CHAPTER 4

Multiple Random Variables

4.1 INTRODUCTION

In many experiments an observation is expressible, not as a single numerical quantity, but as a family of several separate numerical quantities. For example, if a pair of distinguishable dice is tossed, the outcome is a pair (x, y), where x denotes the face value on the first die, and y, the face value on the second die. Similarly, to record the height and weight of every person in a certain community, we need a pair (x, y), where the components represent, respectively, the height and the weight of a particular person. To be able to describe such experiments mathematically, we must study *multidimensional random variables*.

In Section 4.2 we introduce the basic notations involved and study joint, marginal, and conditional distributions. In Section 4.3 we examine independent random variables and investigate some consequences of independence. Section 4.4 deals with functions of several random variables and their induced distributions. In Section 4.5 we consider moments, covariance, and correlation, and in Section 4.6 we study conditional expectation. The last section deals with ordered observations.

4.2 MULTIPLE RANDOM VARIABLES

In this section we study multidimensional RVs. Let $(\Omega, \mathcal{S}, P)$ be a fixed but otherwise arbitrary probability space.

Definition 1. The collection $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ defined on $(\Omega, \mathcal{S}, P)$ into $\mathcal{R}_n$ by

$$\mathbf{X}(\omega) = (X_1(\omega), X_2(\omega), \ldots, X_n(\omega)), \qquad \omega \in \Omega,$$

is called an *n-dimensional* RV if the inverse image of every n-dimensional interval

$$I = \{(x_1, x_2, \ldots, x_n) \colon -\infty < x_i \le a_i, a_i \in \mathcal{R}, i = 1, 2, \ldots, n\}$$

is also in $\mathcal{S}$, that is, if

$$\mathbf{X}^{-1}(I) = \{\omega: X_1(\omega) \le a_1, \dots, X_n(\omega) \le a_n\} \in \mathcal{S} \qquad \text{for } a_i \in \mathcal{R}.$$

Theorem 1. Let $X_1, X_2, \dots, X_n$ be n RVs on $(\Omega, \mathcal{S}, P)$. Then $\mathbf{X} = (X_1, X_2, \dots, X_n)$ is an n-dimensional RV on $(\Omega, \mathcal{S}, P)$.

Proof. Let $I = \{(x_1, x_2, \dots, x_n): -\infty < x_i \le a_i,\ i = 1, 2, \dots, n\}$. Then

$$\begin{aligned}\{(X_1, X_2, \dots, X_n) \in I\} &= \{\omega: X_1(\omega) \le a_1, X_2(\omega) \le a_2, \dots, X_n(\omega) \le a_n\} \\ &= \bigcap_{k=1}^{n} \{\omega: X_k(\omega) \le a_k\} \in \mathcal{S},\end{aligned}$$

as asserted.

From now on we restrict attention to two-dimensional random variables. The discussion for the n-dimensional ($n > 2$) case is similar except when indicated. The development follows closely the one-dimensional case.

Definition 2. The function $F(\cdot, \cdot)$, defined by

(1) $$F(x, y) = P\{X \le x, Y \le y\}, \qquad \text{all } (x, y) \in \mathcal{R}_2,$$

is known as the DF of the RV (X, Y).

Following the discussion in Section 2.3, it is easily shown that

(i) $F(x, y)$ is nondecreasing and continuous from the right with respect to each coordinate, and

(ii) $$\lim_{\substack{x \to +\infty \\ y \to +\infty}} F(x, y) = F(+\infty, +\infty) = 1,$$

$$\lim_{y \to -\infty} F(x, y) = F(x, -\infty) = 0 \qquad \text{for all } x,$$

$$\lim_{x \to -\infty} F(x, y) = F(-\infty, y) = 0 \qquad \text{for all } y.$$

But (i) and (ii) are not sufficient conditions to make any function $F(\cdot, \cdot)$ a DF.

Example 1. Let F be a function (Fig. 1) of two variables defined by

$$F(x, y) = \begin{cases} 0, & x < 0 \text{ or } x + y < 1 \text{ or } y < 0, \\ 1, & \text{otherwise.} \end{cases}$$

Then F satisfies both (i) and (ii) above. However, F is not a DF since

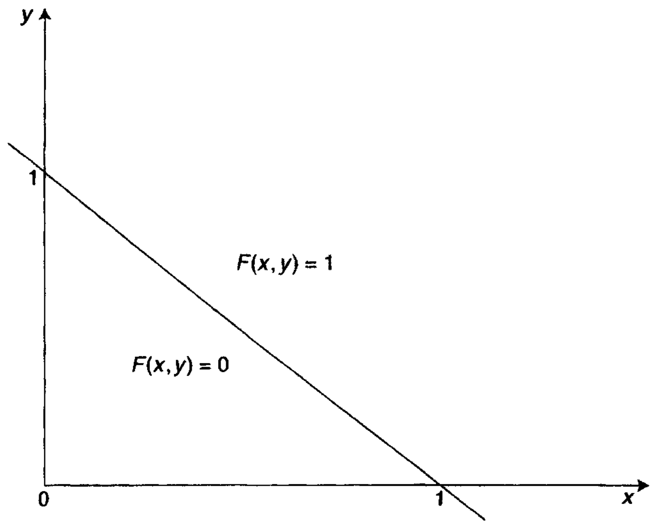

Fig. 1.

$$P\left\{\tfrac{1}{3} < X \le 1, \tfrac{1}{3} < Y \le 1\right\} = F(1, 1) + F\left(\tfrac{1}{3}, \tfrac{1}{3}\right) - F\left(1, \tfrac{1}{3}\right) - F\left(\tfrac{1}{3}, 1\right)$$
$$= 1 + 0 - 1 - 1 = -1 \not\ge 0.$$

Let $x_1 < x_2$ and $y_1 < y_2$. We have

$$\begin{aligned} &P\{x_1 < X \le x_2, y_1 < Y \le y_2\} \\ &\quad = P\{X \le x_2, Y \le y_2\} + P\{X \le x_1, Y \le y_1\} \\ &\qquad - P\{X \le x_1, Y \le y_2\} - P\{X \le x_2, Y \le y_1\} \\ &\quad = F(x_2, y_2) + F(x_1, y_1) - F(x_1, y_2) - F(x_2, y_1) \\ &\quad \ge 0 \end{aligned}$$

for all pairs (x_1, y_1), (x_2, y_2) with $x_1 < x_2$, $y_1 < y_2$, (see Fig. 2).

Theorem 2. A function F of two variables is a DF of some two-dimensional RV if and only if it satisfies the following conditions:

(i) F is nondecreasing and right continuous with respect to both arguments.
(ii) $F(-\infty, y) = F(x, -\infty) = 0$ and $F(+\infty, +\infty) = 1$.

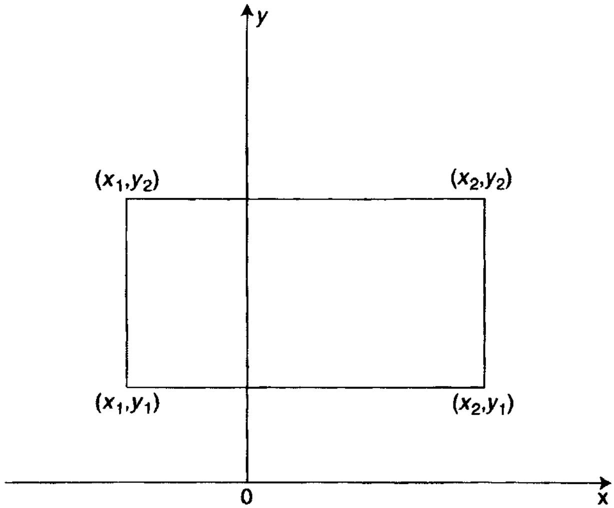

Fig. 2. $\{x_1 < x < x_2,\ y_1 < y \le y_2\}$.

(iii) For every (x_1, y_1), (x_2, y_2) with $x_1 < x_2$ and $y_1 < y_2$ the inequality

$$F(x_2, y_2) - F(x_2, y_1) + F(x_1, y_1) - F(x_1, y_2) \ge 0 \tag{2}$$

holds.

The "if" part of the theorem has already been established. The "only if" part will not be proved here (see Tucker [113, p. 26).

Theorem 2 can be generalized to the n-dimensional case in the following manner.

Theorem 3. A function $F(x_1, x_2, \ldots, x_n)$ is the joint DF of some n-dimensional RV if and only if F is nondecreasing and continuous from the right with respect to all the arguments $x_1, x_2, \ldots, x_n$ and satisfies the following conditions:

(i)
$$\begin{aligned} F(-\infty, x_2, \ldots, x_n) &= F(x_1, -\infty, x_3, \ldots, x_n) \cdots \\ &= F(x_1, \ldots, x_{n-1}, -\infty) = 0, \end{aligned}$$
$$F(+\infty, +\infty, \ldots, +\infty) = 1.$$

(ii) For every $(x_1, x_2, \ldots, x_n) \in \mathcal{R}_n$ and all $\varepsilon_i > 0\,(i = 1, 2, \ldots, n)$, the inequality

$$\begin{aligned} &F(x_1 + \varepsilon_1, x_2 + \varepsilon_2, \ldots, x_n + \varepsilon_n) \\ &\quad - \sum_{i=1}^{n} F(x_1 + \varepsilon_1, \ldots, x_{i-1} + \varepsilon_{i-1}, x_i, x_{i+1} + \varepsilon_{i+1}, \ldots, x_n + \varepsilon_n) \end{aligned} \tag{3}$$

$$
\begin{aligned}
&+\sum_{\substack{i,j=1\\ i<j}}^{n} F(x_1+\varepsilon_1,\dots,x_{i-1}+\varepsilon_{i-1},x_i,x_{i+1}+\varepsilon_{i+1},\dots,\\
&\quad x_{j-1}+\varepsilon_{j-1},x_j,x_{j+1}+\varepsilon_{j+1},\dots,x_n+\varepsilon_n)\\
&+\cdots\\
&+(-1)^n F(x_1,x_2,\dots,x_n)\ge 0
\end{aligned}
$$

holds.

We restrict ourselves here to two-dimensional RVs of the discrete or continuous type, which we now define.

Definition 3. A two-dimensional (or bivariate) RV (X, Y) is said to be of the *discrete* type if it takes on pairs of values belonging to a countable set of pairs A with probability 1. We call every pair (x_i, y_j) that is assumed with positive probability p_{ij} a *jump point* of the DF of (X, Y), and call p_{ij} the *jump* at (x_i, y_j). Here A is the support of the distribution of (X, Y).

Clearly, $\sum_{ij} p_{ij} = 1$. As for the DF of (X, Y), we have

$$F(x, y) = \sum_{B} p_{ij},$$

where $B = \{(i, j)\colon x_i \le x, y_j \le y\}$.

Definition 4. Let (X, Y) be an RV of the discrete type that takes on pairs of values (x_i, y_j), $i = 1, 2, \dots$ and $j = 1, 2, \dots$. We call

$$p_{ij} = P\{X = x_i, Y = y_j\}, \qquad i = 1, 2, \dots, \qquad j = 1, 2, \dots,$$

the *joint probability mass function* (PMF) of (X, Y).

Example 2. A die is rolled, and a coin is tossed independently. Let X be the face value on the die, and let $Y = 0$ if a tail turns up and $Y = 1$ if a head turns up. Then

$$A = \{(1,0), (2,0), \dots, (6,0), (1,1), (2,1), \dots, (6,1)\},$$

and

$$p_{ij} = \frac{1}{12} \qquad \text{for } i = 1, 2, \dots, 6; \quad j = 0, 1.$$

The DF of (X, Y) is given by

$$F(x, y) = \begin{cases} 0, & x < 1, -\infty < y < \infty; -\infty < x < \infty, y < 0, \\ \frac{1}{12}, & 1 \le x < 2, 0 \le y < 1, \\ \frac{1}{6}, & 2 \le x < 3, 0 \le y < 1; 1 \le x < 2, 1 \le y, \\ \frac{1}{4}, & 3 \le x < 4, 0 \le y < 1, \\ \frac{1}{3}, & 4 \le x < 5, 0 \le y < 1; 2 \le x < 3, 1 \le y, \\ \frac{5}{12}, & 5 \le x < 6, 0 \le y < 1, \\ \frac{1}{2}, & 6 \le x, 0 \le y < 1; 3 \le x < 4, 1 \le y, \\ \frac{2}{3}, & 4 \le x < 5, 1 \le y, \\ \frac{5}{6}, & 5 \le x < 6, 1 \le y, \\ 1, & 6 \le x, 1 \le y. \end{cases}$$

Theorem 4. A collection of nonnegative numbers $\{p_{ij}: i = 1, 2, \ldots; j = 1, 2, \ldots\}$ satisfying $\sum_{i,j=1}^{\infty} p_{ij} = 1$ is the PMF of some RV.

The proof of Theorem 4 is easy to construct with the help of Theorem 2.

Definition 5. A two-dimensional RV (X, Y) is said to be of the *continuous* type if there exists a nonnegative function $f(\cdot, \cdot)$ such that for every pair $(x, y) \in \mathcal{R}_2$ we have

(4) $$F(x, y) = \int_{-\infty}^{x} \left[\int_{-\infty}^{y} f(u, v)\, dv \right] du,$$

where F is the DF of (X, Y). The function f is called the (joint) PDF of (X, Y).

Clearly,

$$\begin{aligned} F(+\infty, +\infty) &= \lim_{\substack{x \to +\infty \\ y \to +\infty}} \int_{-\infty}^{x} \int_{-\infty}^{y} f(u, v)\, dv\, du \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(u, v)\, dv\, du = 1. \end{aligned}$$

If f is continuous at (x, y), then

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

Example 3. Let (X, Y) be an RV with joint PDF (Fig. 3) given by

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

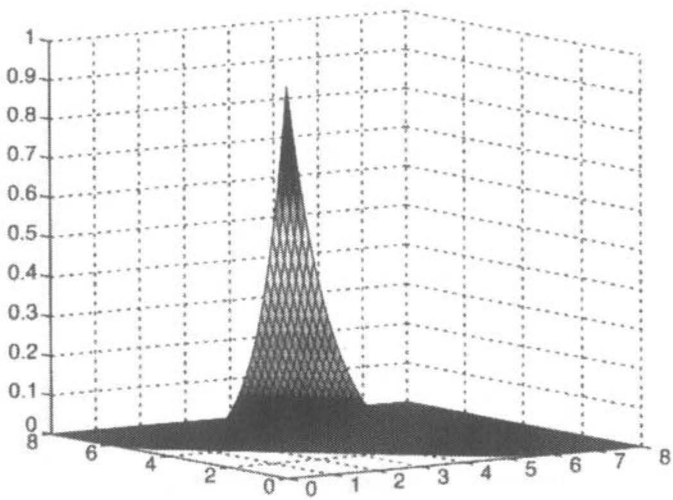

Fig. 3. $f(x, y) = \exp[-(x + y)],\ x > 0,\ y > 0.$

Then

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

Theorem 5. If f is a nonnegative function satisfying $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\, dx\, dy = 1$, then f is the joint density function of some RV.

Proof. For the proof, define

$$F(x, y) = \int_{-\infty}^{x} \left[\int_{-\infty}^{y} f(u, v)\, dv \right] du$$

and use Theorem 2.

Let (X, Y) be a two-dimensional RV with PMF

$$p_{ij} = P\{X = x_i, Y = y_j\}.$$

Then

$$(6) \qquad \sum_{i=1}^{\infty} p_{ij} = \sum_{i=1}^{\infty} P\{X = x_i, Y = y_j\} = P\{Y = y_j\}$$

and

(7) $$\sum_{j=1}^{\infty} p_{ij} = \sum_{j=1}^{\infty} P\{X = x_i, Y = y_j\} = P\{X = x_i\}.$$

Let us write

(8) $$p_{i\cdot} = \sum_{j=1}^{\infty} p_{ij} \quad \text{and} \quad p_{\cdot j} = \sum_{i=1}^{\infty} p_{ij}.$$

Then $p_{i\cdot} \ge 0$ and $\sum_{i=1}^{\infty} p_{i\cdot} = 1$, $p_{\cdot j} \ge 0$ and $\sum_{j=1}^{\infty} p_{\cdot j} = 1$, and $\{p_{i\cdot}\}, \{p_{\cdot j}\}$ represent PMFs.

