P_X(x)$
$x=0$	0.01	0	0	0.01
$x=1$	0.09	0.09	0	0.18
$x=2$	0	0	0.81	0.81
$P_Y(y)$	0.10	0.09	0.81	

(4.18)

Note that the sum of all the entries in the bottom margin is 1 and so is the sum of all the entries in the right margin. This is simply a verification of Theorem 2.1(b), which states that the PMF of any random variable must sum to 1. The complete marginal PMF, $P_Y(y)$, appears in the bottom row of the table, and $P_X(x)$ appears in the last column of the table.

$$P_X(x) = \begin{cases} 0.01 & x=0, \\ 0.18 & x=1, \\ 0.81 & x=2, \\ 0 & \text{otherwise.} \end{cases} \qquad P_Y(y) = \begin{cases} 0.1 & y=0, \\ 0.09 & y=1, \\ 0.81 & y=2, \\ 0 & \text{otherwise.} \end{cases} \tag{4.19}$$

Quiz 4.3 *The probability mass function $P_{H,B}(h,b)$ for the two random variables H and B is given in the following table. Find the marginal PMFs $P_H(h)$ and $P_B(b)$.*

$P_{H,B}(h,b)$	$b=0$	$b=2$	$b=4$
$h=-1$	0	0.4	0.2
$h=0$	0.1	0	0.1
$h=1$	0.1	0.1	0

(4.20)

4.4 Joint Probability Density Function

The most useful probability model of a pair of continuous random variables is a generalization of the PDF of a single random variable (Definition 3.3).

Definition 4.3 ***Joint Probability Density Function (PDF)***
The joint PDF of the continuous random variables X and Y is a function $f_{X,Y}(x,y)$ with the property

$$F_{X,Y}(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(u,v)\, dv\, du.$$

For a single random variable X, the PDF $f_X(x)$ is a measure of probability per unit length. For two random variables X and Y, the joint PDF $f_{X,Y}(x,y)$ measures probability per unit area. In particular, from the definition of the PDF,

$$P[x < X \le x+dx, y < Y \le y+dy] = f_{X,Y}(x,y)\, dx\, dy. \tag{4.21}$$

Given $F_{X,Y}(x,y)$, Definition 4.3 implies that $f_{X,Y}(x,y)$ is a derivative of the CDF.

Theorem 4.4

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

Definition 4.3 and Theorem 4.4 demonstrate that the joint CDF $F_{X,Y}(x, y)$ and the joint PDF $f_{X,Y}(x, y)$ are equivalent probability models for random variables X and Y. In the case of one random variable, we found in Chapter 3 that the PDF is typically more useful for problem solving. This conclusion is even more true for pairs of random variables. Typically, it is very difficult to use $F_{X,Y}(x, y)$ to calculate the probabilities of events. To get an idea of the complication that arises, try proving the following theorem, which expresses the probability of a finite rectangle in the X, Y plane in terms of the joint CDF.

Theorem 4.5

$$\begin{aligned} P[x_1 < X \le x_2, y_1 < Y \le y_2] = {} & F_{X,Y}(x_2, y_2) - F_{X,Y}(x_2, y_1) \\ & - F_{X,Y}(x_1, y_2) + F_{X,Y}(x_1, y_1) \end{aligned}$$

The steps needed to prove the theorem are outlined in Problem 4.1.5. The theorem says that to find the probability that an outcome is in a rectangle, it is necessary to evaluate the joint CDF at all four corners. When the probability of interest corresponds to a nonrectangular area, the joint CDF is much harder to use.

Of course, not every function $f_{X,Y}(x, y)$ is a valid joint PDF. Properties (e) and (f) of Theorem 4.1 for the CDF $F_{X,Y}(x, y)$ imply corresponding properties for the PDF.

Theorem 4.6 *A joint PDF $f_{X,Y}(x, y)$ has the following properties corresponding to first and second axioms of probability (see Section 1.3):*

(a) $f_{X,Y}(x, y) \ge 0$ *for all* (x, y),

(b) $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x, y) \, dx \, dy = 1$.

Given an experiment that produces a pair of continuous random variables X and Y, an event A corresponds to a region of the X, Y plane. The probability of A is the double integral of $f_{X,Y}(x, y)$ over the region of the X, Y plane corresponding to A.

Theorem 4.7 *The probability that the continuous random variables (X, Y) are in A is*

$$P[A] = \iint_A f_{X,Y}(x, y) \, dx \, dy.$$

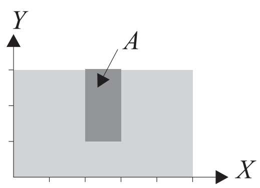
Example 4.4 Random variables X and Y have joint PDF

$$f_{X,Y}(x, y) = \begin{cases} c & 0 \le x \le 5, 0 \le y \le 3, \\ 0 & \text{otherwise.} \end{cases} \tag{4.22}$$

Find the constant c and $P[A] = P[2 \le X < 3, 1 \le Y < 3]$.

....

The large rectangle in the diagram is the area of nonzero probability. Theorem 4.6 states that the integral of the joint PDF over this rectangle is 1:



$$1 = \int_0^5 \int_0^3 c \, dy \, dx = 15c. \tag{4.23}$$

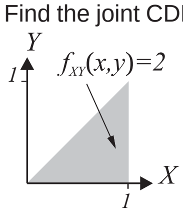
Therefore, $c = 1/15$. The small dark rectangle in the diagram is the event $A = \{2 \le X < 3, 1 \le Y < 3\}$. $P[A]$ is the integral of the PDF over this rectangle, which is

$$P[A] = \int_2^3 \int_1^3 \frac{1}{15} \, dv \, du = 2/15. \tag{4.24}$$

This probability model is an example of a pair of random variables uniformly distributed over a rectangle in the X, Y plane.

The following example derives the CDF of a pair of random variables that has a joint PDF that is easy to write mathematically. The purpose of the example is to introduce techniques for analyzing a more complex probability model than the one in Example 4.4. Typically, we extract interesting information from a model by integrating the PDF or a function of the PDF over some region in the X, Y plane. In performing this integration, the most difficult task is to identify the limits. The PDF in the example is very simple, just a constant over a triangle in the X, Y plane. However, to evaluate its integral over the region in Figure 4.1 we need to consider five different situations depending on the values of (x, y). The solution of the example demonstrates the point that the PDF is usually a more concise probability model that offers more insights into the nature of an experiment than the CDF.

Example 4.5 Find the joint CDF $F_{X,Y}(x, y)$ when X and Y have joint PDF



$$f_{X,Y}(x, y) = \begin{cases} 2 & 0 \le y \le x \le 1, \\ 0 & \text{otherwise.} \end{cases} \tag{4.25}$$

....

The joint CDF can be found using Definition 4.3 in which we integrate the joint PDF $f_{X,Y}(x, y)$ over the area shown in Figure 4.1. To perform the integration it is extremely useful to draw a diagram that clearly shows the area with nonzero probability, and then to use the diagram to derive the limits of the integral in Definition 4.3.

The difficulty with this integral is that the nature of the region of integration depends critically on x and y. In this apparently simple example, there are five cases to consider! The five cases are shown in Figure 4.3. First, we note that with $x < 0$ or $y < 0$, the triangle is completely outside the region of integration as shown in Figure 4.3a. Thus we have $F_{X,Y}(x, y) = 0$ if either $x < 0$ or $y < 0$. Another simple case arises when $x \ge 1$ and $y \ge 1$. In this case, we see in Figure 4.3e that the triangle is completely inside the region of integration and we infer from Theorem 4.6 that $F_{X,Y}(x, y) = 1$. The other cases we must consider are more complicated. In each case, since $f_{X,Y}(x, y) = 2$ over the triangular region, the value of the integral is two times the indicated area.

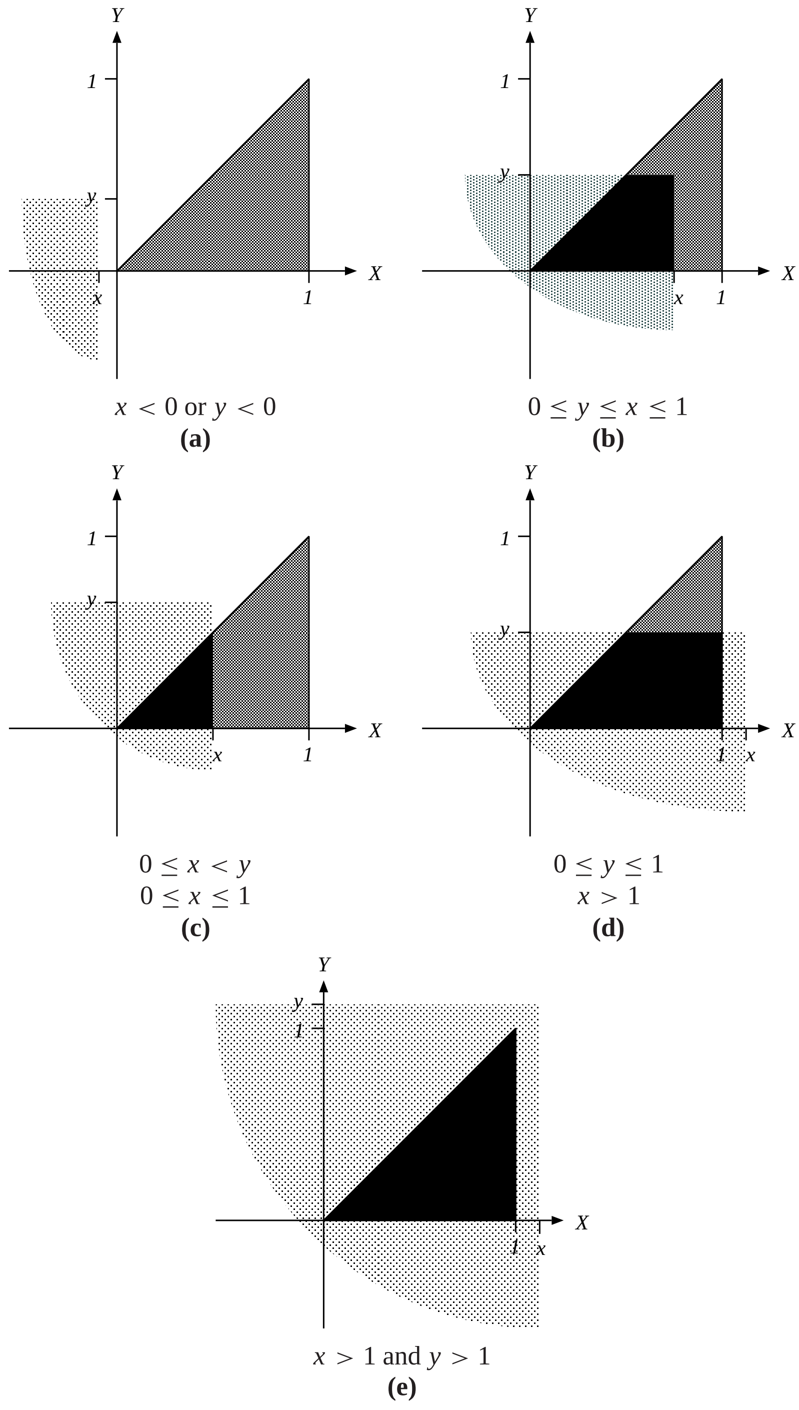


Figure 4.3 Five cases for the CDF $F_{X,Y}(x, y)$ of Example 4.5.

When (x, y) is inside the area of nonzero probability (Figure 4.3b), the integral is

$$F_{X,Y}(x, y) = \int_0^y \int_v^x 2\,du\,dv = 2xy - y^2 \qquad \text{(Figure 4.3b).} \tag{4.26}$$

In Figure 4.3c, (x, y) is above the triangle, and the integral is

$$F_{X,Y}(x, y) = \int_0^x \int_v^x 2\,du\,dv = x^2 \qquad \text{(Figure 4.3c).} \tag{4.27}$$

The remaining situation to consider is shown in Figure 4.3d when (x, y) is to the right of the triangle of nonzero probability, in which case the integral is

$$F_{X,Y}(x, y) = \int_0^y \int_v^1 2\,du\,dv = 2y - y^2 \qquad \text{(Figure 4.3d)} \tag{4.28}$$

The resulting CDF, corresponding to the five cases of Figure 4.3, is

$$F_{X,Y}(x, y) = \begin{cases} 0 & x < 0 \text{ or } y < 0 & \textbf{(a)}, \\ 2xy - y^2 & 0 \le y \le x \le 1 & \textbf{(b)}, \\ x^2 & 0 \le x < y, 0 \le x \le 1 & \textbf{(c)}, \\ 2y - y^2 & 0 \le y \le 1, x > 1 & \textbf{(d)}, \\ 1 & x > 1, y > 1 & \textbf{(e)}. \end{cases} \tag{4.29}$$

In Figure 4.4, the surface plot of $F_{X,Y}(x, y)$ shows that cases (a) through (e) correspond to contours on the "hill" that is $F_{X,Y}(x, y)$. In terms of visualizing the random variables, the surface plot of $F_{X,Y}(x, y)$ is less instructive than the simple triangle characterizing the PDF $f_{X,Y}(x, y)$.

Because the PDF in this example is two over $S_{X,Y}$, each probability is just two times the area of the region shown in one of the diagrams (either a triangle or a trapezoid). You may want to apply some high school geometry to verify that the results obtained from the integrals are indeed twice the areas of the regions indicated. The approach taken in our solution, integrating over $S_{X,Y}$ to obtain the CDF, works for any PDF.

In Example 4.5, it takes careful study to verify that $F_{X,Y}(x, y)$ is a valid CDF that satisfies the properties of Theorem 4.1, or even that it is defined for all values x and y. Comparing the joint PDF with the joint CDF we see that the PDF indicates clearly that X, Y occurs with equal probability in all areas of the same size in the triangular region $0 \le y \le x \le 1$. The joint CDF completely hides this simple, important property of the probability model.

In the previous example, the triangular shape of the area of nonzero probability demanded our careful attention. In the next example, the area of nonzero probability is a rectangle. However, the area corresponding to the event of interest is more complicated.

Example 4.6 As in Example 4.4, random variables X and Y have joint PDF

$$f_{X,Y}(x, y) = \begin{cases} 1/15 & 0 \le x \le 5, 0 \le y \le 3, \\ 0 & \text{otherwise.} \end{cases} \tag{4.30}$$

What is $P[A] = P[Y > X]$?

....

Applying Theorem 4.7, we integrate the density $f_{X,Y}(x, y)$ over the part of the X, Y

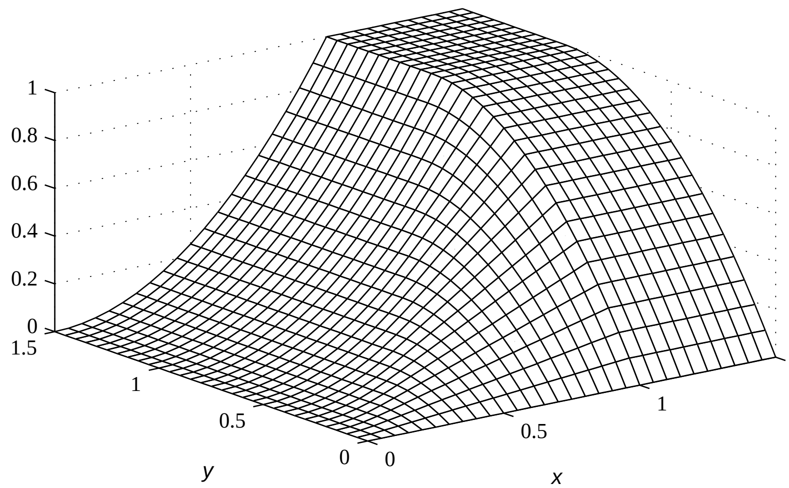
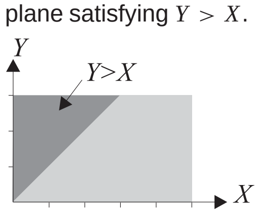


Figure 4.4 A graph of the joint CDF $F_{X,Y}(x, y)$ of Example 4.5.

plane satisfying $Y > X$. In this case,



$$P[A] = \int_0^3 \left(\int_x^3 \frac{1}{15} \right) dy\, dx \tag{4.31}$$

$$= \int_0^3 \frac{3-x}{15}\, dx = -\frac{(3-x)^2}{30}\bigg|_0^3 = \frac{3}{10}. \tag{4.32}$$

In this example, we note that it made little difference whether we integrate first over y and then over x or the other way around. In general, however, an initial effort to decide the simplest way to integrate over a region can avoid a lot of complicated mathematical maneuvering in performing the integration.

Quiz 4.4 *The joint probability density function of random variables X and Y is*

$$f_{X,Y}(x, y) = \begin{cases} cxy & 0 \le x \le 1, 0 \le y \le 2, \\ 0 & \textit{otherwise}. \end{cases} \tag{4.33}$$

Find the constant c. What is the probability of the event $A = X^2 + Y^2 \le 1$?

4.5 Marginal PDF

Suppose we perform an experiment that produces a pair of random variables X and Y with joint PDF $f_{X,Y}(x, y)$. For certain purposes we may be interested only in the random variable X. We can imagine that we ignore Y and observe only X. Since X is a random variable, it has a PDF $f_X(x)$. It should be apparent that there is a relationship between $f_X(x)$ and $f_{X,Y}(x,y)$. In particular, if $f_{X,Y}(x,y)$ completely summarizes our knowledge of joint events of the form $X = x, Y = y$, then we should be able to derive the PDFs of X and Y from $f_{X,Y}(x,y)$. The situation parallels (with integrals replacing sums) the relationship in Theorem 4.3 between the joint PMF $P_{X,Y}(x,y)$, and the marginal PMFs $P_X(x)$ and $P_Y(y)$. Therefore, we refer to $f_X(x)$ and $f_Y(y)$ as the *marginal probability density functions* of $f_{X,Y}(x,y)$.

Theorem 4.8 *If X and Y are random variables with joint PDF $f_{X,Y}(x,y)$,*

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

Proof From the definition of the joint PDF, we can write

$$F_X(x) = P[X \le x] = \int_{-\infty}^{x} \left(\int_{-\infty}^{\infty} f_{X,Y}(u,y)\,dy \right) du. \tag{4.34}$$

Taking the derivative of both sides with respect to x (which involves differentiating an integral with variable limits), we obtain $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy$. A similar argument holds for $f_Y(y)$.

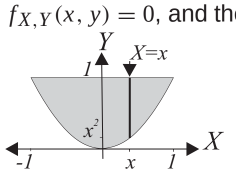
Example 4.7 The joint PDF of X and Y is

$$f_{X,Y}(x,y) = \begin{cases} 5y/4 & -1 \le x \le 1, x^2 \le y \le 1, \\ 0 & \text{otherwise.} \end{cases} \tag{4.35}$$

Find the marginal PDFs $f_X(x)$ and $f_Y(y)$.

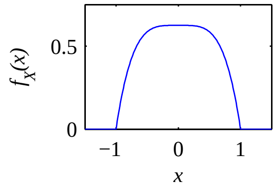
. . . .

