Remark 1.8.1. The terminology of expectation or expected value has its origin in games of chance. For example, consider a game involving a spinner with the numbers 1, 2, 3 and 4 on it. Suppose the corresponding probabilities of spinning these numbers are 0.20, 0.30, 0.35, and 0.15. To begin a game, a player pays \$5 to the "house" to play. The spinner is then spun and the player "wins" the amount in the second line of the table:

Number spun x	1	2	3	4
"Wins"	\$2	\$3	\$4	\$12
$G =$ Gain	$-\$3$	$-\$2$	$-\$1$	\$7
$p_G(x)$	0.20	0.30	0.35	0.15

"Wins" is in quotes, since the player must pay \$5 to play. Of course, the random variable of interest is the gain to the player; i.e., G with the range as given in the third row of the table. Notice that 20% of the time the player gains $-\$3$; 30% of the time the player gains $-\$2$; 35% of the time the player gains $-\$1$; and 15% of the time the player gains \$7. In mathematics this sentence is expressed as

$$(-3) \times 0.20 + (-2) \times 0.30 + (-1) \times 0.35 + 7 \times 0.15 = -0.50,$$

which, of course, is $E(G)$. That is, the expected gain to the player in this game is $-\$0.50$. So the player expects to lose 50 cents per play. We say a game is a **fair game**, if the expected gain is 0. So this spinner game is not a fair game. ■

Let us consider a function of a random variable X. Call this function $Y = g(X)$. Because Y is a random variable, we could obtain its expectation by first finding the distribution of Y. However, as the following theorem states, we can use the distribution of X to determine the expectation of Y.

Theorem 1.8.1. *Let X be a random variable and let $Y = g(X)$ for some function g.*

(a) *Suppose X is continuous with pdf $f_X(x)$. If $\int_{-\infty}^{\infty} |g(x)| f_X(x)\, dx < \infty$, then the expectation of Y exists and it is given by*

$$E(Y) = \int_{-\infty}^{\infty} g(x) f_X(x)\, dx. \tag{1.8.2}$$

(b) *Suppose X is discrete with pmf $p_X(x)$. Suppose the support of X is denoted by $\mathcal{S}_X$. If $\sum_{x \in \mathcal{S}_X} |g(x)| p_X(x) < \infty$, then the expectation of Y exists and it is given by*

$$E(Y) = \sum_{x \in \mathcal{S}_X} g(x) p_X(x). \tag{1.8.3}$$

Proof: We give the proof in the discrete case. The proof for the continuous case requires some advanced results in analysis; see, also, Exercise 1.8.1.

Because $\sum_{x\in\mathcal{S}_X} |g(x)|p_X(x)$ converges, it follows by a theorem in calculus[6] that any rearrangement of the terms of the series converges to the same limit. Thus we have,

$$\begin{aligned}\sum_{x\in\mathcal{S}_X} |g(x)|p_X(x) &= \sum_{y\in\mathcal{S}_Y}\ \sum_{\{x\in\mathcal{S}_X: g(x)=y\}} |g(x)|p_X(x) &(1.8.4)\\ &= \sum_{y\in\mathcal{S}_Y} |y| \sum_{\{x\in\mathcal{S}_X: g(x)=y\}} p_X(x) &(1.8.5)\\ &= \sum_{y\in\mathcal{S}_Y} |y|p_Y(y), &(1.8.6)\end{aligned}$$

where $\mathcal{S}_Y$ denotes the support of Y. So $E(Y)$ exists; i.e., $\sum_{x\in\mathcal{S}_X} g(x)p_X(x)$ converges. Because $\sum_{x\in\mathcal{S}_X} g(x)p_X(x)$ converges and also converges absolutely, the same theorem from calculus can be used to show that the above equations (1.8.4)–(1.8.6) hold without the absolute values. Hence, $E(Y) = \sum_{x\in\mathcal{S}_X} g(x)p_X(x)$, which is the desired result. ■

The following two examples illustrate this theorem.

Example 1.8.4. Let Y be the discrete random variable discussed in Example 1.6.3 and let $Z = e^{-Y}$. Since $(2e)^{-1} < 1$, we have by Theorem 1.8.1 that

$$\begin{aligned}E[Z] &= E\left[e^{-Y}\right] = \sum_{y=0}^{\infty} e^{-y}\left(\frac{1}{2}\right)^{y+1}\\ &= e\sum_{y=0}^{\infty}\left(\frac{1}{2}e^{-1}\right)^{y+1} = \frac{e}{1-(1/(2e))} = \frac{2e^2}{2e-1}.\ ■\end{aligned}$$

Example 1.8.5. Let X be a continuous random variable with the pdf $f(x) = 2x$ which has support on the interval $(0, 1)$. Suppose $Y = 1/(1+X)$. Then by Theorem 1.8.1, we have

$$E(Y) = \int_0^1 \frac{2x}{1+x}\,dx = \int_1^2 \frac{2u-2}{u}\,du = 2(1-\log 2),$$

where we have used the change in variable $u = 1 + x$ in the second integral. ■

Theorem 1.8.2 shows that the expectation operator E is a linear operator.

Theorem 1.8.2. *Let $g_1(X)$ and $g_2(X)$ be functions of a random variable X. Suppose the expectations of $g_1(X)$ and $g_2(X)$ exist. Then for any constants k_1 and k_2, the expectation of $k_1g_1(X) + k_2g_2(X)$ exists and it is given by*

$$E[k_1g_1(X) + k_2g_2(X)] = k_1E[g_1(X)] + k_2E[g_2(X)]. \tag{1.8.7}$$

[6] For example, see Chapter 2 on infinite series in *Mathematical Comments*, referenced in the Preface.

Proof: For the continuous case, existence follows from the hypothesis, the triangle inequality, and the linearity of the integral; i.e.,

$$\begin{aligned}\int_{-\infty}^{\infty} |k_1 g_1(x) + k_2 g_2(x)| f_X(x)\,dx &\leq |k_1| \int_{-\infty}^{\infty} |g_1(x)| f_X(x)\,dx \\ &\quad + |k_2| \int_{-\infty}^{\infty} |g_2(x)| f_X(x)\,dx < \infty.\end{aligned}$$

The result (1.8.7) follows similarly using the linearity of the integral. The proof for the discrete case follows likewise using the linearity of sums. ■

The following examples illustrate these theorems.

Example 1.8.6. Let X have the pdf

$$f(x) = \begin{cases} 2(1-x) & 0 < x < 1 \\ 0 & \text{elsewhere.} \end{cases}$$

Then

$$\begin{aligned} E(X) = \int_{-\infty}^{\infty} x f(x)\,dx &= \int_0^1 (x)2(1-x)\,dx = \frac{1}{3}, \\ E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)\,dx &= \int_0^1 (x^2)2(1-x)\,dx = \frac{1}{6}, \end{aligned}$$

and, of course,

$$E(6X + 3X^2) = 6\left(\frac{1}{3}\right) + 3\left(\frac{1}{6}\right) = \frac{5}{2}. \quad ■$$

Example 1.8.7. Let X have the pmf

$$p(x) = \begin{cases} \frac{x}{6} & x = 1, 2, 3 \\ 0 & \text{elsewhere.} \end{cases}$$

Then

$$E(6X^3 + X) = 6E(X^3) + E(X) = 6\sum_{x=1}^{3} x^3 p(x) + \sum_{x=1}^{3} x p(x) = \frac{301}{3}. \quad ■$$

Example 1.8.8. Let us divide, at random, a horizontal line segment of length 5 into two parts. If X is the length of the left-hand part, it is reasonable to assume that X has the pdf

$$f(x) = \begin{cases} \frac{1}{5} & 0 < x < 5 \\ 0 & \text{elsewhere.} \end{cases}$$

The expected value of the length of X is $E(X) = \frac{5}{2}$ and the expected value of the length $5 - x$ is $E(5 - x) = \frac{5}{2}$. But the expected value of the product of the two lengths is equal to

$$E[X(5 - X)] = \textstyle\int_0^5 x(5-x)(\frac{1}{5})\,dx = \frac{25}{6} \neq (\frac{5}{2})^2.$$

That is, in general, the expected value of a product is not equal to the product of the expected values. ■

1.8.1 R Computation for an Estimation of the Expected Gain

In the following example, we use an R function to estimate the expected gain in a simple game.

Example 1.8.9. Consider the following game. A player pays p_0 to play. He then rolls a fair 6-sided die with the numbers 1 through 6 on it. If the upface is a 1 or a 2, then the game is over. Otherwise, he flips a fair coin. If the coin toss results in a tail, he receives \$1 and the game is over. If, on the other hand, the coin toss results in a head, he draws 2 cards without replacement from a standard deck of 52 cards. If none of the cards is an ace, he receives \$2, while he receives \$10 or \$50 if gets 1 or 2 aces, respectively. In both cases, the game is over. Let G denote the player's gain. To determine the expected gain, we need the distribution of G. The support of G is the set $\{-p_0, 1-p_0, 2-p_0, 10-p_0, 50-p_0\}$. For the associated probabilities we need the distribution of X, where X is the number of aces in a draw of 2 cards from a standard deck of 52 cards without replacement. This is another example of the hypergeometric distribution discussed in Example 1.6.2. For our situation, the distribution is

$$P(X = x) = \frac{\binom{4}{x}\binom{48}{2-x}}{\binom{52}{2}}, \quad x = 0, 1, 2.$$

Using this formula, the probabilities of X, to 4 places, are 0.8507, 0.1448, and 0.0045 for x equal to 0, 1, and 2, respectively. Using these probabilities and independence, the distribution and expected value of G can be determined; see Exercise 1.8.13. Suppose, however, a person does not have this expertise. Such a person would observe the game a number of times and then use the average of the observed gains as his/her estimate of $E(G)$. We will show in Chapter 2 that this estimate, in a probability sense, is close to $E(G)$, as the number of times the game is played increases. To compute this estimation, we use the following R function, `simplegame`, which plays the game and returns the gain. This function can be downloaded at the site given in the Preface. The argument of the function is the amount the player pays to play. Also, the third line of the function computes the distribution of the above random variable X. To draw from a discrete distribution, the code makes use of the R function `sample` which was discussed previously in Example 1.4.12.

```
simplegame <- function(amtpaid){
       gain <- -amtpaid
       x <- 0:2; pace  <- (choose(4,x)*choose(48,2-x))/choose(52,2)
       x <- sample(1:6,1,prob=rep(1/6,6))
       if(x > 2){
            y <- sample(0:1,1,prob=rep(1/2,2))
            if(y==0){
                gain <- gain + 1
            } else {
                z <- sample(0:2,1,prob=pace)
                if(z==0){gain <- gain + 2}
                if(z==1){gain <- gain + 10}
                if(z==2){gain <- gain + 50}
```

```
            }
        }
        return(gain)
}
```

The following R script obtains the average gain for a sample of 10,000 games. For the example, we set the amount the player pays at \$5.

```
amtpaid <- 5; numtimes <- 10000; gains <- c()
for(i in 1:numtimes){gains <- c(gains,simplegame(amtpaid))}
mean(gains)
```

When we ran this script, we obtained -3.5446 as our estimate of $E(G)$. Exercise 1.8.13 shows that $E(G) = -3.54$. ■

EXERCISES

1.8.1. Our proof of Theorem 1.8.1 was for the discrete case. The proof for the continuous case requires some advanced results in in analysis. If, in addition, though, the function $g(x)$ is one-to-one, show that the result is true for the continuous case. *Hint:* First assume that $y = g(x)$ is strictly increasing. Then use the change-of-variable technique with Jacobian dx/dy on the integral $\int_{x\in\mathcal{S}_X} g(x)f_X(x)\,dx$.

1.8.2. Consider the random variable X in Example 1.8.5. As in the example, let $Y = 1/(1+X)$. In the example we found the $E(Y)$ by using Theorem 1.8.1. Verify this result by finding the pdf of Y and use it to obtain the $E(Y)$.

1.8.3. Let X have the pdf $f(x) = (x+2)/18,\ -2 < x < 4$, zero elsewhere. Find $E(X),\ E[(X+2)^3]$, and $E[6X - 2(X+2)^3]$.

1.8.4. Suppose that $p(x) = \frac{1}{5},\ x = 1, 2, 3, 4, 5$, zero elsewhere, is the pmf of the discrete-type random variable X. Compute $E(X)$ and $E(X^2)$. Use these two results to find $E[(X+2)^2]$ by writing $(X+2)^2 = X^2 + 4X + 4$.

1.8.5. Let X be a number selected at random from a set of numbers $\{51, 52, \ldots, 100\}$. Approximate $E(1/X)$.
Hint: Find reasonable upper and lower bounds by finding integrals bounding $E(1/X)$.

1.8.6. Let the pmf $p(x)$ be positive at $x = -1, 0, 1$ and zero elsewhere.

(a) If $p(0) = \frac{1}{4}$, find $E(X^2)$.

(b) If $p(0) = \frac{1}{4}$ and if $E(X) = \frac{1}{4}$, determine $p(-1)$ and $p(1)$.

1.8.7. Let X have the pdf $f(x) = 3x^2,\ 0 < x < 1$, zero elsewhere. Consider a random rectangle whose sides are X and $(1 - X)$. Determine the expected value of the area of the rectangle.

1.8.8. A bowl contains 10 chips, of which 8 are marked \$2 each and 2 are marked \$5 each. Let a person choose, at random and without replacement, three chips from this bowl. If the person is to receive the sum of the resulting amounts, find his expectation.

1.8.9. Let $f(x) = 2x,\ 0 < x < 1$, zero elsewhere, be the pdf of X.

(a) Compute $E(1/X)$.

(b) Find the cdf and the pdf of $Y = 1/X$.

(c) Compute $E(Y)$ and compare this result with the answer obtained in part (a).

1.8.10. Two distinct integers are chosen at random and without replacement from the first six positive integers. Compute the expected value of the absolute value of the difference of these two numbers.

1.8.11. Let X have a Cauchy distribution which has the pdf

$$f(x) = \frac{1}{\pi}\frac{1}{x^2+1}, \quad -\infty < x < \infty. \tag{1.8.8}$$

Then X is symmetrically distributed about 0 (why?). Why isn't $E(X) = 0$?

1.8.12. Let X have the pdf $f(x) = 3x^2$, $0 < x < 1$, zero elsewhere.

(a) Compute $E(X^3)$.

(b) Show that $Y = X^3$ has a uniform$(0, 1)$ distribution.

(c) Compute $E(Y)$ and compare this result with the answer obtained in part (a).

1.8.13. Using the probabilities discussed in Example 1.8.9 and independence, determine the distribution of the random variable G, the gain to a player of the game when he pays p_0 dollars to play. Show that $E(G) = -\$3.54$ if the player pays \$5 to play.

1.8.14. A bowl contains five chips, which cannot be distinguished by a sense of touch alone. Three of the chips are marked \$1 each and the remaining two are marked \$4 each. A player is blindfolded and draws, at random and without replacement, two chips from the bowl. The player is paid an amount equal to the sum of the values of the two chips that he draws and the game is over. Suppose it costs p_0 dollars to play the game. Let the random variable G be the gain to a player of the game. Determine the distribution of G and the $E(G)$. Determine p_0 so that the game is fair. The R code `sample(c(1,1,1,4,4),2)` computes a sample for this game. Expand this into an R function that simulates the game.

1.9 Some Special Expectations

Certain expectations, if they exist, have special names and symbols to represent them. First, let X be a random variable of the discrete type with pmf $p(x)$. Then

$$E(X) = \sum_x xp(x).$$

If the support of X is $\{a_1, a_2, a_3, \ldots\}$, it follows that

$$E(X) = a_1 p(a_1) + a_2 p(a_2) + a_3 p(a_3) + \cdots .$$

This sum of products is seen to be a "weighted average" of the values of $a_1, a_2, a_3, \ldots,$ the "weight" associated with each a_i being $p(a_i)$. This suggests that we call $E(X)$ the arithmetic mean of the values of X, or, more simply, the **mean value** of X (or the mean value of the distribution).

Definition 1.9.1 (Mean)**.** *Let X be a random variable whose expectation exists. The* **mean value** *μ of X is defined to be $\mu = E(X)$.*

The mean is the first moment (about 0) of a random variable. Another special expectation involves the second moment. Let X be a discrete random variable with support $\{a_1, a_2, \ldots\}$ and with pmf $p(x)$, then

$$\begin{aligned} E[(X-\mu)^2] &= \sum_x (x-\mu)^2 p(x) \\ &= (a_1-\mu)^2 p(a_1) + (a_2-\mu)^2 p(a_2) + \cdots . \end{aligned}$$

This sum of products may be interpreted as a "weighted average" of the squares of the deviations of the numbers $a_1, a_2, \ldots$ from the mean value μ of those numbers where the "weight" associated with each $(a_i - \mu)^2$ is $p(a_i)$. It can also be thought of as the second moment of X about μ. This is an important expectation for all types of random variables, and we usually refer to it as the **variance** of X.

Definition 1.9.2 (Variance)**.** *Let X be a random variable with finite mean μ and such that $E[(X-\mu)^2]$ is finite. Then the* **variance** *of X is defined to be $E[(X-\mu)^2]$. It is usually denoted by σ^2 or by Var(X).*

It is worthwhile to observe that Var(X) equals

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

Because E is a linear operator it then follows that

$$\begin{aligned} \sigma^2 &= E(X^2) - 2\mu E(X) + \mu^2 \\ &= E(X^2) - 2\mu^2 + \mu^2 \\ &= E(X^2) - \mu^2. \end{aligned}$$

This frequently affords an easier way of computing the variance of X.

It is customary to call σ (the positive square root of the variance) the **standard deviation** of X (or the standard deviation of the distribution). The number σ is sometimes interpreted as a measure of the dispersion of the points of the space relative to the mean value μ. If the space contains only one point k for which $p(k) > 0$, then $p(k) = 1$, $\mu = k$, and $\sigma = 0$.

While the variance is not a linear operator, it does satisfy the following result:

Theorem 1.9.1. *Let X be a random ravariable with finite mean μ and variance σ^2. Then for all constants a and b,*

$$Var(aX + b) = a^2\, Var(X). \tag{1.9.1}$$

Proof. Because E is linear, $E(aX + b) = a\mu + b$. Hence, by definition

$$\text{Var}(aX + b) = E\left\{[(aX + b) - (a\mu + b)]^2\right\} = E\left\{a^2[X - \mu]^2\right\} = a^2\text{Var}(X).$$

■

Based on this theorem, for standard deviations, $\sigma_{aX+b} = |a|\sigma_X$. The following example illustrates these points.

Example 1.9.1. Suppose the random variable X has a uniform distribution, (1.7.4), with pdf $f_X(x) = 1/(2a)$, $-a < x < a$, zero elsewhere. Then the mean and variance of X are:

$$\begin{aligned}
\mu &= \int_{-a}^{a} x\frac{1}{2a}\,dx = \frac{1}{2a}\,\frac{x^2}{2}\bigg|_{-a}^{a} = 0,\\
\sigma^2 &= \int_{-a}^{a} x^2 = \frac{1}{2a}\,\frac{x^3}{3}\bigg|_{-a}^{a} = \frac{a^2}{3}.
\end{aligned}$$

so that $\sigma_X = a/\sqrt{3}$ is the standard deviation of the distribution of X. Consider the transformation $Y = 2X$. Because the inverse transformation is $x = y/2$ and $dx/dy = 1/2$, it follows from Theorem 1.7.1 that the pdf of Y is $f_Y(y) = 1/4a$, $-2a < y < 2a$, zero elsewhere. Based on the above discussion, $\sigma_Y = (2a)/\sqrt{3}$. Hence, the standard deviation of Y is twice that of X, reflecting the fact that the probability for Y is spread out twice as much (relative to the mean zero) as the probability for X. ■

Example 1.9.2. Let X have the pdf

$$f(x) = \begin{cases} \frac{1}{2}(x+1) & -1 < x < 1 \\ 0 & \text{elsewhere.} \end{cases}$$

Then the mean value of X is

$$\mu = \int_{-\infty}^{\infty} x f(x)\,dx = \int_{-1}^{1} x\frac{x+1}{2}\,dx = \frac{1}{3},$$

while the variance of X is

$$\sigma^2 = \int_{-\infty}^{\infty} x^2 f(x)\,dx - \mu^2 = \int_{-1}^{1} x^2\frac{x+1}{2}\,dx - \left(\frac{1}{3}\right)^2 = \frac{2}{9}. \quad ■$$

Example 1.9.3. If X has the pdf

$$f(x) = \begin{cases} \frac{1}{x^2} & 1 < x < \infty \\ 0 & \text{elsewhere,} \end{cases}$$

then the mean value of X does not exist, because

$$\int_1^\infty |x| \frac{1}{x^2}\, dx = \lim_{b \to \infty} \int_1^b \frac{1}{x}\, dx = \lim_{b \to \infty} (\log b - \log 1) = \infty,$$

which is not finite. ■

We next define a third special expectation.

Definition 1.9.3 (Moment Generating Function)**.** *Let X be a random variable such that for some $h > 0$, the expectation of e^{tX} exists for $-h < t < h$. The* **moment generating function** *of X is defined to be the function $M(t) = E(e^{tX})$, for $-h < t < h$. We use the abbreviation* **mgf** *to denote the moment generating function of a random variable.*

Actually, all that is needed is that the mgf exists in an open neighborhood of 0. Such an interval, of course, includes an interval of the form $(-h, h)$ for some $h > 0$. Further, it is evident that if we set $t = 0$, we have $M(0) = 1$. But note that for an mgf to exist, it must exist in an open interval about 0.

Example 1.9.4. Suppose we have a fair spinner with the numbers $1, 2$, and 3 on it. Let X be the number of spins until the first 3 occurs. Assuming that the spins are independent, the pmf of X is

$$p(x) = \frac{1}{3}\left(\frac{2}{3}\right)^{x-1}, \quad x = 1, 2, 3, \ldots.$$

Then, using the geometric series, the mgf of X is

$$M(t) = E(e^{tX}) = \sum_{x=1}^\infty e^{tx} \frac{1}{3}\left(\frac{2}{3}\right)^{x-1} = \frac{1}{3}e^t \sum_{x=1}^\infty \left(e^t \frac{2}{3}\right)^{x-1} = \frac{1}{3}e^t \left(1 - e^t \frac{2}{3}\right)^{-1},$$

provided that $e^t(2/3) < 1$; i.e., $t < \log(3/2)$. This last interval is an open interval of 0; hence, the mgf of X exists and is given in the final line of the above derivation. ■

If we are discussing several random variables, it is often useful to subscript M as M_X to denote that this is the mgf of X.

Let X and Y be two random variables with mgfs. If X and Y have the same distribution, i.e, $F_X(z) = F_Y(z)$ for all z, then certainly $M_X(t) = M_Y(t)$ in a neighborhood of 0. But one of the most important properties of mgfs is that the converse of this statement is true too. That is, mgfs uniquely identify distributions. We state this as a theorem. The proof of this converse, though, is beyond the scope of this text; see Chung (1974). We verify it for a discrete situation.

Theorem 1.9.2. *Let X and Y be random variables with moment generating functions M_X and M_Y, respectively, existing in open intervals about 0. Then $F_X(z) = F_Y(z)$ for all $z \in R$ if and only if $M_X(t) = M_Y(t)$ for all $t \in (-h, h)$ for some $h > 0$.*

Because of the importance of this theorem, it does seem desirable to try to make the assertion plausible. This can be done if the random variable is of the discrete type. For example, let it be given that

$$M(t) = \tfrac{1}{10}e^t + \tfrac{2}{10}e^{2t} + \tfrac{3}{10}e^{3t} + \tfrac{4}{10}e^{4t}$$

is, for all real values of t, the mgf of a random variable X of the discrete type. If we let $p(x)$ be the pmf of X with support $\{a_1, a_2, a_3, \ldots\}$, then because

$$M(t) = \sum_x e^{tx} p(x),$$

we have

$$\tfrac{1}{10}e^t + \tfrac{2}{10}e^{2t} + \tfrac{3}{10}e^{3t} + \tfrac{4}{10}e^{4t} = p(a_1)e^{a_1 t} + p(a_2)e^{a_2 t} + \cdots.$$

Because this is an identity for all real values of t, it seems that the right-hand member should consist of but four terms and that each of the four should be equal, respectively, to one of those in the left-hand member; hence we may take $a_1 = 1$, $p(a_1) = \frac{1}{10}$; $a_2 = 2$, $p(a_2) = \frac{2}{10}$; $a_3 = 3$, $p(a_3) = \frac{3}{10}$; $a_4 = 4$, $p(a_4) = \frac{4}{10}$. Or, more simply, the pmf of X is

$$p(x) = \begin{cases} \frac{x}{10} & x = 1, 2, 3, 4 \\ 0 & \text{elsewhere.} \end{cases}$$

On the other hand, suppose X is a random variable of the continuous type. Let it be given that

$$M(t) = \frac{1}{1-t}, \quad t < 1,$$

is the mgf of X. That is, we are given

$$\frac{1}{1-t} = \int_{-\infty}^{\infty} e^{tx} f(x)\, dx, \quad t < 1.$$

It is not at all obvious how $f(x)$ is found. However, it is easy to see that a distribution with pdf

$$f(x) = \begin{cases} e^{-x} & 0 < x < \infty \\ 0 & \text{elsewhere} \end{cases}$$

has the mgf $M(t) = (1-t)^{-1}$, $t < 1$. Thus the random variable X has a distribution with this pdf in accordance with the assertion of the uniqueness of the mgf.

Since a distribution that has an mgf $M(t)$ is completely determined by $M(t)$, it would not be surprising if we could obtain some properties of the distribution directly from $M(t)$. For example, the existence of $M(t)$ for $-h < t < h$ implies that derivatives of $M(t)$ of all orders exist at $t = 0$. Also, a theorem in analysis allows us to interchange the order of differentiation and integration (or summation in the discrete case). That is, if X is continuous,

$$M'(t) = \frac{dM(t)}{dt} = \frac{d}{dt}\int_{-\infty}^{\infty} e^{tx} f(x)\,dx = \int_{-\infty}^{\infty} \frac{d}{dt}e^{tx} f(x)\,dx = \int_{-\infty}^{\infty} xe^{tx} f(x)\,dx.$$

Likewise, if X is a discrete random variable,

$$M'(t) = \frac{dM(t)}{dt} = \sum_x xe^{tx} p(x).$$

Upon setting $t = 0$, we have in either case

$$M'(0) = E(X) = \mu.$$

The second derivative of $M(t)$ is

$$M''(t) = \int_{-\infty}^{\infty} x^2 e^{tx} f(x)\,dx \quad \text{or} \quad \sum_x x^2 e^{tx} p(x),$$

so that $M''(0) = E(X^2)$. Accordingly, $\mathrm{Var}(X)$ equals

$$\sigma^2 = E(X^2) - \mu^2 = M''(0) - [M'(0)]^2.$$

For example, if $M(t) = (1-t)^{-1}$, $t < 1$, as in the illustration above, then

$$M'(t) = (1-t)^{-2} \quad \text{and} \quad M''(t) = 2(1-t)^{-3}.$$

Hence

$$\mu = M'(0) = 1$$

and

$$\sigma^2 = M''(0) - \mu^2 = 2 - 1 = 1.$$

Of course, we could have computed μ and σ^2 from the pdf by

$$\mu = \int_{-\infty}^{\infty} xf(x)\,dx \quad \text{and} \quad \sigma^2 = \int_{-\infty}^{\infty} x^2 f(x)\,dx - \mu^2,$$

respectively. Sometimes one way is easier than the other.