Definition 6. The collection of numbers $\{p_{i\cdot}\}$ is called the *marginal PMF of X*, and the collection $\{p_{\cdot j}\}$, the *marginal PMF of Y*.

Example 4. A fair coin is tossed three times. Let X = number of heads in three tossings, and Y = difference, in absolute value, between number of heads and number of tails. The joint PMF of (X, Y) is given in the following table:

Y \ X	0	1	2	3	$P\{Y = y\}$
1	0	$\frac{3}{8}$	$\frac{3}{8}$	0	$\frac{6}{8}$
3	$\frac{1}{8}$	0	0	$\frac{1}{8}$	$\frac{2}{8}$
$P\{X = x\}$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$	1

The marginal PMF of Y is shown in the column representing row totals, and the marginal PMF of X, in the row representing column totals.

If (X, Y) is an RV of the continuous type with PDF f, then

(9) $$f_1(x) = \int_{-\infty}^{\infty} f(x, y)\, dy$$

and

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

satisfy $f_1(x) \ge 0$, $f_2(y) \ge 0$, and $\int_{-\infty}^{\infty} f_1(x)\, dx = 1$, $\int_{-\infty}^{\infty} f_2(y)\, dy = 1$. It follows that $f_1(x)$ and $f_2(y)$ are PDFs.

Definition 7. The functions $f_1(x)$ and $f_2(y)$, defined in (9) and (10), are called the *marginal PDF of X* and the *marginal PDF of Y*, respectively.

Example 5. Let (X, Y) be jointly distributed with PDF $f(x, y) = 2$, $0 < x < y < 1$, and $= 0$ otherwise (Fig. 4). Then

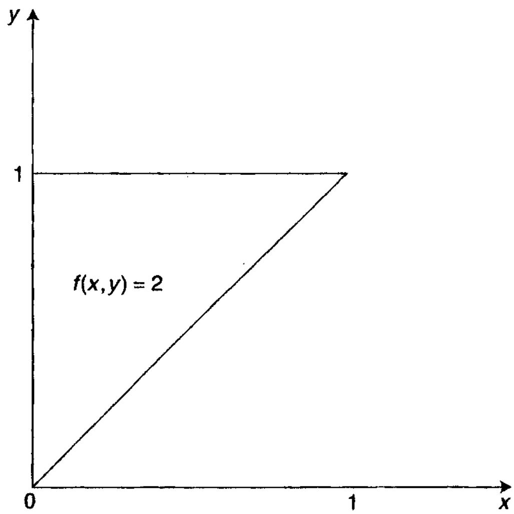

Fig. 4. $f(x, y) = 2,\ 0 < x < y < 1.$

$$f_1(x) = \int_x^1 2\,dy = \begin{cases} 2 - 2x, & 0 < x < 1, \\ 0, & \text{otherwise} \end{cases}$$

and

$$f_2(y) = \int_0^y 2\,dx = \begin{cases} 2y, & 0 < y < 1, \\ 0, & \text{otherwise} \end{cases}$$

are the two marginal density functions.

Definition 8. Let (X, Y) be an RV with DF F. Then the *marginal DF of* X is defined by

$$(11) \qquad F_1(x) = F(x, \infty) = \lim_{y\to\infty} F(x, y)$$

$$= \begin{cases} \sum_{x_i \le x} p_{i\cdot} & \text{if } (X, Y) \text{ is discrete,} \\ \int_{-\infty}^{x} f_1(t)\,dt & \text{if } (X, Y) \text{ is continuous.} \end{cases}$$

A similar definition is given for the *marginal DF of* Y.

In general, given a DF $F(x_1, x_2, \ldots, x_n)$ of an n-dimensional RV $(X_1, X_2, \ldots, X_n)$, one can obtain any k-dimensional $(1 \le k \le n - 1)$ marginal DF from it. Thus the marginal DF of $(X_{i_1}, X_{i_2}, \ldots X_{i_k})$, where $1 \le i_1 < i_2 < \cdots < i_k \le n$, is given by

$$\lim_{\substack{x_i \to \infty \\ i \neq i_1, i_2, \ldots, i_k}} F(x_1, x_2, \ldots, x_n)$$
$$= F(+\infty, \ldots, +\infty, x_{i_1}, +\infty, \ldots, +\infty, \ldots, x_{i_k}, +\infty, \ldots, +\infty).$$

We now consider the concept of *conditional distributions*. Let (X, Y) be an RV of the discrete type with PMF $p_{ij} = P\{X = x_i, Y = y_j\}$. The marginal PMFs are $p_{i\cdot} = \sum_{j=1}^{\infty}$ and $p_{\cdot j} = \sum_{i=1}^{\infty} p_{ij}$. Recall that if $A, B \in \mathcal{S}$ and $PB > 0$, the conditional probability of A, given B, is defined by

$$P\{A \mid B\} = \frac{P(AB)}{P(B)}.$$

Take $A = \{X = x_i\} = \{(x_i, y) \colon -\infty < y < \infty\}$ and $B = \{Y = y_j\} = \{(x, y_j); -\infty < x < \infty\}$, and assume that $PB = P\{Y = y_j\} = p_{\cdot j} > 0$. Then $A \cap B = \{X = x_i, Y = y_j\}$, and

$$P\{A \mid B\} = P\{X = x_i \mid Y = y_j\} = \frac{p_{ij}}{p_{\cdot j}}.$$

For fixed j, the function $P\{X = x_i \mid Y = y_j\} \geq 0$ and $\sum_{i=1}^{\infty} P\{X = x_i \mid Y = y_j\} = 1$. Thus $P\{X = x_i \mid Y = y_j\}$, for fixed j, defines a PMF.

Definition 9. Let (X, Y) be an RV of the discrete type. If $P\{Y = y_j\} > 0$, the function

$$P\{X = x_i \mid Y = y_j\} = \frac{P\{X = x_i, Y = y_j\}}{P\{Y = y_j\}} \tag{12}$$

for fixed j is known as the *conditional PMF of X*, given $Y = y_j$. A similar definition is given for $P\{Y = y_j \mid X = x_i\}$, the *conditional PMF of Y*, given $X = x_i$, provided that $P\{X = x_i\} > 0$.

Example 6. For the joint PMF of Example 4, we have for $Y = 1$,

$$P\{X = i \mid Y = 1\} = \begin{cases} 0, & i = 0, 3, \\ \frac{1}{2}, & i = 1, 2. \end{cases}$$

Similarly,

$$P\{X = i \mid Y = 3\} = \begin{cases} \frac{1}{2}, & \text{if } i = 0, 3, \\ 0, & \text{if } i = 1, 2, \end{cases}$$

$$P\{Y = j \mid X = 0\} = \begin{cases} 0, & \text{if } j = 1, \\ 1, & \text{if } j = 3, \end{cases}$$

and so on.

Next suppose that (X, Y) is an RV of the continuous type with joint PDF f. Since $P\{X = x\} = 0$, $P\{Y = y\} = 0$ for any x, y, the probability $P\{X \le x \mid Y = y\}$, or $P\{Y \le y \mid X = x\}$, is not defined. Let $\varepsilon > 0$, and suppose that $P\{y - \varepsilon < Y \le y + \varepsilon\} > 0$. For every x and every interval $(y - \varepsilon, y + \varepsilon]$, consider the conditional probability of the event $\{X \le x\}$, given that $Y \in (y - \varepsilon, y + \varepsilon]$. We have

$$P\{X \le x \mid y - \varepsilon < Y \le y + \varepsilon\} = \frac{P\{X \le x, y - \varepsilon < Y \le y + \varepsilon\}}{P\{Y \in (y - \varepsilon, y + \varepsilon]\}}.$$

For any fixed interval $(y - \varepsilon, y + \varepsilon]$, the expression above defines the conditional DF of X given that $Y \in (y - \varepsilon, y + \varepsilon]$, provided that $P\{Y \in (y - \varepsilon, y + \varepsilon]\} > 0$. We shall be interested in the case where the limit

$$\lim_{\varepsilon \to 0+} P\{X \le x \mid Y \in (y - \varepsilon, y + \varepsilon]\}$$

exists.

Definition 10. The *conditional DF* of an RV X, given $Y = y$, is defined as the limit

$$\lim_{\varepsilon \to 0+} P\{X \le x \mid Y \in (y - \varepsilon, y + \varepsilon]\}, \tag{13}$$

provided that the limit exists. If the limit exists, we denote it by $F_{X|Y}(x \mid y)$, and define the *conditional density function* of X, given $Y = y$, $f_{X|Y}(x \mid y)$, as a nonnegative function satisfying

$$F_{X|Y}(x \mid y) = \int_{-\infty}^{x} f_{X|Y}(t \mid y)\, dt \qquad \text{for all } x \in \mathcal{R}. \tag{14}$$

For fixed y we see that $f_{X|Y}(x \mid y) \ge 0$ and $\int_{-\infty}^{\infty} f_{X|Y}(x \mid y)\, dx = 1$. Thus $f_{X|Y}(x \mid y)$ is a PDF for fixed y.

Suppose that (X, Y) is an RV of the continuous type with PDF f. At every point (x, y) where f is continuous and the marginal PDF $f_2(y) > 0$ and is continuous, we have

$$\begin{aligned}
F_{X|Y}(x \mid y) &= \lim_{\varepsilon \to 0+} \frac{P\{X \le x, Y \in (y - \varepsilon, y + \varepsilon]\}}{P\{Y \in (y - \varepsilon, y + \varepsilon]\}} \\
&= \lim_{\varepsilon \to 0+} \frac{\int_{-\infty}^{x} \left[\int_{y-\varepsilon}^{y+\varepsilon} f(u, v)\, dv\right] du}{\int_{y-\varepsilon}^{y+\varepsilon} f_2(v)\, dv}.
\end{aligned}$$

Dividing numerator and denominator by 2ε and passing to the limit as $\varepsilon \to 0+$, we have

$$F_{X|Y}(x \mid y) = \frac{\int_{-\infty}^{x} f(u, y)\,du}{f_2(y)}$$
$$= \int_{-\infty}^{x} \left[\frac{f(u, y)}{f_2(y)}\right] du.$$

It follows that there exists a conditional PDF of X, given $Y = y$, that is expressed by

$$f_{X|Y}(x \mid y) = \frac{f(x, y)}{f_2(y)}, \qquad f_2(y) > 0.$$

We have thus proved the following theorem.

Theorem 6. Let f be the PDF of an RV (X, Y) of the continuous type, and let f_2 be the marginal PDF of Y. At every point (x, y) at which f is continuous and $f_2(y) > 0$ and is continuous, the conditional PDF of X, given $Y = y$, exists and is expressed by

(15)
$$f_{X|Y}(x \mid y) = \frac{f(x, y)}{f_2(y)}.$$

Note that

$$\int_{-\infty}^{x} f(u, y)\,du = f_2(y) F_{X|Y}(x \mid y),$$

so that

(16)
$$F_1(x) = \int_{-\infty}^{\infty} \left[\int_{-\infty}^{x} f(u, y)\,du\right] dy = \int_{-\infty}^{\infty} f_2(y) F_{X|Y}(x \mid y)\,dy,$$

where F_1 is the marginal DF of X.

It is clear that similar definitions may be made for the conditional DF and conditional PDF of the RV Y, given $X = x$, and an analog of Theorem 6 holds.

In the general case, let $(X_1, X_2, \ldots, X_n)$ be an n-dimensional RV of the continuous type with PDF $f_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n)$. Also, let $\{i_1 < i_2 < \cdots < i_k, j_1 < j_2 < \cdots < j_l\}$ be a subset of $\{1, 2, \ldots, n\}$. Then

$$F(x_{i_1}, x_{i_2}, \ldots, x_{i_k} \mid x_{j_1}, x_{j_2}, \ldots, x_{j_l})$$
$$= \frac{\int_{-\infty}^{x_{i_1}} \cdots \int_{-\infty}^{x_{i_k}} f_{X_{i_1},\ldots,X_{i_k},X_{j_1},\ldots,X_{j_l}}(u_{i_1}, \ldots, u_{i_k}, x_{j_1}, \ldots, x_{j_l}) \prod_{p=1}^{k} du_{i_p}}{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_{i_1},\ldots,X_{i_k},X_{j_1},\ldots,X_{j_l}}(u_{i_1}, \ldots, u_{i_k}, x_{j_1}, \ldots, x_{j_l}) \prod_{p=1}^{k} du_{i_p}},$$

(17)

provided that the denominator exceeds 0. Here $f_{X_{i_1},\dots,X_{i_k},X_{j_1},\dots,X_{j_l}}$ is the joint marginal PDF of $(X_{i_1}, X_{i_2}, \dots, X_{i_k}, X_{j_1}, X_{j_2}, \dots, X_{j_l})$. The conditional densities are obtained in a similar manner.

The case in which $(X_1, X_2, \dots, X_n)$ is of the discrete type is treated similarly.

Example 7. For the joint PDF of Example 5, we have

$$f_{Y|X}(y \mid x) = \frac{f(x, y)}{f_1(x)} = \frac{1}{1-x}, \qquad x < y < 1,$$

so that the conditional PDF $f_{Y|X}$ is uniform on $(x, 1)$. Also,

$$f_{X|Y}(x \mid y) = \frac{1}{y}, \qquad 0 < x < y,$$

which is uniform on $(0, y)$. Thus

$$P\{Y \ge \tfrac{1}{2} | x = \tfrac{1}{2}\} = \int_{1/2}^{1} \frac{1}{1-\frac{1}{2}}\, dy = 1,$$

$$P\{X \ge \tfrac{1}{3} | y = \tfrac{2}{3}\} = \int_{1/3}^{2/3} \frac{1}{\frac{2}{3}}\, dx = \frac{1}{2}.$$

We conclude this section with a discussion of a technique called *truncation*. We consider two types of truncation, each with a different objective. In probabilistic modeling we use *truncated distributions* when sampling from an incomplete population.

Definition 11. Let X be an RV on $(\Omega, \mathcal{S}, P)$, and $T \in \mathfrak{B}$ such that $0 < P\{X \in T\} < 1$. Then the conditional distribution $P\{X \le x \mid X \in T\}$, defined for any real x, is called the *truncated distribution* of X.

If X is a discrete RV with PMF $p_i = P\{X = x_i\}$, $i = 1, 2, \dots$, the truncated distribution of X is given by

$$(18)\; P\{X = x_i \mid X \in T\} = \frac{P\{X = x_i, X \in T\}}{P\{X \in T\}} = \begin{cases} \dfrac{p_i}{\sum_{x_j \in T} p_j} & \text{if } x_i \in T, \\ 0 & \text{otherwise.} \end{cases}$$

If X is of the continuous type with PDF f, then

$$(19) \qquad P\{X \le x \mid X \in T\} = \frac{P\{X \le x, X \in T\}}{P\{X \in T\}} = \frac{\int_{(-\infty,x]\cap T} f(y)\, dy}{\int_T f(y)\, dy}.$$

The PDF of the truncated distribution is given by

$$h(x) = \begin{cases} \dfrac{f(x)}{\int_T f(y)\,dy}, & x \in T, \\ 0, & x \notin T. \end{cases} \tag{20}$$

Here T is not necessarily a bounded set of real numbers. If we write Y for the RV with distribution function $P\{X \le x \mid X \in T\}$, then Y has support T.

Example 8. Let X be an RV with *standard normal* PDF

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

Let $T = (-\infty, 0]$. Then $P\{X \in T\} = \frac{1}{2}$, since X is symmetric and continuous. For the truncated PDF, we have

$$h(x) = \begin{cases} 2f(x), & -\infty < x \le 0, \\ 0, & x > 0. \end{cases}$$

Some other examples are the truncated Poisson distribution

$$P\{X = k\} = \frac{e^{-\lambda}}{1 - e^{-\lambda}} \frac{x^k}{k!}, \quad k = 1, 2, \dots,$$

where $T = \{X \ge 1\}$, and the truncated uniform distribution

$$f(x) = \frac{1}{\theta}, \quad 0 < x < \theta, \quad \text{and} \quad = 0 \text{ otherwise},$$

where $T = \{X < \theta\}, \theta > 0$.