We use Theorem 4.8 to find the marginal PDF $f_X(x)$. When $x < -1$ or when $x > 1$, $f_{X,Y}(x,y) = 0$, and therefore $f_X(x) = 0$. For $-1 \le x \le 1$,



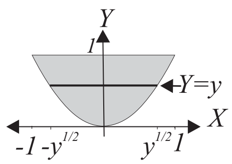
$$f_X(x) = \int_{x^2}^{1} \frac{5y}{4}\,dy = \frac{5(1-x^4)}{8}. \tag{4.36}$$

The complete expression for the marginal PDF of X is



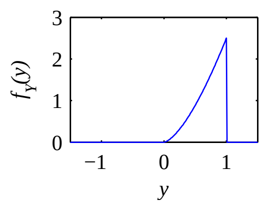
$$f_X(x) = \begin{cases} 5(1-x^4)/8 & -1 \le x \le 1, \\ 0 & \text{otherwise.} \end{cases} \tag{4.37}$$

For the marginal PDF of Y, we note that for $y < 0$ or $y > 1$, $f_Y(y) = 0$. For $0 \le y \le 1$, we integrate over the horizontal bar marked $Y = y$. The boundaries of the bar are $x = -\sqrt{y}$ and $x = \sqrt{y}$. Therefore, for $0 \le y \le 1$,



$$f_Y(y) = \int_{-\sqrt{y}}^{\sqrt{y}} \frac{5y}{4}\,dx = \frac{5y}{4}x\Big|_{x=-\sqrt{y}}^{x=\sqrt{y}} = 5y^{3/2}/2. \tag{4.38}$$

The complete marginal PDF of Y is



$$f_Y(y) = \begin{cases} (5/2)y^{3/2} & 0 \le y \le 1, \\ 0 & \text{otherwise.} \end{cases} \tag{4.39}$$

Quiz 4.5 *The joint probability density function of random variables X and Y is*

$$f_{X,Y}(x,y) = \begin{cases} 6(x+y^2)/5 & 0 \le x \le 1, 0 \le y \le 1, \\ 0 & \textit{otherwise.} \end{cases} \tag{4.40}$$

Find $f_X(x)$ and $f_Y(y)$, the marginal PDFs of X and Y.

4.6 Functions of Two Random Variables

There are many situations in which we observe two random variables and use their values to compute a new random variable. For example, we can describe the amplitude of the signal transmitted by a radio station as a random variable, X. We can describe the attenuation of the signal as it travels to the antenna of a moving car as another random variable, Y. In this case the amplitude of the signal at the radio receiver in the car is the random variable $W = X/Y$. Other practical examples appear in cellular telephone base stations with two antennas. The amplitudes of the signals arriving at the two antennas are modeled as random variables X and Y. The radio receiver connected to the two antennas can use the received signals in a variety of ways.

- It can choose the signal with the larger amplitude and ignore the other one. In this case, the receiver produces the random variable $W = X$ if $|X| > |Y|$ and $W = Y$, otherwise. This is an example of *selection diversity combining*.
- The receiver can add the two signals and use $W = X + Y$. This process is referred to as *equal gain combining* because it treats both signals equally.
- A third alternative is to combine the two signals unequally in order to give less weight to the signal considered to be more distorted. In this case $W = aX + bY$. If a and b are optimized, the receiver performs *maximal ratio combining*.

All three combining processes appear in practical radio receivers.

Formally, we have the following situation. We perform an experiment and observe sample values of two random variables X and Y. Based on our knowledge of the experiment, we have a probability model for X and Y embodied in the joint PMF $P_{X,Y}(x, y)$ or a joint PDF $f_{X,Y}(x, y)$. After performing the experiment, we calculate a sample value of the random variable $W = g(X, Y)$. The mathematical problem is to derive a probability model for W.

When X and Y are discrete random variables, S_W, the range of W, is a countable set corresponding to all possible values of $g(X, Y)$. Therefore, W is a discrete random variable and has a PMF $P_W(w)$. We can apply Theorem 4.2 to find $P_W(w) = P[W = w]$. Observe that $\{W = w\}$ is another name for the event $\{g(X, Y) = w\}$. Thus we obtain $P_W(w)$ by adding the values of $P_{X,Y}(x, y)$ corresponding to the x, y pairs for which $g(x, y) = w$.

Theorem 4.9 *For discrete random variables X and Y, the derived random variable $W = g(X, Y)$ has PMF*

$$P_W(w) = \sum_{(x,y):g(x,y)=w} P_{X,Y}(x, y).$$

Example 4.8 A firm sends out two kinds of promotional facsimiles. One kind contains only text and requires 40 seconds to transmit each page. The other kind contains grayscale pictures that take 60 seconds per page. Faxes can be 1, 2, or 3 pages long. Let the random variable L represent the length of a fax in pages. $S_L = \{1, 2, 3\}$. Let the random variable T represent the time to send each page. $S_T = \{40, 60\}$. After observing many fax transmissions, the firm derives the following probability model:

$P_{L,T}(l, t)$	$t = 40$ sec	$t = 60$ sec
$l = 1$ page	0.15	0.1
$l = 2$ pages	0.3	0.2
$l = 3$ pages	0.15	0.1

(4.41)

Let $D = g(L, T) = LT$ be the total duration in seconds of a fax transmission. Find the range S_D, the PMF $P_D(d)$, and the expected value $E[D]$.

....

By examining the six possible combinations of L and T we find that the possible values of D are $S_D = \{40, 60, 80, 120, 180\}$. For the five elements of S_D, we find the following probabilities:

$$\begin{aligned} &P_D(40) = P_{L,T}(1, 40) = 0.15, && P_D(120) = P_{L,T}(3, 40) + P_{L,T}(2, 60) = 0.35,\\ &P_D(60) = P_{L,T}(1, 60) = 0.1, && P_D(180) = P_{L,T}(3, 60) = 0.1,\\ &P_D(80) = P_{L,T}(2, 40) = 0.3, && P_D(d) = 0; \quad d \neq 40, 60, 80, 120, 180. \end{aligned}$$

The expected duration of a fax transmission is

$$E[D] = \sum_{d \in S_D} d P_D(d) \tag{4.42}$$

$$= (40)(0.15) + 60(0.1) + 80(0.3) + 120(0.35) + 180(0.1) = 96 \text{ sec}. \tag{4.43}$$

When X and Y are continuous random variables and $g(x, y)$ is a continuous function, $W = g(X, Y)$ is a continuous random variable. To find the PDF, $f_W(w)$, it is usually helpful to first find the CDF $F_W(w)$ and then calculate the derivative. Viewing $\{W \leq w\}$ as an event A, we can apply Theorem 4.7.

Theorem 4.10 *For continuous random variables X and Y, the CDF of $W = g(X, Y)$ is*

$$F_W(w) = P[W \leq w] = \iint_{g(x,y) \leq w} f_{X,Y}(x, y)\, dx\, dy.$$

Once we obtain the CDF $F_W(w)$, it is generally straightforward to calculate the derivative $f_W(w) = dF_W(w)/dw$. However, for most functions $g(x, y)$, performing the integration to find $F_W(w)$ can be a tedious process. Fortunately, there are convenient techniques for finding $f_W(w)$ for certain functions that arise in many applications. The most important function, $g(X, Y) = X + Y$, is the subject of Chapter 6. Another interesting function is the maximum of two random variables. The following theorem follows from the observation that $\{W \leq w\} = \{X \leq w\} \cap \{Y \leq w\}$.

Theorem 4.11 *For continuous random variables X and Y, the CDF of $W = \max(X, Y)$ is*

$$F_W(w) = F_{X,Y}(w, w) = \int_{-\infty}^{w} \int_{-\infty}^{w} f_{X,Y}(x, y)\, dx\, dy.$$

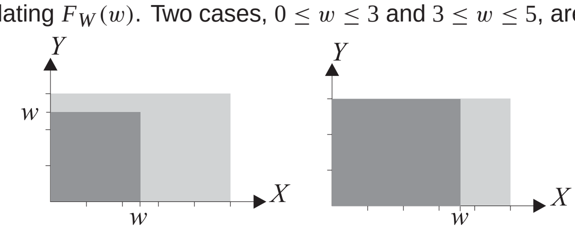
Example 4.9 In Examples 4.4 and 4.6, X and Y have joint PDF

$$f_{X,Y}(x, y) = \begin{cases} 1/15 & 0 \leq x \leq 5, 0 \leq y \leq 3, \\ 0 & \text{otherwise.} \end{cases} \tag{4.44}$$

Find the PDF of $W = \max(X, Y)$.

....

Because $X \geq 0$ and $Y \geq 0$, $W \geq 0$. Therefore, $F_W(w) = 0$ for $w < 0$. Because $X \leq 5$ and $Y \leq 3$, $W \leq 5$. Thus $F_W(w) = 1$ for $w \geq 5$. For $0 \leq w \leq 5$, diagrams provide a guide to calculating $F_W(w)$. Two cases, $0 \leq w \leq 3$ and $3 \leq w \leq 5$, are shown here:



$0 \leq w \leq 3$ $\qquad$ $3 \leq w \leq 5$

When $0 \leq w \leq 3$, Theorem 4.11 yields

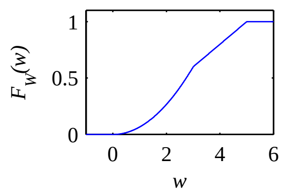
$$F_W(w) = \int_0^w \int_0^w \frac{1}{15}\, dx\, dy = w^2/15. \tag{4.45}$$

Because the joint PDF is uniform, we see this probability is just the area w^2 times the value of the joint PDF over that area. When $3 \leq w \leq 5$, the integral over the region

$\{X \le w, Y \le w\}$ becomes

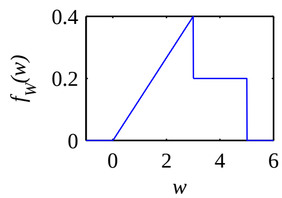
$$F_W(w) = \int_0^w \left(\int_0^3 \frac{1}{15}\, dy \right) dx = \int_0^w \frac{1}{5}\, dx = w/5, \tag{4.46}$$

which is the area $3w$ times the value of the joint PDF over that area. Combining the parts, we can write the joint CDF:



$$F_W(w) = \begin{cases} 0 & w < 0, \\ w^2/15 & 0 \le w \le 3, \\ w/5 & 3 < w \le 5, \\ 1 & w > 5. \end{cases} \tag{4.47}$$

By taking the derivative, we find the corresponding joint PDF:



$$f_W(w) = \begin{cases} 2w/15 & 0 \le w \le 3, \\ 1/5 & 3 < w \le 5, \\ 0 & \text{otherwise.} \end{cases} \tag{4.48}$$

In the following example, W is the quotient of two positive numbers.

Example 4.10 X and Y have the joint PDF

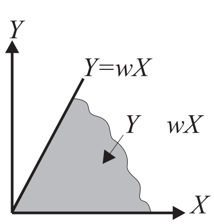
$$f_{X,Y}(x, y) = \begin{cases} \lambda\mu e^{-(\lambda x + \mu y)} & x \ge 0, y \ge 0, \\ 0 & \text{otherwise.} \end{cases} \tag{4.49}$$

Find the PDF of $W = Y/X$.

First we find the CDF:

$$F_W(w) = P[Y/X \le w] = P[Y \le wX]. \tag{4.50}$$

For $w < 0$, $F_W(w) = 0$. For $w \ge 0$, we integrate the joint PDF $f_{X,Y}(x, y)$ over the region of the X, Y plane for which $Y \le wX$, $X \ge 0$, and $Y \ge 0$ as shown:



$$P[Y \le wX] = \int_0^\infty \left(\int_0^{wx} f_{X,Y}(x, y)\, dy \right) dx \tag{4.51}$$

$$= \int_0^\infty \lambda e^{-\lambda x} \left(\int_0^{wx} \mu e^{-\mu y}\, dy \right) dx \tag{4.52}$$

$$= \int_0^\infty \lambda e^{-\lambda x} \left(1 - e^{-\mu w x}\right) dx \tag{4.53}$$

$$= 1 - \frac{\lambda}{\lambda + \mu w} \tag{4.54}$$

Therefore,

$$F_W(w) = \begin{cases} 0 & w < 0, \\ 1 - \frac{\lambda}{\lambda + \mu w} & \omega \ge 0. \end{cases} \tag{4.55}$$

Differentiating with respect to w, we obtain

$$f_W(w) = \begin{cases} \lambda\mu/(\lambda+\mu w)^2 & w \geq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{4.56}$$

Quiz 4.6

(A) *Two computers use modems and a telephone line to transfer e-mail and Internet news every hour. At the start of a data call, the modems at each end of the line negotiate a speed that depends on the line quality. When the negotiated speed is low, the computers reduce the amount of news that they transfer. The number of bits transmitted L and the speed B in bits per second have the joint PMF*

$P_{L,B}(l,b)$	$b = 14,400$	$b = 21,600$	$b = 28,800$
$l = 518,400$	0.2	0.1	0.05
$l = 2,592,000$	0.05	0.1	0.2
$l = 7,776,000$	0	0.1	0.2

(4.57)

Let T denote the number of seconds needed for the transfer. Express T as a function of L and B. What is the PMF of T?

(B) *Find the CDF and the PDF of $W = XY$ when random variables X and Y have joint PDF*

$$f_{X,Y}(x,y) = \begin{cases} 1 & 0 \leq x \leq 1, 0 \leq y \leq 1, \\ 0 & \text{otherwise}. \end{cases} \tag{4.58}$$

4.7 Expected Values

There are many situations in which we are interested only in the expected value of a derived random variable $W = g(X,Y)$, not the entire probability model. In these situations, we can obtain the expected value directly from $P_{X,Y}(x,y)$ or $f_{X,Y}(x,y)$ without taking the trouble to compute $P_W(w)$ or $f_W(w)$. Corresponding to Theorems 2.10 and 3.4, we have:

Theorem 4.12 *For random variables X and Y, the expected value of $W = g(X,Y)$ is*

$$\textit{Discrete:} \quad E[W] = \sum_{x \in S_X} \sum_{y \in S_Y} g(x,y) P_{X,Y}(x,y),$$

$$\textit{Continuous:} \; E[W] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y)\, dx\, dy.$$

Example 4.11 In Example 4.8, compute $E[D]$ directly from $P_{L,T}(l,t)$.

Applying Theorem 4.12 to the discrete random variable D, we obtain

$$E[D] = \sum_{l=1}^{3} \sum_{t=40,60} lt P_{L,T}(l,t) \tag{4.59}$$

$$= (1)(40)(0.15) + (1)60(0.1) + (2)(40)(0.3) + (2)(60)(0.2) \tag{4.60}$$

$$+ (3)(40)(0.15) + (3)(60)(0.1) = 96 \text{ sec}, \tag{4.61}$$

which is the same result obtained in Example 4.8 after calculating $P_D(d)$.

Theorem 4.12 is surprisingly powerful. For example, it lets us calculate easily the expected value of a sum.

Theorem 4.13

$$E[g_1(X,Y) + \cdots + g_n(X,Y)] = E[g_1(X,Y)] + \cdots + E[g_n(X,Y)].$$

Proof Let $g(X,Y) = g_1(X,Y) + \cdots + g_n(X,Y)$. For discrete random variables X, Y, Theorem 4.12 states

$$E[g(X,Y)] = \sum_{x \in S_X} \sum_{y \in S_Y} (g_1(x,y) + \cdots + g_n(x,y)) P_{X,Y}(x,y). \tag{4.62}$$

We can break the double summation into n double summations:

$$E[g(X,Y)] = \sum_{x \in S_X} \sum_{y \in S_Y} g_1(x,y) P_{X,Y}(x,y) + \cdots + \sum_{x \in S_X} \sum_{y \in S_Y} g_n(x,y) P_{X,Y}(x,y). \tag{4.63}$$

By Theorem 4.12, the ith double summation on the right side is $E[g_i(X,Y)]$, thus

$$E[g(X,Y)] = E\left[g_1(X,Y)\right] + \cdots + E[g_n(X,Y)]. \tag{4.64}$$

For continuous random variables, Theorem 4.12 says

$$E[g(X,Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (g_1(x,y) + \cdots + g_n(x,y)) f_{X,Y}(x,y)\, dx\, dy. \tag{4.65}$$

To complete the proof, we express this integral as the sum of n integrals and recognize that each of the new integrals is an expected value, $E[g_i(X,Y)]$.

In words, Theorem 4.13 says that the expected value of a sum equals the sum of the expected values. We will have many occasions to apply this theorem. The following theorem describes the expected sum of two random variables, a special case of Theorem 4.13.

Theorem 4.14 *For any two random variables X and Y,*

$$E[X+Y] = E[X] + E[Y].$$

An important consequence of this theorem is that we can find the expected sum of two random variables from the separate probability models: $P_X(x)$ and $P_Y(y)$ or $f_X(x)$ and $f_Y(y)$. We do not need a complete probability model embodied in $P_{X,Y}(x,y)$ or $f_{X,Y}(x,y)$.

By contrast, the variance of $X + Y$ depends on the entire joint PMF or joint CDF:

Theorem 4.15 *The variance of the sum of two random variables is*

$$\mathrm{Var}[X+Y] = \mathrm{Var}[X] + \mathrm{Var}[Y] + 2E[(X-\mu_X)(Y-\mu_Y)].$$

Proof Since $E[X+Y] = \mu_X + \mu_Y$,

$$\mathrm{Var}[X+Y] = E\left[(X+Y-(\mu_X+\mu_Y))^2\right] \tag{4.66}$$

$$= E\left[((X-\mu_X)+(Y-\mu_Y))^2\right] \tag{4.67}$$

$$= E\left[(X-\mu_X)^2 + 2(X-\mu_X)(Y-\mu_Y) + (Y-\mu_Y)^2\right]. \tag{4.68}$$

We observe that each of the three terms in the preceding expected values is a function of X and Y. Therefore, Theorem 4.13 implies

$$\mathrm{Var}[X+Y] = E\left[(X-\mu_X)^2\right] + 2E[(X-\mu_X)(Y-\mu_Y)] + E\left[(Y-\mu_Y)^2\right]. \tag{4.69}$$

The first and last terms are, respectively, $\mathrm{Var}[X]$ and $\mathrm{Var}[Y]$.

The expression $E[(X-\mu_X)(Y-\mu_Y)]$ in the final term of Theorem 4.15 reveals important properties of the relationship of X and Y. This quantity appears over and over in practical applications, and it has its own name, *covariance*.

Definition 4.4 ***Covariance***
The ***covariance*** *of two random variables X and Y is*

$$\mathrm{Cov}[X,Y] = E[(X-\mu_X)(Y-\mu_Y)].$$

Sometimes, the notation σ_{XY} is used to denote the covariance of X and Y. The *correlation* of two random variables, denoted $r_{X,Y}$, is a close relative of the covariance.

Definition 4.5 ***Correlation***
The ***correlation*** *of X and Y is $r_{X,Y} = E[XY]$*

The following theorem contains useful relationships among three expected values: the covariance of X and Y, the correlation of X and Y, and the variance of $X + Y$.