In general, if m is a positive integer and if $M^{(m)}(t)$ means the mth derivative of $M(t)$, we have, by repeated differentiation with respect to t,

$$M^{(m)}(0) = E(X^m).$$

Now

$$E(X^m) = \int_{-\infty}^{\infty} x^m f(x)\,dx \quad \text{or} \quad \sum_x x^m p(x),$$

and the integrals (or sums) of this sort are, in mechanics, called *moments*. Since $M(t)$ generates the values of $E(X^m)$, $m = 1, 2, 3, \ldots$, it is called the moment-generating function (mgf). In fact, we sometimes call $E(X^m)$ the **mth moment** of the distribution, or the mth moment of X.

The next two examples concern random variables whose distributions do not have mgfs.

Example 1.9.5. It is known that the series

$$\frac{1}{1^2} + \frac{1}{2^2} + \frac{1}{3^2} + \cdots$$

converges to $\pi^2/6$. Then

$$p(x) = \begin{cases} \frac{6}{\pi^2 x^2} & x = 1, 2, 3, \ldots \\ 0 & \text{elsewhere} \end{cases}$$

is the pmf of a discrete type of random variable X. The mgf of this distribution, if it exists, is given by

$$\begin{aligned} M(t) &= E(e^{tX}) = \sum_x e^{tx} p(x) \\ &= \sum_{x=1}^{\infty} \frac{6e^{tx}}{\pi^2 x^2}. \end{aligned}$$

The ratio test of calculus[7] may be used to show that this series diverges if $t > 0$. Thus there does not exist a positive number h such that $M(t)$ exists for $-h < t < h$. Accordingly, the distribution has the pmf $p(x)$ of this example and does not have an mgf. ■

Example 1.9.6. Let X be a continuous random variable with pdf

$$f(x) = \frac{1}{\pi} \frac{1}{x^2 + 1}, \quad -\infty < x < \infty. \tag{1.9.2}$$

This is of course the Cauchy pdf which was introduced in Exercise 1.7.24. Let $t > 0$ be given. If $x > 0$, then by the mean value theorem, for some $0 < \xi_0 < tx$,

$$\frac{e^{tx} - 1}{tx} = e^{\xi_0} \geq 1.$$

Hence, $e^{tx} \geq 1 + tx \geq tx$. This leads to the second inequality in the following derivation:

$$\begin{aligned} \int_{-\infty}^{\infty} e^{tx} \frac{1}{\pi} \frac{1}{x^2 + 1} \, dx &\geq \int_0^{\infty} e^{tx} \frac{1}{\pi} \frac{1}{x^2 + 1} \, dx \\ &\geq \int_0^{\infty} \frac{1}{\pi} \frac{tx}{x^2 + 1} \, dx = \infty. \end{aligned}$$

Because t was arbitrary, the integral does not exist in an open interval of 0. Hence, the mgf of the Cauchy distribution does not exist. ■

Example 1.9.7. Let X have the mgf $M(t) = e^{t^2/2}$, $-\infty < t < \infty$. As discussed in Chapter 3, this is the mgf of a standard normal distribution. We can differentiate $M(t)$ any number of times to find the moments of X. However, it is instructive to

[7] For example, see Chapter 2 of *Mathematical Comments*.

consider this alternative method. The function $M(t)$ is represented by the following Maclaurin's series:[8]

$$
\begin{aligned}
e^{t^2/2} &= 1 + \frac{1}{1!}\left(\frac{t^2}{2}\right) + \frac{1}{2!}\left(\frac{t^2}{2}\right)^2 + \cdots + \frac{1}{k!}\left(\frac{t^2}{2}\right)^k + \cdots \\
&= 1 + \frac{1}{2!}t^2 + \frac{(3)(1)}{4!}t^4 + \cdots + \frac{(2k-1)\cdots(3)(1)}{(2k)!}t^{2k} + \cdots .
\end{aligned}
$$

In general, though, from calculus the Maclaurin's series for $M(t)$ is

$$
\begin{aligned}
M(t) &= M(0) + \frac{M'(0)}{1!}t + \frac{M''(0)}{2!}t^2 + \cdots + \frac{M^{(m)}(0)}{m!}t^m + \cdots \\
&= 1 + \frac{E(X)}{1!}t + \frac{E(X^2)}{2!}t^2 + \cdots + \frac{E(X^m)}{m!}t^m + \cdots .
\end{aligned}
$$

Thus the coefficient of $(t^m/m!)$ in the Maclaurin's series representation of $M(t)$ is $E(X^m)$. So, for our particular $M(t)$, we have

$$
\begin{aligned}
E(X^{2k}) &= (2k-1)(2k-3)\cdots(3)(1) = \frac{(2k)!}{2^k k!}, \quad k = 1, 2, 3, \ldots && (1.9.3) \\
E(X^{2k-1}) &= 0, \quad k = 1, 2, 3, \ldots . && (1.9.4)
\end{aligned}
$$

We make use of this result in Section 3.4. ■

Remark 1.9.1. As Examples 1.9.5 and 1.9.6 show, distributions may not have moment-generating functions. In a more advanced course, we would let i denote the imaginary unit, t an arbitrary real, and we would define $\varphi(t) = E(e^{itX})$. This expectation exists for *every* distribution and it is called the **characteristic function** of the distribution. To see why $\varphi(t)$ exists for all real t, we note, in the continuous case, that its absolute value

$$
|\varphi(t)| = \left|\int_{-\infty}^{\infty} e^{itx} f(x)\, dx\right| \le \int_{-\infty}^{\infty} |e^{itx} f(x)|\, dx.
$$

However, $|f(x)| = f(x)$ since $f(x)$ is nonnegative and

$$
|e^{itx}| = |\cos tx + i \sin tx| = \sqrt{\cos^2 tx + \sin^2 tx} = 1.
$$

Thus

$$
|\varphi(t)| \le \int_{-\infty}^{\infty} f(x)\, dx = 1.
$$

Accordingly, the integral for $\varphi(t)$ exists for all real values of t. In the discrete case, a summation would replace the integral. In reference to Example 1.9.6, it can be shown that the characteristic function of the Cauchy distribution is given by $\varphi(t) = \exp\{-|t|\}$, $-\infty < t < \infty$.

[8]See Chapter 2 of *Mathematical Comments.*

Every distribution has a unique characteristic function; and to each characteristic function there corresponds a unique distribution of probability. If X has a distribution with characteristic function $\varphi(t)$, then, for instance, if $E(X)$ and $E(X^2)$ exist, they are given, respectively, by $iE(X) = \varphi'(0)$ and $i^2E(X^2) = \varphi''(0)$. Readers who are familiar with complex-valued functions may write $\varphi(t) = M(it)$ and, throughout this book, may prove certain theorems in complete generality.

Those who have studied Laplace and Fourier transforms note a similarity between these transforms and $M(t)$ and $\varphi(t)$; it is the uniqueness of these transforms that allows us to assert the uniqueness of each of the moment-generating and characteristic functions. ■

EXERCISES

1.9.1. Find the mean and variance, if they exist, of each of the following distributions.

(a) $p(x) = \frac{3!}{x!(3-x)!}(\frac{1}{2})^3$, $x = 0, 1, 2, 3$, zero elsewhere.

(b) $f(x) = 6x(1-x)$, $0 < x < 1$, zero elsewhere.

(c) $f(x) = 2/x^3$, $1 < x < \infty$, zero elsewhere.

1.9.2. Let $p(x) = (\frac{1}{2})^x$, $x = 1, 2, 3, \ldots$, zero elsewhere, be the pmf of the random variable X. Find the mgf, the mean, and the variance of X.

1.9.3. For each of the following distributions, compute $P(\mu - 2\sigma < X < \mu + 2\sigma)$.

(a) $f(x) = 6x(1-x)$, $0 < x < 1$, zero elsewhere.

(b) $p(x) = (\frac{1}{2})^x$, $x = 1, 2, 3, \ldots$, zero elsewhere.

1.9.4. If the variance of the random variable X exists, show that

$$E(X^2) \geq [E(X)]^2.$$

1.9.5. Let a random variable X of the continuous type have a pdf $f(x)$ whose graph is symmetric with respect to $x = c$. If the mean value of X exists, show that $E(X) = c$.
Hint: Show that $E(X - c)$ equals zero by writing $E(X - c)$ as the sum of two integrals: one from $-\infty$ to c and the other from c to ∞. In the first, let $y = c - x$; and, in the second, $z = x - c$. Finally, use the symmetry condition $f(c-y) = f(c+y)$ in the first.

1.9.6. Let the random variable X have mean μ, standard deviation σ, and mgf $M(t)$, $-h < t < h$. Show that

$$E\left(\frac{X-\mu}{\sigma}\right) = 0, \quad E\left[\left(\frac{X-\mu}{\sigma}\right)^2\right] = 1,$$

and

$$E\left\{\exp\left[t\left(\frac{X-\mu}{\sigma}\right)\right]\right\} = e^{-\mu t/\sigma} M\left(\frac{t}{\sigma}\right), \quad -h\sigma < t < h\sigma.$$

1.9.7. Show that the moment generating function of the random variable X having the pdf $f(x) = \frac{1}{3}$, $-1 < x < 2$, zero elsewhere, is

$$M(t) = \begin{cases} \frac{e^{2t} - e^{-t}}{3t} & t \neq 0 \\ 1 & t = 0. \end{cases}$$

1.9.8. Let X be a random variable such that $E[(X-b)^2]$ exists for all real b. Show that $E[(X-b)^2]$ is a minimum when $b = E(X)$.

1.9.9. Let X be a random variable of the continuous type that has pdf $f(x)$. If m is the unique median of the distribution of X and b is a real constant, show that

$$E(|X-b|) = E(|X-m|) + 2\int_m^b (b-x)f(x)\,dx,$$

provided that the expectations exist. For what value of b is $E(|X-b|)$ a minimum?

1.9.10. Let X denote a random variable for which $E[(X-a)^2]$ exists. Give an example of a distribution of a discrete type such that this expectation is zero. Such a distribution is called a **degenerate distribution**.

1.9.11. Let X denote a random variable such that $K(t) = E(t^X)$ exists for all real values of t in a certain open interval that includes the point $t = 1$. Show that $K^{(m)}(1)$ is equal to the mth **factorial moment** $E[X(X-1)\cdots(X-m+1)]$.

1.9.12. Let X be a random variable. If m is a positive integer, the expectation $E[(X-b)^m]$, if it exists, is called the mth moment of the distribution about the point b. Let the first, second, and third moments of the distribution about the point 7 be 3, 11, and 15, respectively. Determine the mean μ of X, and then find the first, second, and third moments of the distribution about the point μ.

1.9.13. Let X be a random variable such that $R(t) = E(e^{t(X-b)})$ exists for t such that $-h < t < h$. If m is a positive integer, show that $R^{(m)}(0)$ is equal to the mth moment of the distribution about the point b.

1.9.14. Let X be a random variable with mean μ and variance σ^2 such that the third moment $E[(X-\mu)^3]$ about the vertical line through μ exists. The value of the ratio $E[(X-\mu)^3]/\sigma^3$ is often used as a measure of **skewness**. Graph each of the following probability density functions and show that this measure is negative, zero, and positive for these respective distributions (which are said to be skewed to the left, not skewed, and skewed to the right, respectively).

(a) $f(x) = (x+1)/2$, $-1 < x < 1$, zero elsewhere.

(b) $f(x) = \frac{1}{2}$, $-1 < x < 1$, zero elsewhere.

(c) $f(x) = (1-x)/2$, $-1 < x < 1$, zero elsewhere.

1.9.15. Let X be a random variable with mean μ and variance σ^2 such that the fourth moment $E[(X-\mu)^4]$ exists. The value of the ratio $E[(X-\mu)^4]/\sigma^4$ is often used as a measure of **kurtosis**. Graph each of the following probability density functions and show that this measure is smaller for the first distribution.

(a) $f(x) = \frac{1}{2}$, $-1 < x < 1$, zero elsewhere.

(b) $f(x) = 3(1 - x^2)/4$, $-1 < x < 1$, zero elsewhere.

1.9.16. Let the random variable X have pmf

$$p(x) = \begin{cases} p & x = -1, 1 \\ 1 - 2p & x = 0 \\ 0 & \text{elsewhere,} \end{cases}$$

where $0 < p < \frac{1}{2}$. Find the measure of kurtosis as a function of p. Determine its value when $p = \frac{1}{3}$, $p = \frac{1}{5}$, $p = \frac{1}{10}$, and $p = \frac{1}{100}$. Note that the kurtosis increases as p decreases.

1.9.17. Let $\psi(t) = \log M(t)$, where $M(t)$ is the mgf of a distribution. Prove that $\psi'(0) = \mu$ and $\psi''(0) = \sigma^2$. The function $\psi(t)$ is called the **cumulant generating function**.

1.9.18. Find the mean and the variance of the distribution that has the cdf

$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{x}{8} & 0 \le x < 2 \\ \frac{x^2}{16} & 2 \le x < 4 \\ 1 & 4 \le x. \end{cases}$$

1.9.19. Find the moments of the distribution that has mgf $M(t) = (1-t)^{-3}$, $t < 1$. *Hint:* Find the Maclaurin series for $M(t)$.

1.9.20. We say that X has a **Laplace** distribution if its pdf is

$$f(t) = \frac{1}{2}e^{-|t|}, \quad -\infty < t < \infty. \tag{1.9.5}$$

(a) Show that the mgf of X is $M(t) = (1 - t^2)^{-1}$ for $|t| < 1$.

(b) Expand $M(t)$ into a Maclaurin series and use it to find all the moments of X.

1.9.21. Let X be a random variable of the continuous type with pdf $f(x)$, which is positive provided $0 < x < b < \infty$, and is equal to zero elsewhere. Show that

$$E(X) = \int_0^b [1 - F(x)]\, dx,$$

where $F(x)$ is the cdf of X.

1.9.22. Let X be a random variable of the discrete type with pmf $p(x)$ that is positive on the nonnegative integers and is equal to zero elsewhere. Show that

$$E(X) = \sum_{x=0}^{\infty} [1 - F(x)],$$

where $F(x)$ is the cdf of X.

1.9.23. Let X have the pmf $p(x) = 1/k,\ x = 1, 2, 3, \ldots, k$, zero elsewhere. Show that the mgf is

$$M(t) = \begin{cases} \frac{e^t(1-e^{kt})}{k(1-e^t)} & t \neq 0 \\ 1 & t = 0. \end{cases}$$

1.9.24. Let X have the cdf $F(x)$ that is a mixture of the continuous and discrete types, namely

$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{x+1}{4} & 0 \leq x < 1 \\ 1 & 1 \leq x. \end{cases}$$

Determine reasonable definitions of $\mu = E(X)$ and $\sigma^2 = \text{var}(X)$ and compute each. *Hint:* Determine the parts of the pmf and the pdf associated with each of the discrete and continuous parts, and then sum for the discrete part and integrate for the continuous part.

1.9.25. Consider k continuous-type distributions with the following characteristics: pdf $f_i(x)$, mean μ_i, and variance σ_i^2, $i = 1, 2, \ldots, k$. If $c_i \geq 0$, $i = 1, 2, \ldots, k$, and $c_1 + c_2 + \cdots + c_k = 1$, show that the mean and the variance of the distribution having pdf $c_1 f_1(x) + \cdots + c_k f_k(x)$ are $\mu = \sum_{i=1}^k c_i \mu_i$ and $\sigma^2 = \sum_{i=1}^k c_i[\sigma_i^2 + (\mu_i - \mu)^2]$, respectively.

1.9.26. Let X be a random variable with a pdf $f(x)$ and mgf $M(t)$. Suppose f is symmetric about 0; i.e., $f(-x) = f(x)$. Show that $M(-t) = M(t)$.

1.9.27. Let X have the exponential pdf, $f(x) = \beta^{-1}\exp\{-x/\beta\}$, $0 < x < \infty$, zero elsewhere. Find the mgf, the mean, and the variance of X.

1.10 Important Inequalities

In this section, we discuss some famous inequalities involving expectations. We make use of these inequalities in the remainder of the text. We begin with a useful result.

Theorem 1.10.1. *Let X be a random variable and let m be a positive integer. Suppose $E[X^m]$ exists. If k is a positive integer and $k \leq m$, then $E[X^k]$ exists.*

Proof: We prove it for the continuous case; but the proof is similar for the discrete case if we replace integrals by sums. Let $f(x)$ be the pdf of X. Then

$$\begin{aligned}
\int_{-\infty}^{\infty} |x|^k f(x)\,dx &= \int_{|x|\leq 1} |x|^k f(x)\,dx + \int_{|x|>1} |x|^k f(x)\,dx \\
&\leq \int_{|x|\leq 1} f(x)\,dx + \int_{|x|>1} |x|^m f(x)\,dx \\
&\leq \int_{-\infty}^{\infty} f(x)\,dx + \int_{-\infty}^{\infty} |x|^m f(x)\,dx \\
&\leq 1 + E[|X|^m] < \infty, \qquad (1.10.1)
\end{aligned}$$

which is the the desired result. ■

Theorem 1.10.2 (Markov's Inequality). *Let $u(X)$ be a nonnegative function of the random variable X. If $E[u(X)]$ exists, then for every positive constant c,*

$$P[u(X) \geq c] \leq \frac{E[u(X)]}{c}.$$

Proof. The proof is given when the random variable X is of the continuous type; but the proof can be adapted to the discrete case if we replace integrals by sums. Let $A = \{x : u(x) \geq c\}$ and let $f(x)$ denote the pdf of X. Then

$$E[u(X)] = \int_{-\infty}^{\infty} u(x)f(x)\,dx = \int_A u(x)f(x)\,dx + \int_{A^c} u(x)f(x)\,dx.$$

Since each of the integrals in the extreme right-hand member of the preceding equation is nonnegative, the left-hand member is greater than or equal to either of them. In particular,

$$E[u(X)] \geq \int_A u(x)f(x)\,dx.$$

However, if $x \in A$, then $u(x) \geq c$; accordingly, the right-hand member of the preceding inequality is not increased if we replace $u(x)$ by c. Thus

$$E[u(X)] \geq c\int_A f(x)\,dx.$$

Since

$$\int_A f(x)\,dx = P(X \in A) = P[u(X) \geq c],$$

it follows that

$$E[u(X)] \geq cP[u(X) \geq c],$$

which is the desired result. ■

The preceding theorem is a generalization of an inequality that is often called **Chebyshev's Inequality**. This inequality we now establish.

Theorem 1.10.3 (Chebyshev's Inequality). *Let X be a random variable with finite variance σ^2 (by Theorem 1.10.1, this implies that the mean $\mu = E(X)$ exists). Then for every $k > 0$,*

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}, \tag{1.10.2}$$

or, equivalently,

$$P(|X - \mu| < k\sigma) \geq 1 - \frac{1}{k^2}.$$

Proof. In Theorem 1.10.2 take $u(X) = (X-\mu)^2$ and $c = k^2\sigma^2$. Then we have

$$P[(X-\mu)^2 \geq k^2\sigma^2] \leq \frac{E[(X-\mu)^2]}{k^2\sigma^2}.$$

Since the numerator of the right-hand member of the preceding inequality is σ^2, the inequality may be written

$$P(|X-\mu| \geq k\sigma) \leq \frac{1}{k^2},$$

which is the desired result. Naturally, we would take the positive number k to be greater than 1 to have an inequality of interest. ■

Hence, the number $1/k^2$ is an upper bound for the probability $P(|X-\mu| \geq k\sigma)$. In the following example this upper bound and the exact value of the probability are compared in special instances.

Example 1.10.1. Let X have the uniform pdf

$$f(x) = \begin{cases} \frac{1}{2\sqrt{3}} & -\sqrt{3} < x < \sqrt{3} \\ 0 & \text{elsewhere.} \end{cases}$$

Based on Example 1.9.1, for this uniform distribution, we have $\mu = 0$ and $\sigma^2 = 1$. If $k = \frac{3}{2}$, we have the exact probability

$$P(|X-\mu| \geq k\sigma) = P\left(|X| \geq \frac{3}{2}\right) = 1 - \int_{-3/2}^{3/2} \frac{1}{2\sqrt{3}}\, dx = 1 - \frac{\sqrt{3}}{2}.$$

By Chebyshev's inequality, this probability has the upper bound $1/k^2 = \frac{4}{9}$. Since $1 - \sqrt{3}/2 = 0.134$, approximately, the exact probability in this case is considerably less than the upper bound $\frac{4}{9}$. If we take $k = 2$, we have the exact probability $P(|X-\mu| \geq 2\sigma) = P(|X| \geq 2) = 0$. This again is considerably less than the upper bound $1/k^2 = \frac{1}{4}$ provided by Chebyshev's inequality. ■

In each of the instances in Example 1.10.1, the probability $P(|X-\mu| \geq k\sigma)$ and its upper bound $1/k^2$ differ considerably. This suggests that this inequality might be made sharper. However, if we want an inequality that holds for every $k > 0$ and holds for all random variables having a finite variance, such an improvement is impossible, as is shown by the following example.

Example 1.10.2. Let the random variable X of the discrete type have probabilities $\frac{1}{8}, \frac{6}{8}, \frac{1}{8}$ at the points $x = -1, 0, 1$, respectively. Here $\mu = 0$ and $\sigma^2 = \frac{1}{4}$. If $k = 2$, then $1/k^2 = \frac{1}{4}$ and $P(|X-\mu| \geq k\sigma) = P(|X| \geq 1) = \frac{1}{4}$. That is, the probability $P(|X-\mu| \geq k\sigma)$ here attains the upper bound $1/k^2 = \frac{1}{4}$. Hence the inequality cannot be improved without further assumptions about the distribution of X. ■

A convenient form of Chebyshev's Inequality is found by taking $k\sigma = \epsilon$ for $\epsilon > 0$. Then Equation (1.10.2) becomes

$$P(|X-\mu| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2}, \quad \text{for all } \epsilon > 0\,. \tag{1.10.3}$$

The second inequality of this section involves convex functions.

Definition 1.10.1. *A function ϕ defined on an interval (a, b), $-\infty \leq a < b \leq \infty$, is said to be a* **convex** *function if for all x, y in (a, b) and for all $0 < \gamma < 1$,*

$$\phi[\gamma x + (1-\gamma)y] \leq \gamma\phi(x) + (1-\gamma)\phi(y). \tag{1.10.4}$$

We say ϕ is **strictly convex** *if the above inequality is strict.*

Depending on the existence of first or second derivatives of ϕ, the following theorem can be proved.

Theorem 1.10.4. *If ϕ is differentiable on (a, b), then*

(a) ϕ is convex if and only if $\phi'(x) \leq \phi'(y)$, for all $a < x < y < b$,

(b) ϕ is strictly convex if and only if $\phi'(x) < \phi'(y)$, for all $a < x < y < b$.

If ϕ is twice differentiable on (a, b), then

(a) ϕ is convex if and only if $\phi''(x) \geq 0$, for all $a < x < b$,

(b) ϕ is strictly convex if $\phi''(x) > 0$, for all $a < x < b$.

Of course, the second part of this theorem follows immediately from the first part. While the first part appeals to one's intuition, the proof of it can be found in most analysis books; see, for instance, Hewitt and Stromberg (1965). A very useful probability inequality follows from convexity.

Theorem 1.10.5 (Jensen's Inequality)**.** *If ϕ is convex on an open interval I and X is a random variable whose support is contained in I and has finite expectation, then*

$$\phi[E(X)] \leq E[\phi(X)]. \tag{1.10.5}$$

If ϕ is strictly convex, then the inequality is strict unless X is a constant random variable.

Proof: For our proof we assume that ϕ has a second derivative, but in general only convexity is required. Expand $\phi(x)$ into a Taylor series about $\mu = E[X]$ of order 2:

$$\phi(x) = \phi(\mu) + \phi'(\mu)(x-\mu) + \frac{\phi''(\zeta)(x-\mu)^2}{2},$$

where ζ is between x and μ.[9] Because the last term on the right side of the above equation is nonnegative, we have

$$\phi(x) \geq \phi(\mu) + \phi'(\mu)(x-\mu).$$

Taking expectations of both sides leads to the result. The inequality is strict if $\phi''(x) > 0$, for all $x \in (a, b)$, provided X is not a constant. ■

[9] See, for example, the discussion on Taylor series in *Mathematical Comments* referenced in the Preface.

Example 1.10.3. Let X be a nondegenerate random variable with mean μ and a finite second moment. Then $\mu^2 < E(X^2)$. This is obtained by Jensen's inequality using the strictly convex function $\phi(t) = t^2$. ■

The last inequality concerns different means of finite sets of positive numbers.

Example 1.10.4 (Harmonic and Geometric Means)**.** Let $\{a_1, \ldots, a_n\}$ be a set of positive numbers. Create a distribution for a random variable X by placing weight $1/n$ on each of the numbers $a_1, \ldots, a_n$. Then the mean of X is the **arithmetic mean**, (AM), $E(X) = n^{-1}\sum_{i=1}^n a_i$. Then, since $-\log x$ is a convex function, we have by Jensen's inequality that

$$-\log\left(\frac{1}{n}\sum_{i=1}^n a_i\right) \le E(-\log X) = -\frac{1}{n}\sum_{i=1}^n \log a_i = -\log(a_1a_2\cdots a_n)^{1/n}$$

or, equivalently,

$$\log\left(\frac{1}{n}\sum_{i=1}^n a_i\right) \ge \log(a_1a_2\cdots a_n)^{1/n},$$

and, hence,

$$(a_1a_2\cdots a_n)^{1/n} \le \frac{1}{n}\sum_{i=1}^n a_i. \tag{1.10.6}$$

The quantity on the left side of this inequality is called the **geometric mean** (GM). So (1.10.6) is equivalent to saying that GM $\le$ AM for any finite set of positive numbers.

Now in (1.10.6) replace a_i by $1/a_i$ (which is also positive). We then obtain

$$\frac{1}{n}\sum_{i=1}^n \frac{1}{a_i} \ge \left(\frac{1}{a_1}\frac{1}{a_2}\cdots\frac{1}{a_n}\right)^{1/n}$$

or, equivalently,

$$\frac{1}{\frac{1}{n}\sum_{i=1}^n \frac{1}{a_i}} \le (a_1a_2\cdots a_n)^{1/n}. \tag{1.10.7}$$

The left member of this inequality is called the **harmonic mean** (HM). Putting (1.10.6) and (1.10.7) together, we have shown the relationship

$$\text{HM} \le \text{GM} \le \text{AM}, \tag{1.10.8}$$

for any finite set of positive numbers. ■

EXERCISES

1.10.1. Let X be a random variable with mean μ and let $E[(X-\mu)^{2k}]$ exist. Show, with $d > 0$, that $P(|X-\mu| \ge d) \le E[(X-\mu)^{2k}]/d^{2k}$. This is essentially Chebyshev's inequality when $k = 1$. The fact that this holds for all $k = 1, 2, 3, \ldots,$ when those $(2k)$th moments exist, usually provides a much smaller upper bound for $P(|X-\mu| \ge d)$ than does Chebyshev's result.