The second type of truncation is very useful in probability limit theory, especially when the DF F in question does not have a finite mean. Let $a < b$ be finite real numbers. Define the RV X^* by

$$X^* = \begin{cases} X & \text{if } a \le X \le b \\ 0 & \text{if } X < a \quad \text{or} \quad X > b. \end{cases}$$

This method produces an RV for which $P\{a \le X^* \le b\} = 1$ so that X^* has moments of all orders. The special case when $b = c > 0$ and $a = -c$ is quite useful in probability limit theory when we wish to approximate X through bounded RVs. We say that X^c is X *truncated at* c if $X^c = X$ for $|X| \le c$, and $= 0$ for $|X| > c$. Then $E|X^c|^k \le c^k$. Moreover,

$$P\{X \ne X^c\} = P\{|X| > c\},$$

so that c can be selected sufficiently large to make $P\{|X| > c\}$ arbitrarily small. For example, if $E|X|^2 < \infty$, then

$$P\{|X| > c\} \le \frac{E|X|^2}{c^2},$$

and given $\varepsilon > 0$, we can choose c such that $E|X|^2/c^2 < \varepsilon$.

The distribution of X^c is no longer the truncated distribution $P\{X \le x \mid |X| \le c\}$. In fact,

$$F^c(y) = \begin{cases} 0, & y \le -c, \\ F(y) - F(-c), & -c < y < 0, \\ 1 - F(c) + F(y), & 0 \le y < c, \\ 1, & y > c, \end{cases}$$

where F is the DF of X and F^c is that of X^c.

A third type of truncation, sometimes called *Winsorization*, sets

$$X^* = X \quad \text{if } a < X < b, \quad = a \quad \text{if } X \le a, \quad \text{and} \quad = b \quad \text{if } X \ge b.$$

This method also produces an RV for which $P(a \le X^* \le b) = 1$, moments of all orders for X^* exist, but its DF is given by

$$F^*(y) = 0 \quad \text{for } y < a, \quad = F(y) \quad \text{for } a \le y < b, \quad = 1 \quad \text{for } y \ge b.$$

PROBLEMS 4.2

1. Let $F(x, y) = 1$ if $x + 2y \ge 1$, and $= 0$ if $x + 2y < 1$. Does F define a DF in the plane?

2. Let T be a closed triangle in the plane with vertices $(0, 0)$, $(0, \sqrt{2})$, and $(\sqrt{2}, \sqrt{2})$. Let $F(x, y)$ denote the elementary area of the intersection of T with $\{(x_1, x_2)\colon x_1 \le x, x_2 \le y\}$. Show that F defines a DF in the plane, and find its marginal DFs.

3. Let (X, Y) have the joint PDF f defined by $f(x, y) = \frac{1}{2}$ inside the square with corners at the points $(1, 0)$, $(0, 1)$, $(-1, 0)$, and $(0, -1)$ in the (x, y)-plane, and $= 0$ otherwise. Find the marginal PDFs of X and Y and the two conditional PDFs.

4. Let $f(x, y, z) = e^{-x-y-z}$, $x > 0$, $y > 0$, $z > 0$, and $= 0$ otherwise, be the joint PDF of (X, Y, Z). Compute $P\{X < Y < Z\}$ and $P\{X = Y < Z\}$.

5. Let (X, Y) have the joint PDF $f(x, y) = \frac{4}{3}[xy + (x^2/2)]$ if $0 < x < 1$, $0 < y < 2$, and $= 0$ otherwise. Find $P\{Y < 1 \mid X < \frac{1}{2}\}$.

6. For DFs $F, F_1, F_2, \ldots, F_n$ show that

$$1 - \sum_{i=1}^{n}[1 - F_i(x_i)] \le F(x_1, x_2, \ldots, x_n) \le \min_{1\le i\le n} F_i(x_i)$$

for all real numbers $x_1, x_2, \ldots, x_n$ if and only if F_i's are marginal DFs of F.

7. For the *bivariate negative binomial* distribution

$$P\{X = x, Y = y\} = \frac{(x+y+k-1)!}{x!\,y!\,(k-1)!} p_1^x p_2^y (1 - p_1 - p_2)^k,$$

where $x, y = 0, 1, 2, \ldots, k \ge 1$ is an integer, $0 < p_1 < 1, 0 < p_2 < 1$, and $p_1 + p_2 < 1$, find the marginal PMFs of X and Y and the conditional distributions.

In Problems 8 to 10, the bivariate distributions considered are not unique generalizations of the corresponding univariate distributions.

8. For the *bivariate Cauchy* RV (X, Y) with PDF

$$f(x, y) = \frac{c}{2\pi}(c^2 + x^2 + y^2)^{-3/2}, \quad -\infty < x < \infty, \quad -\infty < y < \infty, \quad c > 0,$$

find the marginal PDFs of X and Y. Find the conditional PDF of Y given $X = x$.

9. For the *bivariate beta* RV (X, Y) with PDF

$$f(x, y) = \frac{\Gamma(p_1 + p_2 + p_3)}{\Gamma(p_1)\Gamma(p_2)\Gamma(p_3)} x^{p_1-1} y^{p_2-1} (1 - x - y)^{p_3-1},$$
$$x \ge 0, \quad y \ge 0, \quad x + y \le 1,$$

where p_1, p_2, p_3 are positive real numbers, find the marginal PDFs of X and Y and the conditional PDFs. Find also the conditional PDF of $Y/(1 - X)$, given $X = x$.

10. For the *bivariate gamma* RV (X, Y) with PDF

$$f(x, y) = \frac{\beta^{\alpha+\gamma}}{\Gamma(\alpha)\Gamma(\gamma)} x^{\alpha-1}(y - x)^{\gamma-1} e^{-\beta y}, \qquad 0 < x < y; \quad \alpha, \beta, \gamma > 0,$$

find the marginal PDFs of X and Y and the conditional PDFs. Also, find the conditional PDF of $Y - X$ given $X = x$, and the conditional distribution of X/Y given $Y = y$.

11. For the *bivariate hypergeometric* RV (X, Y) with PMF

$$P\{X = x, Y = y\} = \binom{N}{n}^{-1}\binom{Np_1}{x}\binom{Np_2}{y}\binom{N - Np_1 - Np_2}{n - x - y},$$
$$x, y = 0, 1, 2, \ldots, n,$$

where $x \le Np_1$, $y \le Np_2$, $n - x - y \le N(1 - p_1 - p_2)$, N, n integers with $n \le N$, and $0 < p_1 < 1, 0 < p_2 < 1$ so that $p_1 + p_2 \le 1$, find the marginal PMFs of X and Y and the conditional PMFs.

12. Let X be an RV with PDF $f(x) = 1$ if $0 \le x \le 1$, and $= 0$ otherwise. Let $T = \{x: \frac{1}{3} < x \le \frac{1}{2}\}$. Find the PDF of the truncated distribution of X, its means, and its variance.

13. Let X be an RV with PMF

$$P\{X = x\} = e^{-\lambda}\frac{\lambda^x}{x!}, \quad x = 0, 1, 2, \dots, \lambda > 0.$$

Suppose that the value $x = 0$ cannot be observed. Find the PMF of the truncated RV, its mean, and its variance.

14. Is the function

$$f(x, y, z, u) = \begin{cases} \exp(-u), & 0 < x < y < z < u < \infty \\ 0, & \text{elsewhere} \end{cases}$$

a joint density function? If so, find $P(X \le 7)$ where (X, Y, Z, U) is a random variable with density f.

15. Show that the function defined by

$$f(x, y, z, u) = \frac{24}{(1 + x + y + z + u)^5}, \quad x > 0, \quad y > 0, \quad z > 0, \quad u > 0$$

and zero elsewhere is a joint density function.

(a) Find $P(X > Y > Z > U)$.

(b) Find $P(X + Y + Z + U \ge 1)$.

16. Let (X, Y) have joint density function f and joint distribution function F. Suppose that

$$f(x_1, y_1)f(x_2, y_2) \le f(x_1, y_2)f(x_2, y_1)$$

holds for $x_1 \le a \le x_2$ and $y_1 \le b \le y_2$. Show that

$$F(a, b) \le F_1(a)F_2(b).$$

17. Suppose that (X, Y, Z) are jointly distributed with density

$$f(x, y, z) = \begin{cases} g(x)g(y)g(z), & x > 0, \quad y > 0, \quad z > 0 \\ 0 & \text{elsewhere.} \end{cases}$$

Find $P(X > Y > Z)$. Hence find the probability that $(x, y, z) \notin \{X > Y > Z\}$ or $\{X < Y < Z\}$. (Here g is a density function on $\mathcal{R}$.)

4.3 INDEPENDENT RANDOM VARIABLES

We recall that the joint distribution of a multiple RV uniquely determines the marginal distributions of the component random variables, but in general, knowledge of marginal distributions is not enough to determine the joint distribution. Indeed, it is quite possible to have an infinite collection of joint densities f_α with given marginal densities.

Example 1 (Gumbel [36]). Let f_1, f_2, f_3 be three PDFs with corresponding DFs F_1, F_2, F_3, and let α be a constant, $|\alpha| \le 1$. Define

$$f_\alpha(x_1, x_2, x_3) = f_1(x_1) f_2(x_2) f_3(x_3) \\ \cdot \{1 + \alpha[2F_1(x_1) - 1][2F_2(x_2) - 1][2F_3(x_3) - 1]\}.$$

We show that F_α is a PDF for each α in $[-1, 1]$ and that the collection of densities $\{f_\alpha; -1 \le \alpha \le 1\}$ has the same marginal densities f_1, f_2, f_3. First note that

$$|[2F_1(x_1) - 1][2F_2(x_2) - 1][2F_3(x_3) - 1]| \le 1,$$

so that

$$1 + \alpha[2F_1(x_1) - 1][2F_2(x_2) - 1][2F_3(x_3) - 1] \ge 0.$$

Also,

$$\begin{aligned}
&\iiint f_\alpha(x_1, x_2, x_3)\, dx_1\, dx_2\, dx_3 \\
&= 1 + \alpha\left(\int [2F_1(x_1) - 1] f_1(x_1)\, dx_1\right)\left(\int [2F_2(x_2) - 1] f_2(x_2)\, dx_2\right) \\
&\quad \cdot \left(\int [2F_3(x_3) - 1] f_3(x_3)\, dx_3\right) \\
&= 1 + \alpha\{[F_1^2(x_1)|_{-\infty}^{\infty} - 1][F_2^2(x_2)|_{-\infty}^{\infty} - 1][F_3^2(x_3)|_{-\infty}^{\infty} - 1]\} \\
&= 1.
\end{aligned}$$

It follows that f_α is a density function. That f_1, f_2, f_3 are the marginal densities of f_α follows similarly.

In this section we deal with a very special class of distributions in which the marginal distributions uniquely determine the joint distribution of a multiple RV. First we consider the bivariate case.

Let $F(x, y)$ and $F_1(x)$, $F_2(y)$, respectively, be the joint DF of (X, Y) and the marginal DFs of X and Y.

Definition 1. We say that X and Y are *independent* if and only if

$$F(x, y) = F_1(x) F_2(y) \qquad \text{for all } (x, y) \in \mathcal{R}_2. \tag{1}$$

Lemma 1. If X and Y are independent and $a < c, b < d$ are real numbers, then

(2) $$P\{a < X \le c, b < Y \le d\} = P\{a < X \le c\}P\{b < Y \le d\}.$$

Theorem 1

(a) A necessary and sufficient condition for RVs X, Y of the discrete type to be independent is that

(3) $$P\{X = x_i, Y = y_j\} = P\{X = x_i\}P\{Y = y_j\}$$

for all pairs (x_i, y_j).

(b) Two RVs X and Y, of the continuous type are independent if and only if

(4) $$f(x, y) = f_1(x)f_2(y) \qquad \text{for all } (x, y) \in \mathcal{R}_2,$$

where f, f_1, f_2, respectively, are the joint and marginal densities of X and Y, and f is everywhere continuous.

Proof. (a) Let X, Y be independent. Then from Lemma 1, letting $a \to c$ and $b \to d$, we get

$$P\{X = c, Y = d\} = P\{X = c\}P\{Y = d\}.$$

Conversely,

$$F(x, y) = \sum_B P\{X = x_i, Y = y_j\},$$

where

$$B = \{(i, j)\colon x_i \le x, y_j \le y\}.$$

Then

$$\begin{aligned} F(x, y) &= \sum_B P\{X = x_i\}\, P\{Y = y_j\} \\ &= \sum_{x_i \le x} \left[\sum_{y_j \le y} P\{Y = y_j\} \right] P\{X = x_i\} = F(x)F(y). \end{aligned}$$

The proof of part (b) is left as an exercise.

Corollary. Let X and Y be independent RVs; then $F_{Y|X}(y \mid x) = F_Y(y)$ for all y, and $F_{X|Y}(x \mid y) = F_X(x)$ for all x.

Theorem 1. The RVs X and Y are independent if and only if

(5) $$P\{X \in A_1, Y \in A_2\} = P\{X \in A_1\}\, P\{Y \in A_2\}$$

for all Borel sets A_1 on the x-axis and A_2 on the y-axis.

Theorem 2. Let X and Y be independent RVs and f and g be Borel-measurable functions. Then $f(X)$ and $g(Y)$ are also independent.

Proof. We have

$$\begin{aligned}P\{f(X) \le x, g(Y) \le y\} &= P\{X \in f^{-1}(-\infty, x], Y \in g^{-1}(-\infty, y]\}\\ &= P\{X \in f^{-1}(-\infty, x]\}\, P\{Y \in g^{-1}(-\infty, y]\}\\ &= P\{f(X) \le x\}\, P\{g(Y) \le y\}.\end{aligned}$$

Note that a degenerate RV is independent of any RV.

Example 2. Let X and Y be jointly distributed with PDF

$$f(x, y) = \begin{cases} \dfrac{1+xy}{4}, & |x| < 1, |y| < 1,\\ 0, & \text{otherwise.}\end{cases}$$

Then X and Y are not independent since $f_1(x) = \frac{1}{2}$, $|x| < 1$, and $f_2(y) = \frac{1}{2}$, $|y| < 1$, are the marginal densities of X and Y, respectively. However, the RVs X^2 and Y^2 are independent. Indeed,

$$\begin{aligned}P\{X^2 \le u, Y^2 \le v\} &= \int_{-v^{1/2}}^{v^{1/2}} \int_{-u^{1/2}}^{u^{1/2}} f(x, y)\, dx\, dy\\ &= \frac{1}{4}\int_{-v^{1/2}}^{1/2} \left[\int_{-u^{1/2}}^{u^{1/2}} (1+xy)\, dx\right] dy\\ &= u^{1/2} v^{1/2}\\ &= P\{X^2 \le u\} P\{Y^2 \le v\}.\end{aligned}$$

Note that $\phi(X^2)$ and $\psi(Y^2)$ are independent where ϕ and ψ are Borel-measurable functions. But X is not a Borel-measurable function of X^2.

Example 3. We return to Buffon's needle problem, discussed in Examples 1.2.9 and 1.3.7. Suppose that the RV R, which represents the distance from the center of the needle to the nearest line, is uniformly distributed on $(0, l]$. Suppose further that Θ, the angle that the needle forms with this line, is distributed uniformly on $[0, \pi)$. If R and Θ are assumed to be independent, the joint PDF is given by

$$f_{R,\Theta}(r, \theta) = f_R(r) f_\Theta(\theta) = \begin{cases} \dfrac{1}{l} \cdot \dfrac{1}{\pi} & \text{if } 0 < r \le l, \quad 0 \le \pi,\\ 0 & \text{otherwise.}\end{cases}$$

The needle will intersect the nearest line if and only if

$$\frac{l}{2}\sin\Theta \geq R.$$

Therefore, the required probability is given by

$$P\left\{\sin\Theta \geq \frac{2R}{l}\right\} = \int_0^{\pi}\int_0^{(l/2)\sin\theta} f_{R,\Theta}(r,\theta)\,dr\,d\theta$$
$$= \frac{1}{l\pi}\int_0^{\pi}\frac{l}{2}\sin\theta\,d\theta = \frac{1}{\pi}.$$

Definition 2. A collection of jointly distributed RVs $X_1, X_2, \ldots, X_n$ is said to be *mutually* or *completely independent* if and only if

$$(6) \qquad F(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} F_i(x_i) \qquad \text{for all } (x_1, x_2, \ldots, x_n) \in \mathcal{R}_n,$$

where F is the joint DF of $(X_1, X_2, \ldots, X_n)$, and $F_i (i = 1, 2, \ldots, n)$ is the marginal DF of X_i. $X_1, \ldots, X_n$ are said to be *pairwise independent* if and only if every pair of them are independent.