Theorem 4.16

(a) $\mathrm{Cov}[X,Y] = r_{X,Y} - \mu_X\mu_Y$.

(b) $\mathrm{Var}[X+Y] = \mathrm{Var}[X] + \mathrm{Var}[Y] + 2\,\mathrm{Cov}[X,Y]$.

(c) If $X = Y$, $\mathrm{Cov}[X,Y] = \mathrm{Var}[X] = \mathrm{Var}[Y]$ *and* $r_{X,Y} = E[X^2] = E[Y^2]$.

Proof Cross-multiplying inside the expected value of Definition 4.4 yields

$$\text{Cov}[X, Y] = E[XY - \mu_X Y - \mu_Y X + \mu_X \mu_Y]. \tag{4.70}$$

Since the expected value of the sum equals the sum of the expected values,

$$\text{Cov}[X, Y] = E[XY] - E[\mu_X Y] - E[\mu_Y X] + E[\mu_Y \mu_X]. \tag{4.71}$$

Note that in the expression $E[\mu_Y X]$, μ_Y is a constant. Referring to Theorem 2.12, we set $a = \mu_Y$ and $b = 0$ to obtain $E[\mu_Y X] = \mu_Y E[X] = \mu_Y \mu_X$. The same reasoning demonstrates that $E[\mu_X Y] = \mu_X E[Y] = \mu_X \mu_Y$. Therefore,

$$\text{Cov}[X, Y] = E[XY] - \mu_X \mu_Y - \mu_Y \mu_X + \mu_Y \mu_X = r_{X,Y} - \mu_X \mu_Y. \tag{4.72}$$

The other relationships follow directly from the definitions and Theorem 4.15.

Example 4.12 For the integrated circuits tests in Example 4.1, we found in Example 4.3 that the probability model for X and Y is given by the following matrix.

$P_{X,Y}(x, y)$	$y = 0$	$y = 1$	$y = 2$	$P_X(x)$
$x = 0$	0.01	0	0	0.01
$x = 1$	0.09	0.09	0	0.18
$x = 2$	0	0	0.81	0.81
$P_Y(y)$	0.10	0.09	0.81	

(4.73)

Find $r_{X,Y}$ and $\text{Cov}[X, Y]$.

By Definition 4.5,

$$r_{X,Y} = E[XY] = \sum_{x=0}^{2} \sum_{y=0}^{2} xy P_{X,Y}(x, y) \tag{4.74}$$

$$= (1)(1)0.09 + (2)(2)0.81 = 3.33. \tag{4.75}$$

To use Theorem 4.16(a) to find the covariance, we find

$$E[X] = (1)(0.18) + (2)(0.81) = 1.80, \tag{4.76}$$

$$E[Y] = (1)(0.09) + (2)(0.81) = 1.71. \tag{4.77}$$

Therefore, by Theorem 4.16(a), $\text{Cov}[X, Y] = 3.33 - (1.80)(1.71) = 0.252$.

Associated with the definitions of covariance and correlation are special terms to describe random variables for which $r_{X,Y} = 0$ and random variables for which $\text{Cov}[X, Y] = 0$.

Definition 4.6 ***Orthogonal Random Variables***

*Random variables X and Y are **orthogonal** if $r_{X,Y} = 0$.*

Definition 4.7 ***Uncorrelated Random Variables***

Random variables X and Y are ***uncorrelated*** *if* $\text{Cov}[X, Y] = 0$.

This terminology, while widely used, is somewhat confusing, since orthogonal means zero correlation while uncorrelated means zero covariance.

The correlation coefficient is closely related to the covariance of two random variables.

Definition 4.8 ***Correlation Coefficient***

The ***correlation coefficient*** *of two random variables X and Y is*

$$\rho_{X,Y} = \frac{\text{Cov}[X, Y]}{\sqrt{\text{Var}[X]\,\text{Var}[Y]}} = \frac{\text{Cov}[X, Y]}{\sigma_X \sigma_Y}.$$

Note that the units of the covariance and the correlation are the product of the units of X and Y. Thus, if X has units of kilograms and Y has units of seconds, then $\text{Cov}[X, Y]$ and $r_{X,Y}$ have units of kilogram-seconds. By contrast, $\rho_{X,Y}$ is a dimensionless quantity.

An important property of the correlation coefficient is that it is bounded by -1 and 1:

Theorem 4.17

$$-1 \leq \rho_{X,Y} \leq 1.$$

Proof Let σ_X^2 and σ_Y^2 denote the variances of X and Y and for a constant a, let $W = X - aY$. Then,

$$\text{Var}[W] = E\left[(X - aY)^2\right] - (E[X - aY])^2. \tag{4.78}$$

Since $E[X - aY] = \mu_X - a\mu_Y$, expanding the squares yields

$$\text{Var}[W] = E\left[X^2 - 2aXY + a^2Y^2\right] - \left(\mu_X^2 - 2a\mu_X\mu_Y + a^2\mu_Y^2\right) \tag{4.79}$$

$$= \text{Var}[X] - 2a\,\text{Cov}[X, Y] + a^2\,\text{Var}[Y]. \tag{4.80}$$

Since $\text{Var}[W] \geq 0$ for any a, we have $2a\,\text{Cov}[X, Y] \leq \text{Var}[X] + a^2\,\text{Var}[Y]$. Choosing $a = \sigma_X/\sigma_Y$ yields $\text{Cov}[X, Y] \leq \sigma_Y\sigma_X$, which implies $\rho_{X,Y} \leq 1$. Choosing $a = -\sigma_X/\sigma_Y$ yields $\text{Cov}[X, Y] \geq -\sigma_Y\sigma_X$, which implies $\rho_{X,Y} \geq -1$.

We encounter $\rho_{X,Y}$ in several contexts in this book. We will see that $\rho_{X,Y}$ describes the information we gain about Y by observing X. For example, a positive correlation coefficient, $\rho_{X,Y} > 0$, suggests that when X is high relative to its expected value, Y also tends to be high, and when X is low, Y is likely to be low. A negative correlation coefficient, $\rho_{X,Y} < 0$, suggests that a high value of X is likely to be accompanied by a low value of Y and that a low value of X is likely to be accompanied by a high value of Y. A linear relationship between X and Y produces the extreme values, $\rho_{X,Y} = \pm 1$.

Theorem 4.18 *If X and Y are random variables such that $Y = aX + b$,*

$$\rho_{X,Y} = \begin{cases} -1 & a < 0, \\ 0 & a = 0, \\ 1 & a > 0. \end{cases}$$

The proof is left as an exercise for the reader (Problem 4.7.7). Some examples of positive, negative, and zero correlation coefficients include:

- X is the height of a student. Y is the weight of the same student. $0 < \rho_{X,Y} < 1$.
- X is the distance of a cellular phone from the nearest base station. Y is the power of the received signal at the cellular phone. $-1 < \rho_{X,Y} < 0$.
- X is the temperature of a resistor measured in degrees Celsius. Y is the temperature of the same resistor measured in degrees Kelvin. $\rho_{X,Y} = 1$.
- X is the gain of an electrical circuit measured in decibels. Y is the attenuation, measured in decibels, of the same circuit. $\rho_{X,Y} = -1$.
- X is the telephone number of a cellular phone. Y is the social security number of the phone's owner. $\rho_{X,Y} = 0$.

Quiz 4.7

(A) *Random variables L and T given in Example 4.8 have joint PMF*

$P_{L,T}(l,t)$	$t = 40\,sec$	$t = 60\,sec$
$l = 1$ *page*	0.15	0.1
$l = 2$ *pages*	0.30	0.2
$l = 3$ *pages*	0.15	0.1.

(4.81)

Find the following quantities.

(1) $E[L]$ *and* $\text{Var}[L]$

(2) $E[T]$ *and* $\text{Var}[T]$

(3) *The correlation* $r_{L,T} = E[LT]$

(4) *The covariance* $\text{Cov}[L,T]$

(5) *The correlation coefficient* $\rho_{L,T}$

(B) *The joint probability density function of random variables X and Y is*

$$f_{X,Y}(x,y) = \begin{cases} xy & 0 \le x \le 1, 0 \le y \le 2, \\ 0 & \text{otherwise}. \end{cases} \tag{4.82}$$

Find the following quantities.

(1) $E[X]$ *and* $\text{Var}[X]$

(2) $E[Y]$ *and* $\text{Var}[Y]$

(3) *The correlation* $r_{X,Y} = E[XY]$

(4) *The covariance* $\text{Cov}[X,Y]$

(5) *The correlation coefficient* $\rho_{X,Y}$

4.8 Conditioning by an Event

An experiment produces two random variables, X and Y. We learn that the outcome (x, y) is an element of an event, B. We use the information $(x, y) \in B$ to construct a new probability model. If X and Y are discrete, the new model is a conditional joint PMF, the ratio of the joint PMF to $P[B]$. If X and Y are continuous, the new model is a conditional joint PDF, defined as the ratio of the joint PDF to $P[B]$. The definitions of these functions follow from the same intuition as Definition 1.6 for the conditional probability of an event. Section 4.9 considers the special case of an event that corresponds to an observation of one of the two random variables: either $B = \{X = x\}$, or $B = \{Y = y\}$.

Definition 4.9 ***Conditional Joint PMF***

For discrete random variables X and Y and an event, B with $P[B] > 0$, the ***conditional joint PMF*** *of X and Y given B is*

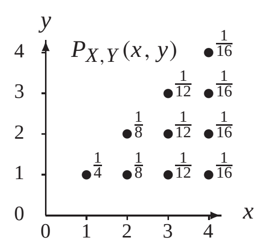
$$P_{X,Y|B}(x, y) = P[X = x, Y = y|B].$$

The following theorem is an immediate consequence of the definition.

Theorem 4.19 *For any event B, a region of the X, Y plane with $P[B] > 0$,*

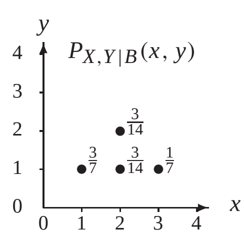
$$P_{X,Y|B}(x, y) = \begin{cases} \dfrac{P_{X,Y}(x, y)}{P[B]} & (x, y) \in B, \\ 0 & \textit{otherwise}. \end{cases}$$

Example 4.13



Random variables X and Y have the joint PMF $P_{X,Y}(x, y)$ as shown. Let B denote the event $X + Y \leq 4$. Find the conditional PMF of X and Y given B.

. . . .



Event $B = \{(1, 1), (2, 1), (2, 2), (3, 1)\}$ consists of all points (x, y) such that $x + y \leq 4$. By adding up the probabilities of all outcomes in B, we find

$$P[B] = P_{X,Y}(1, 1) + P_{X,Y}(2, 1) + P_{X,Y}(2, 2) + P_{X,Y}(3, 1) = \frac{7}{12}.$$

The conditional PMF $P_{X,Y|B}(x, y)$ is shown on the left.

In the case of two continuous random variables, we have the following definition of the conditional probability model.

Definition 4.10 ***Conditional Joint PDF***

Given an event B with $P[B] > 0$, the conditional joint probability density function of X and Y is

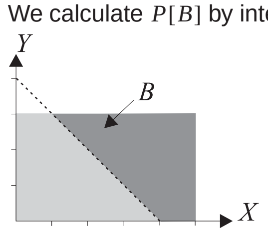
$$f_{X,Y|B}(x,y) = \begin{cases} \dfrac{f_{X,Y}(x,y)}{P[B]} & (x,y) \in B, \\ 0 & \text{otherwise}. \end{cases}$$

Example 4.14 X and Y are random variables with joint PDF

$$f_{X,Y}(x,y) = \begin{cases} 1/15 & 0 \le x \le 5, 0 \le y \le 3, \\ 0 & \text{otherwise.} \end{cases} \tag{4.83}$$

Find the conditional PDF of X and Y given the event $B = \{X + Y \ge 4\}$.

We calculate $P[B]$ by integrating $f_{X,Y}(x,y)$ over the region B.



$$P[B] = \int_0^3 \int_{4-y}^5 \frac{1}{15}\, dx\, dy \tag{4.84}$$

$$= \frac{1}{15} \int_0^3 (1+y)\, dy \tag{4.85}$$

$$= 1/2. \tag{4.86}$$

Definition 4.10 leads to the conditional joint PDF

$$f_{X,Y|B}(x,y) = \begin{cases} 2/15 & 0 \le x \le 5, 0 \le y \le 3, x + y \ge 4, \\ 0 & \text{otherwise.} \end{cases} \tag{4.87}$$

Corresponding to Theorem 4.12, we have

Theorem 4.20 ***Conditional Expected Value***

For random variables X and Y and an event B of nonzero probability, the conditional expected value of $W = g(X,Y)$ given B is

$$\textit{Discrete:} \quad E[W|B] = \sum_{x \in S_X} \sum_{y \in S_Y} g(x,y) P_{X,Y|B}(x,y),$$

$$\textit{Continuous:} \ E[W|B] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f_{X,Y|B}(x,y)\, dx\, dy.$$

Another notation for conditional expected value is $\mu_{W|B}$.

Definition 4.11 ***Conditional variance***

The **conditional variance** *of the random variable $W = g(X, Y)$ is*

$$\mathrm{Var}[W|B] = E\left[\left(W - \mu_{W|B}\right)^2 |B\right].$$

Another notation for conditional variance is $\sigma^2_{W|B}$. The following formula is a convenient computational shortcut.

Theorem 4.21

$$\mathrm{Var}[W|B] = E\left[W^2|B\right] - (\mu_{W|B})^2.$$

Example 4.15 Continuing Example 4.13, find the conditional expected value and the conditional variance of $W = X + Y$ given the event $B = \{X + Y \le 4\}$.

We recall from Example 4.13 that $P_{X,Y|B}(x, y)$ has four points with nonzero probability: $(1, 1)$, $(1, 2)$, $(1, 3)$, and $(2, 2)$. Their probabilities are $3/7$, $3/14$, $1/7$, and $3/14$, respectively. Therefore,

$$E[W|B] = \sum_{x,y}(x + y)P_{X,Y|B}(x, y) \tag{4.88}$$

$$= 2\frac{3}{7} + 3\frac{3}{14} + 4\frac{1}{7} + 4\frac{3}{14} = \frac{41}{14}. \tag{4.89}$$

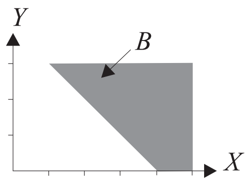
Similarly,

$$E\left[W^2|B\right] = \sum_{x,y}(x + y)^2 P_{X,Y|B}(x, y) \tag{4.90}$$

$$= 2^2\frac{3}{7} + 3^2\frac{3}{14} + 4^2\frac{1}{7} + 4^2\frac{3}{14} = \frac{131}{14}. \tag{4.91}$$

The conditional variance is $\mathrm{Var}[W|B] = E[W^2|B] - (E[W|B])^2 = (131/14) - (41/14)^2 = 153/196$.

Example 4.16 Continuing Example 4.14, find the conditional expected value of $W = XY$ given the event $B = \{X + Y \ge 4\}$.



For the event B shown in the adjacent graph, Example 4.14 showed that the conditional PDF of X, Y given B is

$$f_{X,Y|B}(x, y) = \begin{cases} 2/15 & 0 \le x \le 5, 0 \le y \le 3, (x, y) \in B, \\ 0 & \text{otherwise.} \end{cases}$$

From Theorem 4.20,

$$E[XY|B] = \int_0^3 \int_{4-y}^5 \frac{2}{15} xy\,dx\,dy \tag{4.92}$$

$$= \frac{1}{15}\int_0^3 \left(x^2\Big|_{4-y}^5 \right) y\,dy \tag{4.93}$$

$$= \frac{1}{15}\int_0^3 \left(9y + 8y^2 - y^3 \right) dy = \frac{123}{20}. \tag{4.94}$$

Quiz 4.8

(A) From Example 4.8, random variables L and T have joint PMF

$P_{L,T}(l,t)$	$t = 40\,sec$	$t = 60\,sec$
$l = 1$ *page*	0.15	0.1
$l = 2$ *pages*	0.3	0.2
$l = 3$ *pages*	0.15	0.1

(4.95)

For random variable $V = LT$, we define the event $A = \{V > 80\}$. Find the conditional PMF $P_{L,T|A}(l,t)$ of L and T given A. What are $E[V|A]$ and $\text{Var}[V|A]$?

(B) Random variables X and Y have the joint PDF

$$f_{X,Y}(x,y) = \begin{cases} xy/4000 & 1 \le x \le 3, 40 \le y \le 60, \\ 0 & \textit{otherwise}. \end{cases} \tag{4.96}$$

For random variable $W = XY$, we define the event $B = \{W > 80\}$. Find the conditional joint PDF $f_{X,Y|B}(l,t)$ of X and Y given B. What are $E[W|B]$ and $\text{Var}[W|B]$?

4.9 Conditioning by a Random Variable

In Section 4.8, we use the partial knowledge that the outcome of an experiment $(x, y) \in B$ in order to derive a new probability model for the experiment. Now we turn our attention to the special case in which the partial knowledge consists of the value of one of the random variables: either $B = \{X = x\}$ or $B = \{Y = y\}$. Learning $\{Y = y\}$ changes our knowledge of random variables X, Y. We now have complete knowledge of Y and modified knowledge of X. From this information, we derive a modified probability model for X. The new model is either a *conditional PMF of X given Y* or a *conditional PDF of X given Y*. When X and Y are discrete, the conditional PMF and associated expected values represent a specialized notation for their counterparts, $P_{X,Y|B}(x, y)$ and $E[g(X, Y)|B]$ in Section 4.8. By contrast, when X and Y are continuous, we cannot apply Section 4.8 directly because $P[B] = P[Y = y] = 0$ as discussed in Chapter 3. Instead, we define a conditional PDF as the ratio of the joint PDF to the marginal PDF.

Definition 4.12 ***Conditional PMF***

For any event $Y = y$ such that $P_Y(y) > 0$, the **conditional PMF** *of X given $Y = y$ is*

$$P_{X|Y}(x|y) = P[X = x|Y = y].$$

The following theorem contains the relationship between the joint PMF of X and Y and the two conditional PMFs, $P_{X|Y}(x|y)$ and $P_{Y|X}(y|x)$.

Theorem 4.22 *For random variables X and Y with joint PMF $P_{X,Y}(x, y)$, and x and y such that $P_X(x) > 0$ and $P_Y(y) > 0$,*

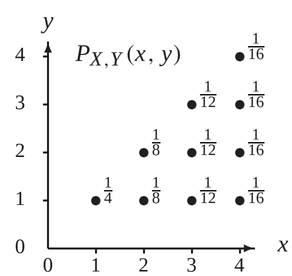
$$P_{X,Y}(x, y) = P_{X|Y}(x|y)\, P_Y(y) = P_{Y|X}(y|x)\, P_X(x).$$

Proof Referring to Definition 4.12, Definition 1.6, and Theorem 4.3, we observe that

$$P_{X|Y}(x|y) = P[X = x|Y = y] = \frac{P[X = x, Y = y]}{P[Y = y]} = \frac{P_{X,Y}(x, y)}{P_Y(y)}. \tag{4.97}$$

Hence, $P_{X,Y}(x, y) = P_{X|Y}(x|y) P_Y(y)$. The proof of the second part is the same with X and Y reversed.