1.10.2. Let X be a random variable such that $P(X \leq 0) = 0$ and let $\mu = E(X)$ exist. Show that $P(X \geq 2\mu) \leq \frac{1}{2}$.

1.10.3. If X is a random variable such that $E(X) = 3$ and $E(X^2) = 13$, use Chebyshev's inequality to determine a lower bound for the probability $P(-2 < X < 8)$.

1.10.4. Suppose X has a Laplace distribution with pdf (1.9.20). Show that the mean and variance of X are 0 and 2, respectively. Using Chebyshev's inequality determine the upper bound for $P(|X| \geq 5)$ and then compare it with the exact probability.

1.10.5. Let X be a random variable with mgf $M(t)$, $-h < t < h$. Prove that

$$P(X \geq a) \leq e^{-at}M(t), \quad 0 < t < h,$$

and that

$$P(X \leq a) \leq e^{-at}M(t), \quad -h < t < 0.$$

Hint: Let $u(x) = e^{tx}$ and $c = e^{ta}$ in Theorem 1.10.2. *Note:* These results imply that $P(X \geq a)$ and $P(X \leq a)$ are less than or equal to their respective least upper bounds for $e^{-at}M(t)$ when $0 < t < h$ and when $-h < t < 0$.

1.10.6. The mgf of X exists for all real values of t and is given by

$$M(t) = \frac{e^t - e^{-t}}{2t}, \quad t \neq 0, \quad M(0) = 1.$$

Use the results of the preceding exercise to show that $P(X \geq 1) = 0$ and $P(X \leq -1) = 0$. Note that here h is infinite.

1.10.7. Let X be a positive random variable; i.e., $P(X \leq 0) = 0$. Argue that

- **(a)** $E(1/X) \geq 1/E(X)$
- **(b)** $E[-\log X] \geq -\log[E(X)]$
- **(c)** $E[\log(1/X)] \geq \log[1/E(X)]$
- **(d)** $E[X^3] \geq [E(X)]^3$.

This page intentionally left blank

Chapter 2

Multivariate Distributions

2.1 Distributions of Two Random Variables

We begin the discussion of a pair of random variables with the following example. A coin is tossed three times and our interest is in the ordered number pair (number of H's on first two tosses, number of H's on all three tosses), where H and T represent, respectively, heads and tails. Let $\mathcal{C} = \{TTT, TTH, THT, HTT, THH, HTH, HHT, HHH\}$ denote the sample space. Let X_1 denote the number of H's on the first two tosses and X_2 denote the number of H's on all three flips. Then our interest can be represented by the pair of random variables (X_1, X_2). For example, $(X_1(HTH), X_2(HTH))$ represents the outcome $(1, 2)$. Continuing in this way, X_1 and X_2 are real-valued functions defined on the sample space $\mathcal{C}$, which take us from the sample space to the space of ordered number pairs.

$$\mathcal{D} = \{(0,0), (0,1), (1,1), (1,2), (2,2), (2,3)\}.$$

Thus X_1 and X_2 are two random variables defined on the space $\mathcal{C}$, and, in this example, the space of these random variables is the two-dimensional set $\mathcal{D}$, which is a subset of two-dimensional Euclidean space R^2. Hence (X_1, X_2) is a vector function from $\mathcal{C}$ to $\mathcal{D}$. We now formulate the definition of a random vector.

Definition 2.1.1 (Random Vector)**.** *Given a random experiment with a sample space $\mathcal{C}$, consider two random variables X_1 and X_2, which assign to each element c of $\mathcal{C}$ one and only one ordered pair of numbers $X_1(c) = x_1$, $X_2(c) = x_2$. Then we say that (X_1, X_2) is a* **random vector***. The* **space** *of (X_1, X_2) is the set of ordered pairs $\mathcal{D} = \{(x_1, x_2) : x_1 = X_1(c), x_2 = X_2(c), c \in \mathcal{C}\}$.*

We often denote random vectors using vector notation $\mathbf{X} = (X_1, X_2)'$, where the $'$ denotes the transpose of the row vector (X_1, X_2). Also, we often use (X, Y) to denote random vectors.

Let $\mathcal{D}$ be the space associated with the random vector (X_1, X_2). Let A be a subset of $\mathcal{D}$. As in the case of one random variable, we speak of the event A. We wish to define the probability of the event A, which we denote by $P_{X_1,X_2}[A]$. As with random variables in Section 1.5 we can uniquely define P_{X_1,X_2} in terms of the **cumulative distribution function** (cdf), which is given by

$$F_{X_1,X_2}(x_1,x_2) = P[\{X_1 \le x_1\} \cap \{X_2 \le x_2\}], \tag{2.1.1}$$

for all $(x_1,x_2) \in R^2$. Because X_1 and X_2 are random variables, each of the events in the above intersection and the intersection of the events are events in the original sample space $\mathcal{C}$. Thus the expression is well defined. As with random variables, we write $P[\{X_1 \le x_1\} \cap \{X_2 \le x_2\}]$ as $P[X_1 \le x_1, X_2 \le x_2]$. As Exercise 2.1.3 shows,

$$\begin{aligned} P[a_1 < X_1 \le b_1, a_2 < X_2 \le b_2] &= F_{X_1,X_2}(b_1,b_2) - F_{X_1,X_2}(a_1,b_2) \\ &\quad - F_{X_1,X_2}(b_1,a_2) + F_{X_1,X_2}(a_1,a_2). \end{aligned} \tag{2.1.2}$$

Hence, all induced probabilities of sets of the form $(a_1,b_1]\times(a_2,b_2]$ can be formulated in terms of the cdf. We often call this cdf the **joint cumulative distribution function** of (X_1,X_2).

As with random variables, we are mainly concerned with two types of random vectors, namely discrete and continuous. We first discuss the discrete type.

A random vector (X_1,X_2) is a **discrete random vector** if its space $\mathcal{D}$ is finite or countable. Hence, X_1 and X_2 are both discrete also. The **joint probability mass function** (pmf) of (X_1,X_2) is defined by

$$p_{X_1,X_2}(x_1,x_2) = P[X_1 = x_1, X_2 = x_2], \tag{2.1.3}$$

for all $(x_1,x_2) \in \mathcal{D}$. As with random variables, the pmf uniquely defines the cdf. It also is characterized by the two properties

$$\text{(i) } 0 \le p_{X_1,X_2}(x_1,x_2) \le 1 \text{ and (ii) } \sum\sum_{\mathcal{D}} p_{X_1,X_2}(x_1,x_2) = 1. \tag{2.1.4}$$

For an event $B \in \mathcal{D}$, we have

$$P[(X_1,X_2) \in B] = \sum\sum_{B} p_{X_1,X_2}(x_1,x_2).$$

Example 2.1.1. Consider the example at the beginning of this section where a fair coin is flipped three times and X_1 and X_2 are the number of heads on the first two flips and all 3 flips, respectively. We can conveniently table the pmf of (X_1,X_2) as

		Support of X_2			
		0	1	2	3
Support of X_1	0	$\frac{1}{8}$	$\frac{1}{8}$	0	0
	1	0	$\frac{2}{8}$	$\frac{2}{8}$	0
	2	0	0	$\frac{1}{8}$	$\frac{1}{8}$

For instance, $P(X_1 \ge 2, X_2 \ge 2) = p(2,2) + p(2,3) = 2/8$. ■

At times it is convenient to speak of the **support** of a discrete random vector (X_1, X_2). These are all the points (x_1, x_2) in the space of (X_1, X_2) such that $p(x_1, x_2) > 0$. In the last example the support consists of the six points $\{(0,0), (0,1), (1,1), (1,2), (2,2), (2,3)\}$.

We say a random vector (X_1, X_2) with space $\mathcal{D}$ is of the **continuous** type if its cdf $F_{X_1,X_2}(x_1, x_2)$ is continuous. For the most part, the continuous random vectors in this book have cdfs that can be represented as integrals of nonnegative functions. That is, $F_{X_1,X_2}(x_1, x_2)$ can be expressed as

$$F_{X_1,X_2}(x_1, x_2) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} f_{X_1,X_2}(w_1, w_2)\, dw_1 dw_2, \tag{2.1.5}$$

for all $(x_1, x_2) \in R^2$. We call the integrand the **joint probability density function** (pdf) of (X_1, X_2). Then

$$\frac{\partial^2 F_{X_1,X_2}(x_1, x_2)}{\partial x_1\, \partial x_2} = f_{X_1,X_2}(x_1, x_2),$$

except possibly on events that have probability zero. A pdf is essentially characterized by the two properties

$$\text{(i) } f_{X_1,X_2}(x_1, x_2) \geq 0 \text{ and (ii) } \int\int_{\mathcal{D}} f_{X_1,X_2}(x_1, x_2)\, dx_1 dx_2 = 1. \tag{2.1.6}$$

For the reader's benefit, Section 4.2 of the accompanying resource *Mathematical Comments*[1] offers a short review of double integration. For an event $A \in \mathcal{D}$, we have

$$P[(X_1, X_2) \in A] = \int\int_A f_{X_1,X_2}(x_1, x_2)\, dx_1 dx_2.$$

Note that the $P[(X_1, X_2) \in A]$ is just the volume under the surface $z = f_{X_1,X_2}(x_1, x_2)$ over the set A.

Remark 2.1.1. As with univariate random variables, we often drop the subscript (X_1, X_2) from joint cdfs, pdfs, and pmfs, when it is clear from the context. We also use notation such as f_{12} instead of f_{X_1,X_2}. Besides (X_1, X_2), we often use (X, Y) to express random vectors. ■

We next present two examples of jointly continuous random variables.

Example 2.1.2. Consider a continuous random vector (X, Y) which is uniformly distributed over the unit circle in R^2. Since the area of the unit circle is π, the joint pdf is

$$f(x, y) = \begin{cases} \frac{1}{\pi} & -1 < y < 1, -\sqrt{1-y^2} < x < \sqrt{1-y^2} \\ 0 & \text{elsewhere.} \end{cases}$$

Probabilities of certain events follow immediately from geometry. For instance, let A be the interior of the circle with radius $1/2$. Then $P[(X, Y) \in A] = \pi(1/2)^2/\pi = 1/4$. Next, let B be the ring formed by the concentric circles with the respective radii of $1/2$ and $\sqrt{2}/2$. Then $P[(X, Y) \in B] = \pi[(\sqrt{2}/2)^2 - (1/2)^2]/\pi = 1/4$. The regions A and B have the same area and hence for this uniform pdf are equilikely. ■

[1] Downloadable at the site listed in the Preface.

In the next example, we use the general fact that double integrals can be expressed as iterated univariate integrals. Thus double integrations can be carried out using iterated univariate integrations. This is discussed in some detail with examples in Section 4.2 of the accompanying resource *Mathematical Comments*.[2] The aid of a simple sketch of the region of integration is valuable in setting up the upper and lower limits of integration for each of the iterated integrals.

Example 2.1.3. Suppose an electrical component has two batteries. Let X and Y denote the lifetimes in standard units of the respective batteries. Assume that the pdf of (X, Y) is

$$f(x,y) = \begin{cases} 4xye^{-(x^2+y^2)} & x > 0, y > 0 \\ 0 & \text{elsewhere.} \end{cases}$$

The surface $z = f(x, y)$ is sketched in Figure 2.1.1 where the grid squares are 0.1 by 0.1. From the figure, the pdf peaks at about $(x, y) = (0.7, 0.7)$. Solving the equations $\partial f/\partial x = 0$ and $\partial f/\partial y = 0$ simultaneously shows that actually the maximum of $f(x, y)$ occurs at $(x, y) = (\sqrt{2}/2, \sqrt{2}/2)$. The batteries are more likely to die in regions near the peak. The surface tapers to 0 as x and y get large in any direction. for instance, the probability that both batteries survive beyond $\sqrt{2}/2$ units is given by

$$\begin{aligned} P\left(X > \frac{\sqrt{2}}{2}, Y > \frac{\sqrt{2}}{2}\right) &= \int_{\sqrt{2}/2}^{\infty}\int_{\sqrt{2}/2}^{\infty} 4xye^{-(x^2+y^2)}\,dxdy \\ &= \int_{\sqrt{2}/2}^{\infty} 2xe^{-x^2}\left[\int_{\sqrt{2}/2}^{\infty} 2ye^{-y^2}\,dy\right]dx \\ &= \int_{1/2}^{\infty} e^{-z}\left[\int_{1/2}^{\infty} e^{-w}dw\right]dz = \left(e^{-1/2}\right)^2 \approx 0.3679, \end{aligned}$$

where we made use of the change-in-variables $z = x^2$ and $w = y^2$. In contrast to the last example, consider the regions $A = \{(x, y) : |x - (1/2)| < 0.3, |y - (1/2)| < 0.3\}$ and $B = \{(x, y) : |x-2| < 0.3, |y-2| < 0.3\}$. The reader should locate these regions on Figure 2.1.1. The areas of A and B are the same, but it is clear from the figure that $P[(X, Y) \in A]$ is much larger than $P[(X, Y) \in B]$. Exercise 2.1.6 confirms this by showing that $P[(X, Y) \in A] = 0.1879$ while $P[(X, Y) \in B] = 0.0026$. ■

For a continuous random vector (X_1, X_2), the **support** of (X_1, X_2) contains all points (x_1, x_2) for which $f(x_1, x_2) > 0$. We denote the support of a random vector by $\mathcal{S}$. As in the univariate case, $\mathcal{S} \subset \mathcal{D}$.

As in the last two examples, we extend the definition of a pdf $f_{X_1,X_2}(x_1, x_2)$ over R^2 by using zero elsewhere. We do this consistently so that tedious, repetitious references to the space $\mathcal{D}$ can be avoided. Once this is done, we replace

$$\int\int_{\mathcal{D}} f_{X_1,X_2}(x_1,x_2)\,dx_1dx_2 \quad \text{by} \quad \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x_1,x_2)\,dx_1\,dx_2.$$

[2]Downloadable at the site listed in the Preface.

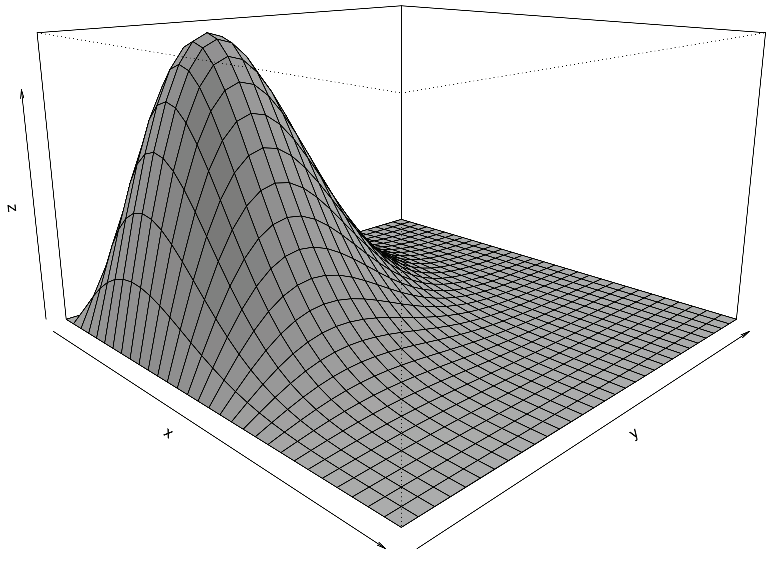

Figure 2.1.1: A sketch of the the surface of the joint pdf discussed in Example 2.1.3. On the figure, the origin is located at the intersection of the x and z axes and the grid squares are 0.1 by 0.1, so points are easily located. As discussed in the text, the peak of the pdf occurs at the point $(\sqrt{2}/2, \sqrt{2}/2)$.

Likewise we may extend the pmf $p_{X_1,X_2}(x_1, x_2)$ over a convenient set by using zero elsewhere. Hence, we replace

$$\sum_{\mathcal{D}}\sum p_{X_1,X_2}(x_1, x_2) \quad \text{by} \quad \sum_{x_2}\sum_{x_1} p(x_1, x_2).$$

2.1.1 Marginal Distributions

Let (X_1, X_2) be a random vector. Then both X_1 and X_2 are random variables. We can obtain their distributions in terms of the joint distribution of (X_1, X_2) as follows. Recall that the event which defined the cdf of X_1 at x_1 is $\{X_1 \leq x_1\}$. However,

$$\{X_1 \leq x_1\} = \{X_1 \leq x_1\} \cap \{-\infty < X_2 < \infty\} = \{X_1 \leq x_1, -\infty < X_2 < \infty\}.$$

Taking probabilities, we have

$$F_{X_1}(x_1) = P[X_1 \leq x_1, -\infty < X_2 < \infty], \tag{2.1.7}$$

Table 2.1.1: Joint and Marginal Distributions for the discrete random vector (X_1, X_2) of Example 2.1.1.

		Support of X_2				
		0	1	2	3	$p_{X_1}(x_1)$
Support of X_1	0	$\frac{1}{8}$	$\frac{1}{8}$	0	0	$\frac{2}{8}$
	1	0	$\frac{2}{8}$	$\frac{2}{8}$	0	$\frac{4}{8}$
	2	0	0	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{2}{8}$
	$p_{X_2}(x_2)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$	

for all $x_1 \in R$. By Theorem 1.3.6 we can write this equation as $F_{X_1}(x_1) = \lim_{x_2 \uparrow \infty} F(x_1, x_2)$. Thus we have a relationship between the cdfs, which we can extend to either the pmf or pdf depending on whether (X_1, X_2) is discrete or continuous.

First consider the discrete case. Let $\mathcal{D}_{X_1}$ be the support of X_1. For $x_1 \in \mathcal{D}_{X_1}$, Equation (2.1.7) is equivalent to

$$F_{X_1}(x_1) = \sum_{w_1 \le x_1, -\infty < x_2 < \infty} \sum p_{X_1,X_2}(w_1, x_2) = \sum_{w_1 \le x_1} \left\{ \sum_{x_2 < \infty} p_{X_1,X_2}(w_1, x_2) \right\}.$$

By the uniqueness of cdfs, the quantity in braces must be the pmf of X_1 evaluated at w_1; that is,

$$p_{X_1}(x_1) = \sum_{x_2 < \infty} p_{X_1,X_2}(x_1, x_2), \tag{2.1.8}$$

for all $x_1 \in \mathcal{D}_{X_1}$. Hence, to find the probability that X_1 is x_1, keep x_1 fixed and sum p_{X_1,X_2} over all of x_2. In terms of a tabled joint pmf with rows comprised of X_1 support values and columns comprised of X_2 support values, this says that the distribution of X_1 can be obtained by the marginal sums of the rows. Likewise, the pmf of X_2 can be obtained by marginal sums of the columns.

Consider the joint discrete distribution of the random vector (X_1, X_2) as presented in Example 2.1.1. In Table 2.1.1, we have added these marginal sums. The final row of this table is the pmf of X_2, while the final column is the pmf of X_1. In general, because these distributions are recorded in the margins of the table, we often refer to them as **marginal** pmfs.

Example 2.1.4. Consider a random experiment that consists of drawing at random one chip from a bowl containing 10 chips of the same shape and size. Each chip has an ordered pair of numbers on it: one with $(1, 1)$, one with $(2, 1)$, two with $(3, 1)$, one with $(1, 2)$, two with $(2, 2)$, and three with $(3, 2)$. Let the random variables X_1 and X_2 be defined as the respective first and second values of the ordered pair. Thus the joint pmf $p(x_1, x_2)$ of X_1 and X_2 can be given by the following table, with $p(x_1, x_2)$ equal to zero elsewhere.

	x_2		
x_1	1	2	$p_1(x_1)$
1	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{2}{10}$
2	$\frac{1}{10}$	$\frac{2}{10}$	$\frac{3}{10}$
3	$\frac{2}{10}$	$\frac{3}{10}$	$\frac{5}{10}$
$p_2(x_2)$	$\frac{4}{10}$	$\frac{6}{10}$	

The joint probabilities have been summed in each row and each column and these sums recorded in the margins to give the marginal probability mass functions of X_1 and X_2, respectively. Note that it is not necessary to have a formula for $p(x_1, x_2)$ to do this. ■

We next consider the continuous case. Let $\mathcal{D}_{X_1}$ be the support of X_1. For $x_1 \in \mathcal{D}_{X_1}$, Equation (2.1.7) is equivalent to

$$F_{X_1}(x_1) = \int_{-\infty}^{x_1} \int_{-\infty}^{\infty} f_{X_1,X_2}(w_1, x_2)\, dx_2 dw_1 = \int_{-\infty}^{x_1} \left\{ \int_{-\infty}^{\infty} f_{X_1,X_2}(w_1, x_2)\, dx_2 \right\} dw_1.$$

By the uniqueness of cdfs, the quantity in braces must be the pdf of X_1, evaluated at w_1; that is,

$$f_{X_1}(x_1) = \int_{-\infty}^{\infty} f_{X_1,X_2}(x_1, x_2)\, dx_2 \tag{2.1.9}$$

for all $x_1 \in \mathcal{D}_{X_1}$. Hence, in the continuous case the marginal pdf of X_1 is found by integrating out x_2. Similarly, the marginal pdf of X_2 is found by integrating out x_1.

Example 2.1.5 (Example 2.1.2, continued)**.** Consider the vector of continuous random variables (X, Y) discussed in Example 2.1.2. The space of the random vector is the unit circle with center at $(0, 0)$ as shown in Figure 2.1.2. To find the marginal distribution of X, fix x between -1 and 1 and then integrate out y from $-\sqrt{1 - x^2}$ to $\sqrt{1 - x^2}$ as the arrow shows on Figure 2.1.2. Hence, the marginal pdf of X is

$$f_X(x) = \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \frac{1}{\pi}\, dy = \frac{2}{\pi}\sqrt{1 - x^2}, \quad -1 < x < 1.$$

Although (X, Y) has a joint uniform distribution, the distribution of X is unimodal with peak at 0. This is not surprising. Since the joint distribution is uniform, from Figure 2.1.2 X is more likely to be near 0 than at either extreme -1 or 1. Because the joint pdf is symmetric in x and y, the marginal pdf of Y is the same as that of X. ■

Example 2.1.6. Let X_1 and X_2 have the joint pdf

$$f(x_1, x_2) = \begin{cases} x_1 + x_2 & 0 < x_1 < 1,\ 0 < x_2 < 1 \\ 0 & \text{elsewhere.} \end{cases}$$

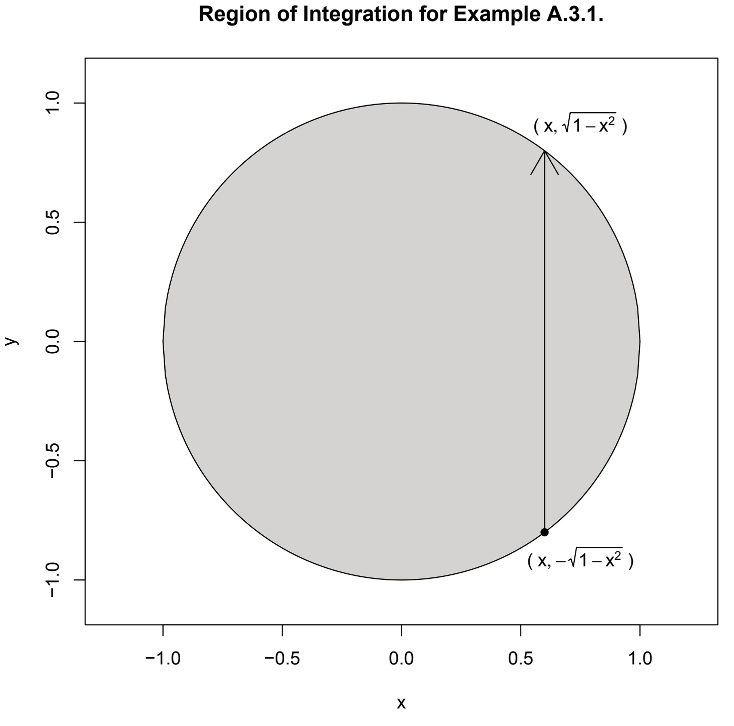

Figure 2.1.2: Region of integration for Example 2.1.5. It depicts the integration with respect to y at a fixed but arbitrary x.

Notice the space of the random vector is the interior of the square with vertices $(0,0),(1,0),(1,1)$ and $(0,1)$. The marginal pdf of X_1 is

$$f_1(x_1) = \int_0^1 (x_1 + x_2)\,dx_2 = x_1 + \tfrac{1}{2}, \quad 0 < x_1 < 1,$$

zero elsewhere, and the marginal pdf of X_2 is

$$f_2(x_2) = \int_0^1 (x_1 + x_2)\,dx_1 = \tfrac{1}{2} + x_2, \quad 0 < x_2 < 1,$$

zero elsewhere. A probability like $P(X_1 \le \frac{1}{2})$ can be computed from either $f_1(x_1)$ or $f(x_1, x_2)$ because

$$\int_0^{1/2}\int_0^1 f(x_1, x_2)\,dx_2 dx_1 = \int_0^{1/2} f_1(x_1)\,dx_1 = \tfrac{3}{8}.$$

Suppose, though, we want to find the probability $P(X_1 + X_2 \le 1)$. Notice that the region of integration is the interior of the triangle with vertices $(0,0),(1,0)$ and

$(0,1)$. The reader should sketch this region on the space of (X_1, X_2). Fixing x_1 and integrating with respect to x_2, we have

$$\begin{aligned} P(X_1 + X_2 \leq 1) &= \int_0^1 \left[\int_0^{1-x_1} (x_1 + x_2)\, dx_2 \right] dx_1 \\ &= \int_0^1 \left[x_1(1-x_1) + \frac{(1-x_1)^2}{2} \right] dx_1 \\ &= \int_0^1 \left(\frac{1}{2} - \frac{1}{2}x_1^2 \right) dx_1 = \frac{1}{3}. \end{aligned}$$

This latter probability is the volume under the surface $f(x_1, x_2) = x_1 + x_2$ above the set $\{(x_1, x_2) : 0 < x_1,\ x_1 + x_2 \leq 1\}$. ■

Example 2.1.7 (Example 2.1.3, Continued)**.** Recall that the random variables X and Y of Example 2.1.3 were the lifetimes of two batteries installed in an electrical component. The joint pdf of (X, Y) is sketched in Figure 2.1.1. Its space is the positive quadrant of R^2 so there are no constraints involving both x and y. Using the change-in-variable $w = y^2$, the marginal pdf of X is

$$f_X(x) = \int_0^\infty 4xye^{-(x^2+y^2)}\, dy = 2xe^{-x^2} \int_0^\infty e^{-w}\, dw = 2xe^{-x^2},$$

for $x > 0$. By the symmetry of x and y in the model, the pdf of Y is the same as that of X. To determine the median lifetime, θ, of these batteries, we need to solve

$$\frac{1}{2} = \int_0^\theta 2xe^{-x^2}\, dx = 1 - e^{-\theta^2},$$

where again we have made use of the change-in-variables $z = x^2$. Solving this equation, we obtain $\theta = \sqrt{\log 2} \approx 0.8326$. So 50% of the batteries have lifetimes exceeding 0.83 units. ■

2.1.2 Expectation

The concept of expectation extends in a straightforward manner. Let (X_1, X_2) be a random vector and let $Y = g(X_1, X_2)$ for some real-valued function; i.e., $g : R^2 \to R$. Then Y is a random variable and we could determine its expectation by obtaining the distribution of Y. But Theorem 1.8.1 is true for random vectors also. Note the proof we gave for this theorem involved the discrete case, and Exercise 2.1.12 shows its extension to the random vector case.