It is clear that an analog of Theorem 1 holds, but we leave it to the reader to construct it.

Example 4. In Example 1 we cannot write

$$f_\alpha(x_1, x_2, x_3) = f_1(x_1)f_2(x_2)f_3(x_3)$$

except when $\alpha = 0$. It follows that X_1, X_2, and X_3 are not independent except when $\alpha = 0$.

The following result is easy to prove.

Theorem 3. If $X_1, X_2, \ldots, X_n$ are independent, every subcollection $X_{i_1}, X_{i_2}, \ldots, X_{i_k}$ of $X_1, X_2, \ldots, X_n$ is also independent.

Remark 1. It is quite possible for RVs $X_1, X_2, \ldots, X_n$ to be *pairwise independent* without being mutually independent. Let (X, Y, Z) have the joint PMF defined by

$$P\{X = x, Y = y, Z = z\} = \begin{cases} \dfrac{3}{16} & \text{if } (x, y, z) \in \{(0,0,0), (0,1,1), \\ & \quad (1,0,1), (1,1,0)\}, \\ \dfrac{1}{16} & \text{if } (x, y, z) \in \{(0,0,1), (0,1,0), \\ & \quad (1,0,0), (1,1,1)\}. \end{cases}$$

Clearly, X, Y, Z are not independent. (Why?) We have

$$P\{X = x, Y = y\} = \tfrac{1}{4}, \qquad (x, y) \in \{(0,0), (0,1), (1,0), (1,1)\},$$

$$P\{Y = y, Z = z\} = \tfrac{1}{4}, \qquad (y, z) \in \{(0,0), (0,1), (1,0), (1,1)\},$$

$$P\{X = x, Z = z\} = \tfrac{1}{4}, \qquad (x, z) \in \{(0,0), (0,1), (1,0), (1,1)\},$$

$$P\{X = x\} = \tfrac{1}{2}, \qquad x = 0, x = 1,$$

$$P\{Y = y\} = \tfrac{1}{2}, \qquad y = 0, y = 1,$$

and

$$P\{Z = z\} = \tfrac{1}{2}, \qquad z = 0, z = 1.$$

It follows that X and Y, Y and Z, and X and Z are pairwise independent.

Definition 3. A sequence $\{X_n\}$ of RVs is said to be *independent* if for every $n = 2, 3, 4, \ldots$ the RVs $X_1, X_2, \ldots, X_n$ are independent.

Similarly, one can speak of an independent family of RVs.

Definition 4. We say that RVs X and Y are *identically distributed* if X and Y have the same DF, that is,

$$F_X(x) = F_Y(x) \qquad \text{for all } x \in \mathcal{R},$$

where F_X and F_Y are the DFs of X and Y, respectively.

Definition 5. We say that $\{X_n\}$ is a sequence of *independent, identically distributed* (iid) RVs with common law $\mathcal{L}(X)$ if $\{X_n\}$ is an independent sequence of RVs and the distribution of $X_n (n = 1, 2, \ldots)$ is the same as that of X.

According to Definition 4, X and Y are identically distributed if and only if they have the same distribution. It does not follow that $X = Y$ with probability 1 (see Problem 7). If $P\{X = Y\} = 1$, we say that X and Y are *equivalent* RVs. All Definition 4 says is that X and Y are identically distributed if and only if

$$P\{X \in A\} = P\{Y \in A\} \qquad \text{for all } A \in \mathfrak{B}.$$

Nothing is said about the equality of events $\{X \in A\}$ and $\{Y \in A\}$.

Definition 6. Two multiple RVs $(X_1, X_2, \ldots, X_m)$ and $(Y_1, Y_2, \ldots, Y_n)$ are said to be *independent* if

$$(7) \quad F(x_1, x_2, \ldots, x_m, y_1, y_2, \ldots, y_n) = F_1(x_1, x_2, \ldots, x_m) F_2(y_1, y_2, \ldots, y_n)$$

for all $(x_1, x_2, \ldots, x_m, y_1, y_2, \ldots, y_n) \in \mathcal{R}_{m+n}$, where F, F_1, F_2 are the joint distribution functions of $(X_1, X_2, \ldots, X_m, Y_1, Y_2, \ldots, Y_n)$, $(X_1, X_2, \ldots, X_m)$, and $(Y_1, Y_2, \ldots, Y_n)$, respectively.

Of course, the independence of $\mathbf{X} = (X_1, X_2, \ldots, X_m)$ and $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)$ does not imply the independence of components $X_1, X_2, \ldots, X_m$ of $\mathbf{X}$ or components $Y_1, Y_2, \ldots, Y_n$ of $\mathbf{Y}$.

Theorem 4. Let $\mathbf{X} = (X_1, X_2, \ldots, X_m)$ and $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)$ be independent RVs. Then the component X_j of $\mathbf{X}(j = 1, 2, \ldots, m)$ and the component Y_k of $\mathbf{Y}(k = 1, 2, \ldots, n)$ are independent RVs. If h and g are Borel-measurable functions, $h(X_1, X_2, \ldots, X_m)$ and $g(Y_1, Y_2, \ldots, Y_n)$ are independent.

Remark 2. It is possible that an RV X may be independent of Y and also of Z, but X may not be independent of the random vector (Y, Z). See the example in Remark 1.

Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed RVs with common DF F. Then the joint DF G of $(X_1, X_2, \ldots, X_n)$ is given by

$$G(x_1, x_2, \ldots, x_n) = \prod_{j=1}^{n} F(x_j).$$

We note that for any of the $n!$ permutations $(x_{i_1}, x_{i_2}, \ldots, x_{i_n})$ of $(x_1, x_2, \ldots, x_n)$

$$G(x_1, x_2, \ldots, x_n) = \prod_{j=1}^{n} F(x_{i_j}) = G(x_{i_1}, x_{i_2}, \ldots, x_{i_n})$$

so that G is a symmetric function of $x_1, x_2, \ldots, x_n$. Thus $(X_1, X_2, \ldots, X_n) \stackrel{d}{=} (X_{i_1}, X_{i_2}, \ldots, X_{i_n})$, where $\mathbf{X} \stackrel{d}{=} \mathbf{Y}$ means that $\mathbf{X}$ and $\mathbf{Y}$ are identically distributed RVs.

Definition 7. The RVs $X_1, X_2, \ldots, X_n$ are said to be *exchangeable* if

$$(X_1, X_2, \ldots, X_n) \stackrel{d}{=} \left(X_{i_1}, X_{i_2}, \ldots, X_{i_n}\right)$$

for all $n!$ permutations $(i_1, i_2, \ldots, i_n)$ of $(1, 2, \ldots, n)$. The RVs in the sequence $\{X_n\}$ are said to be exchangeable if $X_1, X_2, \ldots, X_n$ are exchangeable for each n.

Clearly if $X_1, X_2, \ldots, X_n$ are exchangeable, then X_i are identically distributed but not necessarily independent.

Example 5. Suppose that $X,\ Y,\ Z$ have joint PDF

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

Then $X,\ Y,\ Z$ are exchangeable but not independent.

Example 6. Let $X_1, X_2, \ldots, X_n$ be iid RVs. Let $S_n = \sum_{j=1}^{n} X_j$, $n = 1, 2, \ldots$ and $Y_k = X_k - S_n/n$, $k = 1, 2, \ldots, n-1$. Then $Y_1, Y_2, \ldots, Y_{n-1}$ are exchangeable.

Theorem 5. Let $X,\ Y$ be exchangeable RVs. Then $X - Y$ has a symmetric distribution.

The proof is simple.

Definition 8. Let X be an RV, and let X' be an RV that is independent of X and $X' \stackrel{d}{=} X$. We call the RV

$$X^s = X - X'$$

the *symmetrized* X.

In view of Theorem 5, X^s is symmetric about zero so that

$$P\{X^s \ge 0\} \ge \tfrac{1}{2} \quad \text{and} \quad P\{X^s \le 0\} \ge \tfrac{1}{2}.$$

If $E|X| < \infty$, then $E|X^s| \le 2E|X| < \infty$, and $EX^s = 0$.

The technique of symmetrization is an important tool in the study of probability limit theorems. We will need the following result later. The proof is left to the reader.

Theorem 6. For $\varepsilon > 0$,

(a) $P\{|X^s| > \varepsilon\} \le 2P\{|X| > \varepsilon/2\}$.
(b) If $a \ge 0$ such that $P\{X \ge a\} \le 1 - p$ and $P\{X \le -a\} \le 1 - p$, then

$$P\{|X^s| \ge \varepsilon\} \ge P\{|X| > a + \varepsilon\}$$

for $\varepsilon > 0$.

PROBLEMS 4.3

1. Let A be a set of k numbers and Ω be the set of all ordered samples of size n from A with replacement. Also, let $\mathcal{S}$ be the set of all subsets of Ω and P be a probability defined on $\mathcal{S}$. Let $X_1, X_2, \ldots, X_n$ be RVs defined on $(\Omega, \mathcal{S}, P)$ by setting

$$X_i(a_1, a_2, \ldots, a_n) = a_i \qquad (i = 1, 2, \ldots, n).$$

Show that $X_1, X_2, \ldots, X_n$ are independent if and only if each sample point is equally likely.

2. Let X_1, X_2 be iid RVs with common PMF

$$P\{X = \pm 1\} = \tfrac{1}{2}.$$

Write $X_3 = X_1 X_2$. Show that X_1, X_2, X_3 are pairwise independent but not independent.

3. Let (X_1, X_2, X_3) be an RV with joint PMF

$$\begin{aligned} f(x_1, x_2, x_3) &= \tfrac{1}{4} && \text{if } (x_1, x_2, x_3) \in A, \\ &= 0 && \text{otherwise,} \end{aligned}$$

where

$$A = \{(1, 0, 0), (0, 1, 0), (0, 0, 1), (1, 1, 1)\}.$$

Are X_1, X_2, X_3 independent? Are X_1, X_2, X_3 pairwise independent? Are $X_1 + X_2$ and X_3 independent?

4. Let X and Y be independent RVs such that XY is degenerate at $c \neq 0$. That is, $P(XY = c) = 1$. Show that X and Y are also degenerate.

5. Let $(\Omega, \mathcal{S}, P)$ be a probability space and $A, B \in \mathcal{S}$. Define X and Y so that

$$X(\omega) = I_A(\omega), \qquad Y(\omega) = I_B(\omega) \qquad \text{for all } \omega \in \Omega.$$

Show that X and Y are independent if and only if A and B are independent.

6. Let $X_1, X_2, \ldots, X_n$ be a set of exchangeable RVs. Then

$$E\left(\frac{X_1 + X_2 + \cdots + X_k}{X_1 + X_2 + \cdots + X_n}\right) = \frac{k}{n}, \qquad 1 \leq k \leq n.$$

7. Let X and Y be identically distributed. Construct an example to show that X and Y need not be equal; that is, $P\{X = Y\}$ need not equal 1.

8. Prove Lemma 1.

9. Let $X_1, X_2, \ldots, X_n$ be RVs with joint PDF f, and let f_j be the marginal PDF of $X_j (j = 1, 2, \ldots, n)$. Show that $X_1, X_2, \ldots, X_n$ are independent if and only if

$$f(x_1, x_2, \ldots, x_n) = \prod_{j=1}^{n} f_j(x_j) \qquad \text{for all } (x_1, x_2, \ldots, x_n) \in \mathcal{R}_n.$$

10. Suppose that two buses, A and B, operate on a route. A person arrives at a certain bus stop on this route at time 0. Let X and Y be the arrival times of buses A and B, respectively, at this bus stop. Suppose that X and Y are independent and have density functions given, respectively, by

$$f_1(x) = \frac{1}{a}, \qquad 0 \le x \le a, \qquad \text{and zero elsewhere,}$$

and

$$f_2(y) = \frac{1}{b}, \qquad 0 \le y \le b, \qquad \text{and zero otherwise.}$$

What is the probability that bus A will arrive before bus B?

11. Consider two batteries, one of brand A and the other of brand B. Brand A batteries have a length of life with density function

$$f(x) = 3\lambda x^2 \exp(-\lambda x^3), \qquad x > 0, \qquad \text{and zero elsewhere}$$

whereas brand B batteries have a length of life with density function given by

$$g(x) = 3\mu y^2 \exp(-\mu y^3), \qquad y > 0, \qquad \text{and zero elsewhere.}$$

Brand A and brand B batteries operate independently and are put to a test. What is the probability that brand B battery will outlast brand A? In particular, what is the probability if $\lambda = \mu$?

12. (a) Let (X, Y) have joint density f. Show that X and Y are independent if and only if for some constant $k > 0$ and nonnegative functions f_1 and f_2,

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

for all $x, y \in \mathcal{R}$.

(b) Let $A = \{f_X(x) > 0\}$, $B = \{f_Y(y) > 0\}$, and f_X, f_Y are marginal densities of X and Y, respectively. Show that if X and Y are independent, then $\{f > 0\} = A \times B$.

13. If ϕ is the CF of X, show that the CF of X^s is real and even.

14. Let X, Y be jointly distributed with PDF $f(x, y) = (1 - x^3 y)/4$ for $|x| < 1$, $|y| < 1$, and $= 0$ otherwise. Show that $X \stackrel{d}{=} Y$ and that $X - Y$ has a symmetric distribution.

4.4 FUNCTIONS OF SEVERAL RANDOM VARIABLES

Let $X_1, X_2, \ldots, X_n$ be RVs defined on a probability space $(\Omega, \mathcal{S}, P)$. In practice we deal with functions of $X_1, X_2, \ldots, X_n$ such as $X_1 + X_2$, $X_1 - X_2$, $X_1 X_2$,

$\min(X_1, \ldots, X_n)$, and so on. Are these also RVs? If so, how do we compute their distribution given the joint distribution of $X_1, X_2, \ldots, X_n$?

What functions of $(X_1, X_2, \ldots, X_n)$ are RVs?

Theorem 1. Let $g: \mathcal{R}_n \to \mathcal{R}_m$ be a Borel-measurable function; that is, if $B \in \mathfrak{B}_m$, then $g^{-1}(B) \in \mathfrak{B}_n$. If $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ is an n-dimensional RV ($n \geq 1$), then $g(\mathbf{X})$ is an m-dimensional RV.

Proof. For $B \in \mathfrak{B}_m$,

$$\{g(X_1, X_2, \ldots, X_n) \in B\} = \{(X_1, X_2, \ldots, X_n) \in g^{-1}(B)\},$$

and since $g^{-1}(B) \in \mathfrak{B}_n$, it follows that $\{(X_1, X_2, \ldots, X_n) \in g^{-1}(B)\} \in \mathcal{S}$, which concludes the proof.

In particular, if $g: R_n \to R_m$ is a continuous function, then $g(X_1, X_2, \ldots, X_n)$ is an RV.

How do we compute the distribution of $g(X_1, X_2, \ldots, X_n)$? There are several ways to go about it. We first consider the method of distribution functions. Suppose that $Y = g(X_1, \ldots, X_n)$ is real-valued, and let $y \in \mathcal{R}$. Then

$$P\{Y \leq y\} = P(g(X_1, \ldots, X_n) \leq y)$$
$$= \begin{cases} \displaystyle\sum_{\{(x_1,\ldots,x_n):g(x_1,\ldots,x_n)\leq y\}} P(X_1 = x_1, \ldots, X_n = x_n) & \text{in the discrete case} \\ \displaystyle\int_{\{(x_1,\ldots,x_n):g(x_1,\ldots,x_n)\leq y\}} f(x_1, \ldots, x_n)\, dx_1 \cdots dx_n & \text{in the continuous case} \end{cases}$$

where in the continuous case f is the joint PDF of $(X_1, \ldots, X_n)$.