Example 4.17



Random variables X and Y have the joint PMF $P_{X,Y}(x, y)$, as given in Example 4.13 and repeated in the accompanying graph. Find the conditional PMF of Y given $X = x$ for each $x \in S_X$.

....

To apply Theorem 4.22, we first find the marginal PMF $P_X(x)$. By Theorem 4.3, $P_X(x) = \sum_{y \in S_Y} P_{X,Y}(x, y)$. For a given $X = x$, we sum the nonzero probablities along the vertical line $X = x$. That is,

$$P_X(x) = \begin{cases} 1/4 & x = 1, \\ 1/8 + 1/8 & x = 2, \\ 1/12 + 1/12 + 1/12 & x = 3, \\ 1/16 + 1/16 + 1/16 + 1/16 & x = 4, \\ 0 & \text{otherwise}, \end{cases} = \begin{cases} 1/4 & x = 1, \\ 1/4 & x = 2, \\ 1/4 & x = 3, \\ 1/4 & x = 4, \\ 0 & \text{otherwise}. \end{cases}$$

Theorem 4.22 implies that for $x \in \{1, 2, 3, 4\}$,

$$P_{Y|X}(y|x) = \frac{P_{X,Y}(x, y)}{P_X(x)} = 4P_{X,Y}(x, y). \tag{4.98}$$

For each $x \in \{1, 2, 3, 4\}$, $P_{Y|X}(y|x)$ is a different PMF.

$$P_{Y|X}(y|1) = \begin{cases} 1 & y = 1, \\ 0 & \text{otherwise.} \end{cases} \qquad P_{Y|X}(y|2) = \begin{cases} 1/2 & y \in \{1, 2\}, \\ 0 & \text{otherwise.} \end{cases}$$

$$P_{Y|X}(y|3) = \begin{cases} 1/3 & y \in \{1, 2, 3\}, \\ 0 & \text{otherwise.} \end{cases} \qquad P_{Y|X}(y|4) = \begin{cases} 1/4 & y \in \{1, 2, 3, 4\}, \\ 0 & \text{otherwise.} \end{cases}$$

Given $X = x$, the conditional PMF of Y is the discrete uniform $(1, x)$ random variable.

For each $y \in S_Y$, the conditional probability mass function of X, gives us a new probability model of X. We can use this model in any way that we use $P_X(x)$, the model we have in the absence of knowledge of Y. Most important, we can find expected values with respect to $P_{X|Y}(x|y)$ just as we do in Chapter 2 with respect to $P_X(x)$.

Theorem 4.23 ***Conditional Expected Value of a Function***

X and Y are discrete random variables. For any $y \in S_Y$, the conditional expected value of $g(X, Y)$ given $Y = y$ is

$$E[g(X, Y)|Y = y] = \sum_{x \in S_X} g(x, y) P_{X|Y}(x|y).$$

The conditional expected value of X given $Y = y$ is a special case of Theorem 4.23:

$$E[X|Y = y] = \sum_{x \in S_X} x P_{X|Y}(x|y). \tag{4.99}$$

Theorem 4.22 shows how to obtain the conditional PMF given the joint PMF, $P_{X,Y}(x, y)$. In many practical situations, including the next example, we first obtain information about marginal and conditional probabilities. We can then use that information to build the complete model.

Example 4.18 In Example 4.17, we derived the following conditional PMFs: $P_{Y|X}(y|1)$, $P_{Y|X}(y|2)$, $P_{Y|X}(y|3)$, and $P_{Y|X}(y|4)$. Find $E[Y|X = x]$ for $x = 1, 2, 3, 4$.

Applying Theorem 4.23 with $g(x, y) = x$, we calculate

$$E[Y|X = 1] = 1, \qquad E[Y|X = 2] = 1.5, \tag{4.100}$$

$$E[Y|X = 3] = 2, \qquad E[Y|X = 4] = 2.5. \tag{4.101}$$

Now we consider the case in which X and Y are continuous random variables. We observe $\{Y = y\}$ and define the PDF of X given $\{Y = y\}$. We cannot use $B = \{Y = y\}$ in Definition 4.10 because $P[Y = y] = 0$. Instead, we define a *conditional probability density function*, denoted as $f_{X|Y}(x|y)$.

Definition 4.13 ***Conditional PDF***

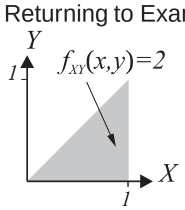
For y such that $f_Y(y) > 0$, the conditional PDF of X given $\{Y = y\}$ is

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}.$$

Definition 4.13 implies

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}. \tag{4.102}$$

Example 4.19 Returning to Example 4.5, random variables X and Y have joint PDF



$$f_{X,Y}(x,y) = \begin{cases} 2 & 0 \le y \le x \le 1, \\ 0 & \text{otherwise.} \end{cases} \tag{4.103}$$

For $0 \le x \le 1$, find the conditional PDF $f_{Y|X}(y|x)$. For $0 \le y \le 1$, find the conditional PDF $f_{X|Y}(x|y)$.

. . . .

For $0 \le x \le 1$, Theorem 4.8 implies

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dy = \int_0^x 2\, dy = 2x. \tag{4.104}$$

The conditional PDF of Y given X is

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \begin{cases} 1/x & 0 \le y \le x, \\ 0 & \text{otherwise.} \end{cases} \tag{4.105}$$

Given $X = x$, we see that Y is the uniform $(0, x)$ random variable. For $0 \le y \le 1$, Theorem 4.8 implies

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dx = \int_y^1 2\, dx = 2(1-y). \tag{4.106}$$

Furthermore, Equation (4.102) implies

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \begin{cases} 1/(1-y) & y \le x \le 1, \\ 0 & \text{otherwise.} \end{cases} \tag{4.107}$$

Conditioned on $Y = y$, we see that X is the uniform $(y, 1)$ random variable.

We can include both expressions for conditional PDFs in the following formulas.

Theorem 4.24

$$f_{X,Y}(x,y) = f_{Y|X}(y|x)\, f_X(x) = f_{X|Y}(x|y)\, f_Y(y).$$

For each y with $f_Y(y) > 0$, the conditional PDF $f_{X|Y}(x|y)$ gives us a new probability model of X. We can use this model in any way that we use $f_X(x)$, the model we have in the absence of knowledge of Y. Most important, we can find expected values with respect to $f_{X|Y}(x|y)$ just as we do in Chapter 3 with respect to $f_X(x)$. More generally, we define the conditional expected value of a function of the random variable X.

Definition 4.14 ***Conditional Expected Value of a Function***

For continuous random variables X and Y and any y such that $f_Y(y) > 0$, the ***conditional expected value*** *of $g(X, Y)$ given $Y = y$ is*

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

The conditional expected value of X given $Y = y$ is a special case of Definition 4.14:

$$E[X|Y = y] = \int_{-\infty}^{\infty} x f_{X|Y}(x|y)\, dx. \tag{4.108}$$

When we introduced the concept of expected value in Chapters 2 and 3, we observed that $E[X]$ is a number derived from the probability model of X. This is also true for $E[X|B]$. The conditional expected value given an event is a number derived from the conditional probability model. The situation is more complex when we consider $E[X|Y = y]$, the conditional expected value given a random variable. In this case, the conditional expected value is a different number for each possible observation $y \in S_Y$. Therefore, $E[X|Y = y]$ is a deterministic function of the observation y. This implies that when we perform an experiment and observe $Y = y$, $E[X|Y = y]$ is a function of the random variable Y. We use the notation $E[X|Y]$ to denote this function of the random variable Y. Since a function of a random variable is another random variable, we conclude that $E[X|Y]$ *is a random variable!* For some readers, the following definition may help to clarify this point.

Definition 4.15 ***Conditional Expected Value***

The conditional expected value $E[X|Y]$ is a function of random variable Y such that if $Y = y$ then $E[X|Y] = E[X|Y = y]$.

Example 4.20 For random variables X and Y in Example 4.5, we found in Example 4.19 that the conditional PDF of X given Y is

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x, y)}{f_Y(y)} = \begin{cases} 1/(1-y) & y \le x \le 1, \\ 0 & \text{otherwise.} \end{cases} \tag{4.109}$$

Find the conditional expected values $E[X|Y = y]$ and $E[X|Y]$.

. . . .

Given the conditional PDF $f_{X|Y}(x|y)$, we perform the integration

$$E[X|Y = y] = \int_{-\infty}^{\infty} x f_{X|Y}(x|y)\, dx \tag{4.110}$$

$$= \int_{y}^{1} \frac{1}{1-y} x\, dx = \left. \frac{x^2}{2(1-y)} \right|_{x=y}^{x=1} = \frac{1+y}{2}. \tag{4.111}$$

Since $E[X|Y=y]=(1+y)/2$, $E[X|Y]=(1+Y)/2$.

An interesting property of the random variable $E[X|Y]$ is its expected value $E[E[X|Y]]$. We find $E[E[X|Y]]$ in two steps: first we calculate $g(y)=E[X|Y=y]$ and then we apply Theorem 3.4 to evaluate $E[g(Y)]$. This two-step process is known as *iterated expectation*.

Theorem 4.25 ***Iterated Expectation***

$$E\left[E\left[X|Y\right]\right]=E\left[X\right].$$

Proof We consider continuous random variables X and Y and apply Theorem 3.4:

$$E\left[E\left[X|Y\right]\right]=\int_{-\infty}^{\infty}E\left[X|Y=y\right]f_Y\left(y\right)\,dy. \tag{4.112}$$

To obtain this formula from Theorem 3.4, we have used $E[X|Y=y]$ in place of $g(x)$ and $f_Y(y)$ in place of $f_X(x)$. Next, we substitute the right side of Equation (4.108) for $E[X|Y=y]$:

$$E\left[E\left[X|Y\right]\right]=\int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty}x f_{X|Y}\left(x|y\right)\,dx\right)f_Y\left(y\right)\,dy. \tag{4.113}$$

Rearranging terms in the double integral and reversing the order of integration, we obtain:

$$E\left[E\left[X|Y\right]\right]=\int_{-\infty}^{\infty}x\int_{-\infty}^{\infty}f_{X|Y}\left(x|y\right)f_Y\left(y\right)\,dy\,dx. \tag{4.114}$$

Next, we apply Theorem 4.24 and Theorem 4.8 to infer that the inner integral is simply $f_X(x)$. Therefore,

$$E\left[E\left[X|Y\right]\right]=\int_{-\infty}^{\infty}x f_X\left(x\right)\,dx. \tag{4.115}$$

The proof is complete because the right side of this formula is the definition of $E[X]$. A similar derivation (using sums instead of integrals) proves the theorem for discrete random variables.

The same derivation can be generalized to any function $g(X)$ of one of the two random variables:

Theorem 4.26

$$E\left[E\left[g(X)|Y\right]\right]=E\left[g(X)\right].$$

The following versions of Theorem 4.26 are instructive. If Y is continuous,

$$E\left[g(X)\right]=E\left[E\left[g(X)|Y\right]\right]=\int_{-\infty}^{\infty}E\left[g(X)|Y=y\right]f_Y\left(y\right)\,dy, \tag{4.116}$$

and if Y is discrete, we have a similar expression,

$$E\left[g(X)\right]=E\left[E\left[g(X)|Y\right]\right]=\sum_{y\in S_Y}E\left[g(X)|Y=y\right]P_Y\left(y\right). \tag{4.117}$$

Theorem 4.26 decomposes the calculation of $E[g(X)]$ into two steps: the calculation of $E[g(X)|Y = y]$, followed by the averaging of $E[g(X)|Y = y]$ over the distribution of Y. This is another example of iterated expectation. In Section 4.11, we will see that the iterated expectation can both facilitate understanding as well as simplify calculations.

Example 4.21 At noon on a weekday, we begin recording new call attempts at a telephone switch. Let X denote the arrival time of the first call, as measured by the number of seconds after noon. Let Y denote the arrival time of the second call. In the most common model used in the telephone industry, X and Y are continuous random variables with joint PDF

$$f_{X,Y}(x, y) = \begin{cases} \lambda^2 e^{-\lambda y} & 0 \le x < y, \\ 0 & \text{otherwise.} \end{cases} \tag{4.118}$$

where $\lambda > 0$ calls/second is the average arrival rate of telephone calls. Find the marginal PDFs $f_X(x)$ and $f_Y(y)$ and the conditional PDFs $f_{X|Y}(x|y)$ and $f_{Y|X}(y|x)$.

For $x < 0$, $f_X(x) = 0$. For $x \ge 0$, Theorem 4.8 gives $f_X(x)$:

$$f_X(x) = \int_x^\infty \lambda^2 e^{-\lambda y}\, dy = \lambda e^{-\lambda x}. \tag{4.119}$$

Referring to Appendix A.2, we see that X is an exponential random variable with expected value $1/\lambda$. Given $X = x$, the conditional PDF of Y is

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x, y)}{f_X(x)} = \begin{cases} \lambda e^{-\lambda(y-x)} & y > x, \\ 0 & \text{otherwise.} \end{cases} \tag{4.120}$$

To interpret this result, let $U = Y - X$ denote the interarrival time, the time between the arrival of the first and second calls. Problem 4.10.15 asks the reader to show that given $X = x$, U has the same PDF as X. That is, U is an exponential (λ) random variable.

Now we can find the marginal PDF of Y. For $y < 0$, $f_Y(y) = 0$. Theorem 4.8 implies

$$f_Y(y) = \begin{cases} \int_0^y \lambda^2 e^{-\lambda y}\, dx = \lambda^2 y e^{-\lambda y} & y \ge 0, \\ 0 & \text{otherwise.} \end{cases} \tag{4.121}$$

Y is the Erlang $(2, \lambda)$ random variable (Appendix A). Given $Y = y$, the conditional PDF of X is

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x, y)}{f_Y(y)} = \begin{cases} 1/y & 0 \le x < y, \\ 0 & \text{otherwise.} \end{cases} \tag{4.122}$$

Under the condition that the second call arrives at time y, the time of arrival of the first call is the uniform $(0, y)$ random variable.

In Example 4.21, we begin with a joint PDF and compute two conditional PDFs. Often in practical situations, we begin with a conditional PDF and a marginal PDF. Then we use this information to compute the joint PDF and the other conditional PDF.

Example 4.22 Let R be the uniform $(0, 1)$ random variable. Given $R = r$, X is the uniform $(0, r)$ random variable. Find the conditional PDF of R given X.

The problem definition states that

$$f_R(r) = \begin{cases} 1 & 0 \le r < 1, \\ 0 & \text{otherwise,} \end{cases} \qquad f_{X|R}(x|r) = \begin{cases} 1/r & 0 \le x < r < 1, \\ 0 & \text{otherwise.} \end{cases} \tag{4.123}$$

It follows from Theorem 4.24 that the joint PDF of R and X is

$$f_{R,X}(r,x) = f_{X|R}(x|r)\, f_R(r) = \begin{cases} 1/r & 0 \le x < r < 1, \\ 0 & \text{otherwise.} \end{cases} \tag{4.124}$$

Now we can find the marginal PDF of X from Theorem 4.8. For $0 < x < 1$,

$$f_X(x) = \int_{-\infty}^{\infty} f_{R,X}(r,x)\, dr = \int_x^1 \frac{dr}{r} = -\ln x. \tag{4.125}$$

By the definition of the conditional PDF,

$$f_{R|X}(r|x) = \frac{f_{R,X}(r,x)}{f_X(x)} = \begin{cases} \frac{1}{-r\ln x} & x \le r \le 1, \\ 0 & \text{otherwise.} \end{cases} \tag{4.126}$$

Quiz 4.9

(A) The probability model for random variable A is

$$P_A(a) = \begin{cases} 0.4 & a = 0, \\ 0.6 & a = 2, \\ 0 & \textit{otherwise.} \end{cases} \tag{4.127}$$

The conditional probability model for random variable B given A is

$$P_{B|A}(b|0) = \begin{cases} 0.8 & b = 0, \\ 0.2 & b = 1, \\ 0 & \textit{otherwise,} \end{cases} \qquad P_{B|A}(b|2) = \begin{cases} 0.5 & b = 0, \\ 0.5 & b = 1, \\ 0 & \textit{otherwise.} \end{cases} \tag{4.128}$$

(1) What is the probability model for A and B? Write the joint PMF $P_{A,B}(a,b)$ as a table.

(2) If $A = 2$, what is the conditional expected value $E[B|A = 2]$?

(3) If $B = 0$, what is the conditional PMF $P_{A|B}(a|0)$?

(4) If $B = 0$, what is the conditional variance $\mathrm{Var}[A|B = 0]$ of A?

(B) The PDF of random variable X and the conditional PDF of random variable Y given X are

$$f_X(x) = \begin{cases} 3x^2 & 0 \le x \le 1, \\ 0 & \textit{otherwise,} \end{cases} \qquad f_{Y|X}(y|x) = \begin{cases} 2y/x^2 & 0 \le y \le x, 0 < x \le 1, \\ 0 & \textit{otherwise.} \end{cases}$$

(1) What is the probability model for X and Y? Find $f_{X,Y}(x,y)$.

(2) If $X = 1/2$, find the conditional PDF $f_{Y|X}(y|1/2)$.

(3) If $Y = 1/2$, what is the conditional PDF $f_{X|Y}(x|1/2)$?

(4) If $Y = 1/2$, what is the conditional variance $\mathrm{Var}[X|Y = 1/2]$?

4.10 Independent Random Variables

Chapter 1 presents the concept of independent events. Definition 1.7 states that events A and B are independent if and only if the probability of the intersection is the product of the individual probabilities, $P[AB] = P[A]P[B]$.

Applying the idea of independence to random variables, we say that X and Y are independent random variables if and only if the events $\{X = x\}$ and $\{Y = y\}$ are independent for all $x \in S_X$ and all $y \in S_Y$. In terms of probability mass functions and probability density functions we have the following definition.

Definition 4.16 ***Independent Random Variables***

*Random variables X and Y are **independent** if and only if*

$$\text{Discrete:} \quad P_{X,Y}(x, y) = P_X(x)\, P_Y(y),$$

$$\text{Continuous:} \quad f_{X,Y}(x, y) = f_X(x)\, f_Y(y).$$

Because Definition 4.16 is an equality of functions, it must be true for all values of x and y. Theorem 4.22 implies that if X and Y are independent discrete random variables, then

$$P_{X|Y}(x|y) = P_X(x), \qquad P_{Y|X}(y|x) = P_Y(y). \tag{4.129}$$

Theorem 4.24 implies that if X and Y are independent continuous random variables, then

$$f_{X|Y}(x|y) = f_X(x) \qquad f_{Y|X}(y|x) = f_Y(y). \tag{4.130}$$

Example 4.23

$$f_{X,Y}(x, y) = \begin{cases} 4xy & 0 \le x \le 1, 0 \le y \le 1, \\ 0 & \text{otherwise.} \end{cases} \tag{4.131}$$

Are X and Y independent?