Suppose (X_1, X_2) is of the continuous type. Then $E(Y)$ exists if

$$\int_{-\infty}^\infty \int_{-\infty}^\infty |g(x_1, x_2)| f_{X_1,X_2}(x_1, x_2)\, dx_1 dx_2 < \infty.$$

Then

$$E(Y) = \int_{-\infty}^\infty \int_{-\infty}^\infty g(x_1, x_2) f_{X_1,X_2}(x_1, x_2)\, dx_1 dx_2. \tag{2.1.10}$$

Likewise if (X_1, X_2) is discrete, then $E(Y)$ exists if

$$\sum_{x_1}\sum_{x_2}|g(x_1, x_2)|p_{X_1,X_2}(x_1, x_2) < \infty.$$

Then

$$E(Y) = \sum_{x_1}\sum_{x_2} g(x_1, x_2)p_{X_1,X_2}(x_1, x_2). \tag{2.1.11}$$

We can now show that E is a linear operator.

Theorem 2.1.1. *Let (X_1, X_2) be a random vector. Let $Y_1 = g_1(X_1, X_2)$ and $Y_2 = g_2(X_1, X_2)$ be random variables whose expectations exist. Then for all real numbers k_1 and k_2,*

$$E(k_1Y_1 + k_2Y_2) = k_1E(Y_1) + k_2E(Y_2). \tag{2.1.12}$$

Proof: We prove it for the continuous case. The existence of the expected value of $k_1Y_1 + k_2Y_2$ follows directly from the triangle inequality and linearity of integrals; i.e.,

$$\begin{aligned} &\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |k_1g_1(x_1, x_2) + k_2g_1(x_1, x_2)|f_{X_1,X_2}(x_1, x_2)\, dx_1dx_2 \\ &\quad\leq\ |k_1|\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |g_1(x_1, x_2)|f_{X_1,X_2}(x_1, x_2)\, dx_1dx_2 \\ &\qquad + |k_2|\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |g_2(x_1, x_2)|f_{X_1,X_2}(x_1, x_2)\, dx_1dx_2 < \infty. \end{aligned}$$

By once again using linearity of the integral, we have

$$\begin{aligned} E(k_1Y_1 + k_2Y_2) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [k_1g_1(x_1, x_2) + k_2g_2(x_1, x_2)]f_{X_1,X_2}(x_1, x_2)\, dx_1dx_2 \\ &= k_1\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g_1(x_1, x_2)f_{X_1,X_2}(x_1, x_2)\, dx_1dx_2 \\ &\quad + k_2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g_2(x_1, x_2)f_{X_1,X_2}(x_1, x_2)\, dx_1dx_2 \\ &= k_1E(Y_1) + k_2E(Y_2), \end{aligned}$$

i.e., the desired result. ■

We also note that the expected value of any function $g(X_2)$ of X_2 can be found in two ways:

$$E(g(X_2)) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x_2)f(x_1, x_2)\, dx_1dx_2 = \int_{-\infty}^{\infty} g(x_2)f_{X_2}(x_2)\, dx_2,$$

the latter single integral being obtained from the double integral by integrating on x_1 first. The following example illustrates these ideas.

Example 2.1.8. Let X_1 and X_2 have the pdf

$$f(x_1, x_2) = \begin{cases} 8x_1x_2 & 0 < x_1 < x_2 < 1 \\ 0 & \text{elsewhere.} \end{cases}$$

Figure 2.1.3 shows the space for (X_1, X_2). Then

$$E(X_1X_2^2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1x_2^2 f(x_1, x_2)\, dx_1dx_2.$$

To compute the integration, as shown by the arrow on Figure 2.1.3, we fix x_2 and then integrate x_1 from 0 to x_2. We then integrate out x_2 from 0 to 1. Hence,

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1x_2^2 f(x_1, x_2) = \int_0^1 \left[\int_0^{x_2} 8x_1^2x_2^3\, dx_1\right] dx_2 = \int_0^1 \tfrac{8}{3}x_2^6\, dx_2 = \tfrac{8}{21}.$$

In addition,

$$E(X_2) = \int_0^1 \left[\int_0^{x_2} x_2(8x_1x_2)\, dx_1\right] dx_2 = \tfrac{4}{5}.$$

Since X_2 has the pdf $f_2(x_2) = 4x_2^3$, $0 < x_2 < 1$, zero elsewhere, the latter expectation can also be found by

$$E(X_2) = \int_0^1 x_2(4x_2^3)\, dx_2 = \tfrac{4}{5}.$$

Using Theorem 2.1.1,

$$\begin{aligned} E(7X_1X_2^2 + 5X_2) &= 7E(X_1X_2^2) + 5E(X_2) \\ &= (7)(\tfrac{8}{21}) + (5)(\tfrac{4}{5}) = \tfrac{20}{3}. \quad \blacksquare \end{aligned}$$

Example 2.1.9. Continuing with Example 2.1.8, suppose the random variable Y is defined by $Y = X_1/X_2$. We determine $E(Y)$ in two ways. The first way is by definition; i.e., find the distribution of Y and then determine its expectation. The cdf of Y, for $0 < y \le 1$, is

$$\begin{aligned} F_Y(y) &= P(Y \le y) = P(X_1 \le yX_2) = \int_0^1 \left[\int_0^{yx_2} 8x_1x_2\, dx_1\right] dx_2 \\ &= \int_0^1 4y^2x_2^3\, dx_2 = y^2. \end{aligned}$$

Hence, the pdf of Y is

$$f_Y(y) = F_Y'(y) = \begin{cases} 2y & 0 < y < 1 \\ 0 & \text{elsewhere,} \end{cases}$$

which leads to

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

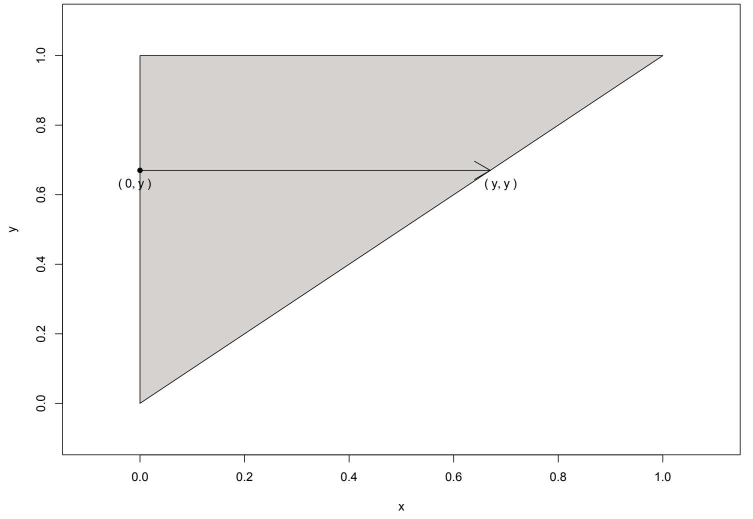

Figure 2.1.3: Region of integration for Example 2.1.8. The arrow depicts the integration with respect to x_1 at a fixed but arbitrary x_2.

For the second way, we make use of expression (2.1.10) and find $E(Y)$ directly by

$$\begin{aligned} E(Y) &= E\left(\frac{X_1}{X_2}\right) = \int_0^1 \left\{ \int_0^{x_2} \left(\frac{x_1}{x_2}\right) 8x_1x_2\, dx_1 \right\} dx_2 \\ &= \int_0^1 \frac{8}{3} x_2^3\, dx_2 = \frac{2}{3}. \quad \blacksquare \end{aligned}$$

We next define the moment generating function of a random vector.

Definition 2.1.2 (Moment Generating Function of a Random Vector)**.** *Let* $\mathbf{X} = (X_1, X_2)'$ *be a random vector. If* $E(e^{t_1X_1+t_2X_2})$ *exists for* $|t_1| < h_1$ *and* $|t_2| < h_2$*, where* h_1 *and* h_2 *are positive, it is denoted by* $M_{X_1,X_2}(t_1,t_2)$ *and is called the* **moment generating function** *(mgf) of* $\mathbf{X}$.

As in the one-variable case, if it exists, the mgf of a random vector uniquely determines the distribution of the random vector.

Let $\mathbf{t} = (t_1, t_2)'$. Then we can write the mgf of $\mathbf{X}$ as

$$M_{X_1,X_2}(\mathbf{t}) = E\left[e^{\mathbf{t}'\mathbf{X}}\right], \tag{2.1.13}$$

so it is quite similar to the mgf of a random variable. Also, the mgfs of X_1 and X_2 are immediately seen to be $M_{X_1,X_2}(t_1, 0)$ and $M_{X_1,X_2}(0, t_2)$, respectively. If there is no confusion, we often drop the subscripts on M.

Example 2.1.10. Let the continuous-type random variables X and Y have the joint pdf

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

The reader should sketch the space of (X, Y). The mgf of this joint distribution is

$$\begin{aligned} M(t_1, t_2) &= \int_0^\infty \left[\int_x^\infty \exp(t_1 x + t_2 y - y)\, dy \right] dx \\ &= \frac{1}{(1 - t_1 - t_2)(1 - t_2)}, \end{aligned}$$

provided that $t_1 + t_2 < 1$ and $t_2 < 1$. Furthermore, the moment-generating functions of the marginal distributions of X and Y are, respectively,

$$\begin{aligned} M(t_1, 0) &= \frac{1}{1 - t_1}, \quad t_1 < 1 \\ M(0, t_2) &= \frac{1}{(1 - t_2)^2}, \quad t_2 < 1. \end{aligned}$$

These moment-generating functions are, of course, respectively, those of the marginal probability density functions,

$$f_1(x) = \int_x^\infty e^{-y}\, dy = e^{-x}, \quad 0 < x < \infty,$$

zero elsewhere, and

$$f_2(y) = e^{-y} \int_0^y dx = y e^{-y}, \quad 0 < y < \infty,$$

zero elsewhere. ■

We also need to define the expected value of the random vector itself, but this is not a new concept because it is defined in terms of componentwise expectation:

Definition 2.1.3 (Expected Value of a Random Vector)**.** *Let* $\mathbf{X} = (X_1, X_2)'$ *be a random vector. Then the* **expected value** *of* $\mathbf{X}$ *exists if the expectations of* X_1 *and* X_2 *exist. If it exists, then the* **expected value** *is given by*

$$E[\mathbf{X}] = \begin{bmatrix} E(X_1) \\ E(X_2) \end{bmatrix}. \tag{2.1.14}$$

EXERCISES

2.1.1. Let $f(x_1, x_2) = 4x_1 x_2,\ 0 < x_1 < 1,\ 0 < x_2 < 1$, zero elsewhere, be the pdf of X_1 and X_2. Find $P(0 < X_1 < \frac{1}{2},\ \frac{1}{4} < X_2 < 1)$, $P(X_1 = X_2)$, $P(X_1 < X_2)$, and $P(X_1 \le X_2)$.
Hint: Recall that $P(X_1 = X_2)$ would be the volume under the surface $f(x_1, x_2) = 4x_1 x_2$ and above the line segment $0 < x_1 = x_2 < 1$ in the $x_1 x_2$-plane.

2.1.2. Let $A_1 = \{(x, y) : x \leq 2,\ y \leq 4\}$, $A_2 = \{(x, y) : x \leq 2,\ y \leq 1\}$, $A_3 = \{(x, y) : x \leq 0,\ y \leq 4\}$, and $A_4 = \{(x, y) : x \leq 0\ y \leq 1\}$ be subsets of the space $\mathcal{A}$ of two random variables X and Y, which is the entire two-dimensional plane. If $P(A_1) = \frac{7}{8}$, $P(A_2) = \frac{4}{8}$, $P(A_3) = \frac{3}{8}$, and $P(A_4) = \frac{2}{8}$, find $P(A_5)$, where $A_5 = \{(x, y) : 0 < x \leq 2,\ 1 < y \leq 4\}$.

2.1.3. Let $F(x, y)$ be the distribution function of X and Y. For all real constants $a < b,\ c < d$, show that $P(a < X \leq b,\ c < Y \leq d) = F(b, d) - F(b, c) - F(a, d) + F(a, c)$.

2.1.4. Show that the function $F(x, y)$ that is equal to 1 provided that $x + 2y \geq 1$, and that is equal to zero provided that $x + 2y < 1$, cannot be a distribution function of two random variables.
Hint: Find four numbers $a < b,\ c < d$, so that

$$F(b, d) - F(a, d) - F(b, c) + F(a, c)$$

is less than zero.

2.1.5. Given that the nonnegative function $g(x)$ has the property that

$$\int_0^\infty g(x)\,dx = 1,$$

show that

$$f(x_1, x_2) = \frac{2g(\sqrt{x_1^2 + x_2^2})}{\pi\sqrt{x_1^2 + x_2^2}},\ 0 < x_1 < \infty, 0 < x_2 < \infty,$$

zero elsewhere, satisfies the conditions for a pdf of two continuous-type random variables X_1 and X_2.
Hint: Use polar coordinates.

2.1.6. Consider Example 2.1.3.

(a) Show that $P(a < X < b, c < Y < d) = (\exp\{-a^2\} - \exp\{-b^2\})(\exp\{-c^2\} - \exp\{-d^2\})$.

(b) Using Part (a) and the notation in Example 2.1.3, show that $P[(X, Y) \in A] = 0.1879$ while $P[(X, Y) \in B] = 0.0026$.

(c) Show that the following R program computes $P(a < X < b, c < Y < d)$. Then use it to compute the probabilities in Part (b).

```
plifetime <- function(a,b,c,d)
     {(exp(-a^2) - exp(-b^2))*(exp(-c^2) - exp(-d^2))}
```

2.1.7. Let $f(x, y) = e^{-x-y}$, $0 < x < \infty$, $0 < y < \infty$, zero elsewhere, be the pdf of X and Y. Then if $Z = X + Y$, compute $P(Z \leq 0)$, $P(Z \leq 6)$, and, more generally, $P(Z \leq z)$, for $0 < z < \infty$. What is the pdf of Z?

2.1.8. Let X and Y have the pdf $f(x,y) = 1,\ 0 < x < 1,\ 0 < y < 1$, zero elsewhere. Find the cdf and pdf of the product $Z = XY$.

2.1.9. Let 13 cards be taken, at random and without replacement, from an ordinary deck of playing cards. If X is the number of spades in these 13 cards, find the pmf of X. If, in addition, Y is the number of hearts in these 13 cards, find the probability $P(X = 2,\ Y = 5)$. What is the joint pmf of X and Y?

2.1.10. Let the random variables X_1 and X_2 have the joint pmf described as follows:

(x_1,x_2)	$(0,0)$	$(0,1)$	$(0,2)$	$(1,0)$	$(1,1)$	$(1,2)$
$p(x_1,x_2)$	$\frac{2}{12}$	$\frac{3}{12}$	$\frac{2}{12}$	$\frac{2}{12}$	$\frac{2}{12}$	$\frac{1}{12}$

and $p(x_1,x_2)$ is equal to zero elsewhere.

(a) Write these probabilities in a rectangular array as in Example 2.1.4, recording each marginal pdf in the "margins."

(b) What is $P(X_1 + X_2 = 1)$?

2.1.11. Let X_1 and X_2 have the joint pdf $f(x_1,x_2) = 15x_1^2x_2,\ 0 < x_1 < x_2 < 1$, zero elsewhere. Find the marginal pdfs and compute $P(X_1 + X_2 \le 1)$.
Hint: Graph the space X_1 and X_2 and carefully choose the limits of integration in determining each marginal pdf.

2.1.12. Let X_1, X_2 be two random variables with the joint pmf $p(x_1,x_2)$, $(x_1,x_2) \in \mathcal{S}$, where $\mathcal{S}$ is the support of X_1, X_2. Let $Y = g(X_1,X_2)$ be a function such that

$$\sum\sum_{(x_1,x_2)\in\mathcal{S}} |g(x_1,x_2)|p(x_1,x_2) < \infty.$$

By following the proof of Theorem 1.8.1, show that

$$E(Y) = \sum\sum_{(x_1,x_2)\in\mathcal{S}} g(x_1,x_2)p(x_1,x_2).$$

2.1.13. Let X_1, X_2 be two random variables with the joint pmf $p(x_1,x_2) = (x_1 + x_2)/12$, for $x_1 = 1,2,\ \ x_2 = 1,2$, zero elsewhere. Compute $E(X_1)$, $E(X_1^2)$, $E(X_2)$, $E(X_2^2)$, and $E(X_1X_2)$. Is $E(X_1X_2) = E(X_1)E(X_2)$? Find $E(2X_1 - 6X_2^2 + 7X_1X_2)$.

2.1.14. Let X_1, X_2 be two random variables with joint pdf $f(x_1,x_2) = 4x_1x_2$, $0 < x_1 < 1$, $0 < x_2 < 1$, zero elsewhere. Compute $E(X_1)$, $E(X_1^2)$, $E(X_2)$, $E(X_2^2)$, and $E(X_1X_2)$. Is $E(X_1X_2) = E(X_1)E(X_2)$? Find $E(3X_2 - 2X_1^2 + 6X_1X_2)$.

2.1.15. Let X_1, X_2 be two random variables with joint pmf $p(x_1,x_2) = (1/2)^{x_1+x_2}$, for $1 \le x_i < \infty, i = 1,2$, where x_1 and x_2 are integers, zero elsewhere. Determine the joint mgf of X_1, X_2. Show that $M(t_1,t_2) = M(t_1,0)M(0,t_2)$.

2.1.16. Let X_1, X_2 be two random variables with joint pdf $f(x_1,x_2) = x_1\exp\{-x_2\}$, for $0 < x_1 < x_2 < \infty$, zero elsewhere. Determine the joint mgf of X_1, X_2. Does $M(t_1,t_2) = M(t_1,0)M(0,t_2)$?

2.1.17. Let X and Y have the joint pdf $f(x,y) = 6(1 - x - y),\ x + y < 1,\ 0 < x$, $0 < y$, zero elsewhere. Compute $P(2X + 3Y < 1)$ and $E(XY + 2X^2)$.

2.2 Transformations: Bivariate Random Variables

Let (X_1, X_2) be a random vector. Suppose we know the joint distribution of (X_1, X_2) and we seek the distribution of a transformation of (X_1, X_2), say, $Y = g(X_1, X_2)$. We may be able to obtain the cdf of Y. Another way is to use a transformation as we did for univariate random variables in Sections 1.6 and 1.7. In this section, we extend this theory to random vectors. It is best to discuss the discrete and continuous cases separately. We begin with the discrete case.

There are no essential difficulties involved in a problem like the following. Let $p_{X_1,X_2}(x_1, x_2)$ be the joint pmf of two discrete-type random variables X_1 and X_2 with $\mathcal{S}$ the (two-dimensional) set of points at which $p_{X_1,X_2}(x_1, x_2) > 0$; i.e., $\mathcal{S}$ is the support of (X_1, X_2). Let $y_1 = u_1(x_1, x_2)$ and $y_2 = u_2(x_1, x_2)$ define a one-to-one transformation that maps $\mathcal{S}$ onto $\mathcal{T}$. The joint pmf of the two new random variables $Y_1 = u_1(X_1, X_2)$ and $Y_2 = u_2(X_1, X_2)$ is given by

$$p_{Y_1,Y_2}(y_1, y_2) = \begin{cases} p_{X_1,X_2}[w_1(y_1, y_2), w_2(y_1, y_2)] & (y_1, y_2) \in \mathcal{T} \\ 0 & \text{elsewhere,} \end{cases}$$

where $x_1 = w_1(y_1, y_2)$, $x_2 = w_2(y_1, y_2)$ is the single-valued inverse of $y_1 = u_1(x_1, x_2)$, $y_2 = u_2(x_1, x_2)$. From this joint pmf $p_{Y_1,Y_2}(y_1, y_2)$ we may obtain the marginal pmf of Y_1 by summing on y_2 or the marginal pmf of Y_2 by summing on y_1.

In using this change of variable technique, it should be emphasized that we need two "new" variables to replace the two "old" variables. An example helps explain this technique.

Example 2.2.1. In a large metropolitan area during flu season, suppose that two strains of flu, A and B, are occurring. For a given week, let X_1 and X_2 be the respective number of reported cases of strains A and B with the joint pmf

$$p_{X_1,X_2}(x_1, x_2) = \frac{\mu_1^{x_1} \mu_2^{x_2} e^{-\mu_1} e^{-\mu_2}}{x_1! x_2!}, \quad x_1 = 0, 1, 2, 3, \ldots, \quad x_2 = 0, 1, 2, 3, \ldots,$$

and is zero elsewhere, where the parameters μ_1 and μ_2 are positive real numbers. Thus the space $\mathcal{S}$ is the set of points (x_1, x_2), where each of x_1 and x_2 is a non-negative integer. Further, repeatedly using the Maclaurin series for the exponential function,[3] we have

$$\begin{aligned} E(X_1) &= e^{-\mu_1} \sum_{x_1=0}^{\infty} x_1 \frac{\mu_1^{x_1}}{x_1!} e^{-\mu_2} \sum_{x_2=0}^{\infty} \frac{\mu_2^{x_2}}{x_2!} \\ &= e^{-\mu_1} \sum_{x_1=1}^{\infty} x_1 \mu_1 \frac{\mu_1^{x_1-1}}{(x_1-1)!} \cdot 1 = \mu_1. \end{aligned}$$

Thus μ_1 is the mean number of cases of Strain A flu reported during a week. Likewise, μ_2 is the mean number of cases of Strain B flu reported during a week.

[3]See for example the discussion on Taylor series in *Mathematical Comments* as referenced in the Preface.

A random variable of interest is $Y_1 = X_1 + X_2$; i.e., the total number of reported cases of A and B flu during a week. By Theorem 2.1.1, we know $E(Y_1) = \mu_1 + \mu_2$; however, we wish to determine the distribution of Y_1. If we use the change of variable technique, we need to define a second random variable Y_2. Because Y_2 is of no interest to us, let us choose it in such a way that we have a simple one-to-one transformation. For this example, we take $Y_2 = X_2$. Then $y_1 = x_1 + x_2$ and $y_2 = x_2$ represent a one-to-one transformation that maps $\mathcal{S}$ onto

$$\mathcal{T} = \{(y_1, y_2) : y_2 = 0, 1, \ldots, y_1 \quad \text{and} \quad y_1 = 0, 1, 2, \ \ldots\}.$$

Note that if $(y_1, y_2) \in \mathcal{T}$, then $0 \leq y_2 \leq y_1$. The inverse functions are given by $x_1 = y_1 - y_2$ and $x_2 = y_2$. Thus the joint pmf of Y_1 and Y_2 is

$$p_{Y_1,Y_2}(y_1, y_2) = \frac{\mu_1^{y_1-y_2}\mu_2^{y_2}e^{-\mu_1-\mu_2}}{(y_1-y_2)!y_2!}, \quad (y_1, y_2) \in \mathcal{T},$$

and is zero elsewhere. Consequently, the marginal pmf of Y_1 is given by

$$\begin{aligned} p_{Y_1}(y_1) &= \sum_{y_2=0}^{y_1} p_{Y_1,Y_2}(y_1, y_2) \\ &= \frac{e^{-\mu_1-\mu_2}}{y_1!}\sum_{y_2=0}^{y_1}\frac{y_1!}{(y_1-y_2)!y_2!}\mu_1^{y_1-y_2}\mu_2^{y_2} \\ &= \frac{(\mu_1+\mu_2)^{y_1}e^{-\mu_1-\mu_2}}{y_1!}, \quad y_1 = 0, 1, 2, \ldots, \end{aligned}$$

and is zero elsewhere, where the third equality follows from the binomial expansion. ■

For the continuous case we begin with an example that illustrates the cdf technique.

Example 2.2.2. Consider an experiment in which a person chooses at random a point (X_1, X_2) from the unit square $\mathcal{S} = \{(x_1, x_2) : 0 < x_1 < 1, \ 0 < x_2 < 1\}$. Suppose that our interest is not in X_1 or in X_2 but in $Z = X_1 + X_2$. Once a suitable probability model has been adopted, we shall see how to find the pdf of Z. To be specific, let the nature of the random experiment be such that it is reasonable to *assume* that the distribution of probability over the unit square is uniform. Then the pdf of X_1 and X_2 may be written

$$f_{X_1,X_2}(x_1, x_2) = \begin{cases} 1 & 0 < x_1 < 1, \ \ 0 < x_2 < 1 \\ 0 & \text{elsewhere,} \end{cases} \tag{2.2.1}$$

and this describes the probability model. Now let the cdf of Z be denoted by $F_Z(z) = P(X_1 + X_2 \leq z)$. Then

$$F_Z(z) = \begin{cases} 0 & z < 0 \\ \int_0^z \int_0^{z-x_1} dx_2 dx_1 = \frac{z^2}{2} & 0 \leq z < 1 \\ 1 - \int_{z-1}^1 \int_{z-x_1}^1 dx_2 dx_1 = 1 - \frac{(2-z)^2}{2} & 1 \leq z < 2 \\ 1 & 2 \leq z. \end{cases}$$

Since $F'_Z(z)$ exists for all values of z, the pmf of Z may then be written

$$f_Z(z) = \begin{cases} z & 0 < z < 1 \\ 2 - z & 1 \le z < 2 \\ 0 & \text{elsewhere.} \quad \blacksquare \end{cases} \tag{2.2.2}$$

In the last example, we used the cdf technique to find the distribution of the transformed random vector. Recall in Chapter 1, Theorem 1.7.1 gave a transformation technique to directly determine the pdf of the transformed random variable for one-to-one transformations. As discussed in Section 4.1 of the accompanying resource *Mathematical Comments*,[4] this is based on the change-in-variable technique for univariate integration. Further Section 4.2 of this resource shows that a similar change-in-variable technique exists for multiple integration. We now discuss in general the transformation technique for the continuous case based on this theory.