In the continuous case we can obtain the PDF of $Y = g(X_1, \ldots, X_n)$ by differentiating the DF $P\{Y \leq y\}$ with respect to y provided that Y is also of the continuous type. In the discrete case it is easier to compute $P\{g(X_1, \ldots, X_n) = y\}$.

We take a few examples,

Example 1. Consider the bivariate negative binomial distribution with PMF

$$P\{X = x, Y = y\} = \frac{(x + y + k - 1)!}{x!\, y!\, (k - 1)!} p_1^x p_2^y (1 - p_1 - p_2)^k,$$

where $x, y = 0, 1, 2, \ldots$; $k \geq 1$ is an integer; $p_1, p_2 \in (0, 1)$; and $p_1 + p_2 < 1$. Let us find the PMF of $U = X + Y$. We introduce an RV $V = Y$ (see Remark 1 below) so that $u = x + y$, $v = y$ represents a one-to-one mapping of $A = \{(x, y) : x, y = 0, 1, 2, \ldots\}$ onto the set $B = \{(u, v)\colon v = 0, 1, 2, \ldots, u;\ u = 0, 1, 2, \ldots\}$ with inverse map $x = u - v$, $y = v$. It follows that the joint PMF of (U, V) is given by

$$P\{U=u, V=v\} = \begin{cases} \dfrac{(u+k-1)!}{(u-v)!\,v!\,(k-1)!} p_1^{u-v} p_2^v (1-p_1-p_2)^k & \text{for } (u,v) \in B, \\ 0 & \text{otherwise.} \end{cases}$$

The marginal PMF of U is given by

$$\begin{aligned} P\{U=u\} &= \frac{(u+k-1)!\,(1-p_1-p_2)^k}{(k-1)!\,u!} \sum_{v=0}^{u} \binom{u}{v} p_1^{u-v} p_2^v \\ &= \frac{(u+k-1)!\,(1-p_1-p_2)^k}{(k-1)!\,u!} (p_1+p_2)^u \\ &= \binom{u+k-1}{u} (p_1+p_2)^u (1-p_1-p_2)^k \qquad (u=0,1,2,\ldots). \end{aligned}$$

Example 2. Let (X_1, X_2) have uniform distribution on the triangle $\{0 \le x_1 \le x_2 \le 1\}$; that is, (X_1, X_2) has joint density function

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

Let $Y = X_1 + X_2$. Then for $y < 0$, $P(Y \le y) = 0$, and for $y > 2$, $P(Y \le y) = 1$. For $0 \le y \le 2$, we have

$$P(Y \le y) = P(X_1 + X_2 \le y) = \iint\limits_{\substack{0 \le x_1 \le x_2 \le 1 \\ x_1 + x_2 \le y}} f(x_1, x_2)\, dx_1\, dx_2.$$

There are two cases to consider according to whether $0 \le y \le 1$ or $1 \le y \le 2$ (Fig. 1a and b). In the former case,

$$P(Y \le y) = \int_{x_1=0}^{y/2} \left(\int_{x_2=x_1}^{y-x_1} 2\, dx_2 \right) dx_1 = 2 \int_0^{y/2} (y - 2x_1)\, dx_1 = \frac{y^2}{2}$$

and in the latter case,

$$\begin{aligned} P(Y \le y) &= 1 - P(Y > y) = 1 - \int_{x_2=y/2}^{1} \left(\int_{x_1=y-x_2}^{x_2} 2\, dx_1 \right) dx_2 \\ &= 1 - 2 \int_{y/2}^{1} (2x_2 - y)\, dx_1 = 1 - \frac{(y-2)^2}{2}. \end{aligned}$$

Hence the density function of Y is given by

$$f_Y(y) = \begin{cases} y, & 0 \le y \le 1, \\ 2 - y, & 1 \le y \le 2, \\ 0, & \text{elsewhere.} \end{cases}$$

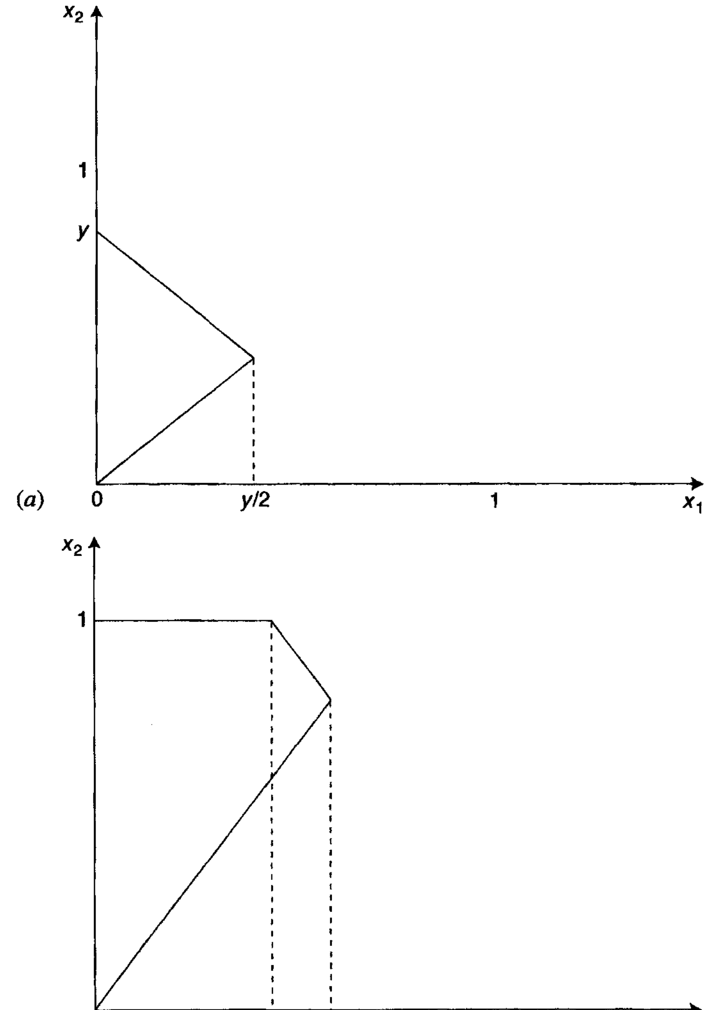

Fig. 1. (*a*) $\{x_1 + x_2 \le y,\ 0 < x_1 \le x_2 \le 1,\ 0 < y \le 1\}$; (*b*) $\{x_1 + x_2 \le y,\ 0 \le x_1 \le x_2 \le 1 \le y \le 2\}$.

The method of distribution functions can also be used in the case when g takes values in $\mathcal{R}_m$, $1 \le m \le n$, but the integration becomes more involved.

Example 3. Let X_1 be the time that a customer takes from getting in line at a service desk in a bank to completion of service, and let X_2 be the time she waits in line before she reaches the service desk. Then $X_1 \ge X_2$ and $X_1 - X_2$ is the service time of the customer. Suppose that the joint density of (X_1, X_2) is given by

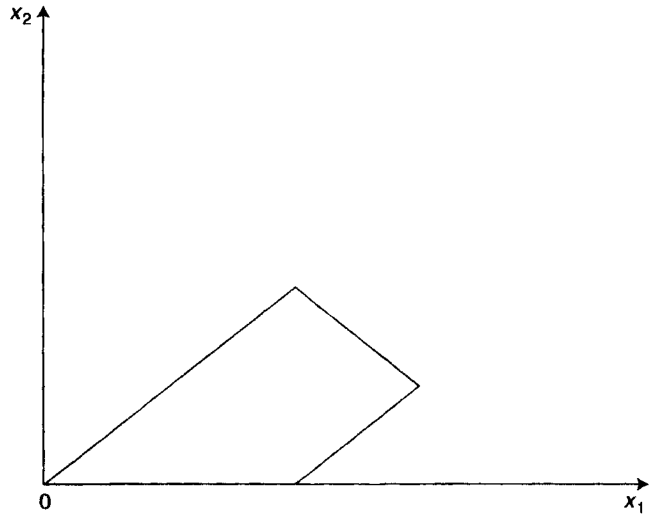

Fig. 2. $A = \{x_1 + x_2 \le y_1,\ x_1 - x_2 \le y_2,\ 0 \le x_2 \le x_1 < \infty\}$.

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

Let $Y_1 = X_1 + X_2$ and $Y_2 = X_1 - X_2$. Then the joint distribution of (Y_1, Y_2) is given by

$$P(Y_1 \le y_1,\ Y_2 \le y_2) = \int\int_A f(x_1, x_2)\, dx_1\, dx_2,$$

where $A = \{(x_1, x_2)\colon x_1 + x_2 \le y_1,\ x_1 - x_2 \le y_2,\ 0 \le x_2 \le x_1 < \infty\}$. Clearly, $x_1 + x_2 \ge x_1 - x_2$, so that the set A is as shown in Fig. 2. It follows that

$$\begin{aligned} P(Y_1 \le y_1,\ y_2 \le y_2) &= \int_{x_2=0}^{(y_1-y_2)/2} \left(\int_{x_1=x_2}^{x_2+y_2} e^{-x_1}\, dx_1 \right) dx_2 \\ &\quad + \int_{x_2=(y_1-y_2)/2}^{y_1/2} \left(\int_{x_1=x_2}^{y_1-x_2} e^{-x_1}\, dx_1 \right) dx_2 \\ &= \int_0^{(y_1-y_2)/2} e^{-x_2}(1 - e^{-y_2})\, dx_2 \\ &\quad + \int_{(y_1-y_2)/2}^{y_1/2} (e^{-x_2} - e^{-y_1+x_2})\, dx_2 \end{aligned}$$

$$
\begin{aligned}
&= (1 - e^{-y_2})(1 - e^{-(y_1 - y_2)/2}) \\
&\quad + (e^{-(y_1 - y_2)/2} - e^{-y_1/2}) - e^{-y_1}(e^{y_1/2} - e^{(y_1 - y_2)/2}) \\
&= 1 - e^{-y_2} - 2e^{-y_1/2} + 2e^{-(y_1 + y_2)/2}.
\end{aligned}
$$

Hence the joint density of Y_1, Y_2 is given by

$$
f_{Y_1,Y_2}(y_1, y_2) = \begin{cases} \frac{1}{2} e^{-(y_1 + y_2)/2}, & 0 \le y_2 \le y_1 < \infty, \\ 0, & \text{elsewhere.} \end{cases}
$$

The marginal densities of Y_1, Y_2 are easily obtained as

$$
f_{y_1}(y_1) = e^{-y_1} \qquad \text{for } y_1 \ge 0, \text{ and 0 elsewhere;}
$$

and

$$
f_{y_2}(y_2) = e^{-y_2/2}(1 - e^{-y_2/2}) \qquad \text{for } y_2 \ge 0, \text{ and 0 elsewhere.}
$$

We next consider the method of transformations. Let $(X_1, \ldots, X_n)$ be jointly distributed with continuous PDF $f(x_1, x_2, \ldots, x_n)$, and let $\mathbf{y} = \mathbf{g}(x_1, x_2, \ldots, x_n) = (y_1, y_2, \ldots, y_n)$, where

$$
y_i = g_i(x_1, x_2, \ldots, x_n), \qquad i = 1, 2, \ldots, n
$$

be a mapping of $\mathcal{R}_n$ to $\mathcal{R}_n$. Then

$$
\begin{aligned}
P\{(Y_1, Y_2, \ldots, Y_n) \in B\} &= P\{(X_1, X_2, \ldots, X_n) \in \mathbf{g}^{-1}(B)\} \\
&= \int_{\mathbf{g}^{-1}(B)} f(x_1, x_2, \ldots, x_n) \prod_{i=1}^{n} dx_i,
\end{aligned}
$$

where $\mathbf{g}^{-1}(B) = \{\mathbf{x} = (x_1, x_2, \ldots, x_n) \in \mathcal{R}_n : \mathbf{g}(\mathbf{x}) \in B\}$. Let us choose B to be the n-dimensional interval

$$
B = B_{\mathbf{y}} = \{(y_1', y_2', \ldots, y_n') : -\infty < y_i' \le y_i, i = 1, 2, \ldots, n\}.
$$

Then the joint DF of $\mathbf{Y}$ is given by

$$
\begin{aligned}
P\{\mathbf{Y} \in B_{\mathbf{y}}\} = G_{\mathbf{Y}}(\mathbf{y}) &= P\{g_1(\mathbf{X}) \le y_1, g_2(\mathbf{X}) \le y_2, \ldots, g_n(\mathbf{X}) \le y_n\} \\
&= \int \cdots \int_{\mathbf{g}^{-1}(B_{\mathbf{y}})} f(x_1, x_2, \ldots, x_n) \prod_{i=1}^{n} dx_i,
\end{aligned}
$$

and (if $G_{\mathbf{Y}}$ is absolutely continuous) the PDF of $\mathbf{Y}$ is given by

$$w(\mathbf{y}) = \frac{\partial^n G_{\mathbf{Y}}(\mathbf{y})}{\partial y_1\, \partial y_2 \cdots \partial y_n}$$

at every continuity point $\mathbf{y}$ of w. Under certain conditions it is possible to write w in terms of f by making a change of variable in the multiple integral.

Theorem 2. Let $(X_1, X_2, \ldots, X_n)$ be an n-dimensional RV of the continuous type with PDF $f(x_1, x_2, \ldots, x_n)$.

(a) Let

$$\begin{aligned} y_1 &= g_1(x_1, x_2, \ldots, x_n), \\ y_2 &= g_2(x_1, x_2, \ldots, x_n), \\ &\vdots \qquad\qquad \vdots \\ y_n &= g_n(x_1, x_2, \ldots, x_n) \end{aligned}$$

be a one-to-one mapping of $\mathcal{R}_n$ into itself; that is, there exists the inverse transformation

$$x_1 = h_1(y_1, y_2, \ldots, y_n), \quad x_2 = h_2(y_1, y_2, \ldots, y_n), \quad \ldots, \quad x_n = h_n(y_1, y_2, \ldots, y_n)$$

defined over the range of the transformation.

(b) Assume that both the mapping and its inverse are continuous.

(c) Assume that the partial derivatives

$$\frac{\partial x_i}{\partial y_j}, \qquad 1 \le i \le n, \quad 1 \le j \le n,$$

exist and are continuous.

(d) Assume that the Jacobian J of the inverse transformation

$$J = \frac{\partial(x_1, \ldots, x_n)}{\partial(y_1, \ldots, y_n)} = \begin{vmatrix} \dfrac{\partial x_1}{\partial y_1} & \dfrac{\partial x_1}{\partial y_2} & \cdots & \dfrac{\partial x_1}{\partial y_n} \\ \dfrac{\partial x_2}{\partial y_1} & \dfrac{\partial x_2}{\partial y_2} & \cdots & \dfrac{\partial x_2}{\partial y_n} \\ \vdots & \vdots & & \vdots \\ \dfrac{\partial x_n}{\partial y_1} & \dfrac{\partial x_n}{\partial y_2} & \cdots & \dfrac{\partial x_n}{\partial y_n} \end{vmatrix}$$

is different from zero for $(y_1, y_2, \ldots, y_n)$ in the range of the transformation.