The marginal PDFs of X and Y are

$$f_X(x) = \begin{cases} 2x & 0 \le x \le 1, \\ 0 & \text{otherwise,} \end{cases} \qquad f_Y(y) = \begin{cases} 2y & 0 \le y \le 1, \\ 0 & \text{otherwise.} \end{cases} \tag{4.132}$$

It is easily verified that $f_{X,Y}(x, y) = f_X(x) f_Y(y)$ for all pairs (x, y) and so we conclude that X and Y are independent.

Example 4.24

$$f_{U,V}(u, v) = \begin{cases} 24uv & u \ge 0, v \ge 0, u + v \le 1, \\ 0 & \text{otherwise.} \end{cases} \tag{4.133}$$

Are U and V independent?

Since $f_{U,V}(u, v)$ looks similar in form to $f_{X,Y}(x, y)$ in the previous example, we might suppose that U and V can also be factored into marginal PDFs $f_U(u)$ and $f_V(v)$.

However, this is not the case. Owing to the triangular shape of the region of nonzero probability, the marginal PDFs are

$$f_U(u) = \begin{cases} 12u(1-u)^2 & 0 \le u \le 1, \\ 0 & \text{otherwise,} \end{cases} \tag{4.134}$$

$$f_V(v) = \begin{cases} 12v(1-v)^2 & 0 \le v \le 1, \\ 0 & \text{otherwise.} \end{cases} \tag{4.135}$$

Clearly, U and V are not independent. Learning U changes our knowledge of V. For example, learning $U = 1/2$ informs us that $P[V \le 1/2] = 1$.

In these two examples, we see that the region of nonzero probability plays a crucial role in determining whether random variables are independent. Once again, we emphasize that to infer that X and Y are independent, it is necessary to verify the functional equalities in Definition 4.16 for all $x \in S_X$ and $y \in S_Y$. There are many cases in which some events of the form $\{X = x\}$ and $\{Y = y\}$ are independent and others are not independent. If this is the case, the random variables X and Y are not independent.

The interpretation of independent random variables is a generalization of the interpretation of independent events. Recall that if A and B are independent, then learning that A has occurred does not change the probability of B occurring. When X and Y are independent random variables, the conditional PMF or the conditional PDF of X given $Y = y$ is the same for all $y \in S_Y$, and the conditional PMF or the conditional PDF of Y given $X = x$ is the same for all $x \in S_X$. Moreover, Equations (4.129) and (4.130) state that when two random variables are indpendent, each conditional PMF or PDF is identical to a corresponding marginal PMF or PDF. In summary, when X and Y are independent, observing $Y = y$ does not alter our probability model for X. Similarly, observing $X = x$ does not alter our probability model for Y. Therefore, learning that $Y = y$ provides no information about X, and learning that $X = x$ provides no information about Y.

The following theorem contains several important properties of expected values of independent random variables.

Theorem 4.27 *For independent random variables X and Y,*

(a) $E[g(X)h(Y)] = E[g(X)]E[h(Y)]$,

(b) $r_{X,Y} = E[XY] = E[X]E[Y]$,

(c) $\text{Cov}[X, Y] = \rho_{X,Y} = 0$,

(d) $\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y]$,

(e) $E[X|Y = y] = E[X]$ *for all* $y \in S_Y$,

(f) $E[Y|X = x] = E[Y]$ *for all* $x \in S_X$.

Proof We present the proof for discrete random variables. By replacing PMFs and sums with PDFs and integrals we arrive at essentially the same proof for continuous random variables. Since

$P_{X,Y}(x,y) = P_X(x)P_Y(y)$,

$$E[g(X)h(Y)] = \sum_{x \in S_X} \sum_{y \in S_Y} g(x)h(y)P_X(x)\,P_Y(y) \tag{4.136}$$

$$= \left(\sum_{x \in S_X} g(x)P_X(x) \right) \left(\sum_{y \in S_Y} h(y)P_Y(y) \right) = E[g(X)]\,E[h(Y)]. \tag{4.137}$$

If $g(X) = X$, and $h(Y) = Y$, this equation implies $r_{X,Y} = E[XY] = E[X]E[Y]$. This equation and Theorem 4.16(a) imply $\text{Cov}[X,Y] = 0$. As a result, Theorem 4.16(b) implies $\text{Var}[X+Y] = \text{Var}[X] + \text{Var}[Y]$. Furthermore, $\rho_{X,Y} = \text{Cov}[X,Y]/(\sigma_X \sigma_Y) = 0$.

Since $P_{X|Y}(x|y) = P_X(x)$,

$$E[X|Y=y] = \sum_{x \in S_X} x P_{X|Y}(x|y) = \sum_{x \in S_X} x P_X(x) = E[X]. \tag{4.138}$$

Since $P_{Y|X}(y|x) = P_Y(y)$,

$$E[Y|X=x] = \sum_{y \in S_Y} y P_{Y|X}(y|x) = \sum_{y \in S_Y} y P_Y(y) = E[Y]. \tag{4.139}$$

These results all follow directly from the joint PMF for independent random variables. We observe that Theorem 4.27(c) states that *independent random variables are uncorrelated*. We will have many occasions to refer to this property. It is important to know that while $\text{Cov}[X,Y] = 0$ is a necessary property for independence, it is not sufficient. There are many pairs of uncorrelated random variables that are *not* independent.

Example 4.25 Random variables X and Y have a joint PMF given by the following matrix

$P_{X,Y}(x,y)$	$y=-1$	$y=0$	$y=1$
$x=-1$	0	0.25	0
$x=1$	0.25	0.25	0.25

(4.140)

Are X and Y independent? Are X and Y uncorrelated?

. . . .

For the marginal PMFs, we have $P_X(-1) = 0.25$ and $P_Y(-1) = 0.25$. Thus

$$P_X(-1)\,P_Y(-1) = 0.0625 \neq P_{X,Y}(-1,-1) = 0, \tag{4.141}$$

and we conclude that X and Y are not independent.

To find $\text{Cov}[X,Y]$, we calculate

$$E[X] = 0.5, \qquad E[Y] = 0, \qquad E[XY] = 0. \tag{4.142}$$

Therefore, Theorem 4.16(a) implies

$$\text{Cov}[X,Y] = E[XY] - E[X]\,E[Y] = \rho_{X,Y} = 0, \tag{4.143}$$

and by definition X and Y are uncorrelated.

Quiz 4.10

(A) Random variables X and Y in Example 4.1 and random variables Q and G in Quiz 4.2 have joint PMFs:

$P_{X,Y}(x,y)$	$y=0$	$y=1$	$y=2$
$x=0$	0.01	0	0
$x=1$	0.09	0.09	0
$x=2$	0	0	0.81

$P_{Q,G}(q,g)$	$g=0$	$g=1$	$g=2$	$g=3$
$q=0$	0.06	0.18	0.24	0.12
$q=1$	0.04	0.12	0.16	0.08

(1) Are X and Y independent? *(2) Are Q and G independent?*

(B) Random variables X_1 and X_2 are independent and identically distributed with probability density function

$$f_X(x) = \begin{cases} 1 - x/2 & 0 \le x \le 2, \\ 0 & \textit{otherwise}. \end{cases} \tag{4.144}$$

(1) What is the joint PDF $f_{X_1,X_2}(x_1,x_2)$? *(2) Find the CDF of $Z = \max(X_1, X_2)$.*

4.11 Bivariate Gaussian Random Variables

The *bivariate Gaussian* disribution is a probability model for X and Y with the property that X and Y are each Gaussian random variables.

Definition 4.17 ***Bivariate Gaussian Random Variables***

Random variables X and Y have a ***bivariate Gaussian PDF*** *with parameters μ_1, σ_1, μ_2, σ_2, and ρ if*

$$f_{X,Y}(x,y) = \frac{\exp\left[-\dfrac{\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - \frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \left(\frac{y-\mu_2}{\sigma_2}\right)^2}{2(1-\rho^2)}\right]}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}},$$

where μ_1 and μ_2 can be any real numbers, $\sigma_1 > 0$, $\sigma_2 > 0$, and $-1 < \rho < 1$.

Figure 4.5 illustrates the bivariate Gaussian PDF for $\mu_1 = \mu_2 = 0$, $\sigma_1 = \sigma_2 = 1$, and three values of ρ. When $\rho = 0$, the joint PDF has the circular symmetry of a sombrero. When $\rho = 0.9$, the joint PDF forms a ridge over the line $x = y$, and when $\rho = -0.9$ there is a ridge over the line $x = -y$. The ridge becomes increasingly steep as $\rho \to \pm 1$.

To examine mathematically the properties of the bivariate Gaussian PDF, we define

$$\tilde{\mu}_2(x) = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x - \mu_1), \qquad \tilde{\sigma}_2 = \sigma_2\sqrt{1-\rho^2}, \tag{4.145}$$

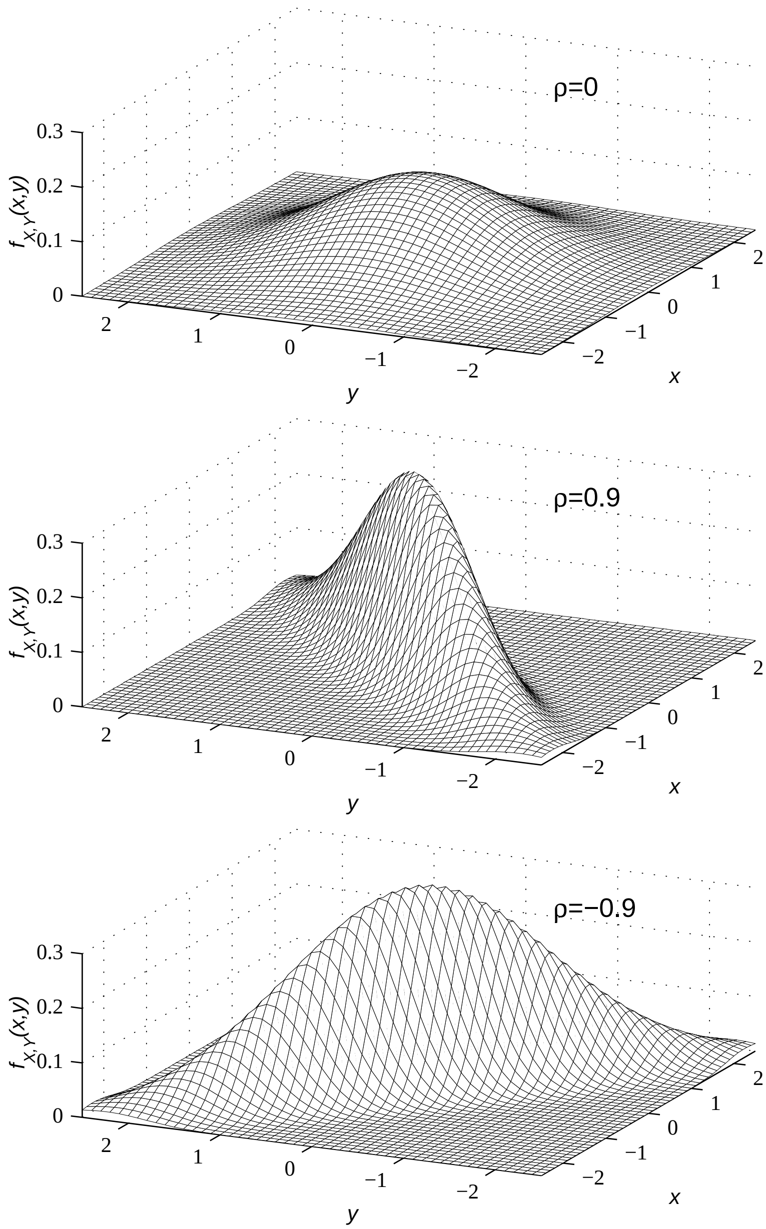


Figure 4.5 The Joint Gaussian PDF $f_{X,Y}(x, y)$ for $\mu_1 = \mu_2 = 0$, $\sigma_1 = \sigma_2 = 1$, and three values of ρ.

and manipulate the formula in Definition 4.17 to obtain the following expression for the joint Gaussian PDF:

$$f_{X,Y}(x, y) = \frac{1}{\sigma_1\sqrt{2\pi}} e^{-(x-\mu_1)^2/2\sigma_1^2} \frac{1}{\tilde{\sigma}_2\sqrt{2\pi}} e^{-(y-\tilde{\mu}_2(x))^2/2\tilde{\sigma}_2^2}. \tag{4.146}$$

Equation (4.146) expresses $f_{X,Y}(x, y)$ as the product of two Gaussian PDFs, one with parameters μ_1 and σ_1 and the other with parameters $\tilde{\mu}_2$ and $\tilde{\sigma}_2$. This formula plays a key role in the proof of the following theorem.

Theorem 4.28 *If X and Y are the bivariate Gaussian random variables in Definition 4.17, X is the Gaussian (μ_1, σ_1) random variable and Y is the Gaussian (μ_2, σ_2) random variable:*

$$f_X(x) = \frac{1}{\sigma_1\sqrt{2\pi}} e^{-(x-\mu_1)^2/2\sigma_1^2} \qquad f_Y(y) = \frac{1}{\sigma_2\sqrt{2\pi}} e^{-(y-\mu_2)^2/2\sigma_2^2}.$$

Proof Integrating $f_{X,Y}(x, y)$ in Equation (4.146) over all y, we have

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dy \tag{4.147}$$

$$= \frac{1}{\sigma_1\sqrt{2\pi}} e^{-(x-\mu_1)^2/2\sigma_1^2} \underbrace{\int_{-\infty}^{\infty} \frac{1}{\tilde{\sigma}_2\sqrt{2\pi}} e^{-(y-\tilde{\mu}_2(x))^2/2\tilde{\sigma}_2^2}\, dy}_{1} \tag{4.148}$$

The integral above the bracket equals 1 because it is the integral of a Gaussian PDF. The remainder of the formula is the PDF of the Gaussian (μ_1, σ_1) random variable. The same reasoning with the roles of X and Y reversed leads to the formula for $f_Y(y)$.

Given the marginal PDFs of X and Y, we use Definition 4.13 to find the conditional PDFs.

Theorem 4.29 *If X and Y are the bivariate Gaussian random variables in Definition 4.17, the conditional PDF of Y given X is*

$$f_{Y|X}(y|x) = \frac{1}{\tilde{\sigma}_2\sqrt{2\pi}} e^{-(y-\tilde{\mu}_2(x))^2/2\tilde{\sigma}_2^2},$$

where, given $X = x$, the conditional expected value and variance of Y are

$$\tilde{\mu}_2(x) = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x - \mu_1), \qquad \tilde{\sigma}_2^2 = \sigma_2^2(1 - \rho^2).$$

Theorem 4.29 is the result of dividing $f_{X,Y}(x, y)$ in Equation (4.146) by $f_X(x)$ to obtain $f_{Y|X}(y|x)$. The cross sections of Figure 4.6 illustrate the conditional PDF. The figure is a graph of $f_{X,Y}(x, y) = f_{Y|X}(y|x) f_X(x)$. Since X is a constant on each cross section, the cross section is a scaled picture of $f_{Y|X}(y|x)$. As Theorem 4.29 indicates, the cross section has the Gaussian bell shape. Corresponding to Theorem 4.29, the conditional PDF of X

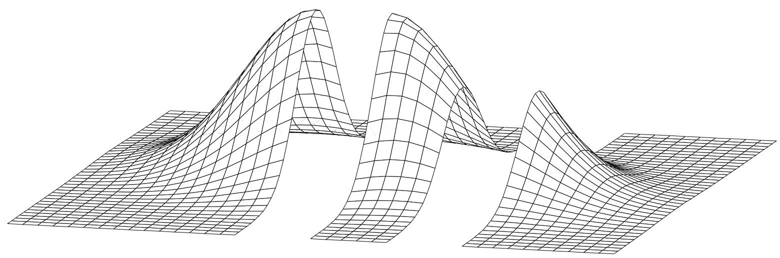


Figure 4.6 Cross-sectional view of the joint Gaussian PDF with $\mu_1 = \mu_2 = 0$, $\sigma_1 = \sigma_2 = 1$, and $\rho = 0.9$. Theorem 4.29 confirms that the bell shape of the cross section occurs because the conditional PDF $f_{Y|X}(y|x)$ is Gaussian.

given Y is also Gaussian. This conditional PDF is found by dividing $f_{X,Y}(x, y)$ by $f_Y(y)$ to obtain $f_{X|Y}(x|y)$.

Theorem 4.30 *If X and Y are the bivariate Gaussian random variables in Definition 4.17, the conditional PDF of X given Y is*

$$f_{X|Y}(x|y) = \frac{1}{\tilde{\sigma}_1\sqrt{2\pi}} e^{-(x-\tilde{\mu}_1(y))^2/2\tilde{\sigma}_1^2},$$

where, given $Y = y$, the conditional expected value and variance of X are

$$\tilde{\mu}_1(y) = \mu_1 + \rho\frac{\sigma_1}{\sigma_2}(y - \mu_2) \qquad \tilde{\sigma}_1^2 = \sigma_1^2(1 - \rho^2).$$

The next theorem identifies ρ in Definition 4.17 as the correlation coefficient of X and Y, $\rho_{X,Y}$.

Theorem 4.31 *Bivariate Gaussian random variables X and Y in Definition 4.17 have correlation coefficient*

$$\rho_{X,Y} = \rho.$$

Proof Substituting μ_1, σ_1, μ_2, and σ_2 for μ_X, σ_X, μ_Y, and σ_Y in Definition 4.4 and Definition 4.8, we have

$$\rho_{X,Y} = \frac{E\left[(X - \mu_1)(Y - \mu_2)\right]}{\sigma_1\sigma_2}. \tag{4.149}$$

To evaluate this expected value, we use the substitution $f_{X,Y}(x, y) = f_{Y|X}(y|x) f_X(x)$ in the double integral of Theorem 4.12. The result can be expressed as

$$\rho_{X,Y} = \frac{1}{\sigma_1\sigma_2} \int_{-\infty}^{\infty} (x - \mu_1) \left(\int_{-\infty}^{\infty} (y - \mu_2)\, f_{Y|X}(y|x)\, dy \right) f_X(x)\, dx \tag{4.150}$$

$$= \frac{1}{\sigma_1\sigma_2} \int_{-\infty}^{\infty} (x - \mu_1)\, E\left[Y - \mu_2 | X = x\right] f_X(x)\, dx \tag{4.151}$$

Because $E[Y|X = x] = \tilde{\mu}_2(x)$ in Theorem 4.29, it follows that

$$E\left[Y - \mu_2 | X = x\right] = \tilde{\mu}_2(x) - \mu_2 = \rho\frac{\sigma_2}{\sigma_1}(x - \mu_1) \tag{4.152}$$

Therefore,

$$\rho_{X,Y} = \frac{\rho}{\sigma_1^2} \int_{-\infty}^{\infty} (x - \mu_1)^2\, f_X(x)\, dx = \rho, \tag{4.153}$$

because the integral in the final expression is $\text{Var}[X] = \sigma_1^2$.