Let (X_1, X_2) have a jointly continuous distribution with pdf $f_{X_1,X_2}(x_1, x_2)$ and support set $\mathcal{S}$. Consider the transformed random vector $(Y_1, Y_2) = T(X_1, X_2)$ where T is a one-to-one continuous transformation. Let $\mathcal{T} = T(\mathcal{S})$ denote the support of (Y_1, Y_2). The transformation is depicted in Figure 2.2.1. Rewrite the transformation in terms of its components as $(Y_1, Y_2) = T(X_1, X_2) = (u_1(X_1, X_2), u_2(X_1, X_2))$, where the functions $y_1 = u_1(x_1, x_2)$ and $y_2 = u_2(x_1, x_2)$ define T. Since the transformation is one-to-one, the inverse transformation T^{-1} exists. We write it as $x_1 = w_1(y_1, y_2),\ x_2 = w_2(y_1, y_2)$. Finally, we need the **Jacobian** of the transformation which is the determinant of order 2 given by

$$J = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{vmatrix}.$$

Note that J plays the role of dx/dy in the univariate case. We assume that these first-order partial derivatives are continuous and that the Jacobian J is not identically equal to zero in $\mathcal{T}$.

Let B be any region[5] in $\mathcal{T}$ and let $A = T^{-1}(B)$ as shown in Figure 2.2.1. Because the transformation T is one-to-one, $P[(X_1, X_2) \in A] = P[T(X_1, X_2) \in T(A)] = P[(Y_1, Y_2) \in B]$. Then based on the change-in-variable technique, cited above, we have

$$\begin{aligned} P[(X_1, X_2) \in A] &= \iint\limits_A f_{X_1,X_2}(x_1, x_2)\, dx_1 dy_2 \\ &= \iint\limits_{T(A)} f_{X_1,X_2}[T^{-1}(y_1, y_2)]|J|\, dy_1 dy_2 \\ &= \iint\limits_B f_{X_1,X_2}[w_1(y_1, y_2), w_2(y_1, y_2)]|J|\, dy_1 dy_2. \end{aligned}$$

[4]See the reference for *Mathematical Comments* in the Preface.
[5]Technically an event in the support of (Y_1, Y_2).

Since B is arbitrary, the last integrand must be the joint pdf of (Y_1, Y_2). That is the pdf of (Y_1, Y_2) is

$$f_{Y_1,Y_2}(y_1, y_2) = \begin{cases} f_{X_1,X_2}[w_1(y_1, y_2), w_2(y_1, y_2)]|J| & (y_1, y_2) \in \mathcal{T} \\ 0 & \text{elsewhere.} \end{cases} \tag{2.2.3}$$

Several examples of this result are given next.

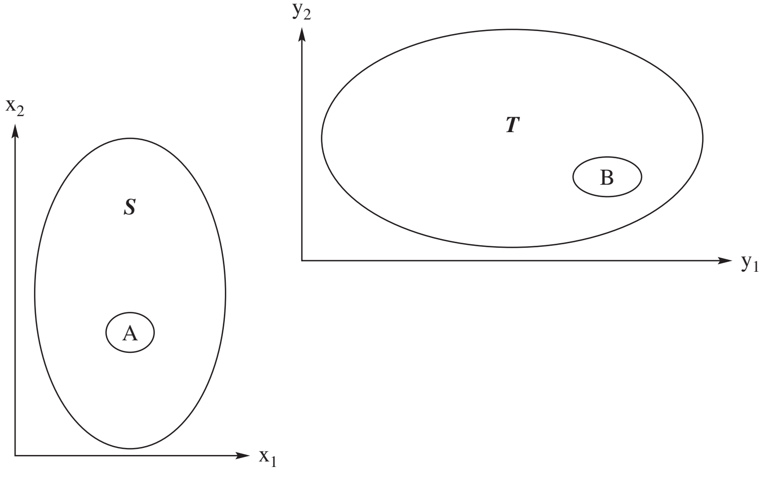

Figure 2.2.1: A general sketch of the supports of (X_1, X_2), $(\mathcal{S})$, and (Y_1, Y_2), $(\mathcal{T})$.

Example 2.2.3. Reconsider Example 2.2.2, where (X_1, X_2) have the uniform distribution over the unit square with the pdf given in expression (2.2.1). The support of (X_1, X_2) is the set $\mathcal{S} = \{(x_1, x_2) : 0 < x_1 < 1, 0 < x_2 < 1\}$ as depicted in Figure 2.2.2.

Suppose $Y_1 = X_1 + X_2$ and $Y_2 = X_1 - X_2$. The transformation is given by

$$\begin{aligned} y_1 &= u_1(x_1, x_2) = x_1 + x_2 \\ y_2 &= u_2(x_1, x_2) = x_1 - x_2. \end{aligned}$$

This transformation is one-to-one. We first determine the set $\mathcal{T}$ in the y_1y_2-plane that is the mapping of $\mathcal{S}$ under this transformation. Now

$$\begin{aligned} x_1 &= w_1(y_1, y_2) = \tfrac{1}{2}(y_1 + y_2) \\ x_2 &= w_2(y_1, y_2) = \tfrac{1}{2}(y_1 - y_2). \end{aligned}$$

To determine the set $\mathcal{S}$ in the y_1y_2-plane onto which $\mathcal{T}$ is mapped under the transformation, note that the boundaries of $\mathcal{S}$ are transformed as follows into the boundaries

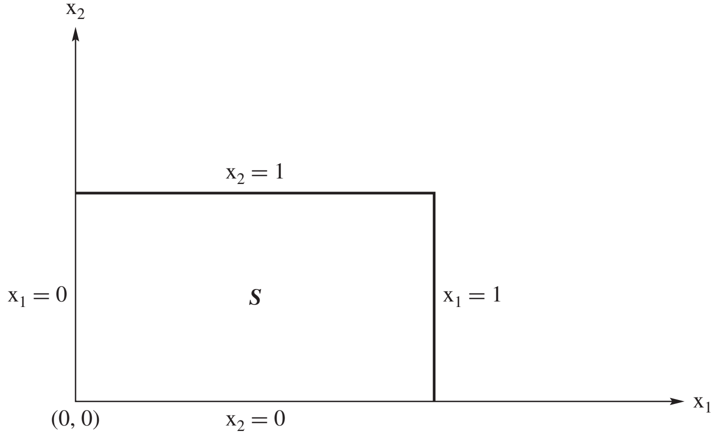

Figure 2.2.2: The support of (X_1, X_2) of Example 2.2.3.

of $\mathcal{T}$:

$$\begin{array}{ccc} x_1 = 0 & \text{into} & 0 = \frac{1}{2}(y_1 + y_2) \\ x_1 = 1 & \text{into} & 1 = \frac{1}{2}(y_1 + y_2) \\ x_2 = 0 & \text{into} & 0 = \frac{1}{2}(y_1 - y_2) \\ x_2 = 1 & \text{into} & 1 = \frac{1}{2}(y_1 - y_2). \end{array}$$

Accordingly, $\mathcal{T}$ is shown in Figure 2.2.3. Next, the Jacobian is given by

$$J = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{vmatrix} = \begin{vmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{vmatrix} = -\frac{1}{2}.$$

Although we suggest transforming the boundaries of $\mathcal{S}$, others might want to use the inequalities

$$0 < x_1 < 1 \quad \text{and} \quad 0 < x_2 < 1$$

directly. These four inequalities become

$$0 < \tfrac{1}{2}(y_1 + y_2) < 1 \quad \text{and} \quad 0 < \tfrac{1}{2}(y_1 - y_2) < 1.$$

It is easy to see that these are equivalent to

$$-y_1 < y_2, \quad y_2 < 2 - y_1, \quad y_2 < y_1 \quad y_1 - 2 < y_2;$$

and they define the set $\mathcal{T}$.

Hence, the joint pdf of (Y_1, Y_2) is given by

$$f_{Y_1,Y_2}(y_1, y_2) = \begin{cases} f_{X_1,X_2}[\frac{1}{2}(y_1 + y_2), \frac{1}{2}(y_1 - y_2)]|J| = \frac{1}{2} & (y_1, y_2) \in \mathcal{T} \\ 0 & \text{elsewhere.} \end{cases}$$

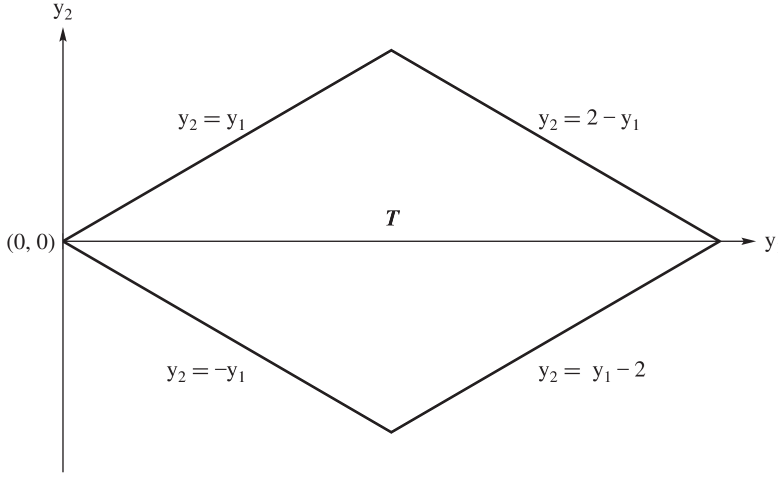

Figure 2.2.3: The support of (Y_1, Y_2) of Example 2.2.3.

The marginal pdf of Y_1 is given by

$$f_{Y_1}(y_1) = \int_{-\infty}^{\infty} f_{Y_1,Y_2}(y_1, y_2)\, dy_2.$$

If we refer to Figure 2.2.3, we can see that

$$f_{Y_1}(y_1) = \begin{cases} \int_{-y_1}^{y_1} \frac{1}{2}\, dy_2 = y_1 & 0 < y_1 \le 1 \\ \int_{y_1-2}^{2-y_1} \frac{1}{2}\, dy_2 = 2 - y_1 & 1 < y_1 < 2 \\ 0 & \text{elsewhere,} \end{cases}$$

which agrees with expression (2.2.2) of Example 2.2.2. In a similar manner, the marginal pdf $f_{Y_2}(y_2)$ is given by

$$f_{Y_2}(y_2) = \begin{cases} \int_{-y_2}^{y_2+2} \frac{1}{2}\, dy_1 = y_2 + 1 & -1 < y_2 \le 0 \\ \int_{y_2}^{2-y_2} \frac{1}{2}\, dy_1 = 1 - y_2 & 0 < y_2 < 1 \\ 0 & \text{elsewhere.} \end{cases} \blacksquare$$

Example 2.2.4. Let $Y_1 = \frac{1}{2}(X_1 - X_2)$, where X_1 and X_2 have the joint pdf

$$f_{X_1,X_2}(x_1, x_2) = \begin{cases} \frac{1}{4} \exp\left(-\frac{x_1+x_2}{2}\right) & 0 < x_1 < \infty,\ \ 0 < x_2 < \infty \\ 0 & \text{elsewhere.} \end{cases}$$

Let $Y_2 = X_2$ so that $y_1 = \frac{1}{2}(x_1 - x_2),\ y_2 = x_2$ or, equivalently, $x_1 = 2y_1 + y_2,\ x_2 = y_2$, define a one-to-one transformation from $\mathcal{S} = \{(x_1, x_2) : 0 < x_1 < \infty, 0 < x_2 < \infty\}$ onto $\mathcal{T} = \{(y_1, y_2) : -2y_1 < y_2 \text{ and } 0 < y_2 < \infty,\ -\infty < y_1 < \infty\}$. The Jacobian of the transformation is

$$J = \begin{vmatrix} 2 & 1 \\ 0 & 1 \end{vmatrix} = 2;$$

hence the joint pdf of Y_1 and Y_2 is

$$f_{Y_1,Y_2}(y_1,y_2) = \begin{cases} \frac{|2|}{4}\, e^{-y_1-y_2} & (y_1,y_2) \in \mathcal{T} \\ 0 & \text{elsewhere.} \end{cases}$$

Thus the pdf of Y_1 is given by

$$f_{Y_1}(y_1) = \begin{cases} \int_{-2y_1}^{\infty} \frac{1}{2} e^{-y_1-y_2}\, dy_2 = \frac{1}{2}\, e^{y_1} & -\infty < y_1 < 0 \\ \\ \int_{0}^{\infty} \frac{1}{2}\, e^{-y_1-y_2}\, dy_2 = \frac{1}{2}\, e^{-y_1} & 0 \le y_1 < \infty, \end{cases}$$

or

$$f_{Y_1}(y_1) = \tfrac{1}{2}\, e^{-|y_1|}, \quad -\infty < y_1 < \infty. \tag{2.2.4}$$

Recall from expression (1.9.20) of Chapter 1 that Y_1 has the Laplace distribution. This pdf is also frequently called the **double exponential** pdf. ■

Example 2.2.5. Let X_1 and X_2 have the joint pdf

$$f_{X_1,X_2}(x_1,x_2) = \begin{cases} 10x_1x_2^2 & 0 < x_1 < x_2 < 1 \\ 0 & \text{elsewhere.} \end{cases}$$

Suppose $Y_1 = X_1/X_2$ and $Y_2 = X_2$. Hence, the inverse transformation is $x_1 = y_1y_2$ and $x_2 = y_2$, which has the Jacobian

$$J = \begin{vmatrix} y_2 & y_1 \\ 0 & 1 \end{vmatrix} = y_2.$$

The inequalities defining the support $\mathcal{S}$ of (X_1, X_2) become

$$0 < y_1y_2, \;\; y_1y_2 < y_2, \;\text{ and }\; y_2 < 1.$$

These inequalities are equivalent to

$$0 < y_1 < 1 \;\text{ and }\; 0 < y_2 < 1,$$

which defines the support set $\mathcal{T}$ of (Y_1, Y_2). Hence, the joint pdf of (Y_1, Y_2) is

$$f_{Y_1,Y_2}(y_1,y_2) = 10y_1y_2y_2^2|y_2| = 10y_1y_2^4\,, \quad (y_1,y_2) \in \mathcal{T}.$$

The marginal pdfs are

$$f_{Y_1}(y_1) = \int_0^1 10y_1y_2^4\, dy_2 = 2y_1, \quad 0 < y_1 < 1,$$

zero elsewhere, and

$$f_{Y_2}(y_2) = \int_0^1 10y_1y_2^4\, dy_1 = 5y_2^4, \quad 0 < y_1 < 1,$$

zero elsewhere. ■

In addition to the change-of-variable and cdf techniques for finding distributions of functions of random variables, there is another method, called the **moment generating function (mgf) technique**, which works well for linear functions of random variables. In Subsection 2.1.2, we pointed out that if $Y = g(X_1, X_2)$, then $E(Y)$, if it exists, could be found by

$$E(Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x_1, x_2) f_{X_1,X_2}(x_1, x_2)\, dx_1 dx_2$$

in the continuous case, with summations replacing integrals in the discrete case. Certainly, that function $g(X_1, X_2)$ could be $\exp\{tu(X_1, X_2)\}$, so that in reality we would be finding the mgf of the function $Z = u(X_1, X_2)$. If we could then recognize this mgf as belonging to a certain distribution, then Z would have that distribution. We give two illustrations that demonstrate the power of this technique by reconsidering Examples 2.2.1 and 2.2.4.

Example 2.2.6 (Continuation of Example 2.2.1)**.** Here X_1 and X_2 have the joint pmf

$$p_{X_1,X_2}(x_1, x_2) = \begin{cases} \frac{\mu_1^{x_1} \mu_2^{x_2} e^{-\mu_1} e^{-\mu_2}}{x_1! x_2!} & x_1 = 0, 1, 2, 3, \ldots, \quad x_2 = 0, 1, 2, 3, \ldots \\ 0 & \text{elsewhere,} \end{cases}$$

where μ_1 and μ_2 are fixed positive real numbers. Let $Y = X_1 + X_2$ and consider

$$\begin{aligned} E(e^{tY}) &= \sum_{x_1=0}^{\infty} \sum_{x_2=0}^{\infty} e^{t(x_1+x_2)} p_{X_1,X_2}(x_1, x_2) \\ &= \sum_{x_1=0}^{\infty} e^{tx_1} \frac{\mu^{x_1} e^{-\mu_1}}{x_1!} \sum_{x_2=0}^{\infty} e^{tx_2} \frac{\mu^{x_2} e^{-\mu_2}}{x_2!} \\ &= \left[e^{-\mu_1} \sum_{x_1=0}^{\infty} \frac{(e^t \mu_1)^{x_1}}{x_1!} \right] \left[e^{-\mu_2} \sum_{x_2=0}^{\infty} \frac{(e^t \mu_2)^{x_2}}{x_2!} \right] \\ &= \left[e^{\mu_1(e^t-1)} \right] \left[e^{\mu_2(e^t-1)} \right] \\ &= e^{(\mu_1+\mu_2)(e^t-1)}. \end{aligned}$$

Notice that the factors in the brackets in the next-to-last equality are the mgfs of X_1 and X_2, respectively. Hence, the mgf of Y is the same as that of X_1 except μ_1 has been replaced by $\mu_1 + \mu_2$. Therefore, by the uniqueness of mgfs, the pmf of Y must be

$$p_Y(y) = e^{-(\mu_1+\mu_2)} \frac{(\mu_1 + \mu_2)^y}{y!}, \quad y = 0, 1, 2, \ldots,$$

which is the same pmf that was obtained in Example 2.2.1. ■

Example 2.2.7 (Continuation of Example 2.2.4)**.** Here X_1 and X_2 have the joint pdf

$$f_{X_1,X_2}(x_1, x_2) = \begin{cases} \frac{1}{4} \exp\left(-\frac{x_1+x_2}{2}\right) & 0 < x_1 < \infty, \quad 0 < x_2 < \infty \\ 0 & \text{elsewhere.} \end{cases}$$

So the mgf of $Y = (1/2)(X_1 - X_2)$ is given by

$$\begin{aligned} E(e^{tY}) &= \int_0^\infty \int_0^\infty e^{t(x_1-x_2)/2} \frac{1}{4} e^{-(x_1+x_2)/2}\, dx_1 dx_2 \\ &= \left[\int_0^\infty \frac{1}{2} e^{-x_1(1-t)/2}\, dx_1\right] \left[\int_0^\infty \frac{1}{2} e^{-x_2(1+t)/2}\, dx_2\right] \\ &= \left[\frac{1}{1-t}\right]\left[\frac{1}{1+t}\right] = \frac{1}{1-t^2} \end{aligned}$$

provided that $1 - t > 0$ and $1 + t > 0$; i.e., $-1 < t < 1$. However, the mgf of a Laplace distribution with pdf (1.9.20) is

$$\begin{aligned} \int_{-\infty}^\infty e^{tx} \frac{e^{-|x|}}{2}\, dx &= \int_{-\infty}^0 \frac{e^{(1+t)x}}{2}\, dx + \int_0^\infty \frac{e^{(t-1)x}}{2}\, dx \\ &= \frac{1}{2(1+t)} + \frac{1}{2(1-t)} = \frac{1}{1-t^2}, \end{aligned}$$

provided $-1 < t < 1$. Thus, by the uniqueness of mgfs, Y has a Laplace distribution with pdf (1.9.20). ■

EXERCISES

2.2.1. If $p(x_1, x_2) = (\frac{2}{3})^{x_1+x_2}(\frac{1}{3})^{2-x_1-x_2}$, $(x_1, x_2) = (0,0), (0,1), (1,0), (1,1)$, zero elsewhere, is the joint pmf of X_1 and X_2, find the joint pmf of $Y_1 = X_1 - X_2$ and $Y_2 = X_1 + X_2$.

2.2.2. Let X_1 and X_2 have the joint pmf $p(x_1, x_2) = x_1x_2/36$, $x_1 = 1, 2, 3$ and $x_2 = 1, 2, 3$, zero elsewhere. Find first the joint pmf of $Y_1 = X_1X_2$ and $Y_2 = X_2$, and then find the marginal pmf of Y_1.

2.2.3. Let X_1 and X_2 have the joint pdf $h(x_1, x_2) = 2e^{-x_1-x_2}$, $0 < x_1 < x_2 < \infty$, zero elsewhere. Find the joint pdf of $Y_1 = 2X_1$ and $Y_2 = X_2 - X_1$.

2.2.4. Let X_1 and X_2 have the joint pdf $h(x_1, x_2) = 8x_1x_2$, $0 < x_1 < x_2 < 1$, zero elsewhere. Find the joint pdf of $Y_1 = X_1/X_2$ and $Y_2 = X_2$.
Hint: Use the inequalities $0 < y_1y_2 < y_2 < 1$ in considering the mapping from $\mathcal{S}$ onto $\mathcal{T}$.

2.2.5. Let X_1 and X_2 be continuous random variables with the joint probability density function $f_{X_1,X_2}(x_1, x_2)$, $-\infty < x_i < \infty$, $i = 1, 2$. Let $Y_1 = X_1 + X_2$ and $Y_2 = X_2$.

(a) Find the joint pdf f_{Y_1,Y_2}.

(b) Show that

$$f_{Y_1}(y_1) = \int_{-\infty}^\infty f_{X_1,X_2}(y_1 - y_2, y_2)\, dy_2, \tag{2.2.5}$$

which is sometimes called the **convolution formula**.

2.2.6. Suppose X_1 and X_2 have the joint pdf $f_{X_1,X_2}(x_1, x_2) = e^{-(x_1+x_2)}$, $0 < x_i < \infty$, $i = 1, 2$, zero elsewhere.

(a) Use formula (2.2.5) to find the pdf of $Y_1 = X_1 + X_2$.

(b) Find the mgf of Y_1.

2.2.7. Use the formula (2.2.5) to find the pdf of $Y_1 = X_1 + X_2$, where X_1 and X_2 have the joint pdf $f_{X_1,X_2}(x_1, x_2) = 2e^{-(x_1+x_2)}$, $0 < x_1 < x_2 < \infty$, zero elsewhere.

2.2.8. Suppose X_1 and X_2 have the joint pdf

$$f(x_1, x_2) = \begin{cases} e^{-x_1}e^{-x_2} & x_1 > 0, x_2 > 0 \\ 0 & \text{elsewhere.} \end{cases}$$

For constants $w_1 > 0$ and $w_2 > 0$, let $W = w_1X_1 + w_2X_2$.

(a) Show that the pdf of W is

$$f_W(w) = \begin{cases} \frac{1}{w_1-w_2}(e^{-w/w_1} - e^{-w/w_2}) & w > 0 \\ 0 & \text{elsewhere.} \end{cases}$$

(b) Verify that $f_W(w) > 0$ for $w > 0$.

(c) Note that the pdf $f_W(w)$ has an indeterminate form when $w_1 = w_2$. Rewrite $f_W(w)$ using h defined as $w_1 - w_2 = h$. Then use l'Hôpital's rule to show that when $w_1 = w_2$, the pdf is given by $f_W(w) = (w/w_1^2)\exp\{-w/w_1\}$ for $w > 0$ and zero elsewhere.

2.3 Conditional Distributions and Expectations

In Section 2.1 we introduced the joint probability distribution of a pair of random variables. We also showed how to recover the individual (marginal) distributions for the random variables from the joint distribution. In this section, we discuss conditional distributions, i.e., the distribution of one of the random variables when the other has assumed a specific value. We discuss this first for the discrete case, which follows easily from the concept of conditional probability presented in Section 1.4.

Let X_1 and X_2 denote random variables of the discrete type, which have the joint pmf $p_{X_1,X_2}(x_1, x_2)$ that is positive on the support set $\mathcal{S}$ and is zero elsewhere. Let $p_{X_1}(x_1)$ and $p_{X_2}(x_2)$ denote, respectively, the marginal probability mass functions of X_1 and X_2. Let x_1 be a point in the support of X_1; hence, $p_{X_1}(x_1) > 0$. Using the definition of conditional probability, we have

$$P(X_2 = x_2|X_1 = x_1) = \frac{P(X_1 = x_1, X_2 = x_2)}{P(X_1 = x_1)} = \frac{p_{X_1,X_2}(x_1, x_2)}{p_{X_1}(x_1)}$$

for all x_2 in the support $\mathcal{S}_{X_2}$ of X_2. Define this function as

$$p_{X_2|X_1}(x_2|x_1) = \frac{p_{X_1,X_2}(x_1, x_2)}{p_{X_1}(x_1)}, \quad x_2 \in S_{X_2}. \tag{2.3.1}$$

For any fixed x_1 with $p_{X_1}(x_1) > 0$, this function $p_{X_2|X_1}(x_2|x_1)$ satisfies the conditions of being a pmf of the discrete type because $p_{X_2|X_1}(x_2|x_1)$ is nonnegative and

$$\begin{aligned}\sum_{x_2} p_{X_2|X_1}(x_2|x_1) &= \sum_{x_2} \frac{p_{X_1,X_2}(x_1,x_2)}{p_{X_1}(x_1)} \\ &= \frac{1}{p_{X_1}(x_1)} \sum_{x_2} p_{X_1,X_2}(x_1,x_2) = \frac{p_{X_1}(x_1)}{p_{X_1}(x_1)} = 1.\end{aligned}$$

We call $p_{X_2|X_1}(x_2|x_1)$ the **conditional pmf** of the discrete type of random variable X_2, given that the discrete type of random variable $X_1 = x_1$. In a similar manner, provided $x_2 \in \mathcal{S}_{X_2}$, we define the symbol $p_{X_1|X_2}(x_1|x_2)$ by the relation

$$p_{X_1|X_2}(x_1|x_2) = \frac{p_{X_1,X_2}(x_1,x_2)}{p_{X_2}(x_2)}, \quad x_1 \in S_{X_1},$$

and we call $p_{X_1|X_2}(x_1|x_2)$ the conditional pmf of the discrete type of random variable X_1, given that the discrete type of random variable $X_2 = x_2$. We often abbreviate $p_{X_1|X_2}(x_1|x_2)$ by $p_{1|2}(x_1|x_2)$ and $p_{X_2|X_1}(x_2|x_1)$ by $p_{2|1}(x_2|x_1)$. Similarly, $p_1(x_1)$ and $p_2(x_2)$ are used to denote the respective marginal pmfs.

Now let X_1 and X_2 denote random variables of the continuous type and have the joint pdf $f_{X_1,X_2}(x_1,x_2)$ and the marginal probability density functions $f_{X_1}(x_1)$ and $f_{X_2}(x_2)$, respectively. We use the results of the preceding paragraph to motivate a definition of a conditional pdf of a continuous type of random variable. When $f_{X_1}(x_1) > 0$, we define the symbol $f_{X_2|X_1}(x_2|x_1)$ by the relation

$$f_{X_2|X_1}(x_2|x_1) = \frac{f_{X_1,X_2}(x_1,x_2)}{f_{X_1}(x_1)}. \tag{2.3.2}$$

In this relation, x_1 is to be thought of as having a fixed (but any fixed) value for which $f_{X_1}(x_1) > 0$. It is evident that $f_{X_2|X_1}(x_2|x_1)$ is nonnegative and that

$$\begin{aligned}\int_{-\infty}^{\infty} f_{X_2|X_1}(x_2|x_1)\,dx_2 &= \int_{-\infty}^{\infty} \frac{f_{X_1,X_2}(x_1,x_2)}{f_{X_1}(x_1)}\,dx_2 \\ &= \frac{1}{f_{X_1}(x_1)} \int_{-\infty}^{\infty} f_{X_1,X_2}(x_1,x_2)\,dx_2 \\ &= \frac{1}{f_{X_1}(x_1)} f_{X_1}(x_1) = 1.\end{aligned}$$

That is, $f_{X_2|X_1}(x_2|x_1)$ has the properties of a pdf of one continuous type of random variable. It is called the **conditional pdf** of the continuous type of random variable X_2, given that the continuous type of random variable X_1 has the value x_1. When $f_{X_2}(x_2) > 0$, the conditional pdf of the continuous random variable X_1, given that the continuous type of random variable X_2 has the value x_2, is defined by

$$f_{X_1|X_2}(x_1|x_2) = \frac{f_{X_1,X_2}(x_1,x_2)}{f_{X_2}(x_2)}, \quad f_{X_2}(x_2) > 0.$$

We often abbreviate these conditional pdfs by $f_{1|2}(x_1|x_2)$ and $f_{2|1}(x_2|x_1)$, respectively. Similarly, $f_1(x_1)$ and $f_2(x_2)$ are used to denote the respective marginal pdfs.