Then $(Y_1, Y_2, \dots, Y_n)$ has a joint absolutely continuous DF with PDF given by

$$(1) \quad w(y_1, y_2, \dots, y_n) = |J| f(h_1(y_1, \dots y_n), \dots, h_n(y_1, \dots, y_n)).$$

Proof. For $(y_1, y_2, \dots, y_n) \in \mathcal{R}_n$, let

$$B = \{(y_1', y_2', \dots, y_n') \in \mathcal{R}_n: -\infty < y_i' \le y_i, \quad i = 1, 2, \dots, n\}.$$

Then

$$\mathbf{g}^{-1}(B) = \{\mathbf{x} \in \mathcal{R}_n: \mathbf{g}(\mathbf{x}) \in B\} = \{(x_1, x_2, \dots, x_n): g_i(\mathbf{x}) \le y_i, \quad i = 1, 2, \dots, n\}$$

and

$$\begin{aligned} G_{\mathbf{Y}}(\mathbf{y}) &= P\{\mathbf{Y} \in B\} = P\{\mathbf{X} \in \mathbf{g}^{-1}(B)\} \\ &= \int \cdots \int_{\mathbf{g}^{-1}(B)} f(x_1, x_2, \dots, x_n)\, dx_1\, dx_2 \cdots dx_n \\ &= \int_{-\infty}^{y_1} \cdots \int_{-\infty}^{y_n} f(h_1(\mathbf{y}), \dots, h_n(\mathbf{y})) \left| \frac{\partial(x_1, x_2, \dots, x_n)}{\partial(y_1, y_2, \dots, y_n)} \right| dy_1 \cdots dy_n. \end{aligned}$$

Result (1) now follows on differentiation of DF $G_{\mathbf{Y}}$.

Remark 1. In actual applications we will not know the mapping from $x_1, x_2, \dots, x_n$ to $y_1, y_2, \dots, y_n$ completely, but one or more of the functions g_i will be known. If only k, $1 \le k < n$, of the g_i's are known, we introduce arbitrarily $n - k$ functions such that the conditions of the theorem are satisfied. To find the joint marginal density of these k variables, we simply integrate the w function over all the $n - k$ variables that were introduced arbitrarily.

Remark 2. An analog of Theorem 2.5.4 holds, which we state without proof.

Let $\mathbf{X} = (X_1, X_2, \dots, X_n)$ be an RV of the continuous type with joint PDF f, and let $y_i = g_i(x_1, x_2, \dots, x_n)$, $i = 1, 2, \dots, n$, be a mapping of $\mathcal{R}_n$ into itself. Suppose that for each $\mathbf{y}$ the transformation $\mathbf{g}$ has a finite number $k = k(\mathbf{y})$ of inverses. Suppose further that $\mathcal{R}_n$ can be partitioned into k disjoint sets $A_1, A_2, \dots, A_k$, such that the transformation $\mathbf{g}$ from $A_i (i = 1, 2, \dots, n)$ into $\mathcal{R}_n$ is one-to-one with inverse transformation

$$x_1 = h_{1_i}(y_1, y_2, \dots, y_n), \quad \dots, \quad x_n = h_{n_i}(y_1, y_2, \dots, y_n), \qquad i = 1, 2, \dots, k.$$

Suppose that the first partial derivatives are continuous and that each Jacobian

$$J_i = \begin{vmatrix} \dfrac{\partial h_{1i}}{\partial y_1} & \dfrac{\partial h_{1i}}{\partial y_2} & \cdots & \dfrac{\partial h_{1i}}{\partial y_n} \\ \dfrac{\partial h_{2i}}{\partial y_1} & \dfrac{d h_{2i}}{\partial y_2} & \cdots & \dfrac{\partial h_{2i}}{\partial y_n} \\ \vdots & \vdots & & \vdots \\ x\dfrac{d h_{ni}}{d y_1} & \dfrac{\partial h_{ni}}{\partial y_2} & \cdots & \dfrac{\partial h_{ni}}{\partial y_n} \end{vmatrix}$$

is different from zero in the range of the transformation. Then the joint PDF of $\mathbf{Y}$ is given by

$$w(y_1, y_2, \ldots, y_n) = \sum_{i=1}^{k} |J_i| f(h_{1i}(y_1, y_2, \ldots, y_n), \ldots, h_{ni}(y_1, y_2, \ldots, y_n)).$$

Example 4. Let X_1, X_2, X_3 be iid RVs with common *exponential* density function

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

Also, let

$$Y_1 = X_1 + X_2 + X_3, \quad Y_2 = \frac{X_1 + X_2}{X_1 + X_2 + X_3}, \quad \text{and} \quad Y_3 = \frac{X_1}{X_1 + X_2}.$$

Then

$$x_1 = y_1 y_2 y_3, \quad x_2 = y_1 y_2 - x_1 = y_1 y_2 (1 - y_3), \quad \text{and}$$
$$x_3 = y_1 - y_1 y_2 = y_1 (1 - y_2).$$

The Jacobian of transformation is given by

$$J = \begin{vmatrix} y_2 y_3 & y_1 y_3 & y_1 y_2 \\ y_2(1 - y_3) & y_1(1 - y_3) & -y_1 y_2 \\ 1 - y_2 & -y_1 & 0 \end{vmatrix} = -y_1^2 y_2.$$

Note that $0 < y_1 < \infty$, $0 < y_2 < 1$, and $0 < y_3 < 1$. Thus the joint PDF of Y_1, Y_2, Y_3 is given by

$$\begin{aligned} w(y_1, y_2, y_3) &= y_1^2 y_2 e^{-y_1} \\ &= (2y_2)\left(\tfrac{1}{2} y_1^2 e^{-y_1}\right), \qquad 0 < y_1 < \infty, \quad 0 < y_2, \quad y_3 < 1. \end{aligned}$$

It follows that Y_1, Y_2, and Y_3 are independent.

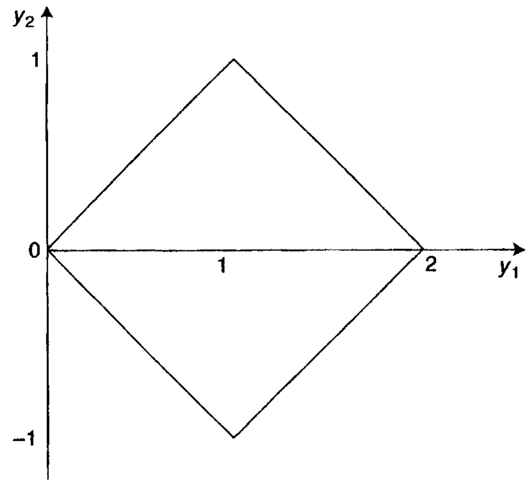

Fig. 3. $\{0 < y_1 + y_2 < 2,\ 0 < y_1 - y_2 < 2\}$.

Example 5. Let X_1, X_2 be independent RVs with common density given by

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

Let $Y_1 = X_1 + X_2$, $Y_2 = X_1 - X_2$. Then the Jacobian of the transformation is given by

$$J = \begin{vmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{vmatrix} = -\tfrac{1}{2},$$

and the joint density of Y_1, Y_2 (Fig. 3) is given by

$$\begin{aligned} f_{Y_1,Y_2}(y_1, y_2) &= \frac{1}{2} f\left(\frac{y_1 + y_2}{2}\right) f\left(\frac{y_1 - y_2}{2}\right) \\ &\qquad \text{if } 0 < \frac{y_1 + y_2}{2} < 1, \quad 0 < \frac{y_1 - y_2}{2} < 1, \\ &= \frac{1}{2} \qquad \text{if } (y_1, y_2) \in \{0 < y_1 + y_2 < 2, 0 < y_1 - y_2 < 2\}. \end{aligned}$$

The marginal PDFs of Y_1 and Y_2 are given by

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

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

Example 6. Let X_1, X_2, X_3 be iid RVs with common PDF

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \qquad -\infty < x < \infty.$$

Let $Y_1 = (X_1 - X_2)/\sqrt{2}$, $Y_2 = (X_1 + X_2 - 2X_3)/\sqrt{6}$, and $Y_3 = (X_1 + X_2 + X_3)/\sqrt{3}$. Then

$$x_1 = \frac{y_1}{\sqrt{2}} + \frac{y_2}{\sqrt{6}} + \frac{y_3}{\sqrt{3}},$$

$$x_2 = -\frac{y_1}{\sqrt{2}} + \frac{y_2}{\sqrt{6}} + \frac{y_3}{\sqrt{3}},$$

and

$$x_3 = -\frac{\sqrt{2}y_2}{\sqrt{3}} + \frac{y_3}{\sqrt{3}}.$$

The Jacobian of transformation is given by

$$J = \begin{vmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ 0 & \frac{-\sqrt{2}}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{vmatrix} = 1.$$

The joint PDF of X_1, X_2, X_3 is given by

$$g(x_1, x_2, x_3) = \frac{1}{(\sqrt{2\pi})^3} \exp\left(-\frac{x_1^2 + x_2^2 + x_3^2}{2} \right), \qquad x_1, x_2, x_3 \in \mathcal{R}.$$

It is easily checked that

$$x_1^2 + x_2^2 + x_3^2 = y_1^2 + y_2^2 + y_3^2,$$

so that the joint PDF of Y_1, Y_2, Y_3 is given by

$$w(y_1, y_2, y_3) = \frac{1}{(\sqrt{2\pi})^3} \exp\left(-\frac{y_1^2 + y_2^2 + y_3^2}{2} \right).$$

It follows that Y_1, Y_2, Y_3 are also iid RVs with common PDF f.

In Example 6 the transformation used is orthogonal and is known as *Helmert's transformation*. In fact, we will show in Section 7.6 that under orthogonal transformations iid RVs with PDF f defined above are transformed into iid RVs with the same PDF.

In Example 6 it is easily verified that

$$y_1^2 + y_2^2 = \sum_{j=1}^{3} \left(x_j - \frac{x_1 + x_2 + x_3}{3} \right)^2 .$$

We have therefore proved that $(X_1 + X_2 + X_3)$ is independent of $\sum_{j=1}^{3} \{X_j - [(X_1 + X_2 + X_3)/3]\}^2$. This is a very important result in mathematical statistics, and we will return to it in Section 7.5.

Example 7. Let (X, Y) be a *bivariate normal* RV with joint PDF

$$\begin{aligned} f(x, y) &= \frac{1}{2\pi \sigma_1 \sigma_2 (1 - \rho^2)^{1/2}} \\ &\cdot \exp\left\{ -\frac{1}{2(1-\rho^2)} \left[\frac{(x-\mu_1)^2}{\sigma_1^2} - \frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1 \sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2} \right] \right\}, \\ &-\infty < x < \infty, \quad -\infty < y < \infty; \ \mu_1 \in \mathcal{R}, \mu_2 \in \mathcal{R}; \\ &\text{and} \quad \sigma_1 > 0, \quad \sigma_2 > 0, \quad |\rho| < 1. \end{aligned}$$

Let

$$U_1 = \sqrt{X^2 + Y^2} \quad \text{and} \quad U_2 = \frac{X}{Y}.$$

For $u_1 > 0$,

$$\sqrt{x^2 + y^2} = u_1 \quad \text{and} \quad \frac{x}{y} = u_2$$

have two solutions:

$$x_1 = \frac{u_1 u_2}{\sqrt{1 + u_2^2}}, \quad y_1 = \frac{u_1}{\sqrt{1 + u_2^2}}, \quad \text{and} \quad x_2 = -x_1, \quad y_2 = -y_1$$

for any $u_2 \in \mathcal{R}$. The Jacobians are given by

$$J_1 = J_2 = \begin{vmatrix} \dfrac{u_2}{\sqrt{1 + u_2^2}} & \dfrac{u_1}{(1 + u_2^2)^{3/2}} \\ \dfrac{1}{\sqrt{1 + u_2^2}} & -\dfrac{u_1 u_2}{(1 + u_2^2)^{3/2}} \end{vmatrix} = -\frac{u_1}{1 + u_2^2}.$$

It follows from the result in Remark 2 that the joint PDF of (U_1, U_2) is given by

$$w(u_1, u_2) = \begin{cases} \dfrac{u_1}{1+u_2^2}\left[f\left(\dfrac{u_1 u_2}{\sqrt{1+u_2^2}}, \dfrac{u_1}{\sqrt{1+u_2^2}}\right)\right. \\ \left. \quad + f\left(\dfrac{-u_1 u_2}{\sqrt{1+u_2^2}}, \dfrac{-u_1}{\sqrt{1+u_2^2}}\right)\right] & \text{if } u_1 > 0, u_2 \in \mathcal{R}, \\ 0 & \text{otherwise.} \end{cases}$$

In the special case where $\mu_1 = \mu_2 = 0$, $\rho = 0$, and $\sigma_1 = \sigma_2 = \sigma$, we have

$$f(x, y) = \frac{1}{2\pi\sigma^2} e^{-[(x^2+y^2)/2\sigma^2]}$$

so that X and Y are independent. Moreover,

$$f(x, y) = f(-x, -y),$$

and it follows that when X and Y are independent,

$$w(u_1, u_2) = \begin{cases} \dfrac{1}{2\pi\sigma^2}\dfrac{2u_1}{1+u_2^2} e^{-u_1^2/2\sigma^2}, & u_1 > 0, \quad -\infty < u_2 < \infty, \\ 0, & \text{otherwise.} \end{cases}$$

Since

$$w(u_1, u_2) = \frac{1}{\pi(1+u_2^2)} \frac{u_1}{\sigma^2} e^{-u_1^2/2\sigma^2},$$

it follows that U_1 and U_2 are independent with marginal PDFs given by

$$w_1(u_1) = \begin{cases} \dfrac{u_1}{\sigma^2} e^{-u_1^2/2\sigma^2}, & u_1 > 0, \\ 0, & u_1 \le 0, \end{cases}$$

and

$$w_2(u_2) = \frac{1}{\pi(1+u_2^2)}, \qquad -\infty < u_2 < \infty,$$

respectively.

An important application of the result in Remark 2 will appear in Theorem 4.7.2.

Theorem 3. Let (X, Y) be an RV of the continuous type with PDF f. Let

$$Z = X + Y, \quad U = X - Y, \quad \text{and} \quad V = XY;$$

and let $W = X/Y$. Then the PDFs of Z, V, U, and W are, respectively, given by

$$f_Z(z) = \int_{-\infty}^{\infty} f(x, z - x)\,dx, \tag{2}$$

$$f_U(u) = \int_{-\infty}^{\infty} f(u + y, y)\,dy, \tag{3}$$

$$f_V(v) = \int_{-\infty}^{\infty} f\left(x, \frac{v}{x}\right)\frac{1}{|x|}\,dx, \tag{4}$$

and

$$f_W(w) = \int_{-\infty}^{\infty} f(xw, x)|x|\,dx. \tag{5}$$

The proof is left as an exercise.

Corollary. If X and Y are independent with PDFs f_1 and f_2, respectively, then

$$f_Z(z) = \int_{-\infty}^{\infty} f_1(x) f_2(z - x)\,dx, \tag{6}$$

$$f_U(u) = \int_{-\infty}^{\infty} f_1(u + y) f_2(y)\,dy, \tag{7}$$

$$f_V(v) = \int_{-\infty}^{\infty} f_1(x) f_2\left(\frac{v}{x}\right)\frac{1}{|x|}\,dx, \tag{8}$$

and

$$f_W(w) = \int_{-\infty}^{\infty} f_1(xw) f_2(x)|x|\,dx. \tag{9}$$

Remark 3. Let F and G be two absolutely continuous DFs; then

$$H(x) = \int_{-\infty}^{\infty} F(x - y)G'(y)\,dy = \int_{-\infty}^{\infty} G(x - y)F'(y)\,dy$$

is also an absolutely continuous DF with PDF

$$H'(x) = \int_{-\infty}^{\infty} F'(x - y)G'(y)\,dy = \int_{-\infty}^{\infty} G'(x - y)F'(y)\,dy.$$

If

$$F(x) = \sum_k p_k \varepsilon(x - x_k) \quad \text{and} \quad G(x) = \sum_j q_j \varepsilon(x - y_j)$$

are two DFs, then

$$H(x) = \sum_k \sum_j p_k q_j \varepsilon(x - x_k - y_j)$$

is also a DF of an RV of the discrete type. The DF H is called the *convolution* of F and G, and we write $H = F * G$. Clearly, the operation is commutative and associative; that is, if F_1, F_2, F_3 are DFs, $F_1 * F_2 \doteq F_2 * F_1$ and $(F_1 * F_2) * F_3 = F_1 * (F_2 * F_3)$. In this terminology, if X and Y are independent RVs with DFs F and G, respectively, $X + Y$ has the convolution DF $H = F * G$. Extension to an arbitrary number of independent RVs is obvious.