From Theorem 4.31, we observe that if X and Y are uncorrelated, then $\rho = 0$ and, from Theorems 4.29 and 4.30, $f_{Y|X}(y|x) = f_Y(y)$ and $f_{X|Y}(x|y) = f_X(x)$. Thus we have the following theorem.

Theorem 4.32 *Bivariate Gaussian random variables X and Y are uncorrelated if and only if they are independent.*

Theorem 4.31 identifies the parameter ρ in the bivariate gaussian PDF as the correlation coefficient $\rho_{X,Y}$ of bivariate Gaussian random variables X and Y. Theorem 4.17 states that for any pair of random variables, $|\rho_{X,Y}| < 1$, which explains the restriction $|\rho| < 1$ in Definition 4.17. Introducing this inequality to the formulas for conditional variance in Theorem 4.29 and Theorem 4.30 leads to the following inequalities:

$$\text{Var}[Y|X = x] = \sigma_2^2(1 - \rho^2) \leq \sigma_2^2, \tag{4.154}$$

$$\text{Var}[X|Y = y] = \sigma_1^2(1 - \rho^2) \leq \sigma_1^2. \tag{4.155}$$

These formulas state that for $\rho \neq 0$, learning the value of one of the random variables leads to a model of the other random variable with reduced variance. This suggests that learning the value of Y reduces our uncertainty regarding X.

Quiz 4.11 *Let X and Y be jointly Gaussian $(0, 1)$ random variables with correlation coefficient $1/2$.*

(1) What is the joint PDF of X and Y?

(2) What is the conditional PDF of X given $Y = 2$?

4.12 MATLAB

MATLAB is a useful tool for studying experiments that produce a pair of random variables X, Y. For discrete random variables X and Y with joint PMF $P_{X,Y}(x, y)$, we use MATLAB to to calculate probabilities of events and expected values of derived random variables $W = g(X, Y)$ using Theorem 4.9. In addition, simulation experiments often depend on the generation of sample pairs of random variables with specific probability models. That is, given a joint PMF $P_{X,Y}(x, y)$ or PDF $f_{X,Y}(x, y)$, we need to produce a collection $\{(x_1, y_1), (x_2, y_2), \ldots, (x_m, y_m)\}$. For finite discrete random variables, we are able to develop some general techniques. For continuous random variables, we give some specific examples.

Discrete Random Variables

We start with the case when X and Y are finite random variables with ranges

$$S_X = \{x_1, \ldots, x_n\} \qquad S_Y = \{y_1, \ldots, y_m\}. \tag{4.156}$$

In this case, we can take advantage of MATLAB techniques for surface plots of $g(x, y)$ over the x, y plane. In MATLAB, we represent S_X and S_Y by the n element vector `sx` and m element vector `sy`. The function `[SX,SY]=ndgrid(sx,sy)` produces the pair of $n \times m$ matrices,

$$\mathtt{SX} = \begin{bmatrix} x_1 & \cdots & x_1 \\ \vdots & & \vdots \\ x_n & \cdots & x_n \end{bmatrix}, \qquad \mathtt{SY} = \begin{bmatrix} y_1 & \cdots & y_m \\ \vdots & & \vdots \\ y_1 & \cdots & y_m \end{bmatrix}. \tag{4.157}$$

We refer to matrices `SX` and `SY` as a *sample space grid* because they are a grid representation of the joint sample space

$$S_{X,Y} = \{(x, y)|x \in S_X, y \in S_Y\}. \tag{4.158}$$

That is, `[SX(i,j) SY(i,j)]` is the pair (x_i, y_j).

To complete the probability model, for X and Y, in MATLAB, we employ the $n \times m$ matrix `PXY` such that `PXY(i,j)` $= P_{X,Y}(x_i, y_j)$. To make sure that probabilities have been generated properly, we note that `[SX(:) SY(:) PXY(:)]` is a matrix whose rows list all possible pairs x_i, y_j and corresponding probabilities $P_{X,Y}(x_i, y_j)$.

Given a function $g(x, y)$ that operates on the elements of vectors `x` and `y`, the advantage of this grid approach is that the MATLAB function `g(SX,SY)` will calculate $g(x, y)$ for each $x \in S_X$ and $y \in S_Y$. In particular, `g(SX,SY)` produces an $n \times m$ matrix with i, jth element $g(x_i, y_j)$.

Example 4.26 An Internet photo developer Web site prints compressed photo images. Each image file contains a variable-sized image of $X \times Y$ pixels described by the joint PMF

$P_{X,Y}(x, y)$	$y = 400$	$y = 800$	$y = 1200$
$x = 800$	0.2	0.05	0.1
$x = 1200$	0.05	0.2	0.1
$x = 1600$	0	0.1	0.2.

(4.159)

For random variables X, Y, write a script `imagepmf.m` that defines the sample space grid matrices `SX`, `SY`, and `PXY`.

In the script `imagepmf.m`, the matrix `SX` has $\begin{bmatrix}800 & 1200 & 1600\end{bmatrix}'$ for each column while `SY` has $\begin{bmatrix}400 & 800 & 1200\end{bmatrix}$ for each row. After running `imagepmf.m`, we can inspect the variables:

```
%imagepmf.m
PXY=[0.2  0.05 0.1; ...
     0.05 0.2  0.1; ...
     0    0.1  0.2];
[SX,SY]=ndgrid([800 1200 1600],...
    [400 800 1200]);
```

```
» imagepmf
» SX
SX =
     800    800   800
    1200   1200  1200
    1600   1600  1600
» SY
SY =
     400    800  1200
     400    800  1200
     400    800  1200
```

Example 4.27 At 24 bits (3 bytes) per pixel, a 10:1 image compression factor yields image files with $B = 0.3XY$ bytes. Find the expected value $E[B]$ and the PMF $P_B(b)$.

The script `imagesize.m` produces the expected value as `eb`, and the PMF, represented by the vectors `sb` and `pb`.

```
%imagesize.m
imagepmf;
SB=0.3*(SX.*SY);
eb=sum(sum(SB.*PXY))
sb=unique(SB)'
pb=finitepmf(SB,PXY,sb)'
```

The 3×3 matrix `SB` has i, jth element $g(x_i, y_j) = 0.3x_i y_j$. The calculation of `eb` is simply a MATLAB implementation of Theorem 4.12. Since some elements of `SB` are identical, `sb=unique(SB)` extracts the unique elements.

Although `SB` and `PXY` are both 3×3 matrices, each is stored internally by MATLAB as a 9-element vector. Hence, we can pass `SB` and `PXY` to the `finitepmf()` function which was designed to handle a finite random variable described by a pair of column vectors. Figure 4.7 shows one result of running the program `imagesize`. The vectors `sb` and `pb` comprise $P_B(b)$. For example, $P_B(288000) = 0.3$.

We note that `ndgrid` is very similar to another MATLAB function `meshgrid` that is more commonly used for graphing scalar functions of two variables. For our purposes, `ndgrid` is more convenient. In particular, as we can observe from Example 4.27, the matrix `PXY` has the same row and column structure as our table representation of $P_{X,Y}(x, y)$.

Random Sample Pairs

For finite random variable pairs X, Y described by S_X, S_Y and joint PMF $P_{X,Y}(x, y)$, or equivalently `SX`, `SY`, and `PXY` in MATLAB, we can generate random sample pairs using the function `finiterv(s,p,m)` defined in Chapter 2. Recall that `x=finiterv(s,p,m)` returned m samples (arranged as a column vector `x`) of a random variable X such that a sample value is `s(i)` with probability `p(i)`. In fact, to support random variable pairs

```
» imagesize
eb =
      319200
sb =
    96000  144000  192000  288000  384000  432000  576000
pb =
   0.2000  0.0500  0.0500  0.3000  0.1000  0.1000  0.2000
```

Figure 4.7 Output resulting from `imagesize.m` in Example 4.27.

X, Y, the function `w=finiterv(s,p,m)` permits `s` to be a $k \times 2$ matrix where the rows of `s` enumerate all pairs (x, y) with nonzero probability. Given the grid representation `SX`, `SY`, and `PXY`, we generate m sample pairs via

```
xy=finiterv([SX(:) SY(:)],PXY(:),m)
```

In particular, the ith pair, `SX(i),SY(i)`, will occur with probability `PXY(i)`. The output `xy` will be an $m \times 2$ matrix such that each row represents a sample pair x, y.

Example 4.28 Write a function `xy=imagerv(m)` that generates m sample pairs of the image size random variables X, Y of Example 4.27.

The function `imagerv` uses the `imagesize.m` script to define the matrices `SX`, `SY`, and `PXY`. It then calls the `finiterv.m` function. Here is the code `imagerv.m` and a sample run:

```
function xy = imagerv(m);
imagepmf;
S=[SX(:) SY(:)];
xy=finiterv(S,PXY(:),m);
```

```
» xy=imagerv(3)
xy =
         800         400
        1200         800
        1600         800
```

Example 4.28 can be generalized to produce sample pairs for any discrete random variable pair X, Y. However, given a collection of, for example, $m = 10,000$ samples of X, Y, it is desirable to be able to check whether the code generates the sample pairs properly. In particular, we wish to check for each $x \in S_X$ and $y \in S_Y$ whether the relative frequency of x, y in m samples is close to $P_{X,Y}(x, y)$. In the following example, we develop a program to calculate a matrix of relative frequencies that corresponds to the matrix `PXY`.

Example 4.29 Given a list `xy` of sample pairs of random variables X, Y with MATLAB range grids `SX` and `SY`, write a MATLAB function `fxy=freqxy(xy,SX,SY)` that calculates the relative frequency of every pair x, y. The output `fxy` should correspond to the matrix `[SX(:)  SY(:)  PXY(:)]`.

```
function fxy = freqxy(xy,SX,SY)
xy=[xy; SX(:) SY(:)];
[U,I,J]=unique(xy,'rows');
N=hist(J,1:max(J))-1;
N=N/sum(N);
fxy=[U N(:)];
fxy=sortrows(fxy,[2 1 3]);
```

In `freqxy.m`, the rows of the matrix `[SX(:) SY(:)]` list all possible pairs x, y. We append this matrix to `xy` to ensure that the new `xy` has every possible pair x, y. Next, the `unique` function copies all unique rows of `xy` to the matrix `U` and also provides the vector `J` that indexes the rows of `xy` in `U`; that is, `xy=U(J)`.

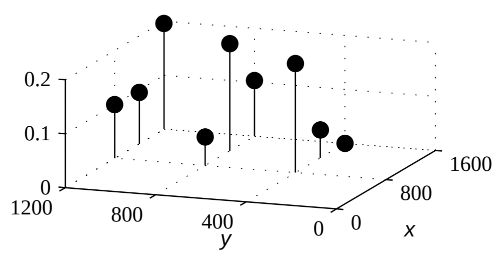
In addition, the number of occurrences of `j` in `J` indicates the number of occurrences in `xy` of row `j` in `U`. Thus we use the `hist` function on `J` to calculate the relative frequencies. We include the correction factor `-1` because we had appended `[SX(:) SY(:)]` to `xy` at the start. Lastly, we reorder the rows of `fxy` because the output of `unique` produces the rows of `U` in a different order from `[SX(:) SY(:) PXY(:)]`.

MATLAB provides the function `stem3(x,y,z)`, where `x`, `y`, and `z` are length n vectors, for visualizing a bivariate PMF $P_{X,Y}(x, y)$ or for visualizing relative frequencies of sample values of a pair of random variables. At each position `x(i)`, `y(i)` on the xy plane, the function draws a stem of height `z(i)`.

Example 4.30 Generate $m = 10,000$ samples of random variables X, Y of Example 4.27. Calculate the relative frequencies and use `stem3` to graph them.

The script `imagestem.m` generates the following relative frequency stem plot.

```
%imagestem.m
imagepmf;
xy=imagerv(10000);
fxy=freqxy(xy,SX,SY);
stem3(fxy(:,1),...
    fxy(:,2),fxy(:,3));
xlabel('\it x');
ylabel('\it y');
```



Continuous Random Variables

Finally, we turn to the subject of generating sample pairs of continuous random variables. In this case, there are no general techniques such as the sample space grids for discrete random variables. In general, a joint PDF $f_{X,Y}(x, y)$ or CDF $F_{X,Y}(x, y)$ can be viewed using the function `plot3`. For example, Figure 4.4 was generated this way. In addition, one can calculate $E[g(X, Y)]$ in Theorem 4.12 using MATLAB's numerical integration methods; however, such methods tend to be slow and not particularly instructive.

There exist a wide variety of techniques for generating sample values of pairs of continuous random variables of specific types. This is particularly true for bivariate Gaussian random variables. In the general case of an arbitrary joint PDF $f_{X,Y}(x, y)$, a basic approach is to generate sample values $x_1, \ldots, x_m$ for X using the marginal PDF $f_X(x)$. Then for each sample x_i, we generate y_i using the conditional PDF $f_{X|Y}(x|y_i)$. MATLAB can do this efficiently provided the samples $y_1, \ldots, y_m$ can be generated from $x_1, \ldots, x_m$ using vector processing techniques, as in the following example.

Example 4.31 Write a function `xy= xytrianglerv(m)` that generates m sample pairs of X and Y in Example 4.19.

In Example 4.19, we found that

$$f_X(x) = \begin{cases} 2x & 0 \le x \le 1, \\ 0 & \text{otherwise,} \end{cases} \qquad f_{Y|X}(y|x) = \begin{cases} 1/x & 0 \le y \le x, \\ 0 & \text{otherwise.} \end{cases} \tag{4.160}$$

For $0 \le x \le 1$, we have that $F_X(x) = x^2$. Using Theorem 3.22 to generate sample values of X, we define $u = F_X(x) = x^2$. Then, for $0 < u < 1$, $x = \sqrt{u}$. By Theorem 3.22, if U is uniform $(0, 1)$, then $\sqrt{U}$ has PDF $f_X(x)$. Next, we observe that given $X = x_i$, Y is a uniform $(0, x_i)$ random variable. Given another uniform $(0, 1)$ random variable U', Theorem 3.20(a) states that $Y = x_i U'$ is a uniform $(0, x_i)$ random variable.

```
function xy = xytrianglerv(m);
x=sqrt(rand(m,1));
y=x.*rand(m,1);
xy=[x y];
```

We implement these ideas in the function `xytrianglerv.m`.

Quiz 4.12 *For random variables X and Y with joint PMF $P_{X,Y}(x, y)$ given in Example 4.13, write a* MATLAB *function* `xy=dtrianglerv(m)` *that generates m sample pairs.*

Chapter Summary

This chapter introduces experiments that produce two or more random variables.

- *The joint CDF* $F_{X,Y}(x, y) = P[X \le x, Y \le y]$ is a complete probability model of the random variables X and Y. However, it is much easier to use the joint PMF $P_{X,Y}(x, y)$ for discrete random variables and the joint PDF $f_{X,Y}(x, y)$ for continuous random variables.
- *The marginal PMFs* $P_X(x)$ and $P_Y(y)$ for discrete random variables and *the marginal PDFs* $f_X(x)$ and $f_Y(y)$ for continuous random variables are probability models for the individual random variables X and Y.
- *Expected values* $E[g(X, Y)]$ of functions $g(X, Y)$ summarize properties of the entire probability model of an experiment. $\text{Cov}[X, Y]$ and $r_{X,Y}$ convey valuable insights into the relationship of X and Y.
- *Conditional probability models* occur when we obtain partial information about the random variables X and Y. We derive new probability models, including the conditional joint PMF $P_{X,Y|A}(x, y)$ and the conditional PMFs $P_{X|Y}(x|y)$ and $P_{Y|X}(y|x)$ for discrete

random variables, as well as the conditional joint PDF $f_{X,Y|A}(x, y)$ and the conditional PDFs $f_{X|Y}(x|y)$ and $f_{Y|X}(y|x)$ for continuous random variables.

- *Random variables X and Y are independent* if the events $\{X = x\}$ and $\{Y = y\}$ are independent for all x, y in $S_{X,Y}$. If X and Y are discrete, they are independent if and only if $P_{X,Y}(x, y) = P_X(x)P_Y(y)$ for all x and y. If X and Y are continuous, they are independent if and only if $f_{X,Y}(x, y) = f_X(x) f_Y(y)$ for all x and y.

Problems

Difficulty: ● Easy ■ Moderate ♦ Difficult ♦♦ Experts Only

4.1.1 ● Random variables X and Y have the joint CDF

$$F_{X,Y}(x, y) = \begin{cases} (1 - e^{-x})(1 - e^{-y}) & x \geq 0; \\ & y \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

(a) What is $P[X \leq 2, Y \leq 3]$?

(b) What is the marginal CDF, $F_X(x)$?

(c) What is the marginal CDF, $F_Y(y)$?

4.1.2 ● Express the following extreme values of $F_{X,Y}(x, y)$ in terms of the marginal cumulative distribution functions $F_X(x)$ and $F_Y(y)$.

(a) $F_{X,Y}(x, -\infty)$

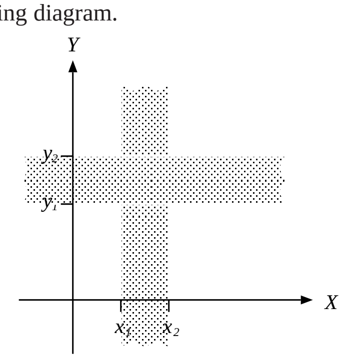
(b) $F_{X,Y}(x, \infty)$

(c) $F_{X,Y}(-\infty, \infty)$

(d) $F_{X,Y}(-\infty, y)$

(e) $F_{X,Y}(\infty, y)$

4.1.3 ■ For continuous random variables X, Y with joint CDF $F_{X,Y}(x, y)$ and marginal CDFs $F_X(x)$ and $F_Y(y)$, find $P[x_1 \leq X < x_2 \cup y_1 \leq Y < y_2]$. This is the probability of the shaded "cross" region in the following diagram.



4.1.4 ■ Random variables X and Y have CDF $F_X(x)$ and $F_Y(y)$. Is $F(x, y) = F_X(x)F_Y(y)$ a valid CDF? Explain your answer.

4.1.5 ♦ In this problem, we prove Theorem 4.5.

(a) Sketch the following events on the X, Y plane:

$$A = \{X \leq x_1, y_1 < Y \leq y_2\},$$
$$B = \{x_1 < X \leq x_2, Y \leq y_1\},$$
$$C = \{x_1 < X \leq x_2, y_1 < Y \leq y_2\}.$$

(b) Express the probability of the events A, B, and $A \cup B \cup C$ in terms of the joint CDF $F_{X,Y}(x, y)$.

(c) Use the observation that events A, B, and C are mutually exclusive to prove Theorem 4.5.