Since each of $f_{2|1}(x_2|x_1)$ and $f_{1|2}(x_1|x_2)$ is a pdf of one random variable, each has all the properties of such a pdf. Thus we can compute probabilities and mathematical expectations. If the random variables are of the continuous type, the probability

$$P(a < X_2 < b|X_1 = x_1) = \int_a^b f_{2|1}(x_2|x_1)\,dx_2$$

is called "the conditional probability that $a < X_2 < b$, given that $X_1 = x_1$." If there is no ambiguity, this may be written in the form $P(a < X_2 < b|x_1)$. Similarly, the conditional probability that $c < X_1 < d$, given $X_2 = x_2$, is

$$P(c < X_1 < d|X_2 = x_2) = \int_c^d f_{1|2}(x_1|x_2)\,dx_1.$$

If $u(X_2)$ is a function of X_2, the **conditional expectation** of $u(X_2)$, given that $X_1 = x_1$, if it exists, is given by

$$E[u(X_2)|x_1] = \int_{-\infty}^{\infty} u(x_2) f_{2|1}(x_2|x_1)\,dx_2.$$

Note that $E[u(X_2)|x_1]$ is a function of x_1. If they do exist, then $E(X_2|x_1)$ is the mean and $E\{[X_2 - E(X_2|x_1)]^2|x_1\}$ is the **conditional variance** of the conditional distribution of X_2, given $X_1 = x_1$, which can be written more simply as $\mathrm{Var}(X_2|x_1)$. It is convenient to refer to these as the "conditional mean" and the "conditional variance" of X_2, given $X_1 = x_1$. Of course, we have

$$\mathrm{Var}(X_2|x_1) = E(X_2^2|x_1) - [E(X_2|x_1)]^2$$

from an earlier result. In a like manner, the conditional expectation of $u(X_1)$, given $X_2 = x_2$, if it exists, is given by

$$E[u(X_1)|x_2] = \int_{-\infty}^{\infty} u(x_1) f_{1|2}(x_1|x_2)\,dx_1.$$

With random variables of the discrete type, these conditional probabilities and conditional expectations are computed by using summation instead of integration. An illustrative example follows.

Example 2.3.1. Let X_1 and X_2 have the joint pdf

$$f(x_1, x_2) = \begin{cases} 2 & 0 < x_1 < x_2 < 1 \\ 0 & \text{elsewhere.} \end{cases}$$

Then the marginal probability density functions are, respectively,

$$f_1(x_1) = \begin{cases} \int_{x_1}^1 2\,dx_2 = 2(1 - x_1) & 0 < x_1 < 1 \\ 0 & \text{elsewhere,} \end{cases}$$

and

$$f_2(x_2) = \begin{cases} \int_0^{x_2} 2\,dx_1 = 2x_2 & 0 < x_2 < 1 \\ 0 & \text{elsewhere.} \end{cases}$$

The conditional pdf of X_1, given $X_2 = x_2$, $0 < x_2 < 1$, is

$$f_{1|2}(x_1|x_2) = \begin{cases} \frac{2}{2x_2} = \frac{1}{x_2} & 0 < x_1 < x_2 < 1 \\ 0 & \text{elsewhere.} \end{cases}$$

Here the conditional mean and the conditional variance of X_1, given $X_2 = x_2$, are respectively,

$$\begin{aligned} E(X_1|x_2) &= \int_{-\infty}^{\infty} x_1 f_{1|2}(x_1|x_2)\,dx_1 \\ &= \int_0^{x_2} x_1 \left(\frac{1}{x_2}\right) dx_1 \\ &= \frac{x_2}{2}, \quad 0 < x_2 < 1, \end{aligned}$$

and

$$\begin{aligned} \text{Var}(X_1|x_2) &= \int_0^{x_2} \left(x_1 - \frac{x_2}{2}\right)^2 \left(\frac{1}{x_2}\right) dx_1 \\ &= \frac{x_2^2}{12}, \quad 0 < x_2 < 1. \end{aligned}$$

Finally, we compare the values of

$$P(0 < X_1 < \tfrac{1}{2}|X_2 = \tfrac{3}{4}) \quad \text{and } P(0 < X_1 < \tfrac{1}{2}).$$

We have

$$P(0 < X_1 < \tfrac{1}{2}|X_2 = \tfrac{3}{4}) = \int_0^{1/2} f_{1|2}(x_1|\tfrac{3}{4})\,dx_1 = \int_0^{1/2} (\tfrac{4}{3})\,dx_1 = \tfrac{2}{3},$$

but

$$P(0 < X_1 < \tfrac{1}{2}) = \int_0^{1/2} f_1(x_1)\,dx_1 = \int_0^{1/2} 2(1 - x_1)\,dx_1 = \tfrac{3}{4}. \quad \blacksquare$$

Since $E(X_2|x_1)$ is a function of x_1, then $E(X_2|X_1)$ is a random variable with its own distribution, mean, and variance. Let us consider the following illustration of this.

Example 2.3.2. Let X_1 and X_2 have the joint pdf

$$f(x_1, x_2) = \begin{cases} 6x_2 & 0 < x_2 < x_1 < 1 \\ 0 & \text{elsewhere.} \end{cases}$$

Then the marginal pdf of X_1 is

$$f_1(x_1) = \int_0^{x_1} 6x_2\,dx_2 = 3x_1^2, \quad 0 < x_1 < 1,$$

zero elsewhere. The conditional pdf of X_2, given $X_1 = x_1$, is

$$f_{2|1}(x_2|x_1) = \frac{6x_2}{3x_1^2} = \frac{2x_2}{x_1^2}, \quad 0 < x_2 < x_1,$$

zero elsewhere, where $0 < x_1 < 1$. The conditional mean of X_2, given $X_1 = x_1$, is

$$E(X_2|x_1) = \int_0^{x_1} x_2 \left(\frac{2x_2}{x_1^2}\right) dx_2 = \frac{2}{3}x_1, \quad 0 < x_1 < 1.$$

Now $E(X_2|X_1) = 2X_1/3$ is a random variable, say Y. The cdf of $Y = 2X_1/3$ is

$$G(y) = P(Y \le y) = P\left(X_1 \le \frac{3y}{2}\right), \quad 0 \le y < \frac{2}{3}.$$

From the pdf $f_1(x_1)$, we have

$$G(y) = \int_0^{3y/2} 3x_1^2\, dx_1 = \frac{27y^3}{8}, \quad 0 \le y < \frac{2}{3}.$$

Of course, $G(y) = 0$ if $y < 0$, and $G(y) = 1$ if $\frac{2}{3} < y$. The pdf, mean, and variance of $Y = 2X_1/3$ are

$$g(y) = \frac{81y^2}{8}, \quad 0 \le y < \frac{2}{3},$$

zero elsewhere,

$$E(Y) = \int_0^{2/3} y \left(\frac{81y^2}{8}\right) dy = \frac{1}{2},$$

and

$$\text{Var}(Y) = \int_0^{2/3} y^2 \left(\frac{81y^2}{8}\right) dy - \frac{1}{4} = \frac{1}{60}.$$

Since the marginal pdf of X_2 is

$$f_2(x_2) = \int_{x_2}^{1} 6x_2\, dx_1 = 6x_2(1 - x_2), \quad 0 < x_2 < 1,$$

zero elsewhere, it is easy to show that $E(X_2) = \frac{1}{2}$ and $\text{Var}(X_2) = \frac{1}{20}$. That is, here

$$E(Y) = E[E(X_2|X_1)] = E(X_2)$$

and

$$\text{Var}(Y) = \text{Var}[E(X_2|X_1)] \le \text{Var}(X_2). \quad \blacksquare$$

Example 2.3.2 is excellent, as it provides us with the opportunity to apply many of these new definitions as well as review the cdf technique for finding the distribution of a function of a random variable, namely $Y = 2X_1/3$. Moreover, the two observations at the end of this example are no accident because they are true in general.

Theorem 2.3.1. *Let (X_1, X_2) be a random vector such that the variance of X_2 is finite. Then,*

(a) $E[E(X_2|X_1)] = E(X_2)$.

(b) $Var[E(X_2|X_1)] \leq Var(X_2)$.

Proof: The proof is for the continuous case. To obtain it for the discrete case, exchange summations for integrals. We first prove (a). Note that

$$\begin{aligned}
E(X_2) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_2 f(x_1, x_2)\, dx_2 dx_1 \\
&= \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} x_2 \frac{f(x_1, x_2)}{f_1(x_1)}\, dx_2\right] f_1(x_1)\, dx_1 \\
&= \int_{-\infty}^{\infty} E(X_2|x_1) f_1(x_1)\, dx_1 \\
&= E[E(X_2|X_1)],
\end{aligned}$$

which is the first result.

Next we show (b). Consider with $\mu_2 = E(X_2)$,

$$\begin{aligned}
\mathrm{Var}(X_2) &= E[(X_2 - \mu_2)^2] \\
&= E\{[X_2 - E(X_2|X_1) + E(X_2|X_1) - \mu_2]^2\} \\
&= E\{[X_2 - E(X_2|X_1)]^2\} + E\{[E(X_2|X_1) - \mu_2]^2\} \\
&\quad + 2E\{[X_2 - E(X_2|X_1)][E(X_2|X_1) - \mu_2]\}.
\end{aligned}$$

We show that the last term of the right-hand member of the immediately preceding equation is zero. It is equal to

$$\begin{aligned}
&2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [x_2 - E(X_2|x_1)][E(X_2|x_1) - \mu_2] f(x_1, x_2)\, dx_2 dx_1 \\
&= 2\int_{-\infty}^{\infty} [E(X_2|x_1) - \mu_2]\left\{\int_{-\infty}^{\infty} [x_2 - E(X_2|x_1)] \frac{f(x_1, x_2)}{f_1(x_1)}\, dx_2\right\} f_1(x_1)\, dx_1.
\end{aligned}$$

But $E(X_2|x_1)$ is the conditional mean of X_2, given $X_1 = x_1$. Since the expression in the inner braces is equal to

$$E(X_2|x_1) - E(X_2|x_1) = 0,$$

the double integral is equal to zero. Accordingly, we have

$$\mathrm{Var}(X_2) = E\{[X_2 - E(X_2|X_1)]^2\} + E\{[E(X_2|X_1) - \mu_2]^2\}.$$

The first term in the right-hand member of this equation is nonnegative because it is the expected value of a nonnegative function, namely $[X_2 - E(X_2|X_1)]^2$. Since $E[E(X_2|X_1)] = \mu_2$, the second term is $\mathrm{Var}[E(X_2|X_1)]$. Hence we have

$$\mathrm{Var}(X_2) \geq \mathrm{Var}[E(X_2|X_1)],$$

which completes the proof. ■

Intuitively, this result could have this useful interpretation. Both the random variables X_2 and $E(X_2|X_1)$ have the same mean μ_2. If we did not know μ_2, we could use either of the two random variables to guess at the unknown μ_2. Since, however, $\text{Var}(X_2) \geq \text{Var}[E(X_2|X_1)]$, we would put more reliance in $E(X_2|X_1)$ as a guess. That is, if we observe the pair (X_1, X_2) to be (x_1, x_2), we could prefer to use $E(X_2|x_1)$ to x_2 as a guess at the unknown μ_2. When studying the use of sufficient statistics in estimation in Chapter 7, we make use of this famous result, attributed to C. R. Rao and David Blackwell.

We finish this section with an example illustrating Theorem 2.3.1.

Example 2.3.3. Let X_1 and X_2 be discrete random variables. Suppose the conditional pmf of X_1 given X_2 and the marginal distribution of X_2 are given by

$$\begin{aligned} p(x_1|x_2) &= \binom{x_2}{x_1}\left(\frac{1}{2}\right)^{x_2}, \quad x_1 = 0, 1, \ldots, x_2 \\ p(x_2) &= \frac{2}{3}\left(\frac{1}{3}\right)^{x_2-1}, \quad x_2 = 1, 2, 3 \ldots. \end{aligned}$$

Let us determine the mgf of X_1. For fixed x_2, by the binomial theorem,

$$\begin{aligned} E\left(e^{tX_1}|x_2\right) &= \sum_{x_1=0}^{x_2}\binom{x_2}{x_1}e^{tx_1}\left(\frac{1}{2}\right)^{x_2-x_1}\left(\frac{1}{2}\right)^{x_1} \\ &= \left(\frac{1}{2}+\frac{1}{2}e^t\right)^{x_2}. \end{aligned}$$

Hence, by the geometric series and Theorem 2.3.1,

$$\begin{aligned} E\left(e^{tX_1}\right) &= E\left[E\left(e^{tX_1}|X_2\right)\right] \\ &= \sum_{x_2=1}^{\infty}\left(\frac{1}{2}+\frac{1}{2}e^t\right)^{x_2}\frac{2}{3}\left(\frac{1}{3}\right)^{x_2-1} \\ &= \frac{2}{3}\left(\frac{1}{2}+\frac{1}{2}e^t\right)\sum_{x_2=1}^{\infty}\left(\frac{1}{6}+\frac{1}{6}e^t\right)^{x_2-1} \\ &= \frac{2}{3}\left(\frac{1}{2}+\frac{1}{2}e^t\right)\frac{1}{1-[(1/6)+(1/6)e^t]}, \end{aligned}$$

provided $(1/6) + (1/6)e^t < 1$ or $t < \log 5$ (which includes $t = 0$). ■

EXERCISES

2.3.1. Let X_1 and X_2 have the joint pdf $f(x_1, x_2) = x_1 + x_2,\ 0 < x_1 < 1,\ 0 < x_2 < 1$, zero elsewhere. Find the conditional mean and variance of X_2, given $X_1 = x_1,\ 0 < x_1 < 1$.

2.3.2. Let $f_{1|2}(x_1|x_2) = c_1x_1/x_2^2,\ 0 < x_1 < x_2,\ 0 < x_2 < 1$, zero elsewhere, and $f_2(x_2) = c_2x_2^4,\ 0 < x_2 < 1$, zero elsewhere, denote, respectively, the conditional pdf of X_1, given $X_2 = x_2$, and the marginal pdf of X_2. Determine:

(a) The constants c_1 and c_2.

(b) The joint pdf of X_1 and X_2.

(c) $P(\frac{1}{4} < X_1 < \frac{1}{2}|X_2 = \frac{5}{8})$.

(d) $P(\frac{1}{4} < X_1 < \frac{1}{2})$.

2.3.3. Let $f(x_1, x_2) = 21x_1^2x_2^3,\ 0 < x_1 < x_2 < 1$, zero elsewhere, be the joint pdf of X_1 and X_2.

(a) Find the conditional mean and variance of X_1, given $X_2 = x_2,\ 0 < x_2 < 1$.

(b) Find the distribution of $Y = E(X_1|X_2)$.

(c) Determine $E(Y)$ and $\text{Var}(Y)$ and compare these to $E(X_1)$ and $\text{Var}(X_1)$, respectively.

2.3.4. Suppose X_1 and X_2 are random variables of the discrete type that have the joint pmf $p(x_1, x_2) = (x_1 + 2x_2)/18,\ (x_1, x_2) = (1, 1), (1, 2), (2, 1), (2, 2)$, zero elsewhere. Determine the conditional mean and variance of X_2, given $X_1 = x_1$, for $x_1 = 1$ or 2. Also, compute $E(3X_1 - 2X_2)$.

2.3.5. Let X_1 and X_2 be two random variables such that the conditional distributions and means exist. Show that:

(a) $E(X_1 + X_2 \,|\, X_2) = E(X_1 \,|\, X_2) + X_2$,

(b) $E(u(X_2) \,|\, X_2) = u(X_2)$.

2.3.6. Let the joint pdf of X and Y be given by

$$f(x, y) = \begin{cases} \frac{2}{(1+x+y)^3} & 0 < x < \infty,\ 0 < y < \infty \\ 0 & \text{elsewhere.} \end{cases}$$

(a) Compute the marginal pdf of X and the conditional pdf of Y, given $X = x$.

(b) For a fixed $X = x$, compute $E(1 + x + Y|x)$ and use the result to compute $E(Y|x)$.

2.3.7. Suppose X_1 and X_2 are discrete random variables which have the joint pmf $p(x_1, x_2) = (3x_1 + x_2)/24,\ (x_1, x_2) = (1, 1), (1, 2), (2, 1), (2, 2)$, zero elsewhere. Find the conditional mean $E(X_2|x_1)$, when $x_1 = 1$.

2.3.8. Let X and Y have the joint pdf $f(x, y) = 2\exp\{-(x + y)\},\ 0 < x < y < \infty$, zero elsewhere. Find the conditional mean $E(Y|x)$ of Y, given $X = x$.

2.3.9. Five cards are drawn at random and without replacement from an ordinary deck of cards. Let X_1 and X_2 denote, respectively, the number of spades and the number of hearts that appear in the five cards.

(a) Determine the joint pmf of X_1 and X_2.

(b) Find the two marginal pmfs.

(c) What is the conditional pmf of X_2, given $X_1 = x_1$?

2.3.10. Let X_1 and X_2 have the joint pmf $p(x_1, x_2)$ described as follows:

(x_1, x_2)	$(0,0)$	$(0,1)$	$(1,0)$	$(1,1)$	$(2,0)$	$(2,1)$
$p(x_1, x_2)$	$\frac{1}{18}$	$\frac{3}{18}$	$\frac{4}{18}$	$\frac{3}{18}$	$\frac{6}{18}$	$\frac{1}{18}$

and $p(x_1, x_2)$ is equal to zero elsewhere. Find the two marginal probability mass functions and the two conditional means.
Hint: Write the probabilities in a rectangular array.

2.3.11. Let us choose at random a point from the interval $(0, 1)$ and let the random variable X_1 be equal to the number that corresponds to that point. Then choose a point at random from the interval $(0, x_1)$, where x_1 is the experimental value of X_1; and let the random variable X_2 be equal to the number that corresponds to this point.

(a) Make assumptions about the marginal pdf $f_1(x_1)$ and the conditional pdf $f_{2|1}(x_2|x_1)$.

(b) Compute $P(X_1 + X_2 \geq 1)$.

(c) Find the conditional mean $E(X_1|x_2)$.

2.3.12. Let $f(x)$ and $F(x)$ denote, respectively, the pdf and the cdf of the random variable X. The conditional pdf of X, given $X > x_0$, x_0 a fixed number, is defined by $f(x|X > x_0) = f(x)/[1-F(x_0)]$, $x_0 < x$, zero elsewhere. This kind of conditional pdf finds application in a problem of time until death, given survival until time x_0.

(a) Show that $f(x|X > x_0)$ is a pdf.

(b) Let $f(x) = e^{-x}$, $0 < x < \infty$, and zero elsewhere. Compute $P(X > 2|X > 1)$.

2.4 Independent Random Variables

Let X_1 and X_2 denote the random variables of the continuous type that have the joint pdf $f(x_1, x_2)$ and marginal probability density functions $f_1(x_1)$ and $f_2(x_2)$, respectively. In accordance with the definition of the conditional pdf $f_{2|1}(x_2|x_1)$, we may write the joint pdf $f(x_1, x_2)$ as

$$f(x_1, x_2) = f_{2|1}(x_2|x_1)f_1(x_1).$$

Suppose that we have an instance where $f_{2|1}(x_2|x_1)$ does not depend upon x_1. Then the marginal pdf of X_2 is, for random variables of the continuous type,

$$\begin{aligned} f_2(x_2) &= \int_{-\infty}^{\infty} f_{2|1}(x_2|x_1) f_1(x_1)\, dx_1 \\ &= f_{2|1}(x_2|x_1) \int_{-\infty}^{\infty} f_1(x_1)\, dx_1 \\ &= f_{2|1}(x_2|x_1). \end{aligned}$$

Accordingly,

$$f_2(x_2) = f_{2|1}(x_2|x_1) \quad \text{and} \quad f(x_1, x_2) = f_1(x_1) f_2(x_2),$$

when $f_{2|1}(x_2|x_1)$ does not depend upon x_1. That is, if the conditional distribution of X_2, given $X_1 = x_1$, is independent of any assumption about x_1, then $f(x_1, x_2) = f_1(x_1) f_2(x_2)$.

The same discussion applies to the discrete case too, which we summarize in parentheses in the following definition.

Definition 2.4.1 (Independence). *Let the random variables X_1 and X_2 have the joint pdf $f(x_1, x_2)$ [joint pmf $p(x_1, x_2)$] and the marginal pdfs [pmfs] $f_1(x_1)$ [$p_1(x_1)$] and $f_2(x_2)$ [$p_2(x_2)$], respectively. The random variables X_1 and X_2 are said to be* **independent** *if, and only if, $f(x_1, x_2) \equiv f_1(x_1) f_2(x_2)$ [$p(x_1, x_2) \equiv p_1(x_1) p_2(x_2)$]. Random variables that are not independent are said to be* **dependent**.

Remark 2.4.1. Two comments should be made about the preceding definition. First, the product of two positive functions $f_1(x_1) f_2(x_2)$ means a function that is positive on the product space. That is, if $f_1(x_1)$ and $f_2(x_2)$ are positive on, and only on, the respective spaces $\mathcal{S}_1$ and $\mathcal{S}_2$, then the product of $f_1(x_1)$ and $f_2(x_2)$ is positive on, and only on, the product space $\mathcal{S} = \{(x_1, x_2) : x_1 \in \mathcal{S}_1,\ x_2 \in \mathcal{S}_2\}$. For instance, if $\mathcal{S}_1 = \{x_1 : 0 < x_1 < 1\}$ and $\mathcal{S}_2 = \{x_2 : 0 < x_2 < 3\}$, then $\mathcal{S} = \{(x_1, x_2) : 0 < x_1 < 1,\ 0 < x_2 < 3\}$. The second remark pertains to the identity. The identity in Definition 2.4.1 should be interpreted as follows. There may be certain points $(x_1, x_2) \in \mathcal{S}$ at which $f(x_1, x_2) \neq f_1(x_1) f_2(x_2)$. However, if A is the set of points (x_1, x_2) at which the equality does not hold, then $P(A) = 0$. In subsequent theorems and the subsequent generalizations, a product of nonnegative functions and an identity should be interpreted in an analogous manner. ■

Example 2.4.1. Suppose an urn contains 10 blue, 8 red, and 7 yellow balls that are the same except for color. Suppose 4 balls are drawn without replacement. Let X and Y be the number of red and blue balls drawn, respectively. The joint pmf of (X, Y) is

$$p(x, y) = \frac{\binom{10}{x}\binom{8}{y}\binom{7}{4-x-y}}{\binom{25}{4}}, \quad 0 \leq x, y \leq 4; x + y \leq 4.$$

Since $X + Y \leq 4$, it would seem that X and Y are dependent. To see that this is true by definition, we first find the marginal pmf's which are:

$$\begin{aligned} p_X(x) &= \frac{\binom{10}{x}\binom{15}{4-x}}{\binom{25}{4}}, \quad 0 \leq x \leq 4; \\ p_Y(y) &= \frac{\binom{8}{y}\binom{17}{4-y}}{\binom{25}{4}}, \quad 0 \leq y \leq 4. \end{aligned}$$

To show dependence, we need to find only one point in the support of (X_1, X_2) where the joint pmf does not factor into the product of the marginal pmf's. Suppose we select the point $x = 1$ and $y = 1$. Then, using R for calculation, we compute (to 4 places):

$$\begin{aligned} p(1,1) &= 10 \cdot 8 \cdot \binom{7}{2} \Big/ \binom{25}{4} = 0.1328 \\ p_X(1) &= 10\binom{15}{3} \Big/ \binom{25}{4} = 0.3597 \\ p_Y(1) &= 8\binom{17}{3} \Big/ \binom{25}{4} = 0.4300. \end{aligned}$$

Since $0.1328 \neq 0.1547 = 0.3597 \cdot 0.4300$, X and Y are dependent random variables. ■

Example 2.4.2. Let the joint pdf of X_1 and X_2 be

$$f(x_1, x_2) = \begin{cases} x_1 + x_2 & 0 < x_1 < 1, \ 0 < x_2 < 1 \\ 0 & \text{elsewhere.} \end{cases}$$

We show that X_1 and X_2 are dependent. Here the marginal probability density functions are

$$f_1(x_1) = \begin{cases} \int_{-\infty}^{\infty} f(x_1, x_2)\, dx_2 = \int_0^1 (x_1 + x_2)\, dx_2 = x_1 + \frac{1}{2} & 0 < x_1 < 1 \\ 0 & \text{elsewhere,} \end{cases}$$

and

$$f_2(x_2) = \begin{cases} \int_{-\infty}^{\infty} f(x_1, x_2)\, dx_1 = \int_0^1 (x_1 + x_2)\, dx_1 = \frac{1}{2} + x_2 & 0 < x_2 < 1 \\ 0 & \text{elsewhere.} \end{cases}$$

Since $f(x_1, x_2) \not\equiv f_1(x_1) f_2(x_2)$, the random variables X_1 and X_2 are dependent. ■

The following theorem makes it possible to assert that the random variables X_1 and X_2 of Example 2.4.2 are dependent, without computing the marginal probability density functions.

Theorem 2.4.1. *Let the random variables X_1 and X_2 have supports $\mathcal{S}_1$ and $\mathcal{S}_2$, respectively, and have the joint pdf $f(x_1, x_2)$. Then X_1 and X_2 are independent if* *and only if $f(x_1, x_2)$ can be written as a product of a nonnegative function of x_1 and a nonnegative function of x_2. That is,*

$$f(x_1, x_2) \equiv g(x_1)h(x_2),$$

where $g(x_1) > 0$, $x_1 \in \mathcal{S}_1$, zero elsewhere, and $h(x_2) > 0$, $x_2 \in \mathcal{S}_2$, zero elsewhere.

Proof. If X_1 and X_2 are independent, then $f(x_1, x_2) \equiv f_1(x_1)f_2(x_2)$, where $f_1(x_1)$ and $f_2(x_2)$ are the marginal probability density functions of X_1 and X_2, respectively. Thus the condition $f(x_1, x_2) \equiv g(x_1)h(x_2)$ is fulfilled.