Finally, we consider a technique based on MGF or CF which can be used in certain situations to determine the distribution of a function $g(X_1, X_2, \ldots, X_n)$ of $X_1, X_2, \ldots, X_n$.

Let $(X_1, X_2, \ldots, X_n)$ be an n-variate RV, and g be a Borel-measurable function from $\mathcal{R}_n$ to $\mathcal{R}_1$.

Definition 1. If $(X_1, X_2, \ldots, X_n)$ is discrete type and

$$\sum_{x_1,\ldots,x_n} |g(x_1, x_2, \ldots, x_n)| P\{X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n\} < \infty,$$

then the series

$$\begin{aligned} &Eg(X_1, X_2, \ldots, X_n) \\ &= \sum_{x_1,\ldots,x_n} g(x_1, x_2, \ldots, x_n) P\{X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n\} \end{aligned}$$

is called the *expected value* of $g(X_1, X_2, \ldots, X_n)$. If $(X_1, X_2, \ldots, X_n)$ is a continuous RV with joint PDF f, and if

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} |g(x_1, x_2, \ldots, x_n)| f(x_1, x_2, \ldots, x_n) \prod_{i=1}^{n} dx_i < \infty,$$

then

$$\begin{aligned} &Eg(X_1, X_2, \ldots, X_n) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1, x_2, \ldots, x_n) f(x_1, x_2, \ldots, x_n) \prod_{i=1}^{n} dx_i \end{aligned}$$

is called the expected value of $g(X_1, X_2, \ldots, X_n)$.

Let $Y = g(X_1, X_2, \ldots, X_n)$, and let $h(y)$ be its PDF. If $E|Y| < \infty$, then

$$EY = \int_{-\infty}^{\infty} y h(y)\, dy.$$

An analog of Theorem 3.2.1 holds. That is,

$$\int_{-\infty}^{\infty} y h(y)\, dy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} g(x_1, x_2, \ldots, x_n) f(x_1, x_2, \ldots, x_n) \prod_{i=1}^{n} dx_i,$$

in the sense that if either integral exists, so does the other, and the two are equal. The result also holds in the discrete case.

Some special functions of interest are $\sum_{j=1}^n x_j$, $\prod_{j=1}^n x_j^{k_j}$ where $k_1, k_2, \ldots, k_n$ are nonnegative integers, $e^{\sum_{j=1}^n t_j x_j}$, where $t_1, t_2, \ldots, t_n$ are real numbers, and $e^{i\sum_{j=1}^n t_j x_j}$, where $i = \sqrt{-1}$.

Definition 2. Let $X_1, X_2, \ldots, X_n$ be jointly distributed. If $E(e^{\sum_{j=1}^n t_j X_j})$ exists for $|t_j| \le h_j$, $j = 1, 2, \ldots, n$, for some $h_j > 0$, $j = 1, 2, \ldots, n$, we write

$$(10) \qquad M(t_1, t_2, \ldots, t_n) = E e^{t_1 X_1 + t_2 X_2 + \cdots + t_n X_n}$$

and call it the *MGF of the joint distribution* of $(X_1, X_2, \ldots, X_n)$ or, simply, the MGF of $(X_1, X_2, \ldots, X_n)$.

Definition 3. Let $t_1, t_2, \ldots, t_n$ be real numbers and $i = \sqrt{-1}$. Then the CF of $(X_1, X_2, \ldots, X_n)$ is defined by

$$(11) \qquad \begin{aligned}\phi(t_1, t_2, \ldots, t_n) &= E\left[\exp\left(i\sum_{j=1}^n t_j X_j\right)\right] \\ &= E\left[\cos\left(\sum_{j=1}^n t_j X_j\right)\right] + iE\left[\sin\left(\sum_{j=1}^n t_j X_j\right)\right].\end{aligned}$$

As in the univariate case $\phi(t_1, t_2, \ldots, t_n)$ always exists.

We will deal mostly with MGF even though the condition that it exist for $|t_j| \le h_j$, $j = 1, 2, \ldots, n$ restricts its application considerably. The multivariate MGF (CF) has properties similar to the univariate MGF discussed earlier. We state some of these without proof. For notational convenience we restrict ourselves to the bivariate case.

Theorem 4. The MGF $M(t_1, t_2)$ uniquely determines the joint distribution of (X, Y), and conversely, if the MGF exists, it is unique.

Corollary. The MGF $M(t_1, t_2)$ completely determines the marginal distributions of X and Y. Indeed,

$$M(t_1, 0) = Ee^{t_1 X} = M_X(t_1), \tag{12}$$

and

$$M(0, t_2) = Ee^{t_2 Y} = M_Y(t_2). \tag{13}$$

Theorem 5. If $M(t_1, t_2)$ exists, the moments of all orders of (X, Y) exist and may be obtained from

$$\left. \frac{\partial^{m+n} M(t_1, t_2)}{\partial t_1^m \, \partial t_2^n} \right|_{t_1=t_2=0} = E(X^m Y^n). \tag{14}$$

Thus

$$\frac{\partial M(0,0)}{\partial t_1} = EX, \qquad \frac{\partial M(0,0)}{\partial t_2} = EY,$$

$$\frac{\partial^2 M(0,0)}{\partial t_1^2} = EX^2, \qquad \frac{\partial^2 M(0,0)}{\partial t_2^2} = EY^2,$$

$$\frac{\partial^2 M(0,0)}{\partial t_1 \partial t_2} = E(XY),$$

and so on.

A formal definition of moments in the multivariate case will be given in Section 4.5.

Theorem 6. X and Y are independent RVs if and only if

$$M(t_1, t_2) = M(t_1, 0)\, M(0, t_2) \qquad \text{for all } t_1, t_2 \in \mathcal{R}. \tag{15}$$

Proof. Let X and Y be independent. Then

$$M(t_1, t_2) = Ee^{t_1 X + t_2 Y} = (Ee^{t_1 X})(Ee^{t_2 Y}) = M(t_1, 0) M(0, t_2).$$

Conversely, if

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

then in the continuous case,

$$\iint e^{t_1 x + t_2 y} f(x, y)\, dx\, dy = \left[\int e^{t_1 x} f_1(x)\, dx\right] \left[\int e^{t_2 y} f_2(y)\, dy\right],$$

that is,

$$\iint e^{t_1 x + t_2 y} f(x, y)\, dx\, dy = \iint e^{t_1 x + t_2 y} f_1(x) f_2(y)\, dx\, dy.$$

By the uniqueness of the MGF (Theorem 4) we must have

$$f(x,y) = f_1(x)f_2(y) \qquad \text{for all } (x,y) \in \mathcal{R}_2.$$

It follows that X and Y are independent. A similar proof is given in the case where (X, Y) is of the discrete type.

The MGF technique uses the uniqueness property of Theorem 4. To find the distribution (DF, PDF, or PMF) of $Y = g(X_1, X_2, \dots, X_n)$ we compute the MGF of Y using the definition. If this MGF is one of the known kind, Y must have this kind of distribution. Although the technique applies to the case when Y is an m-dimensional RV, $1 \le k \le n$, we will use it mostly for the $m = 1$ case.

Example 8. Let us first consider a simple case when X is normal PDF

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \qquad -\infty < x < -\infty.$$

Let $Y = X^2$. Then

$$\begin{aligned} M_Y(s) = Ee^{sX^2} &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{(1/2)(1-2s)x^2}\,dx \\ &= \frac{1}{\sqrt{1-2s}} \qquad \text{for } x < \tfrac{1}{2}. \end{aligned}$$

It follows (see Section 5.3 and Example 2.5.7) that Y has a chi-square PDF

$$w(y) = \frac{e^{-y/2}}{\sqrt{y\pi}}, \qquad y > 0.$$

Example 9. Suppose that X_1 and X_2 are independent with common PDF f of Example 8. Let $Y_1 = X_1 - X_2$. There are three equivalent ways to use MGF technique here. Let $Y_2 = X_2$. Then rather than compute

$$M(s_1, s_2) = Ee^{s_1 Y_1 + s_2 Y_2},$$

it is simpler to recognize that Y_1 is univariate, so

$$\begin{aligned} M_{Y_1}(s) &= Ee^{s(X_1 - X_2)} \\ &= (Ee^{sX_1})(Ee^{-sX_2}) \\ &= e^{s^2/2}e^{s^2/2} = e^{s^2}. \end{aligned}$$

It follows that Y_1 has PDF

$$f(x) = \frac{1}{\sqrt{4\pi}} e^{-x^2/4}, \qquad -\infty < x < \infty.$$

Note that $M_{Y_1}(s) = M(s, 0)$.

Let $Y_3 = X_1 + X_2$. Let us find the joint distribution of Y_1 and Y_3. Indeed,

$$\begin{aligned} Ee^{s_1 Y_1 + s_2 Y_3} &= E\left(e^{(s_1+s_2)X_1} \cdot e^{(s_1-s_2)X_2}\right) \\ &= \left(Ee^{(s_1+s_2)X_1}\right)\left(Ee^{(s_1-s_2)X_2}\right) \\ &= e^{(s_1+s_2)^2/2} \cdot e^{(s_1-s_2)^2/2} = e^{s_1^2} \cdot e^{s_2^2} \end{aligned}$$

and it follows that Y_1 and Y_3 are independent RVs with common PDF f defined above.

The following result has many applications, as we will see. Example 9 is a special case.

Theorem 7. Let $X_1, X_2, \dots, X_n$ be independent RVs with respective MGFs $M_i(s)$, $i = 1, 2, \dots, n$. Then the MGF of $Y = \sum_{i=1}^n a_i X_i$ for real numbers $a_1, a_2, \dots, a_n$ is given by

$$M_Y(s) = \prod_{i=1}^n M_i(a_i s).$$

Proof. If M_i exists for $|s| \le h_i, h_i > 0$, then M_Y exists for $|s| \le \min(h_1, \dots, h_n)$ and

$$M_Y(s) = Ee^{s\sum_{i=1}^n a_i X_i} = \prod_{i=1}^n Ee^{sa_i X_i} = \prod_{i=1}^n M_i(a_i s).$$

Corollary. If X_i's are iid, the MGF of $Y = \sum_1^n X_i$ is given by $M_Y(s) = [M(s)]^n$.

Remark 4. The converse of Theorem 7 does not hold. We leave the reader to construct an example illustrating this fact.

Example 10. Let $X_1, X_2, \dots, X_m$ be iid RVs with common PMF

$$P\{X = k\} = \binom{n}{k} p^k (1-p)^{n-k}, \qquad k = 0, 1, 2, \dots, n; \quad 0 < p < 1.$$

Then the MGF of X_i is given by

$$M(t) = (1 - p + pe^t)^n.$$

It follows that the MGF of $S_m = X_1 + X_2 + \cdots + X_m$ is

$$M_{S_m}(t) = \prod_1^m (1 - p + pe^t)^n = (1 - p + pe^t)^{nm},$$

and we see that S_m has the PMF

$$P\{S_m = s\} = \binom{mn}{s} p^s (1-p)^{mn-s}, \qquad s = 0, 1, 2, \ldots, mn.$$

From these examples it is clear that to use this technique effectively one must be able to recognize the MGF of the function under consideration. In Chapter 5 we study a number of commonly occurring probability distributions and derive their MGFs (whenever they exist). We will have occasion to use Theorem 7 quite frequently.

For integer-valued RVs one can sometimes use PGFs to compute the distribution of certain functions of a multiple RV.

We emphasize the fact that a CF always exists and analogs of Theorems 4 to 7 can be stated in terms of CF's.

PROBLEMS 4.4

1. Let F be a DF and ε be a positive real number. Show that

$$\Psi_1(x) = \frac{1}{\varepsilon} \int_x^{x+\varepsilon} F(x)\, dx$$

and

$$\Psi_2(x) = \frac{1}{2\varepsilon} \int_{x-\varepsilon}^{x+\varepsilon} F(x)\, dx$$

are also distribution functions.

2. Let X, Y be iid RVs with common PDF

$$f(x) = \begin{cases} e^{-x} & \text{if } x > 0, \\ 0 & \text{if } x \le 0. \end{cases}$$

(a) Find the PDF of RVs $X+Y$, $X-Y$, XY, X/Y, $\min\{X,Y\}$, $\max\{X,Y\}$, $\min\{X,Y\}/\max\{X,Y\}$, and $X/(X+Y)$.

(b) Let $U = X+Y$ and $V = X-Y$. Find the conditional PDF of V, given $U = u$, for some fixed $u > 0$.

(c) Show that U and $Z = X/(X+Y)$ are independent.

3. Let X and Y be independent RVs defined on the space $(\Omega, \mathcal{S}, P)$. Let X be uniformly distributed on $(-a, a)$, $a > 0$, and Y be an RV of the continuous type with density f, where f is continuous and positive on $\mathcal{R}$. Let F be the DF of Y. If $u_0 \in (-a, a)$ is a fixed number, show that

$$f_{Y|X+Y}(y \mid u_0) = \begin{cases} \dfrac{f(y)}{F(u_0+a) - F(u_0-a)} & \text{if } u_0 - a < y < u_0 + a, \\ 0 & \text{otherwise.} \end{cases}$$

where $f_{Y|X+Y}(y \mid u_0)$ is the conditional density function of Y, given $X+Y = u_0$.

4. Let X and Y be iid RVs with common PDF

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

Find the PDFs of RVs XY, X/Y, $\min\{X, Y\}$, $\max\{X, Y\}$, $\min\{X, Y\}/\max\{X, Y\}$.

5. Let X_1, X_2, X_3 be iid RVs with common density function

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

Show that the PDF of $U = X_1 + X_2 + X_3$ is given by

$$g(u) = \begin{cases} \dfrac{u^2}{2}, & 0 \le u < 1, \\ 3u - u^2 - \dfrac{3}{2}, & 1 \le u < 2, \\ \dfrac{(u-3)^2}{2}, & 2 \le u \le 3, \\ 0, & \text{elsewhere.} \end{cases}$$

An extension to the n-variate case holds.

6. Let X and Y be independent RVs with common geometric PMF

$$P\{X = k\} = \pi(1-\pi)^k, \qquad k = 0, 1, 2, \ldots; \quad 0 < \pi < 1.$$

Also, let $M = \max\{X, Y\}$. Find the joint distribution of M and X, the marginal distribution of M, and the conditional distribution of X, given M.

7. Let X be a nonnegative RV of the continuous type. The integral part, Y, of X is distributed with PMF $P\{Y = k\} = \lambda^k e^{-\lambda}/k!$, $k = 0, 1, 2, \ldots, \lambda > 0$; and the fractional part, Z, of X has PDF $f_Z(z) = 1$ if $0 \le z \le 1$, and $= 0$ otherwise. Find the PDF of X, assuming that Y and Z are independent.

8. Let X and Y be independent RVs. If at least one of X and Y is of the continuous type, show that $X + Y$ is also continuous. What if X and Y are not independent?

9. Let X and Y be independent integral RVs. Show that

$$P(t) = P_X(t)P_Y(t),$$

where P, P_X, and P_Y, respectively, are the PGFs of $X+Y$, X, and Y.

10. Let X and Y be independent nonnegative RVs of the continuous type with PDFs f and g, respectively. Let $f(x) = e^{-x}$ if $x > 0$, and $= 0$ if $x \le 0$, and let g be arbitrary. Show that the MGF $M(t)$ of Y, which is assumed to exist, has the property that the DF of X/Y is $1 - M(-t)$.

11. Let X, Y, Z have the joint PDF

$$f(x,y,z) = \begin{cases} 6(1+x+y+z)^{-4} & \text{if } 0 < x, 0 < y, 0 < z, \\ 0 & \text{otherwise.} \end{cases}$$

Find the PDF of $U = X + Y + Z$.

12. Let X and Y be iid RVs with common PDF

$$f(x) = \begin{cases} (x\sqrt{2\pi})^{-1} e^{-(1/2)(\log x)^2}, & x > 0, \\ 0, & x \le 0. \end{cases}$$

Find the PDF of $Z = XY$.