4.1.6 ♦ Can the following function be the joint CDF of random variables X and Y?

$$F(x, y) = \begin{cases} 1 - e^{-(x+y)} & x \geq 0, y \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

4.2.1 ● Random variables X and Y have the joint PMF

$$P_{X,Y}(x, y) = \begin{cases} cxy & x = 1, 2, 4; \quad y = 1, 3, \\ 0 & \text{otherwise.} \end{cases}$$

(a) What is the value of the constant c?

(b) What is $P[Y < X]$?

(c) What is $P[Y > X]$?

(d) What is $P[Y = X]$?

(e) What is $P[Y = 3]$?

4.2.2 ● Random variables X and Y have the joint PMF

$$P_{X,Y}(x, y) = \begin{cases} c|x + y| & x = -2, 0, 2; \\ & y = -1, 0, 1, \\ 0 & \text{otherwise.} \end{cases}$$

(a) What is the value of the constant c?

(b) What is $P[Y < X]$?

(c) What is $P[Y > X]$?

(d) What is $P[Y = X]$?

(e) What is $P[X < 1]$?

4.2.3 Test two integrated circuits. In each test, the probability of rejecting the circuit is p. Let X be the number of rejects (either 0 or 1) in the first test and let Y be the number of rejects in the second test. Find the joint PMF $P_{X,Y}(x, y)$.

4.2.4 For two flips of a fair coin, let X equal the total number of tails and let Y equal the number of heads on the last flip. Find the joint PMF $P_{X,Y}(x, y)$.

4.2.5 In Figure 4.2, the axes of the figures are labeled X and Y because the figures depict possible values of the random variables X and Y. However, the figure at the end of Example 4.1 depicts $P_{X,Y}(x, y)$ on axes labeled with lowercase x and y. Should those axes be labeled with the uppercase X and Y? Hint: Reasonable arguments can be made for both views.

4.2.6 As a generalization of Example 4.1, consider a test of n circuits such that each circuit is acceptable with probability p, independent of the outcome of any other test. Show that the joint PMF of X, the number of acceptable circuits, and Y, the number of acceptable circuits found before observing the first reject, is

$$P_{X,Y}(x, y) = \begin{cases} \binom{n-y-1}{x-y} p^x (1-p)^{n-x} & 0 \le y \le x < n, \\ p^n & x = y = n, \\ 0 & \text{otherwise.} \end{cases}$$

Hint: For $0 \le y \le x < n$, show that

$$\{X = x, Y = y\} = A \cap B \cap C,$$

where

A: The first y tests are acceptable.

B: Test $y + 1$ is a rejection.

C: The remaining $n - y - 1$ tests yield $x - y$ acceptable circuits

4.2.7 Each test of an integrated circuit produces an acceptable circuit with probability p, independent of the outcome of the test of any other circuit. In testing n circuits, let K denote the number of circuits rejected and let X denote the number of acceptable circuits (either 0 or 1) in the last test. Find the joint PMF $P_{K,X}(k, x)$.

4.2.8 Each test of an integrated circuit produces an acceptable circuit with probability p, independent of the outcome of the test of any other circuit. In testing n circuits, let K denote the number of circuits rejected and let X denote the number of acceptable circuits that appear before the first reject is found. Find the joint PMF $P_{K,X}(k, x)$.

4.3.1 Given the random variables X and Y in Problem 4.2.1, find

(a) The marginal PMFs $P_X(x)$ and $P_Y(y)$,

(b) The expected values $E[X]$ and $E[Y]$,

(c) The standard deviations σ_X and σ_Y.

4.3.2 Given the random variables X and Y in Problem 4.2.2, find

(a) The marginal PMFs $P_X(x)$ and $P_Y(y)$,

(b) The expected values $E[X]$ and $E[Y]$,

(c) The standard deviations σ_X and σ_Y.

4.3.3 For $n = 0, 1, \ldots$ and $0 \le k \le 100$, the joint PMF of random variables N and K is

$$P_{N,K}(n, k) = \frac{100^n e^{-100}}{n!} \binom{100}{k} p^k (1-p)^{100-k}.$$

Otherwise, $P_{N,K}(n, k) = 0$. Find the marginal PMFs $P_N(n)$ and $P_K(k)$.

4.3.4 Random variables N and K have the joint PMF

$$P_{N,K}(n, k) = \begin{cases} (1-p)^{n-1} p/n & k = 1, \ldots, n; \\ & n = 1, 2, \ldots, \\ 0 & \text{otherwise.} \end{cases}$$

Find the marginal PMFs $P_N(n)$ and $P_K(k)$.

4.3.5 Random variables N and K have the joint PMF

$$P_{N,K}(n, k) = \begin{cases} \frac{100^n e^{-100}}{(n+1)!} & k = 0, 1, \ldots, n; \\ & n = 0, 1, \ldots, \\ 0 & \text{otherwise.} \end{cases}$$

Find the marginal PMF $P_N(n)$. Show that the marginal PMF $P_K(k)$ satisfies $P_K(k) = P[N > k]/100$.

4.4.1 Random variables X and Y have the joint PDF

$$f_{X,Y}(x, y) = \begin{cases} c & x + y \le 1, x \ge 0, y \ge 0, \\ 0 & \text{otherwise.} \end{cases}$$

(a) What is the value of the constant c?

(b) What is $P[X \le Y]$?

(c) What is $P[X + Y \le 1/2]$?

4.4.2 Random variables X and Y have joint PDF

$$f_{X,Y}(x, y) = \begin{cases} cxy^2 & 0 \le x \le 1, 0 \le y \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

(a) Find the constant c.

(b) Find $P[X > Y]$ and $P[Y < X^2]$.

(c) Find $P[\min(X, Y) \le 1/2]$.

(d) Find $P[\max(X, Y) \le 3/4]$.

4.4.3 Random variables X and Y have joint PDF

$$f_{X,Y}(x, y) = \begin{cases} 6e^{-(2x+3y)} & x \ge 0, y \ge 0, \\ 0 & \text{otherwise.} \end{cases}$$

(a) Find $P[X > Y]$ and $P[X + Y \le 1]$.

(b) Find $P[\min(X, Y) \ge 1]$.

(c) Find $P[\max(X, Y) \le 1]$.

4.4.4 Random variables X and Y have joint PDF

$$f_{X,Y}(x, y) = \begin{cases} 8xy & 0 \le y \le x \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

Following the method of Example 4.5, find the joint CDF $F_{X,Y}(x, y)$.

4.5.1 Random variables X and Y have the joint PDF

$$f_{X,Y}(x, y) = \begin{cases} 1/2 & -1 \le x \le y \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

(a) Sketch the region of nonzero probability.

(b) What is $P[X > 0]$?

(c) What is $f_X(x)$?

(d) What is $E[X]$?

4.5.2 X and Y are random variables with the joint PDF

$$f_{X,Y}(x, y) = \begin{cases} 2 & x + y \le 1, x \ge 0, y \ge 0, \\ 0 & \text{otherwise.} \end{cases}$$

(a) What is the marginal PDF $f_X(x)$?

(b) What is the marginal PDF $f_Y(y)$?

4.5.3 Over the circle $X^2 + Y^2 \le r^2$, random variables X and Y have the uniform PDF

$$f_{X,Y}(x, y) = \begin{cases} 1/(\pi r^2) & x^2 + y^2 \le r^2, \\ 0 & \text{otherwise.} \end{cases}$$

(a) What is the marginal PDF $f_X(x)$?

(b) What is the marginal PDF $f_Y(y)$?

4.5.4 X and Y are random variables with the joint PDF

$$f_{X,Y}(x, y) = \begin{cases} 5x^2/2 & -1 \le x \le 1; \\ & 0 \le y \le x^2, \\ 0 & \text{otherwise.} \end{cases}$$

(a) What is the marginal PDF $f_X(x)$?

(b) What is the marginal PDF $f_Y(y)$?

4.5.5 Over the circle $X^2 + Y^2 \le r^2$, random variables X and Y have the PDF

$$f_{X,Y}(x, y) = \begin{cases} 2\,|xy|\,/r^4 & x^2 + y^2 \le r^2, \\ 0 & \text{otherwise.} \end{cases}$$

(a) What is the marginal PDF $f_X(x)$?

(b) What is the marginal PDF $f_Y(y)$?

4.5.6 Random variables X and Y have the joint PDF

$$f_{X,Y}(x, y) = \begin{cases} cy & 0 \le y \le x \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

(a) Draw the region of nonzero probability.

(b) What is the value of the constant c?

(c) What is $F_X(x)$?

(d) What is $F_Y(y)$?

(e) What is $P[Y \le X/2]$?

4.6.1 Given random variables X and Y in Problem 4.2.1 and the function $W = X - Y$, find

(a) The probability mass function $P_W(w)$,

(b) The expected value $E[W]$,

(c) $P[W > 0]$.

4.6.2 Given random variables X and Y in Problem 4.2.2 and the function $W = X + 2Y$, find

(a) The probability mass function $P_W(w)$,

(b) The expected value $E[W]$,

(c) $P[W > 0]$.

4.6.3 Let X and Y be discrete random variables with joint PMF $P_{X,Y}(x, y)$ that is zero except when x and y are integers. Let $W = X + Y$ and show that the PMF of W satisfies

$$P_W(w) = \sum_{x=-\infty}^{\infty} P_{X,Y}(x, w - x).$$

4.6.4 Let X and Y be discrete random variables with joint PMF

$$P_{X,Y}(x, y) = \begin{cases} 0.01 & x = 1, 2 \ldots, 10, \\ & y = 1, 2 \ldots, 10, \\ 0 & \text{otherwise.} \end{cases}$$

What is the PMF of $W = \min(X, Y)$?

4.6.5 For random variables X and Y given in Problem 4.6.4, what is the PMF of $V = \max(X, Y)$?

4.6.6 Random variables X and Y have joint PDF

$$f_{X,Y}(x, y) = \begin{cases} x + y & 0 \le x \le 1, 0 \le y \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

Let $W = \max(X, Y)$.

(a) What is S_W, the range of W?

(b) Find $F_W(w)$ and $f_W(w)$.

4.6.7 Random variables X and Y have joint PDF

$$f_{X,Y}(x, y) = \begin{cases} 6y & 0 \le y \le x \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

Let $W = Y - X$.

(a) What is S_W, the range of W?

(b) Find $F_W(w)$ and $f_W(w)$.

4.6.8 Random variables X and Y have joint PDF

$$f_{X,Y}(x, y) = \begin{cases} 2 & 0 \le y \le x \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

Let $W = Y/X$.

(a) What is S_W, the range of W?

(b) Find $F_W(w)$, $f_W(w)$, and $E[W]$.

4.6.9 Random variables X and Y have joint PDF

$$f_{X,Y}(x, y) = \begin{cases} 2 & 0 \le y \le x \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

Let $W = X/Y$.

(a) What is S_W, the range of W?

(b) Find $F_W(w)$, $f_W(w)$, and $E[W]$.

4.6.10 In a simple model of a cellular telephone system, a portable telephone is equally likely to be found anywhere in a circular cell of radius 4 km. (See Problem 4.5.3.) Find the CDF $F_R(r)$ and PDF $f_R(r)$ of R, the distance (in km) between the telephone and the base station at the center of the cell.

4.6.11 For a constant $a > 0$, random variables X and Y have joint PDF

$$f_{X,Y}(x, y) = \begin{cases} 1/a^2 & 0 \le x \le a, 0 \le y \le a \\ 0 & \text{otherwise} \end{cases}$$

Find the CDF and PDF of random variable

$$W = \max\left(\frac{X}{Y}, \frac{Y}{X}\right).$$

Hint: Is it possible to observe $W < 1$?

4.7.1 For the random variables X and Y in Problem 4.2.1, find

(a) The expected value of $W = Y/X$,

(b) The correlation, $E[XY]$,

(c) The covariance, $\text{Cov}[X, Y]$,

(d) The correlation coefficient, $\rho_{X,Y}$,

(e) The variance of $X + Y$, $\text{Var}[X + Y]$.

(Refer to the results of Problem 4.3.1 to answer some of these questions.)

4.7.2 For the random variables X and Y in Problem 4.2.2 find

(a) The expected value of $W = 2^{XY}$,

(b) The correlation, $E[XY]$,

(c) The covariance, $\text{Cov}[X, Y]$,

(d) The correlation coefficient, $\rho_{X,Y}$,

(e) The variance of $X + Y$, $\text{Var}[X + Y]$.

(Refer to the results of Problem 4.3.2 to answer some of these questions.)

4.7.3 Let H and B be the random variables in Quiz 4.3. Find $r_{H,B}$ and $\text{Cov}[H, B]$.

4.7.4 For the random variables X and Y in Example 4.13, find

(a) The expected values $E[X]$ and $E[Y]$,

(b) The variances $\text{Var}[X]$ and $\text{Var}[Y]$,

(c) The correlation, $E[XY]$,

(d) The covariance, $\text{Cov}[X, Y]$,

(e) The correlation coefficient, $\rho_{X,Y}$.

4.7.5 Random variables X and Y have joint PMF

$$P_{X,Y}(x, y) = \begin{cases} 1/21 & x = 0, 1, 2, 3, 4, 5; \\ & y = 0, 1, \ldots, x, \\ 0 & \text{otherwise.} \end{cases}$$

Find the marginal PMFs $P_X(x)$ and $P_Y(y)$. Also find the covariance $\text{Cov}[X, Y]$.

4.7.6 For the random variables X and Y in Example 4.13, let $W = \min(X, Y)$ and $V = \max(X, Y)$. Find

(a) The expected values, $E[W]$ and $E[V]$,

(b) The variances, $\mathrm{Var}[W]$ and $\mathrm{Var}[V]$,

(c) The correlation, $E[WV]$,

(d) The covariance, $\mathrm{Cov}[W, V]$,

(e) The correlation coefficient, $\mathrm{Cov}[W, V]$.

4.7.7 For a random variable X, let $Y = aX + b$. Show that if $a > 0$ then $\rho_{X,Y} = 1$. Also show that if $a < 0$, then $\rho_{X,Y} = -1$.

4.7.8 Random variables X and Y have joint PDF

$$f_{X,Y}(x, y) = \begin{cases} (x+y)/3 & 0 \le x \le 1; \\ & 0 \le y \le 2, \\ 0 & \text{otherwise.} \end{cases}$$

(a) What are $E[X]$ and $\mathrm{Var}[X]$?

(b) What are $E[Y]$ and $\mathrm{Var}[Y]$?

(c) What is $\mathrm{Cov}[X, Y]$?

(d) What is $E[X + Y]$?

(e) What is $\mathrm{Var}[X + Y]$?

4.7.9 Random variables X and Y have joint PDF

$$f_{X,Y}(x, y) = \begin{cases} 4xy & 0 \le x \le 1, 0 \le y \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

(a) What are $E[X]$ and $\mathrm{Var}[X]$?

(b) What are $E[Y]$ and $\mathrm{Var}[Y]$?

(c) What is $\mathrm{Cov}[X, Y]$?

(d) What is $E[X + Y]$?

(e) What is $\mathrm{Var}[X + Y]$?

4.7.10 Random variables X and Y have joint PDF

$$f_{X,Y}(x, y) = \begin{cases} 5x^2/2 & -1 \le x \le 1; \\ & 0 \le y \le x^2, \\ 0 & \text{otherwise.} \end{cases}$$

(a) What are $E[X]$ and $\mathrm{Var}[X]$?

(b) What are $E[Y]$ and $\mathrm{Var}[Y]$?

(c) What is $\mathrm{Cov}[X, Y]$?

(d) What is $E[X + Y]$?

(e) What is $\mathrm{Var}[X + Y]$?

4.7.11 Random variables X and Y have joint PDF

$$f_{X,Y}(x, y) = \begin{cases} 2 & 0 \le y \le x \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

(a) What are $E[X]$ and $\mathrm{Var}[X]$?

(b) What are $E[Y]$ and $\mathrm{Var}[Y]$?

(c) What is $\mathrm{Cov}[X, Y]$?

(d) What is $E[X + Y]$?

(e) What is $\mathrm{Var}[X + Y]$?

4.7.12 Random variables X and Y have joint PDF

$$f_{X,Y}(x, y) = \begin{cases} 1/2 & -1 \le x \le y \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

Find $E[XY]$ and $E[e^{X+Y}]$.

4.7.13 Random variables N and K have the joint PMF

$$P_{N,K}(n, k) = \begin{cases} (1-p)^{n-1}p/n & k = 1, \ldots, n; \\ & n = 1, 2, \ldots, \\ 0 & \text{otherwise.} \end{cases}$$

Find the marginal PMF $P_N(n)$ and the expected values $E[N]$, $\mathrm{Var}[N]$, $E[N^2]$, $E[K]$, $\mathrm{Var}[K]$, $E[N + K]$, $E[NK]$, $\mathrm{Cov}[N, K]$.

4.8.1 Let random variables X and Y have the joint PMF $P_{X,Y}(x, y)$ given in Problem 4.6.4. Let A denote the event that $\min(X, Y) > 5$. Find the conditional PMF $P_{X,Y|A}(x, y)$.

4.8.2 Let random variables X and Y have the joint PMF $P_{X,Y}(x, y)$ given in Problem 4.6.4. Let B denote the event that $\max(X, Y) \le 5$. Find the conditional PMF $P_{X,Y|B}(x, y)$.

4.8.3 Random variables X and Y have joint PDF

$$f_{X,Y}(x, y) = \begin{cases} 6e^{-(2x+3y)} & x \ge 0, y \ge 0, \\ 0 & \text{otherwise.} \end{cases}$$

Let A be the event that $X + Y \le 1$. Find the conditional PDF $f_{X,Y|A}(x, y)$.

4.8.4 For $n = 1, 2, \ldots$ and $k = 1, \ldots, n$, the joint PMF of N and K satisfies

$$P_{N,K}(n, k) = (1-p)^{n-1}p/n.$$

Otherwise, $P_{N,K}(n, k) = 0$. Let B denote the event that $N \ge 10$. Find the conditional PMFs $P_{N,K|B}(n, k)$ and $P_{N|B}(n)$. In addition, find the conditional expected values $E[N|B]$, $E[K|B]$, $E[N + K|B]$, $\mathrm{Var}[N|B]$, $\mathrm{Var}[K|B]$, $E[NK|B]$.

4.8.5 Random variables X and Y have joint PDF

$$f_{X,Y}(x,y) = \begin{cases} (x+y)/3 & 0 \le x \le 1; \\ & 0 \le y \le 2, \\ 0 & \text{otherwise.} \end{cases}$$

Let $A = \{Y \le 1\}$.

(a) What is $P[A]$?

(b) Find $f_{X,Y|A}(x,y)$, $f_{X|A}(x)$, and $f_{Y|A}(y)$.

4.8.6 Random variables X and Y have joint PDF

$$f_{X,Y}(x,y) = \begin{cases} (4x+2y)/3 & 0 \le x \le 1; \\ & 0 \le y \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

Let $A = \{Y \le 1/2\}$.

(a) What is $P[A]$?

(b) Find $f_{X,Y|A}(x,y)$, $f_{X|A}(x)$, and $f_{Y|A}(y)$.