Conversely, if $f(x_1, x_2) \equiv g(x_1)h(x_2)$, then, for random variables of the continuous type, we have

$$f_1(x_1) = \int_{-\infty}^{\infty} g(x_1)h(x_2)\,dx_2 = g(x_1)\int_{-\infty}^{\infty} h(x_2)\,dx_2 = c_1 g(x_1)$$

and

$$f_2(x_2) = \int_{-\infty}^{\infty} g(x_1)h(x_2)\,dx_1 = h(x_2)\int_{-\infty}^{\infty} g(x_1)\,dx_1 = c_2 h(x_2),$$

where c_1 and c_2 are constants, not functions of x_1 or x_2. Moreover, $c_1c_2 = 1$ because

$$1 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x_1)h(x_2)\,dx_1dx_2 = \left[\int_{-\infty}^{\infty} g(x_1)\,dx_1\right]\left[\int_{-\infty}^{\infty} h(x_2)\,dx_2\right] = c_2c_1.$$

These results imply that

$$f(x_1, x_2) \equiv g(x_1)h(x_2) \equiv c_1g(x_1)c_2h(x_2) \equiv f_1(x_1)f_2(x_2).$$

Accordingly, X_1 and X_2 are independent. ■

This theorem is true for the discrete case also. Simply replace the joint pdf by the joint pmf. For instance, the discrete random variables X and Y of Example 2.4.1 are immediately seen to be dependent because the support of (X, Y) is not a product space.

Next, consider the joint distribution of the continuous random vector (X, Y) given in Example 2.1.3. The joint pdf is

$$f(x, y) = 4xe^{-x^2}ye^{-y^2}, \quad x > 0, y > 0.$$

which is a product of a nonnegative function of x and a nonnegative function of y. Further, the joint support is a product space. Hence, X and Y are independent random variables.

Example 2.4.3. Let the pdf of the random variable X_1 and X_2 be $f(x_1, x_2) = 8x_1x_2$, $0 < x_1 < x_2 < 1$, zero elsewhere. The formula $8x_1x_2$ might suggest to some that X_1 and X_2 are independent. However, if we consider the space $\mathcal{S} = \{(x_1, x_2) : 0 < x_1 < x_2 < 1\}$, we see that it is not a product space. This should make it clear that, in general, X_1 and X_2 must be dependent if the space of positive probability density of X_1 and X_2 is bounded by a curve that is neither a horizontal nor a vertical line. ■

Instead of working with pdfs (or pmfs) we could have presented independence in terms of cumulative distribution functions. The following theorem shows the equivalence.

Theorem 2.4.2. *Let (X_1, X_2) have the joint cdf $F(x_1, x_2)$ and let X_1 and X_2 have the marginal cdfs $F_1(x_1)$ and $F_2(x_2)$, respectively. Then X_1 and X_2 are independent if and only if*

$$F(x_1, x_2) = F_1(x_1)F_2(x_2) \quad \text{for all } (x_1, x_2) \in R^2 \; . \tag{2.4.1}$$

Proof: We give the proof for the continuous case. Suppose expression (2.4.1) holds. Then the mixed second partial is

$$\frac{\partial^2}{\partial x_1 \partial x_2} F(x_1, x_2) = f_1(x_1)f_2(x_2).$$

Hence, X_1 and X_2 are independent. Conversely, suppose X_1 and X_2 are independent. Then by the definition of the joint cdf,

$$\begin{aligned} F(x_1, x_2) &= \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} f_1(w_1)f_2(w_2)\, dw_2 dw_1 \\ &= \int_{-\infty}^{x_1} f_1(w_1)\, dw_1 \cdot \int_{-\infty}^{x_2} f_2(w_2)\, dw_2 = F_1(x_1)F_2(x_2). \end{aligned}$$

Hence, condition (2.4.1) is true. ■

We now give a theorem that frequently simplifies the calculations of probabilities of events that involves independent variables.

Theorem 2.4.3. *The random variables X_1 and X_2 are independent random variables if and only if the following condition holds,*

$$P(a < X_1 \le b, c < X_2 \le d) = P(a < X_1 \le b)P(c < X_2 \le d) \tag{2.4.2}$$

for every $a < b$ and $c < d$, where a, b, c, and d are constants.

Proof: If X_1 and X_2 are independent, then an application of the last theorem and expression (2.1.2) shows that

$$\begin{aligned} P(a < X_1 \le b, c < X_2 \le d) &= F(b, d) - F(a, d) - F(b, c) + F(a, c) \\ &= F_1(b)F_2(d) - F_1(a)F_2(d) - F_1(b)F_2(c) \\ &\quad + F_1(a)F_2(c) \\ &= [F_1(b) - F_1(a)][F_2(d) - F_2(c)], \end{aligned}$$

which is the right side of expression (2.4.2). Conversely, condition (2.4.2) implies that the joint cdf of (X_1, X_2) factors into a product of the marginal cdfs, which in turn by Theorem 2.4.2 implies that X_1 and X_2 are independent. ■

Example 2.4.4 (Example 2.4.2, Continued)**.** Independence is necessary for condition (2.4.2). For example, consider the dependent variables X_1 and X_2 of Example 2.4.2. For these random variables, we have

$$P(0 < X_1 < \tfrac{1}{2}, 0 < X_2 < \tfrac{1}{2}) = \int_0^{1/2}\int_0^{1/2}(x_1 + x_2)\,dx_1dx_2 = \tfrac{1}{8},$$

whereas

$$P(0 < X_1 < \tfrac{1}{2}) = \int_0^{1/2}(x_1 + \tfrac{1}{2})\,dx_1 = \tfrac{3}{8}$$

and

$$P(0 < X_2 < \tfrac{1}{2}) = \int_0^{1/2}(\tfrac{1}{2} + x_1)\,dx_2 = \tfrac{3}{8}.$$

Hence, condition (2.4.2) does not hold. ■

Not merely are calculations of some probabilities usually simpler when we have independent random variables, but many expectations, including certain moment generating functions, have comparably simpler computations. The following result proves so useful that we state it in the form of a theorem.

Theorem 2.4.4. *Suppose X_1 and X_2 are independent and that $E(u(X_1))$ and $E(v(X_2))$ exist. Then*

$$E[u(X_1)v(X_2)] = E[u(X_1)]E[v(X_2)].$$

Proof. We give the proof in the continuous case. The independence of X_1 and X_2 implies that the joint pdf of X_1 and X_2 is $f_1(x_1)f_2(x_2)$. Thus we have, by definition of expectation,

$$\begin{aligned}
E[u(X_1)v(X_2)] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} u(x_1)v(x_2)f_1(x_1)f_2(x_2)\,dx_1dx_2 \\
&= \left[\int_{-\infty}^{\infty} u(x_1)f_1(x_1)\,dx_1\right]\left[\int_{-\infty}^{\infty} v(x_2)f_2(x_2)\,dx_2\right] \\
&= E[u(X_1)]E[v(X_2)].
\end{aligned}$$

Hence, the result is true. ■

Upon taking the functions $u(\cdot)$ and $v(\cdot)$ to be the identity functions in Theorem 2.4.4, we have that for independent random variables X_1 and X_2,

$$E(X_1X_2) = E(X_1)E(X_2). \tag{2.4.3}$$

We next prove a very useful theorem about independent random variables. The proof of the theorem relies heavily upon our assertion that an mgf, when it exists, is unique and that it uniquely determines the distribution of probability.

Theorem 2.4.5. *Suppose the joint mgf, $M(t_1, t_2)$, exists for the random variables X_1 and X_2. Then X_1 and X_2 are independent if and only if*

$$M(t_1, t_2) = M(t_1, 0)M(0, t_2);$$

that is, the joint mgf is identically equal to the product of the marginal mgfs.

Proof. If X_1 and X_2 are independent, then

$$\begin{aligned} M(t_1, t_2) &= E(e^{t_1 X_1 + t_2 X_2}) \\ &= E(e^{t_1 X_1} e^{t_2 X_2}) \\ &= E(e^{t_1 X_1}) E(e^{t_2 X_2}) \\ &= M(t_1, 0) M(0, t_2). \end{aligned}$$

Thus the independence of X_1 and X_2 implies that the mgf of the joint distribution factors into the product of the moment-generating functions of the two marginal distributions.

Suppose next that the mgf of the joint distribution of X_1 and X_2 is given by $M(t_1, t_2) = M(t_1, 0)M(0, t_2)$. Now X_1 has the unique mgf, which, in the continuous case, is given by

$$M(t_1, 0) = \int_{-\infty}^{\infty} e^{t_1 x_1} f_1(x_1)\, dx_1.$$

Similarly, the unique mgf of X_2, in the continuous case, is given by

$$M(0, t_2) = \int_{-\infty}^{\infty} e^{t_2 x_2} f_2(x_2)\, dx_2.$$

Thus we have

$$\begin{aligned} M(t_1, 0)M(0, t_2) &= \left[\int_{-\infty}^{\infty} e^{t_1 x_1} f_1(x_1)\, dx_1\right] \left[\int_{-\infty}^{\infty} e^{t_2 x_2} f_2(x_2)\, dx_2\right] \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{t_1 x_1 + t_2 x_2} f_1(x_1) f_2(x_2)\, dx_1 dx_2. \end{aligned}$$

We are given that $M(t_1, t_2) = M(t_1, 0)M(0, t_2)$; so

$$M(t_1, t_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{t_1 x_1 + t_2 x_2} f_1(x_1) f_2(x_2)\, dx_1 dx_2.$$

But $M(t_1, t_2)$ is the mgf of X_1 and X_2. Thus

$$M(t_1, t_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{t_1 x_1 + t_2 x_2} f(x_1, x_2)\, dx_1 dx_2.$$

The uniqueness of the mgf implies that the two distributions of probability that are described by $f_1(x_1)f_2(x_2)$ and $f(x_1, x_2)$ are the same. Thus

$$f(x_1, x_2) \equiv f_1(x_1) f_2(x_2).$$

That is, if $M(t_1, t_2) = M(t_1, 0)M(0, t_2)$, then X_1 and X_2 are independent. This completes the proof when the random variables are of the continuous type. With random variables of the discrete type, the proof is made by using summation instead of integration. ■

Example 2.4.5 (Example 2.1.10, Continued)**.** Let (X, Y) be a pair of random variables with the joint pdf

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

In Example 2.1.10, we showed that the mgf of (X, Y) is

$$\begin{aligned} M(t_1, t_2) &= \int_0^\infty \int_x^\infty \exp(t_1 x + t_2 y - y)\, dy dx \\ &= \frac{1}{(1 - t_1 - t_2)(1 - t_2)}, \end{aligned}$$

provided that $t_1 + t_2 < 1$ and $t_2 < 1$. Because $M(t_1, t_2) \neq M(t_1, 0)M(t_1, 0)$, the random variables are dependent. ∎

Example 2.4.6 (Exercise 2.1.15, Continued)**.** For the random variable X_1 and X_2 defined in Exercise 2.1.15, we showed that the joint mgf is

$$M(t_1, t_2) = \left[\frac{\exp\{t_1\}}{2 - \exp\{t_1\}}\right]\left[\frac{\exp\{t_2\}}{2 - \exp\{t_2\}}\right], \quad t_i < \log 2\ , i = 1, 2.$$

We showed further that $M(t_1, t_2) = M(t_1, 0)M(0, t_2)$. Hence, X_1 and X_2 are independent random variables. ∎

EXERCISES

2.4.1. Show that the random variables X_1 and X_2 with joint pdf

$$f(x_1, x_2) = \begin{cases} 12x_1x_2(1 - x_2) & 0 < x_1 < 1,\ 0 < x_2 < 1 \\ 0 & \text{elsewhere} \end{cases}$$

are independent.

2.4.2. If the random variables X_1 and X_2 have the joint pdf $f(x_1, x_2) = 2e^{-x_1 - x_2}$, $0 < x_1 < x_2$, $0 < x_2 < \infty$, zero elsewhere, show that X_1 and X_2 are dependent.

2.4.3. Let $p(x_1, x_2) = \frac{1}{16}$, $x_1 = 1, 2, 3, 4$, and $x_2 = 1, 2, 3, 4$, zero elsewhere, be the joint pmf of X_1 and X_2. Show that X_1 and X_2 are independent.

2.4.4. Find $P(0 < X_1 < \frac{1}{3}, 0 < X_2 < \frac{1}{3})$ if the random variables X_1 and X_2 have the joint pdf $f(x_1, x_2) = 4x_1(1 - x_2)$, $0 < x_1 < 1$, $0 < x_2 < 1$, zero elsewhere.

2.4.5. Find the probability of the union of the events $a < X_1 < b$, $-\infty < X_2 < \infty$, and $-\infty < X_1 < \infty$, $c < X_2 < d$ if X_1 and X_2 are two independent variables with $P(a < X_1 < b) = \frac{2}{3}$ and $P(c < X_2 < d) = \frac{5}{8}$.

2.4.6. If $f(x_1, x_2) = e^{-x_1-x_2}$, $0 < x_1 < \infty$, $0 < x_2 < \infty$, zero elsewhere, is the joint pdf of the random variables X_1 and X_2, show that X_1 and X_2 are independent and that $M(t_1, t_2) = (1-t_1)^{-1}(1-t_2)^{-1}$, $t_2 < 1$, $t_1 < 1$. Also show that

$$E(e^{t(X_1+X_2)}) = (1-t)^{-2}, \quad t < 1.$$

Accordingly, find the mean and the variance of $Y = X_1 + X_2$.

2.4.7. Let the random variables X_1 and X_2 have the joint pdf $f(x_1, x_2) = 1/\pi$, for $(x_1 - 1)^2 + (x_2 + 2)^2 < 1$, zero elsewhere. Find $f_1(x_1)$ and $f_2(x_2)$. Are X_1 and X_2 independent?

2.4.8. Let X and Y have the joint pdf $f(x, y) = 3x$, $0 < y < x < 1$, zero elsewhere. Are X and Y independent? If not, find $E(X|y)$.

2.4.9. Suppose that a man leaves for work between 8:00 a.m. and 8:30 a.m. and takes between 40 and 50 minutes to get to the office. Let X denote the time of departure and let Y denote the time of travel. If we assume that these random variables are independent and uniformly distributed, find the probability that he arrives at the office before 9:00 a.m.

2.4.10. Let X and Y be random variables with the space consisting of the four points $(0,0), (1,1), (1,0), (1,-1)$. Assign positive probabilities to these four points so that the correlation coefficient is equal to zero. Are X and Y independent?

2.4.11. Two line segments, each of length two units, are placed along the x-axis. The midpoint of the first is between $x = 0$ and $x = 14$ and that of the second is between $x = 6$ and $x = 20$. Assuming independence and uniform distributions for these midpoints, find the probability that the line segments overlap.

2.4.12. Cast a fair die and let $X = 0$ if $1, 2,$ or 3 spots appear, let $X = 1$ if 4 or 5 spots appear, and let $X = 2$ if 6 spots appear. Do this two independent times, obtaining X_1 and X_2. Calculate $P(|X_1 - X_2| = 1)$.

2.4.13. For X_1 and X_2 in Example 2.4.6, show that the mgf of $Y = X_1 + X_2$ is $e^{2t}/(2-e^t)^2$, $t < \log 2$, and then compute the mean and variance of Y.

2.5 The Correlation Coefficient

Let (X, Y) denote a random vector. In the last section, we discussed the concept of independence between X and Y. What if, though, X and Y are dependent and, if so, how are they related? There are many measures of dependence. In this section, we introduce a parameter ρ of the joint distribution of (X, Y) which measures linearity between X and Y. In this section, we assume the existence of all expectations under discussion.

Definition 2.5.1. *Let (X, Y) have a joint distribution. Denote the means of X and Y respectively by μ_1 and μ_2 and their respective variances by σ_1^2 and σ_2^2. The* **covariance** *of (X, Y) is denoted by $cov(X, Y)$ and is defined by the expectation*

$$cov(X, Y) = E[(X - \mu_1)(Y - \mu_2)]. \quad \blacksquare \tag{2.5.1}$$

It follows by the linearity of expectation, Theorem 2.1.1, that the covariance of X and Y can also be expressed as

$$\begin{aligned} \text{cov}(X,Y) &= E(XY - \mu_2 X - \mu_1 Y + \mu_1\mu_2) \\ &= E(XY) - \mu_2 E(X) - \mu_1 E(Y) + \mu_1\mu_2 \\ &= E(XY) - \mu_1\mu_2, \end{aligned} \tag{2.5.2}$$

which is often easier to compute than using the definition, (2.5.1).

The measure that we seek is a standardized (unitless) version of the covariance.

Definition 2.5.2. *If each of σ_1 and σ_2 is positive, then the* **correlation coefficient** *between X and Y is defined by*

$$\rho = \frac{E[(X-\mu_1)(Y-\mu_2)]}{\sigma_1\sigma_2} = \frac{cov(X,Y)}{\sigma_1\sigma_2}. \ \blacksquare \tag{2.5.3}$$

It should be noted that the expected value of the product of two random variables is equal to the product of their expectations plus their covariance; that is, $E(XY) = \mu_1\mu_2 + \text{cov}(X,Y) = \mu_1\mu_2 + \rho\sigma_1\sigma_2$.

As illustrations, we present two examples. The first is for a discrete model while the second concerns a continuous model.

Example 2.5.1. Reconsider the random vector (X_1, X_2) of Example 2.1.1 where a fair coin is flipped three times and X_1 is the number of heads on the first two flips while X_2 is the number of heads on all three flips. Recall that Table 2.1.1 contains the marginal distributions of X_1 and X_2. By symmetry of these pmfs, we have $E(X_1) = 1$ and $E(X_2) = 3/2$. To compute the correlation coefficient of (X_1, X_2), we next sketch the computation of the required moments:

$$\begin{aligned} E(X_1^2) &= \frac{1}{2} + 2^2\cdot\frac{1}{4} = \frac{3}{2} \Rightarrow \sigma_1^2 = \frac{3}{2} - 1^2 = \frac{1}{2}; \\ E(X_2^2) &= \frac{3}{8} + 4\cdot\frac{3}{8} + 9\cdot\frac{1}{8} = 3 \Rightarrow \sigma_2^2 = 3 - \left(\frac{3}{2}\right)^2 1^2 = \frac{1}{2}; \\ E(X_1X_2) &= \frac{2}{8} + 1\cdot 2\cdot\frac{2}{8} + 2\cdot 2\cdot\frac{1}{8} + 2\cdot 3\cdot\frac{1}{8} = 2 \Rightarrow \text{cov}(X_1,X_2) = 2 - 1\cdot\frac{3}{2} = \frac{1}{2} \end{aligned}$$

From which it follows that $\rho = (1/2)/(\sqrt{(1/2)}\sqrt{3/4}) = 0.816$. $\blacksquare$

Example 2.5.2. Let the random variables X and Y have the joint pdf

$$f(x,y) = \begin{cases} x+y & 0<x<1, \ \ 0<y<1 \\ 0 & \text{elsewhere.} \end{cases}$$

We next compute the correlation coefficient ρ of X and Y. Now

$$\mu_1 = E(X) = \int_0^1\int_0^1 x(x+y)\,dxdy = \frac{7}{12}$$

and

$$\sigma_1^2 = E(X^2) - \mu_1^2 = \int_0^1 \int_0^1 x^2(x+y)\,dxdy - \left(\frac{7}{12}\right)^2 = \frac{11}{144}.$$

Similarly,

$$\mu_2 = E(Y) = \frac{7}{12} \quad \text{and} \quad \sigma_2^2 = E(Y^2) - \mu_2^2 = \frac{11}{144}.$$

The covariance of X and Y is

$$E(XY) - \mu_1\mu_2 = \int_0^1 \int_0^1 xy(x+y)\,dxdy - \left(\frac{7}{12}\right)^2 = -\frac{1}{144}.$$

Accordingly, the correlation coefficient of X and Y is

$$\rho = \frac{-\frac{1}{144}}{\sqrt{(\frac{11}{144})(\frac{11}{144})}} = -\frac{1}{11}. \quad \blacksquare$$

We next establish that, in general, $|\rho| \leq 1$.

Theorem 2.5.1. *For all jointly distributed random variables (X,Y) whose correlation coefficient ρ exists, $-1 \leq \rho \leq 1$.*

Proof: Consider the polynomial in v given by

$$h(v) = E\left\{[(X-\mu_1) + v(Y-\mu_2)]^2\right\}.$$

Then $h(v) \geq 0$, for all v. Hence, the discriminant of $h(v)$ is less than or equal to 0. To obtain the discriminant, we expand $h(v)$ as

$$h(v) = \sigma_1^2 + 2v\rho\sigma_1\sigma_2 + v^2\sigma_2^2.$$

Hence, the discriminant of $h(v)$ is $4\rho^2\sigma_1^2\sigma_2^2 - 4\sigma_2^2\sigma_1^2$. Since this is less than or equal to 0, we have

$$4\rho^2\sigma_1^2\sigma_2^2 \leq 4\sigma_2^2\sigma_1^2 \quad \text{or} \quad \rho^2 \leq 1,$$

which is the result sought. $\blacksquare$

Theorem 2.5.2. *If X and Y are independent random variables then cov$(X,Y) = 0$ and, hence, $\rho = 0$.*

Proof: Because X and Y are independent, it follows from expression (2.4.3) that $E(XY) = E(X)E(Y)$. Hence, by (2.5.2) the covariance of X and Y is 0; i.e., $\rho = 0$. $\blacksquare$

As the following example shows, the converse of this theorem is not true:

Example 2.5.3. Let X and Y be jointly discrete random variables whose distribution has mass $1/4$ at each of the four points $(-1,0), (0,-1), (1,0)$ and $(0,1)$. It follows that both X and Y have the same marginal distribution with range $\{-1,0,1\}$ and respective probabilities $1/4, 1/2$, and $1/4$. Hence, $\mu_1 = \mu_2 = 0$ and a quick calculation shows that $E(XY) = 0$. Thus, $\rho = 0$. However, $P(X = 0, Y = 0) = 0$ while $P(X = 0)P(Y = 0) = (1/2)(1/2) = 1/4$. Thus, X and Y are dependent but the correlation coefficient of X and Y is 0. $\blacksquare$

Although the converse of Theorem 2.5.2 is not true, the contrapositive is; i.e., if $\rho \neq 0$ then X and Y are dependent. For instance, in Example 2.5.1, since $\rho = 0.816$, we know that the random variables X_1 and X_2 discussed in this example are dependent. As discussed in Section 10.8, this contrapositive is often used in Statistics.

Exercise 2.5.7 points out that in the proof of Theorem 2.5.1, the discriminant of the polynomial $h(v)$ is 0 if and only if $\rho = \pm 1$. In that case X and Y are linear functions of one another with probability one; although, as shown, the relationship is degenerate. This suggests the following interesting question: When ρ does not have one of its extreme values, is there a line in the xy-plane such that the probability for X and Y tends to be concentrated in a band about this line? Under certain restrictive conditions this is, in fact, the case, and under those conditions we can look upon ρ as a measure of the intensity of the concentration of the probability for X and Y about that line.

We summarize these thoughts in the next theorem. For notation, let $f(x, y)$ denote the joint pdf of two random variables X and Y and let $f_1(x)$ denote the marginal pdf of X. Recall from Section 2.3 that the conditional pdf of Y, given $X = x$, is

$$f_{2|1}(y|x) = \frac{f(x, y)}{f_1(x)}$$

at points where $f_1(x) > 0$, and the conditional mean of Y, given $X = x$, is given by

$$E(Y|x) = \int_{-\infty}^{\infty} y f_{2|1}(y|x)\, dy = \frac{\displaystyle\int_{-\infty}^{\infty} y f(x, y)\, dy}{f_1(x)},$$

when dealing with random variables of the continuous type. This conditional mean of Y, given $X = x$, is, of course, a function of x, say $u(x)$. In a like vein, the conditional mean of X, given $Y = y$, is a function of y, say $v(y)$.

In case $u(x)$ is a linear function of x, say $u(x) = a + bx$, we say the conditional mean of Y is linear in x; or that Y has a linear conditional mean. When $u(x) = a + bx$, the constants a and b have simple values which we show in the following theorem.

Theorem 2.5.3. *Suppose (X, Y) have a joint distribution with the variances of X and Y finite and positive. Denote the means and variances of X and Y by μ_1, μ_2 and σ_1^2, σ_2^2, respectively, and let ρ be the correlation coefficient between X and Y. If $E(Y|X)$ is linear in X then*

$$E(Y|X) = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(X - \mu_1) \tag{2.5.4}$$

and

$$E(\mathit{Var}(Y|X)) = \sigma_2^2(1 - \rho^2). \tag{2.5.5}$$

Proof: The proof is given in the continuous case. The discrete case follows similarly by changing integrals to sums. Let $E(Y|x) = a + bx$. From

$$E(Y|x) = \frac{\int_{-\infty}^{\infty} yf(x,y)\,dy}{f_1(x)} = a + bx,$$

we have

$$\int_{-\infty}^{\infty} yf(x,y)\,dy = (a+bx)f_1(x). \tag{2.5.6}$$

If both members of Equation (2.5.6) are integrated on x, it is seen that

$$E(Y) = a + bE(X)$$

or

$$\mu_2 = a + b\mu_1, \tag{2.5.7}$$

where $\mu_1 = E(X)$ and $\mu_2 = E(Y)$. If both members of Equation (2.5.6) are first multiplied by x and then integrated on x, we have

$$E(XY) = aE(X) + bE(X^2),$$

or

$$\rho\sigma_1\sigma_2 + \mu_1\mu_2 = a\mu_1 + b(\sigma_1^2 + \mu_1^2), \tag{2.5.8}$$

where $\rho\sigma_1\sigma_2$ is the covariance of X and Y. The simultaneous solution of equations (2.5.7) and (2.5.8) yields

$$a = \mu_2 - \rho\frac{\sigma_2}{\sigma_1}\mu_1 \quad \text{and} \quad b = \rho\frac{\sigma_2}{\sigma_1}.$$

These values give the first result (2.5.4).

Next, the conditional variance of Y is given by

$$\begin{aligned}
\operatorname{Var}(Y|x) &= \int_{-\infty}^{\infty}\left[y - \mu_2 - \rho\frac{\sigma_2}{\sigma_1}(x-\mu_1)\right]^2 f_{2|1}(y|x)\,dy \\
&= \frac{\int_{-\infty}^{\infty}\left[(y - \mu_2) - \rho\frac{\sigma_2}{\sigma_1}(x-\mu_1)\right]^2 f(x,y)\,dy}{f_1(x)}.
\end{aligned} \tag{2.5.9}$$

This variance is nonnegative and is at most a function of x alone. If it is multiplied by $f_1(x)$ and integrated on x, the result obtained is nonnegative. This result is

$$\begin{aligned}
&\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[(y - \mu_2) - \rho\frac{\sigma_2}{\sigma_1}(x-\mu_1)\right]^2 f(x,y)\,dydx \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[(y-\mu_2)^2 - 2\rho\frac{\sigma_2}{\sigma_1}(y-\mu_2)(x-\mu_1) + \rho^2\frac{\sigma_2^2}{\sigma_1^2}(x-\mu_1)^2\right] f(x,y)\,dydx \\
&= E[(Y-\mu_2)^2] - 2\rho\frac{\sigma_2}{\sigma_1}E[(X-\mu_1)(Y-\mu_2)] + \rho^2\frac{\sigma_2^2}{\sigma_1^2}E[(X-\mu_1)^2] \\
&= \sigma_2^2 - 2\rho\frac{\sigma_2}{\sigma_1}\rho\sigma_1\sigma_2 + \rho^2\frac{\sigma_2^2}{\sigma_1^2}\sigma_1^2 \\
&= \sigma_2^2 - 2\rho^2\sigma_2^2 + \rho^2\sigma_2^2 = \sigma_2^2(1-\rho^2),
\end{aligned}$$

which is the desired result. ■

Note that if the variance, Equation (2.5.9), is denoted by $k(x)$, then $E[k(X)] = \sigma_2^2(1-\rho^2) \geq 0$. Accordingly, $\rho^2 \leq 1$, or $-1 \leq \rho \leq 1$. This verifies Theorem 2.5.1 for the special case of linear conditional means.

As a corollary to Theorem 2.5.3, suppose that the variance, Equation (2.5.9), is positive but not a function of x; that is, the variance is a constant $k > 0$. Now if k is multiplied by $f_1(x)$ and integrated on x, the result is k, so that $k = \sigma_2^2(1-\rho^2)$. Thus, in this case, the variance of each conditional distribution of Y, given $X = x$, is $\sigma_2^2(1-\rho^2)$. If $\rho = 0$, the variance of each conditional distribution of Y, given $X = x$, is σ_2^2, the variance of the marginal distribution of Y. On the other hand, if ρ^2 is near 1, the variance of each conditional distribution of Y, given $X = x$, is relatively small, and there is a high concentration of the probability for this conditional distribution near the mean $E(Y|x) = \mu_2 + \rho(\sigma_2/\sigma_1)(x-\mu_1)$. Similar comments can be made about $E(X|y)$ if it is linear. In particular, $E(X|y) = \mu_1 + \rho(\sigma_1/\sigma_2)(y-\mu_2)$ and $E[\mathrm{Var}(X|Y)] = \sigma_1^2(1-\rho^2)$.

Example 2.5.4. Let the random variables X and Y have the linear conditional means $E(Y|x) = 4x + 3$ and $E(X|y) = \frac{1}{16}y - 3$. In accordance with the general formulas for the linear conditional means, we see that $E(Y|x) = \mu_2$ if $x = \mu_1$ and $E(X|y) = \mu_1$ if $y = \mu_2$. Accordingly, in this special case, we have $\mu_2 = 4\mu_1 + 3$ and $\mu_1 = \frac{1}{16}\mu_2 - 3$ so that $\mu_1 = -\frac{15}{4}$ and $\mu_2 = -12$. The general formulas for the linear conditional means also show that the product of the coefficients of x and y, respectively, is equal to ρ^2 and that the quotient of these coefficients is equal to σ_2^2/σ_1^2. Here $\rho^2 = 4(\frac{1}{16}) = \frac{1}{4}$ with $\rho = \frac{1}{2}$ (not $-\frac{1}{2}$), and $\sigma_2^2/\sigma_1^2 = 64$. Thus, from the two linear conditional means, we are able to find the values of μ_1, μ_2, ρ, and σ_2/σ_1, but not the values of σ_1 and σ_2. ■

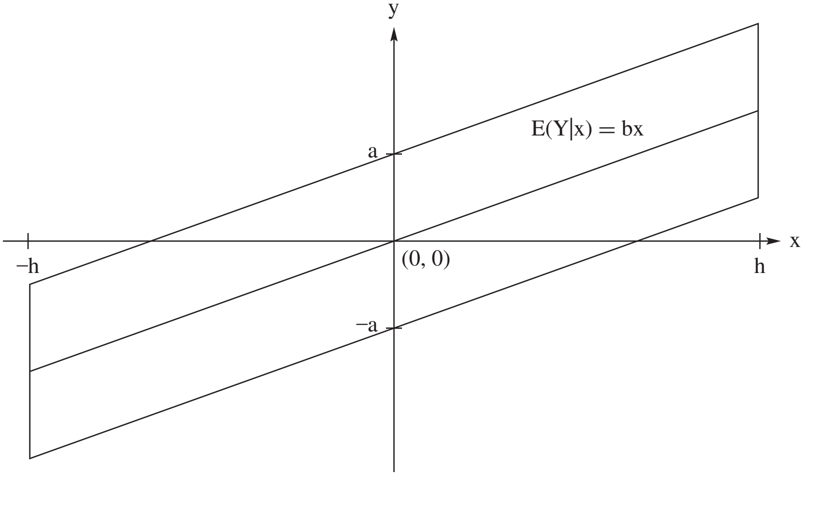

Figure 2.5.1: Illustration for Example 2.5.5.

Example 2.5.5. To illustrate how the correlation coefficient measures the intensity of the concentration of the probability for X and Y about a line, let these random variables have a distribution that is uniform over the area depicted in Figure 2.5.1. That is, the joint pdf of X and Y is

$$f(x,y) = \begin{cases} \frac{1}{4ah} & -a+bx < y < a+bx, \quad -h < x < h \\ 0 & \text{elsewhere.} \end{cases}$$

We assume here that $b \geq 0$, but the argument can be modified for $b \leq 0$. It is easy to show that the pdf of X is uniform, namely

$$f_1(x) = \begin{cases} \int_{-a+bx}^{a+bx} \frac{1}{4ah}\, dy = \frac{1}{2h} & -h < x < h \\ 0 & \text{elsewhere.} \end{cases}$$

The conditional mean and variance are

$$E(Y|x) = bx \quad \text{and} \quad \operatorname{var}(Y|x) = \frac{a^2}{3}.$$

From the general expressions for those characteristics we know that

$$b = \rho\frac{\sigma_2}{\sigma_1} \quad \text{and} \quad \frac{a^2}{3} = \sigma_2^2(1-\rho^2).$$

Additionally, we know that $\sigma_1^2 = h^2/3$. If we solve these three equations, we obtain an expression for the correlation coefficient, namely

$$\rho = \frac{bh}{\sqrt{a^2 + b^2h^2}}.$$

Referring to Figure 2.5.1, we note

1. As a gets small (large), the straight-line effect is more (less) intense and ρ is closer to 1 (0).

2. As h gets large (small), the straight-line effect is more (less) intense and ρ is closer to 1 (0).

3. As b gets large (small), the straight-line effect is more (less) intense and ρ is closer to 1 (0). ■

Recall that in Section 2.1 we introduced the mgf for the random vector (X, Y). As for random variables, the joint mgf also gives explicit formulas for certain moments. In the case of random variables of the continuous type,

$$\frac{\partial^{k+m} M(t_1, t_2)}{\partial t_1^k \partial t_2^m} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^k y^m e^{t_1 x + t_2 y} f(x,y)\, dxdy,$$

so that

$$\left.\frac{\partial^{k+m} M(t_1, t_2)}{\partial t_1^k \partial t_2^m}\right|_{t_1=t_2=0} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^k y^m f(x,y)\, dxdy = E(X^k Y^m).$$

For instance, in a simplified notation that appears to be clear,

$$\begin{aligned}
\mu_1 &= E(X) = \frac{\partial M(0,0)}{\partial t_1} \\
\mu_2 &= E(Y) = \frac{\partial M(0,0)}{\partial t_2} \\
\sigma_1^2 &= E(X^2) - \mu_1^2 = \frac{\partial^2 M(0,0)}{\partial t_1^2} - \mu_1^2 \\
\sigma_2^2 &= E(Y^2) - \mu_2^2 = \frac{\partial^2 M(0,0)}{\partial t_2^2} - \mu_2^2 \\
E[(X-\mu_1)(Y-\mu_2)] &= \frac{\partial^2 M(0,0)}{\partial t_1 \partial t_2} - \mu_1\mu_2,
\end{aligned} \tag{2.5.10}$$

and from these we can compute the correlation coefficient ρ.

It is fairly obvious that the results of equations (2.5.10) hold if X and Y are random variables of the discrete type. Thus the correlation coefficients may be computed by using the mgf of the joint distribution if that function is readily available. An illustrative example follows.

Example 2.5.6 (Example 2.1.10, Continued)**.** In Example 2.1.10, we considered the joint density

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

and showed that the mgf was

$$M(t_1,t_2) = \frac{1}{(1-t_1-t_2)(1-t_2)},$$

for $t_1 + t_2 < 1$ and $t_2 < 1$. For this distribution, equations (2.5.10) become

$$\begin{aligned}
\mu_1 = 1, &\quad \mu_2 = 2 \\
\sigma_1^2 = 1, &\quad \sigma_2^2 = 2 \\
E[(X-\mu_1)(Y-\mu_2)] &= 1.
\end{aligned} \tag{2.5.11}$$

Verification of (2.5.11) is left as an exercise; see Exercise 2.5.5. If, momentarily, we accept these results, the correlation coefficient of X and Y is $\rho = 1/\sqrt{2}$. ■

EXERCISES

2.5.1. Let the random variables X and Y have the joint pmf

(a) $p(x,y) = \frac{1}{3}$, $(x,y) = (0,0), (1,1), (2,2)$, zero elsewhere.

(b) $p(x,y) = \frac{1}{3}$, $(x,y) = (0,2), (1,1), (2,0)$, zero elsewhere.

(c) $p(x,y) = \frac{1}{3}$, $(x,y) = (0,0), (1,1), (2,0)$, zero elsewhere.

In each case compute the correlation coefficient of X and Y.

2.5.2. Let X and Y have the joint pmf described as follows:

(x,y)	$(1,1)$	$(1,2)$	$(1,3)$	$(2,1)$	$(2,2)$	$(2,3)$
$p(x,y)$	$\frac{2}{15}$	$\frac{4}{15}$	$\frac{3}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{4}{15}$

and $p(x,y)$ is equal to zero elsewhere.

(a) Find the means μ_1 and μ_2, the variances σ_1^2 and σ_2^2, and the correlation coefficient ρ.

(b) Compute $E(Y|X=1)$, $E(Y|X=2)$, and the line $\mu_2 + \rho(\sigma_2/\sigma_1)(x-\mu_1)$. Do the points $[k, E(Y|X=k)]$, $k=1,2$, lie on this line?

2.5.3. Let $f(x,y) = 2$, $0 < x < y$, $0 < y < 1$, zero elsewhere, be the joint pdf of X and Y. Show that the conditional means are, respectively, $(1+x)/2$, $0 < x < 1$, and $y/2$, $0 < y < 1$. Show that the correlation coefficient of X and Y is $\rho = \frac{1}{2}$.

2.5.4. Show that the variance of the conditional distribution of Y, given $X = x$, in Exercise 2.5.3, is $(1-x)^2/12$, $0 < x < 1$, and that the variance of the conditional distribution of X, given $Y = y$, is $y^2/12$, $0 < y < 1$.

2.5.5. Verify the results of equations (2.5.11) of this section.

2.5.6. Let X and Y have the joint pdf $f(x,y) = 1$, $-x < y < x$, $0 < x < 1$, zero elsewhere. Show that, on the set of positive probability density, the graph of $E(Y|x)$ is a straight line, whereas that of $E(X|y)$ is not a straight line.

2.5.7. In the proof of Theorem 2.5.1, consider the case when the discriminant of the polynomial $h(v)$ is 0. Show that this is equivalent to $\rho = \pm 1$. Consider the case when $\rho = 1$. Find the unique root of $h(v)$ and then use the fact that $h(v)$ is 0 at this root to show that Y is a linear function of X with probability 1.

2.5.8. Let $\psi(t_1,t_2) = \log M(t_1,t_2)$, where $M(t_1,t_2)$ is the mgf of X and Y. Show that

$$\frac{\partial \psi(0,0)}{\partial t_i}, \quad \frac{\partial^2 \psi(0,0)}{\partial t_i^2}, \quad i = 1,2,$$

and

$$\frac{\partial^2 \psi(0,0)}{\partial t_1 \partial t_2}$$

yield the means, the variances, and the covariance of the two random variables. Use this result to find the means, the variances, and the covariance of X and Y of Example 2.5.6.

2.5.9. Let X and Y have the joint pmf $p(x,y) = \frac{1}{7}$, $(0,0), (1,0), (0,1), (1,1), (2,1), (1,2), (2,2)$, zero elsewhere. Find the correlation coefficient ρ.

2.5.10. Let X_1 and X_2 have the joint pmf described by the following table:

(x_1, x_2)	$(0,0)$	$(0,1)$	$(0,2)$	$(1,1)$	$(1,2)$	$(2,2)$
$p(x_1, x_2)$	$\frac{1}{12}$	$\frac{2}{12}$	$\frac{1}{12}$	$\frac{3}{12}$	$\frac{4}{12}$	$\frac{1}{12}$

Find $p_1(x_1), p_2(x_2), \mu_1, \mu_2, \sigma_1^2, \sigma_2^2$, and ρ.

2.5.11. Let $\sigma_1^2 = \sigma_2^2 = \sigma^2$ be the common variance of X_1 and X_2 and let ρ be the correlation coefficient of X_1 and X_2. Show for $k > 0$ that

$$P[|(X_1 - \mu_1) + (X_2 - \mu_2)| \geq k\sigma] \leq \frac{2(1+\rho)}{k^2}.$$

2.6 Extension to Several Random Variables

The notions about two random variables can be extended immediately to n random variables. We make the following definition of the space of n random variables.

Definition 2.6.1. *Consider a random experiment with the sample space $\mathcal{C}$. Let the random variable X_i assign to each element $c \in \mathcal{C}$ one and only one real number $X_i(c) = x_i$, $i = 1, 2, \ldots, n$. We say that $(X_1, \ldots, X_n)$ is an n-dimensional* **random vector**. *The* **space** *of this random vector is the set of ordered n-tuples $\mathcal{D} = \{(x_1, x_2, \ldots, x_n) : x_1 = X_1(c), \ldots, x_n = X_n(c), c \in \mathcal{C}\}$. Furthermore, let A be a subset of the space $\mathcal{D}$. Then $P[(X_1, \ldots, X_n) \in A] = P(C)$, where $C = \{c : c \in \mathcal{C}$ and $(X_1(c), X_2(c), \ldots, X_n(c)) \in A\}$.*

In this section, we often use vector notation. We denote $(X_1, \ldots, X_n)'$ by the n-dimensional column vector $\mathbf{X}$ and the observed values $(x_1, \ldots, x_n)'$ of the random variables by $\mathbf{x}$. The joint cdf is defined to be

$$F_{\mathbf{X}}(\mathbf{x}) = P[X_1 \leq x_1, \ldots, X_n \leq x_n]. \tag{2.6.1}$$

We say that the n random variables $X_1, X_2, \ldots, X_n$ are of the discrete type or of the continuous type and have a distribution of that type according to whether the joint cdf can be expressed as

$$F_{\mathbf{X}}(\mathbf{x}) = \sum \underset{w_1 \leq x_1, \ldots, w_n \leq x_n}{\cdots} \sum p(w_1, \ldots, w_n),$$

or as

$$F_{\mathbf{X}}(\mathbf{x}) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \cdots \int_{-\infty}^{x_n} f(w_1, \ldots, w_n)\, dw_n \cdots dw_1.$$

For the continuous case,

$$\frac{\partial^n}{\partial x_1 \cdots \partial x_n} F_{\mathbf{X}}(\mathbf{x}) = f(\mathbf{x}), \tag{2.6.2}$$

except possibly on points that have probability zero.

In accordance with the convention of extending the definition of a joint pdf, it is seen that a continuous function f essentially satisfies the conditions of being a pdf if (a) f is defined and is nonnegative for all real values of its argument(s) and (b) its integral over all real values of its argument(s) is 1. Likewise, a point function p essentially satisfies the conditions of being a joint pmf if (a) p is defined and is nonnegative for all real values of its argument(s) and (b) its sum over all real values of its argument(s) is 1. As in previous sections, it is sometimes convenient to speak of the support set of a random vector. For the discrete case, this would be all points in $\mathcal{D}$ that have positive mass, while for the continuous case these would be all points in $\mathcal{D}$ that can be embedded in an open set of positive probability. We use $\mathcal{S}$ to denote support sets.

Example 2.6.1. Let

$$f(x,y,z) = \begin{cases} e^{-(x+y+z)} & 0 < x,y,z < \infty \\ 0 & \text{elsewhere} \end{cases}$$

be the pdf of the random variables X, Y, and Z. Then the distribution function of X, Y, and Z is given by

$$\begin{aligned} F(x,y,z) &= P(X \le x, Y \le y, Z \le z) \\ &= \int_0^z \int_0^y \int_0^x e^{-u-v-w}\, dudvdw \\ &= (1-e^{-x})(1-e^{-y})(1-e^{-z}), \quad 0 \le x,y,z < \infty, \end{aligned}$$

and is equal to zero elsewhere. The relationship (2.6.2) can easily be verified. ■

Let $(X_1, X_2, \ldots, X_n)$ be a random vector and let $Y = u(X_1, X_2, \ldots, X_n)$ for some function u. As in the bivariate case, the expected value of the random variable exists if the n-fold integral

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} |u(x_1, x_2, \ldots, x_n)| f(x_1, x_2, \ldots, x_n)\, dx_1 dx_2 \cdots dx_n$$

exists when the random variables are of the continuous type, or if the n-fold sum

$$\sum_{x_n} \cdots \sum_{x_1} |u(x_1, x_2, \ldots, x_n)| p(x_1, x_2, \ldots, x_n)$$

exists when the random variables are of the discrete type. If the expected value of Y exists, then its expectation is given by

$$E(Y) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} u(x_1, x_2, \ldots, x_n) f(x_1, x_2, \ldots, x_n)\, dx_1 dx_2 \cdots dx_n \qquad (2.6.3)$$

for the continuous case, and by

$$E(Y) = \sum_{x_n} \cdots \sum_{x_1} u(x_1, x_2, \ldots, x_n) p(x_1, x_2, \ldots, x_n) \qquad (2.6.4)$$

for the discrete case. The properties of expectation discussed in Section 2.1 hold for the n-dimensional case also. In particular, E is a linear operator. That is, if

$Y_j = u_j(X_1, \ldots, X_n)$ for $j = 1, \ldots, m$ and each $E(Y_i)$ exists, then

$$E\left[\sum_{j=1}^{m} k_j Y_j\right] = \sum_{j=1}^{m} k_j E\left[Y_j\right], \tag{2.6.5}$$

where $k_1, \ldots, k_m$ are constants.

We next discuss the notions of marginal and conditional probability density functions from the point of view of n random variables. All of the preceding definitions can be directly generalized to the case of n variables in the following manner. Let the random variables $X_1, X_2, \ldots, X_n$ be of the continuous type with the joint pdf $f(x_1, x_2, \ldots, x_n)$. By an argument similar to the two-variable case, we have for every b,

$$F_{X_1}(b) = P(X_1 \le b) = \int_{-\infty}^{b} f_1(x_1)\, dx_1,$$

where $f_1(x_1)$ is defined by the $(n-1)$-fold integral

$$f_1(x_1) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, x_2, \ldots, x_n)\, dx_2 \cdots dx_n.$$

Therefore, $f_1(x_1)$ is the pdf of the random variable X_1 and $f_1(x_1)$ is called the marginal pdf of X_1. The marginal probability density functions $f_2(x_2), \ldots, f_n(x_n)$ of $X_2, \ldots, X_n$, respectively, are similar $(n-1)$-fold integrals.

Up to this point, each marginal pdf has been a pdf of one random variable. It is convenient to extend this terminology to joint probability density functions, which we do now. Let $f(x_1, x_2, \ldots, x_n)$ be the joint pdf of the n random variables $X_1, X_2, \ldots, X_n$, just as before. Now, however, take any group of $k < n$ of these random variables and find the joint pdf of them. This joint pdf is called the marginal pdf of this particular group of k variables. To fix the ideas, take $n = 6$, $k = 3$, and let us select the group X_2, X_4, X_5. Then the marginal pdf of X_2, X_4, X_5 is the joint pdf of this particular group of three variables, namely,

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x_1, x_2, x_3, x_4, x_5, x_6)\, dx_1 dx_3 dx_6,$$

if the random variables are of the continuous type.

Next we extend the definition of a conditional pdf. Suppose $f_1(x_1) > 0$. Then we define the symbol $f_{2,\ldots,n|1}(x_2, \ldots, x_n|x_1)$ by the relation

$$f_{2,\ldots,n|1}(x_2, \ldots, x_n|x_1) = \frac{f(x_1, x_2, \ldots, x_n)}{f_1(x_1)},$$

and $f_{2,\ldots,n|1}(x_2, \ldots, x_n|x_1)$ is called the **joint conditional pdf** of $X_2, \ldots, X_n$, given $X_1 = x_1$. The joint conditional pdf of any $n-1$ random variables, say $X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n$, given $X_i = x_i$, is defined as the joint pdf of $X_1, \ldots, X_n$ divided by the marginal pdf $f_i(x_i)$, provided that $f_i(x_i) > 0$. More generally, the joint conditional pdf of $n-k$ of the random variables, for given values of the remaining k variables, is defined as the joint pdf of the n variables divided by the marginal pdf of the particular group of k variables, provided that the latter pdf is positive. We remark that there are many other conditional probability density functions; for instance, see Exercise 2.3.12.

Because a conditional pdf is the pdf of a certain number of random variables, the expectation of a function of these random variables has been defined. To emphasize the fact that a conditional pdf is under consideration, such expectations are called conditional expectations. For instance, the conditional expectation of $u(X_2, \ldots, X_n)$, given $X_1 = x_1$, is, for random variables of the continuous type, given by

$$E[u(X_2, \ldots, X_n)|x_1] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} u(x_2, \ldots, x_n) f_{2,\ldots,n|1}(x_2, \ldots, x_n|x_1)\, dx_2 \cdots dx_n$$

provided $f_1(x_1) > 0$ and the integral converges (absolutely). A useful random variable is given by $h(X_1) = E[u(X_2, \ldots, X_n)|X_1)]$.

The above discussion of marginal and conditional distributions generalizes to random variables of the discrete type by using pmfs and summations instead of integrals.

Let the random variables $X_1, X_2, \ldots, X_n$ have the joint pdf $f(x_1, x_2, \ldots, x_n)$ and the marginal probability density functions $f_1(x_1), f_2(x_2), \ldots, f_n(x_n)$, respectively. The definition of the independence of X_1 and X_2 is generalized to the mutual independence of $X_1, X_2, \ldots, X_n$ as follows: The random variables $X_1, X_2, \ldots, X_n$ are said to be **mutually independent** if and only if

$$f(x_1, x_2, \ldots, x_n) \equiv f_1(x_1) f_2(x_2) \cdots f_n(x_n),$$

for the continuous case. In the discrete case, $X_1, X_2, \ldots, X_n$ are said to be **mutually independent** if and only if

$$p(x_1, x_2, \ldots, x_n) \equiv p_1(x_1) p_2(x_2) \cdots p_n(x_n).$$

Suppose $X_1, X_2, \ldots, X_n$ are mutually independent. Then

$$\begin{aligned} & P(a_1 < X_1 < b_1,\ a_2 < X_2 < b_2, \ldots, a_n < X_n < b_n) \\ &\quad = \quad P(a_1 < X_1 < b_1) P(a_2 < X_2 < b_2) \cdots P(a_n < X_n < b_n) \\ &\quad = \quad \prod_{i=1}^{n} P(a_i < X_i < b_i), \end{aligned}$$

where the symbol $\prod_{i=1}^{n} \varphi(i)$ is defined to be

$$\prod_{i=1}^{n} \varphi(i) = \varphi(1)\varphi(2) \cdots \varphi(n).$$

The theorem that

$$E[u(X_1)v(X_2)] = E[u(X_1)]E[v(X_2)]$$

for independent random variables X_1 and X_2 becomes, for mutually independent random variables $X_1, X_2, \ldots, X_n$,

$$E[u_1(X_1)u_2(X_2)\cdots u_n(X_n)] = E[u_1(X_1)]E[u_2(X_2)]\cdots E[u_n(X_n)],$$

or

$$E\left[\prod_{i=1}^{n} u_i(X_i)\right] = \prod_{i=1}^{n} E[u_i(X_i)].$$

The moment-generating function (mgf) of the joint distribution of n random variables $X_1, X_2, \ldots, X_n$ is defined as follows. Suppose that

$$E[\exp(t_1X_1 + t_2X_2 + \cdots + t_nX_n)]$$

exists for $-h_i < t_i < h_i,\ i = 1, 2, \ldots, n$, where each h_i is positive. This expectation is denoted by $M(t_1, t_2, \ldots, t_n)$ and it is called the mgf of the joint distribution of $X_1, \ldots, X_n$ (or simply the mgf of $X_1, \ldots, X_n$). As in the cases of one and two variables, this mgf is unique and uniquely determines the joint distribution of the n variables (and hence all marginal distributions). For example, the mgf of the marginal distributions of X_i is $M(0, \ldots, 0, t_i, 0, \ldots, 0),\ i = 1, 2, \ldots, n$; that of the marginal distribution of X_i and X_j is $M(0, \ldots, 0, t_i, 0, \ldots, 0, t_j, 0, \ldots, 0)$; and so on. Theorem 2.4.5 of this chapter can be generalized, and the factorization

$$M(t_1, t_2, \ldots, t_n) = \prod_{i=1}^{n} M(0, \ldots, 0, t_i, 0, \ldots, 0) \tag{2.6.6}$$

is a necessary and sufficient condition for the mutual independence of $X_1, X_2, \ldots, X_n$. Note that we can write the joint mgf in vector notation as

$$M(\mathbf{t}) = E[\exp(\mathbf{t}'\mathbf{X})], \quad \text{for } \mathbf{t} \in B \subset R^n,$$

where $B = \{\mathbf{t} : -h_i < t_i < h_i, i = 1, \ldots, n\}$.

The following is a theorem that proves useful in the sequel. It gives the mgf of a linear combination of independent random variables.

Theorem 2.6.1. *Suppose $X_1, X_2, \ldots, X_n$ are n mutually independent random variables. Suppose, for all $i = 1, 2, \ldots, n$, X_i has mgf $M_i(t)$, for $-h_i < t < h_i$, where $h_i > 0$. Let $T = \sum_{i=1}^{n} k_iX_i$, where $k_1, k_2, \ldots, k_n$ are constants. Then T has the mgf given by*

$$M_T(t) = \prod_{i=1}^{n} M_i(k_it), \quad -\min_i\{h_i\} < t < \min_i\{h_i\}. \tag{2.6.7}$$

Proof. Assume t is in the interval $(-\min_i\{h_i\}, \min_i\{h_i\})$. Then, by independence,

$$\begin{aligned} M_T(t) &= E\left[e^{\sum_{i=1}^{n} tk_iX_i}\right] = E\left[\prod_{i=1}^{n} e^{(tk_i)X_i}\right] \\ &= \prod_{i=1}^{n} E\left[e^{tk_iX_i}\right] = \prod_{i=1}^{n} M_i(k_it), \end{aligned}$$

which completes the proof. ■