4.8.7 Random variables X and Y have joint PDF

$$f_{X,Y}(x,y) = \begin{cases} 5x^2/2 & -1 \le x \le 1; \\ & 0 \le y \le x^2, \\ 0 & \text{otherwise.} \end{cases}$$

Let $A = \{Y \le 1/4\}$.

(a) What is the conditional PDF $f_{X,Y|A}(x,y)$?

(b) What is $f_{Y|A}(y)$?

(c) What is $E[Y|A]$?

(d) What is $f_{X|A}(x)$?

(e) What is $E[X|A]$?

4.9.1 A business trip is equally likely to take 2, 3, or 4 days. After a d-day trip, the change in the traveler's weight, measured as an integer number of pounds, is uniformly distributed between $-d$ and d pounds. For one such trip, denote the number of days by D and the change in weight by W. Find the joint PMF $P_{D,W}(d,w)$.

4.9.2 Flip a coin twice. On each flip, the probability of heads equals p. Let X_i equal the number of heads (either 0 or 1) on flip i. Let $W = X_1 - X_2$ and $Y = X_1 + X_2$. Find $P_{W,Y}(w,y)$, $P_{W|Y}(w|y)$, and $P_{Y|W}(y|w)$.

4.9.3 X and Y have joint PDF

$$f_{X,Y}(x,y) = \begin{cases} (4x+2y)/3 & 0 \le x \le 1; \\ & 0 \le y \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

(a) For which values of y is $f_{X|Y}(x|y)$ defined? What is $f_{X|Y}(x|y)$?

(b) For which values of x is $f_{Y|X}(y|x)$ defined? What is $f_{Y|X}(y|x)$?

4.9.4 Random variables X and Y have joint PDF

$$f_{X,Y}(x,y) = \begin{cases} 2 & 0 \le y \le x \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

Find the PDF $f_Y(y)$, the conditional PDF $f_{X|Y}(x|y)$, and the conditional expected value $E[X|Y=y]$.

4.9.5 Let random variables X and Y have joint PDF $f_{X,Y}(x,y)$ given in Problem 4.9.4. Find the PDF $f_X(x)$, the conditional PDF $f_{Y|X}(y|x)$, and the conditional expected value $E[Y|X=x]$.

4.9.6 A student's final exam grade depends on how close the student sits to the center of the classroom during lectures. If a student sits r feet from the center of the room, the grade is a Gaussian random variable with expected value $80 - r$ and standard deviation r. If r is a sample value of random variable R, and X is the exam grade, what is $f_{X|R}(x|r)$?

4.9.7 The probability model for random variable A is

$$P_A(a) = \begin{cases} 1/3 & a = -1, \\ 2/3 & a = 1, \\ 0 & \text{otherwise.} \end{cases}$$

The conditional probability model for random variable B given A is:

$$P_{B|A}(b|-1) = \begin{cases} 1/3 & b = 0, \\ 2/3 & b = 1, \\ 0 & \text{otherwise,} \end{cases}$$

$$P_{B|A}(b|1) = \begin{cases} 1/2 & b = 0, \\ 1/2 & b = 1, \\ 0 & \text{otherwise.} \end{cases}$$

(a) What is the probability model for random variables A and B? Write the joint PMF $P_{A,B}(a,b)$ as a table.

(b) If $A = 1$, what is the conditional expected value $E[B|A=1]$?

(c) If $B = 1$, what is the conditional PMF $P_{A|B}(a|1)$?

(d) If $B = 1$, what is the conditional variance $\text{Var}[A|B=1]$ of A?

(e) What is the covariance $\text{Cov}[A,B]$ of A and B?

4.9.8 For random variables A and B given in Problem 4.9.7, let $U = E[B|A]$. Find the PMF $P_U(u)$. What is $E[U] = E[E[B|A]]$?

4.9.9 Random variables N and K have the joint PMF

$$P_{N,K}(n,k) = \begin{cases} \frac{100^n e^{-100}}{(n+1)!} & k = 0, 1, \ldots, n; \\ & n = 0, 1, \ldots, \\ 0 & \text{otherwise.} \end{cases}$$

Find the marginal PMF $P_N(n)$, the conditional PMF $P_{K|N}(k|n)$, and the conditional expected value $E[K|N = n]$. Express the random variable $E[K|N]$ as a function of N and use the iterated expectation to find $E[K]$.

4.9.10 At the One Top Pizza Shop, mushrooms are the only topping. Curiously, a pizza sold before noon has mushrooms with probability $p = 1/3$ while a pizza sold after noon never has mushrooms. Also, an arbitrary pizza is equally likely to be sold before noon as after noon. On a day in which 100 pizzas are sold, let N equal the number of pizzas sold before noon and let M equal the number of mushroom pizzas sold during the day. What is the joint PMF $P_{M,N}(m,n)$? Hint: Find the conditional PMF of M given N.

4.9.11 Random variables X and Y have joint PDF

$$f_{X,Y}(x,y) = \begin{cases} 1/2 & -1 \le x \le y \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

(a) What is $f_Y(y)$?

(b) What is $f_{X|Y}(x|y)$?

(c) What is $E[X|Y = y]$?

4.9.12 Over the circle $X^2 + Y^2 \le r^2$, random variables X and Y have the uniform PDF

$$f_{X,Y}(x,y) = \begin{cases} 1/(\pi r^2) & x^2 + y^2 \le r^2, \\ 0 & \text{otherwise.} \end{cases}$$

(a) What is $f_{Y|X}(y|x)$?

(b) What is $E[Y|X = x]$?

4.9.13 Calls arriving at a telephone switch are either voice calls (v) or data calls (d). Each call is a voice call with probability p, independent of any other call. Observe calls at a telephone switch until you see two voice calls. Let M equal the number of calls up to and including the first voice call. Let N equal the number of calls observed up to and including the second voice call. Find the conditional PMFs $P_{M|N}(m|n)$ and $P_{N|M}(n|m)$. Interpret your results.

4.9.14 Suppose you arrive at a bus stop at time 0 and at the end of each minute, with probability p, a bus arrives, or with probability $1 - p$, no bus arrives. Whenever a bus arrives, you board that bus with probability q and depart. Let T equal the number of minutes you stand at a bus stop. Let N be the number of buses that arrive while you wait at the bus stop.

(a) Identify the set of points (n, t) for which $P[N = n, T = t] > 0$.

(b) Find $P_{N,T}(n,t)$.

(c) Find the marginal PMFs $P_N(n)$ and $P_T(t)$.

(d) Find the conditional PMFs $P_{N|T}(n|t)$ and $P_{T|N}(t|n)$.

4.9.15 Each millisecond at a telephone switch, a call independently arrives with probability p. Each call is either a data call (d) with probability q or a voice call (v). Each data call is a fax call with probability r. Let N equal the number of milliseconds required to observe the first 100 fax calls. Let T equal the number of milliseconds you observe the switch waiting for the first fax call. Find the marginal PMF $P_T(t)$ and the conditional PMF $P_{N|T}(n|t)$. Lastly, find the conditional PMF $P_{T|N}(t|n)$.

4.10.1 Flip a fair coin 100 times. Let X equal the number of heads in the first 75 flips. Let Y equal the number of heads in the remaining 25 flips. Find $P_X(x)$ and $P_Y(y)$. Are X and Y independent? Find $P_{X,Y}(x,y)$.

4.10.2 X and Y are independent, identically distributed random variables with PMF

$$P_X(k) = P_Y(k) = \begin{cases} 3/4 & k = 0, \\ 1/4 & k = 20, \\ 0 & \text{otherwise.} \end{cases}$$

Find the following quantities:

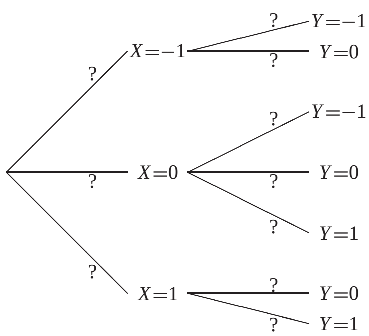
$$E[X], \qquad \mathrm{Var}[X],$$
$$E[X+Y], \qquad \mathrm{Var}[X+Y], \qquad E\left[XY2^{XY}\right].$$

4.10.3 Random variables X and Y have a joint PMF described by the following table.

$P_{X,Y}(x,y)$	$y=-1$	$y=0$	$y=1$
$x=-1$	$3/16$	$1/16$	0
$x=0$	$1/6$	$1/6$	$1/6$
$x=1$	0	$1/8$	$1/8$

(a) Are X and Y independent?

(b) In fact, the experiment from which X and Y are derived is performed sequentially. First, X is found, then Y is found. In this context, label the conditional branch probabilities of the following tree:



4.10.4 For the One Top Pizza Shop of Problem 4.9.10, are M and N independent?

4.10.5 Flip a fair coin until heads occurs twice. Let X_1 equal the number of flips up to and including the first H. Let X_2 equal the number of additional flips up to and including the second H. What are $P_{X_1}(x_1)$ and $P_{X_2}(x_2)$. Are X_1 and X_2 independent? Find $P_{X_1,X_2}(x_1,x_2)$.

4.10.6 Flip a fair coin until heads occurs twice. Let X_1 equal the number of flips up to and including the first H. Let X_2 equal the number of additional flips up to and including the second H. Let $Y = X_1 - X_2$. Find $E[Y]$ and $\text{Var}[Y]$. Hint: Don't try to find $P_Y(y)$.

4.10.7 X and Y are independent random variables with PDFs

$$f_X(x) = \begin{cases} \frac{1}{3}e^{-x/3} & x \geq 0, \\ 0 & \text{otherwise}, \end{cases}$$

$$f_Y(y) = \begin{cases} \frac{1}{2}e^{-y/2} & y \geq 0, \\ 0 & \text{otherwise}. \end{cases}$$

(a) What is $P[X > Y]$?

(b) What is $E[XY]$?

(c) What is $\text{Cov}[X, Y]$?

4.10.8 X_1 and X_2 are independent identically distributed random variables with expected value $E[X]$ and variance $\text{Var}[X]$.

(a) What is $E[X_1 - X_2]$?

(b) What is $\text{Var}[X_1 - X_2]$?

4.10.9 Let X and Y be independent discrete random variables such that $P_X(k) = P_Y(k) = 0$ for all non-integer k. Show that the PMF of $W = X + Y$ satisfies

$$P_W(w) = \sum_{k=-\infty}^{\infty} P_X(k) P_Y(w-k).$$

4.10.10 An ice cream company orders supplies by fax. Depending on the size of the order, a fax can be either

1 page for a short order,
2 pages for a long order.

The company has three different suppliers:

The vanilla supplier is 20 miles away.
The chocolate supplier is 100 miles away.
The strawberry supplier is 300 miles away.

An experiment consists of monitoring an order and observing N, the number of pages, and D, the distance the order is transmitted.

The following probability model describes the experiment:

	van.	choc.	straw.
short	0.2	0.2	0.2
long	0.1	0.2	0.1

(a) What is the joint PMF $P_{N,D}(n,d)$ of the number of pages and the distance?

(b) What is $E[D]$, the expected distance of an order?

(c) Find $P_{D|N}(d|2)$, the conditional PMF of the distance when the order requires 2 pages.

(d) Write $E[D|N=2]$, the expected distance given that the order requires 2 pages.

(e) Are the random variables D and N independent?

(f) The price per page of sending a fax is one cent per mile transmitted. C cents is the price of one fax. What is $E[C]$, the expected price of one fax?

4.10.11 A company receives shipments from two factories. Depending on the size of the order, a shipment can be in

1 box for a small order,
2 boxes for a medium order,
3 boxes for a large order.

The company has two different suppliers:

Factory Q is 60 miles from the company.
Factory R is 180 miles from the company.

An experiment consists of monitoring a shipment and observing B, the number of boxes, and M, the number of miles the shipment travels. The following probability model describes the experiment:

	Factory Q	Factory R
small order	0.3	0.2
medium order	0.1	0.2
large order	0.1	0.1

(a) Find $P_{B,M}(b, m)$, the joint PMF of the number of boxes and the distance. (You may present your answer as a matrix if you like.)

(b) What is $E[B]$, the expected number of boxes?

(c) What is $P_{M|B}(m|2)$, the conditional PMF of the distance when the order requires two boxes?

(d) Find $E[M|B = 2]$, the expected distance given that the order requires 2 boxes.

(e) Are the random variables B and M independent?

(f) The price per mile of sending each box is one cent per mile the box travels. C cents is the price of one shipment. What is $E[C]$, the expected price of one shipment?

4.10.12 X_1 and X_2 are independent, identically distributed random variables with PDF

$$f_X(x) = \begin{cases} x/2 & 0 \le x \le 2, \\ 0 & \text{otherwise.} \end{cases}$$

(a) Find the CDF, $F_X(x)$.

(b) What is $P[X_1 \le 1, X_2 \le 1]$, the probability that X_1 and X_2 are both less than or equal to 1?

(c) Let $W = \max(X_1, X_2)$. What is $F_W(1)$, the CDF of W evaluated at $w = 1$?

(d) Find the CDF $F_W(w)$.

4.10.13 X and Y are independent random variables with PDFs

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

$$f_Y(y) = \begin{cases} 3y^2 & 0 \le y \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

Let $A = \{X > Y\}$.

(a) What are $E[X]$ and $E[Y]$?

(b) What are $E[X|A]$ and $E[Y|A]$?

4.10.14 Prove that random variables X and Y are independent if and only if

$$F_{X,Y}(x, y) = F_X(x) F_Y(y).$$

4.10.15 Following Example 4.21, let X and Y denote the arrival times of the first two calls at a telephone switch. The joint PDF of X and Y is

$$f_{X,Y}(x, y) = \begin{cases} \lambda^2 e^{-\lambda y} & 0 \le x < y, \\ 0 & \text{otherwise.} \end{cases}$$

What is the PDF of $W = Y - X$?

4.10.16 Consider random variables X, Y, and W from Problem 4.10.15.

(a) Are W and X independent?

(b) Are W and Y independent?

4.10.17 X and Y are independent random variables with CDFs $F_X(x)$ and $F_Y(y)$. Let $U = \min(X, Y)$ and $V = \max(X, Y)$.

(a) What is $F_{U,V}(u, v)$?

(b) What is $f_{U,V}(u, v)$?

Hint: To find the joint CDF, let $A = \{U \le u\}$ and $B = \{V \le v\}$ and note that $P[AB] = P[B] - P[A^c B]$.

4.11.1 Random variables X and Y have joint PDF

$$f_{X,Y}(x, y) = ce^{-(x^2/8)-(y^2/18)}.$$

What is the constant c? Are X and Y independent?

4.11.2 Random variables X and Y have joint PDF

$$f_{X,Y}(x, y) = ce^{-(2x^2-4xy+4y^2)}.$$

(a) What are $E[X]$ and $E[Y]$?

(b) Find ρ, the correlation coefficient of X and Y.

(c) What are Var$[X]$ and Var$[Y]$?

(d) What is the constant c?

(e) Are X and Y independent?

4.11.3 X and Y are jointly Gaussian random variables with $E[X] = E[Y] = 0$ and Var$[X] =$ Var$[Y] = 1$. Furthermore, $E[Y|X] = X/2$. What is the joint PDF of X and Y?

4.11.4 An archer shoots an arrow at a circular target of radius 50cm. The arrow pierces the target at a random position (X, Y), measured in centimeters from the center of the disk at position $(X, Y) = (0, 0)$. The "bullseye" is a solid black circle of radius 2cm, at the center of the target. Calculate the probability $P[B]$ of the event that the archer hits the bullseye under each of the following models:

(a) X and Y are iid continuous uniform $(-50, 50)$ random variables.

(b) The PDF $f_{X,Y}(x, y)$ is uniform over the 50cm circular target.

(c) X and Y are iid Gaussian $(\mu = 0, \sigma = 10)$ random variables.

4.11.5 A person's white blood cell (WBC) count W (measured in thousands of cells per microliter of blood) and body temperature T (in degrees Celsius) can be modeled as bivariate Gaussian random variables such that W is Gaussian $(7, 2)$ and T is Gaussian $(37, 1)$. To determine whether a person is sick, first the person's temperature T is measured. If $T > 38$, then the person's WBC count is measured. If $W > 10$, the person is declared ill (event I).

(a) Suppose W and T are uncorrelated. What is $P[I]$? Hint: Draw a tree diagram for the experiment.

(b) Now suppose W and T have correlation coefficient $\rho = 1/\sqrt{2}$. Find the conditional probability $P[I|T = t]$ that a person is declared ill given that the person's temperature is $T = t$.

4.11.6 Under what conditions on the constants a, b, c, and d is

$$f(x, y) = de^{-(a^2x^2+bxy+c^2y^2)}$$

a joint Gaussian PDF?

4.11.7 Show that the joint Gaussian PDF $f_{X,Y}(x, y)$ given by Definition 4.17 satisfies

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

Hint: Use Equation (4.146) and the result of Problem 3.5.9.

4.11.8 Let X_1 and X_2 have a bivariate Gaussian PDF with correlation coefficient ρ_{12} such that each X_i is a Gaussian (μ_i, σ_i) random variable. Show that $Y = X_1X_2$ has variance

$$\text{Var}[Y] = \sigma_1^2\sigma_2^2(1+\rho_{12}^2)+\sigma_1^2\mu_2^2+\mu_1^2\sigma_2^2-\mu_1^2\mu_2^2$$

Hints: Use the iterated expectation to calculate

$$E\left[X_1^2X_2^2\right] = E\left[E\left[X_1^2X_2^2|X_2\right]\right].$$

You may also need to look ahead to Problem 6.3.4.

4.12.1 For random variables X and Y in Example 4.27, use MATLAB to generate a list of the form

$$\begin{array}{ccc} x_1 & y_1 & P_{X,Y}(x_1, y_1) \\ x_2 & y_2 & P_{X,Y}(x_2, y_2) \\ \vdots & \vdots & \vdots \end{array}$$

that includes all possible pairs (x, y).

4.12.2 For random variables X and Y in Example 4.27, use MATLAB to calculate $E[X]$, $E[Y]$, the correlation $E[XY]$, and the covariance $\text{Cov}[X, Y]$.

4.12.3 Write a script `trianglecdfplot.m` that generates the graph of $F_{X,Y}(x, y)$ of Figure 4.4.

4.12.4 Problem 4.2.6 extended Example 4.1 to a test of n circuits and identified the joint PDF of X, the number of acceptable circuits, and Y, the number of successful tests before the first reject. Write a MATLAB function

```
[SX,SY,PXY]=circuits(n,p)
```

that generates the sample space grid for the n circuit test. Check your answer against Equation (4.6) for the $p = 0.9$ and $n = 2$ case. For $p = 0.9$ and $n = 50$, calculate the correlation coefficient $\rho_{X,Y}$.

4.12.5 For random variable W of Example 4.10, we can generate random samples in two different ways:

1. Generate samples of X and Y and calculate $W = Y/X$.
2. Find the CDF $F_W(w)$ and generate samples using Theorem 3.22.

Write MATLAB functions `w=wrv1(m)` and `w=wrv2(m)` to implement these methods. Does one method run much faster? If so, why? (Use `cputime` to make run-time comparisons.)