13. Let X and Y be iid RVs with common PDF f defined in Example 8. Find the joint PDF of U and V in the following cases:

(a) $U = \sqrt{X^2 + Y^2}$, $V = \tan^{-1}(X/Y)$, $-\pi/2 < V \le \pi/2$.

(b) $U = (X+Y)/2$, $V = (X-Y)^2/2$.

14. Construct an example to show that even when the MGF of $X + Y$ can be written as a product of the MGF of X and the MGF of Y, X and Y need not be independent.

15. Let $X_1, X_2, \dots, X_n$ be iid with common PDF

$$f(x) = \frac{1}{b-a}, \qquad a < x < b, \qquad = 0 \text{ otherwise.}$$

Using the distribution function technique, show that:

(a) The joint PDF of $X_{(n)} = \max(X_1, X_2, \dots, X_n)$, and $X_{(1)} = \min(X_1, X_2, \dots, X_n)$ is given by

$$u(x,y) = \frac{n(n-1)(x-y)^{n-2}}{(b-a)^n}, \qquad a < y < x < b,$$

and $= 0$ otherwise.

(b) The PDF of $X_{(n)}$ is given by

$$g(z) = \frac{n(z-a)^n}{(b-a)^n}, \qquad a < z < b, \quad = 0 \text{ otherwise}$$

and that of $X_{(1)}$ by

$$h(z) = \frac{n(b-z)^{n-1}}{(b-a)^n}, \qquad a < z < b, \quad = 0 \text{ otherwise.}$$

16. Let X_1, X_2 be iid with common Poisson PMF

$$P(X_i = x) = e^{-\lambda}\frac{\lambda^x}{x!}, \qquad x = 0, 1, 2, \dots, \qquad i = 1, 2,$$

where $\lambda > 0$ is a constant. Let $X_{(2)} = \max(X_1, X_2)$ and $X_{(1)} = \min(X_1, X_2)$. Find the PMF of $X_{(2)}$.

17. Let X have the binomial PMF

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \qquad k = 0, 1, \dots, n; \quad 0 < p < 1.$$

Let Y be independent of X and $Y \stackrel{d}{=} X$. Find the PMF of $U = X + Y$ and $W = X - Y$.

4.5 COVARIANCE, CORRELATION, AND MOMENTS

Let X and Y be jointly distributed on $(\Omega, \mathcal{S}, P)$. In Section 4.4 we defined $Eg(X, Y)$ for Borel functions g on $\mathcal{R}_2$. Functions of the form $g(x, y) = x^j y^k$, where j and k are nonnegative integers, are of interest in probability and statistics.

Definition 1. If $E|X^j Y^k| < \infty$ for nonnegative integers j and k, we call $E(X^j Y^k)$ a *moment of order* $(j + k)$ of (X, Y) and write

(1) $$m_{jk} = E(X^j Y^k).$$

Clearly,

(2) $$\begin{aligned} &m_{10} = EX, \qquad m_{01} = EY, \\ &m_{20} = EX^2, \quad m_{11} = E(XY), \quad \text{and } m_{02} = EY^2. \end{aligned}$$

Definition 2. If $E\left|(X - EX)^j (Y - EY)^k\right| < \infty$ for nonnegative integers j and k, we call $E\left\{(X - EX)^j (Y - EY)^k\right\}$ a *central moment of order* $(j + k)$ and write

(3) $$\mu_{jk} = E\left\{(X - EX)^j (Y - EY)^k\right\}.$$

Clearly,

(4) $$\mu_{10} = \mu_{01} = 0, \qquad \mu_{20} = \operatorname{var}(X), \qquad \mu_{02} = \operatorname{var}(Y), \qquad \text{and}$$
$$\mu_{11} = E[(X - m_{10})(Y - m_{01})].$$

We see easily that

(5) $$\mu_{11} = E(XY) - EXEY.$$

Note that if X and Y increase (or decrease) together, then $(X - EX)(Y - EY)$ should be positive, whereas if X decreases while Y increases (and conversely), the product should be negative. Hence the average value of $(X - EX)(Y - EY)$, namely μ_{11}, provides a measure of association or joint variation between X and Y.

Definition 3. If $E[(X - EX)(Y - EY)]$ exists, we call it the *covariance* between X and Y and write

(6) $$\operatorname{cov}(X, Y) = E[(X - EX)(Y - EY)] = E(XY) - EXEY.$$

Recall (Theorem 3.2.8) that $E(Y - a)^2$ is minimized when we choose $a = EY$ so that EY may be interpreted as the best constant predictor of Y. If, instead, we choose to predict Y by a linear function of X, say $aX + b$, and measure the error in this prediction by $E(Y - aX - b)^2$, we should choose a and b to minimize this *mean square error*. Clearly, $E(Y - aX - b)^2$ is minimized, for any a, by choosing $b = E(Y - aX) = EY - aEX$. With this choice of b, we find a such that

$$\begin{aligned} E(Y - aX - b)^2 &= E[(Y - EY) - a(X - EX)]^2 \\ &= \sigma_Y^2 - 2a\mu_{11} + a^2\sigma_X^2 \end{aligned}$$

is minimum. An easy computation shows that the minimum occurs if we choose

(7) $$a = \frac{\mu_{11}}{\sigma_X^2},$$

provided that $\sigma_X^2 > 0$. Moreover,

(8) $$\begin{aligned} \min_{a,b} E(Y - aX - b)^2 &= \min_{a} \left\{\sigma_Y^2 - 2a\mu_{11} + a^2\sigma_X^2\right\} \\ &= \sigma_Y^2 - \frac{\mu_{11}^2}{\sigma_X^2} \\ &= \sigma_Y^2\left[1 - \left(\frac{\mu_{11}}{\sigma_X\sigma_Y}\right)^2\right]. \end{aligned}$$

Let us write

$$\rho = \frac{\mu_{11}}{\sigma_X \sigma_Y}. \tag{9}$$

Then (8) shows that predicting Y by a linear function of X reduces the prediction error from σ_Y^2 to $\sigma_Y^2(1-\rho^2)$. We may therefore think of ρ as a measure of the *linear dependence* between RVs X and Y.

Definition 4. If EX^2, EY^2 exist, we define the *correlation coefficient* between X and Y as

$$\rho = \frac{\operatorname{cov}(X,Y)}{\mathrm{SD}(X)\mathrm{SD}(Y)} = \frac{E(XY) - EXEY}{\sqrt{EX^2 - (EX)^2}\sqrt{EY^2 - (EY)^2}}, \tag{10}$$

where $\mathrm{SD}(X)$ denotes the standard deviation of RV X.

We note that for any two real numbers a and b,

$$|ab| \le \frac{a^2 + b^2}{2},$$

so that $E|XY| < \infty$ if $EX^2 < \infty$ and $EY^2 < \infty$.

Definition 5. We say that RVs X and Y are *uncorrelated* if $\rho = 0$, or equivalently, $\operatorname{cov}(X, Y) = 0$.

If X and Y are independent, then from (5) $\operatorname{cov}(X, Y) = 0$ and, X and Y are uncorrelated. If, however, $\rho = 0$, then X and Y may not necessarily be independent.

Example 1. Let U and V be two RVs with common mean and common variance. Let $X = U + V$ and $Y = U - V$. Then

$$\operatorname{cov}(X, Y) = E(U^2 - V^2) - E(U+V)E(U-V) = 0$$

so that X and Y are uncorrelated but not necessarily independent (see Example 4.4.9).

Let us now study some properties of the correlation coefficient. From the definition we see that ρ [and also $\operatorname{cov}(X, Y)$] is symmetric in X and Y.

Theorem 1

(a) The correlation coefficient ρ between two RVs X and Y satisfies

$$|\rho| \le 1. \tag{11}$$

(b) The equality $|\rho| = 1$ holds if and only if there exist constants $a \ne 0$ and b such that $P\{aX + b = 1\} = 1$.

Proof. From (8) since $E(Y - aX - b)^2 \geq 0$, we must have $1 - \rho^2 \geq 0$, or equivalently, (11) holds.

Equality in (11) holds if and only if $\rho^2 = 1$, or equivalently, $E(Y - aX - b)^2 = 0$ holds. This implies and is implied by $P(Y = aX + b) = 1$. Here $a \neq 0$.

Remark 1. From (7) and (9) we note that the signs of a and ρ are the same, so if $\rho = 1$, then $P(Y = aX + b)$ where $a > 0$, and if $\rho = -1$, then $a < 0$.

Theorem 2. Let $EX^2 < \infty$, $EY^2 < \infty$, and let $U = aX + b$, $V = cY + d$. Then

$$\rho_{X,Y} = \pm \rho_{U,V},$$

where $\rho_{X,Y}$ and $\rho_{U,V}$, respectively, are the correlation coefficients between X and Y and U and V.

The proof is simple and is left as an exercise.

Example 2. Let X, Y be identically distributed with common PMF

$$P\{X = k\} = \frac{1}{N}, \qquad k = 1, 2, \ldots, N (N > 1).$$

Then

$$EX = EY = \frac{N+1}{2}, \qquad EX^2 = EY^2 = \frac{(N+1)(2N+1)}{6},$$

so that

$$\operatorname{var}(X) = \operatorname{var}(Y) = \frac{N^2 - 1}{12}.$$

Also,

$$\begin{aligned} E(XY) &= \tfrac{1}{2}[EX^2 + EY^2 - E(X - Y)^2] \\ &= \frac{(N+1)(2N+1)}{6} - \frac{E(X - Y)^2}{2}. \end{aligned}$$

Thus

$$\begin{aligned} \operatorname{cov}(X, Y) &= \frac{(N+1)(2N+1)}{6} - \frac{E(X - Y)^2}{2} - \frac{(N+1)^2}{4} \\ &= \frac{(N+1)(N-1)}{12} - \frac{1}{2}E(X - Y)^2, \end{aligned}$$

and

$$\rho_{X,Y} = \frac{(N^2 - 1)/12 - E(X - Y)^2/2}{(N^2 - 1)/12}$$

$$= 1 - \frac{6E(X-Y)^2}{N^2-1}.$$

If $P\{X = Y\} = 1$, then $\rho = 1$, and conversely. If $P\{Y = N+1-X\} = 1$, then

$$\begin{aligned} E(X-Y)^2 &= E(2X-N-1)^2 \\ &= 4\frac{(N+1)(2N+1)}{6} - 4\frac{(N+1)^2}{2} + (N+1)^2, \end{aligned}$$

and it follows that $\rho_{XY} = -1$. Conversely, if $\rho_{X,Y} = -1$, from Remark 1 it follows that $Y = -aX + b$ with probability 1 for some $a > 0$ and some real number b. To find a and b, we note that $EY = -aEX + b$, so that $b = [(N+1)/2](1+a)$. Also, $EY^2 = E(b-aX)^2$, which yields

$$(1-a^2)EX^2 + 2abEX - b^2 = 0.$$

Substituting for b in terms of a and the values of EX^2 and EX, we see that $a^2 = 1$, so that $a = 1$. Hence $b = N+1$, and it follows that $Y = N+1-X$ with probability 1.

Example 3. Let (X, Y) be jointly distributed with density function

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

Then

$$\begin{aligned} E(X^l Y^m) &= \int_0^1 \int_0^1 x^l y^m (x+y)\,dx\,dy \\ &= \int_0^1 \int_0^1 x^{l+1} y^m \,dx\,dy + \int_0^1 \int_0^1 x^l y^{m+1}\,dx\,dy \\ &= \frac{1}{(l+2)(m+1)} + \frac{1}{(l+1)(m+2)}, \end{aligned}$$

where l and m are positive integers. Thus

$$\begin{aligned} EX = EY &= \tfrac{7}{12}, \\ EX^2 = EY^2 &= \tfrac{5}{12}, \\ \operatorname{var}(X) = \operatorname{var}(Y) &= \tfrac{5}{12} - \tfrac{49}{144} = \tfrac{11}{144}, \end{aligned}$$

and

$$\operatorname{cov}(X,Y) = \tfrac{1}{3} - \tfrac{49}{144} = -\tfrac{1}{144}, \qquad \rho = -\tfrac{1}{11}.$$

Theorem 3. Let $X_1, X_2, \ldots, X_n$ be RVs such that $E|X_i| < \infty$, $i = 1, 2, \ldots, n$. Let $a_1, a_2, \ldots, a_n$ be real numbers, and write

$$S = a_1 X_1 + a_2 X_2 + \cdots + a_n X_n.$$

Then ES exists, and we have

$$ES = \sum_{j=1}^{n} a_j EX_j. \tag{12}$$

Proof. If $(X_1, X_2, \ldots, X_n)$ is of the discrete type, then

$$\begin{aligned}
ES &= \sum_{i_1, i_2, \ldots, i_n} (a_1 x_{i_1} + a_2 x_{i_2} + \ldots + a_n x_{i_n}) P\{X_1 = x_{i_1}, X_2 = x_{i_2}, \ldots, X_n = x_{i_n}\} \\
&= a_1 \sum_{i_1} x_{i_1} \sum_{i_2, \ldots, i_n} P\{X_1 = x_{i_1}, \ldots, X_n = x_{i_n}\} \\
&\qquad + \cdots + a_n \sum_{i_n} x_{i_n} \sum_{i_1, \ldots, i_{n-1}} P\{X_1 = x_{i_1}, \ldots, X_n = x_{i_n}\} \\
&= a_1 \sum_{i_1} x_{i_1} P\{X_1 = x_{i_1}\} + \cdots + a_n \sum_{i_n} P\{X_n = x_{i_n}\} \\
&= a_1 EX_1 + \cdots + a_n EX_n.
\end{aligned}$$

The existence of ES follows easily by replacing each a_j by $|a_j|$ and each x_{ij} by $|x_{ij}|$ and remembering that $E|X_j| < \infty$, $j = 1, 2, \ldots, n$. The case of continuous type $(X_1, X_2, \ldots, X_n)$ is treated similarly.

Corollary. Take $a_1 = a_2 = \cdots = a_n = 1/n$. Then

$$E\left(\frac{X_1 + X_2 + \cdots + X_n}{n}\right) = \frac{1}{n} \sum_{i=1}^{n} EX_i,$$

and if $EX_1 = EX_2 = \cdots = EX_n = \mu$, then

$$E\left(\frac{X_1 + X_2 + \cdots + X_n}{n}\right) = \mu.$$

Theorem 4. Let $X_1, X_2, \ldots, X_n$ be independent RVs such that $E|X_i| < \infty$, $i = 1, 2, \ldots, n$. Then $E(\prod_{i=1}^{n} X_i)$ exists and

$$E\left(\prod_{i=1}^{n} X_i\right) = \prod_{i=1}^{n} EX_i. \tag{13}$$

Let X and Y be independent and $g_1(\cdot)$ and $g_2(\cdot)$ be Borel-measurable functions. Then we know (Theorem 4.3.2) that $g_1(X)$ and $g_2(Y)$ are independent. If $E[g_1(X)]$,

$E[g_2(Y)]$, and $E[g_1(X)\,g_2(Y)]$ exist, it follows from Theorem 4 that

$$E[g_1(X)\,g_2(Y)] = E[g_1(X)]\,E[g_2(Y)]. \tag{14}$$

Conversely, if for any Borel sets A_1 and A_2 we take $g_1(X) = 1$ if $X \in A_1$, and $= 0$ otherwise, and $g_2(Y) = 1$ if $Y \in A_2$, and $= 0$ otherwise, then

$$E[g_1(X)g_2(Y)] = P\{X \in A_1, Y \in A_2\}$$

and $E[g_1(X)] = P\{X \in A_1\}$, $E[g_2(Y)] = P\{Y \in A_2\}$. Relation (14) implies that for any Borel sets A_1 and A_2 of real numbers

$$P\{X \in A_1,\ Y \in A_2\} = P\{X \in A_1\}P\{Y \in A_2\}.$$

It follows that X and Y are independent if (14) holds. We have thus proved the following theorem.

Theorem 5. Two RVs X and Y are independent if and only if for every pair of Borel-measurable functions g_1 and g_2 the relation

$$E[g_1(X)g_2(Y)] = E[g_1(X)]\,E[g_2(Y)] \tag{15}$$

holds, provided that the expectations on both sides of (15) exist.

Theorem 6. Let $X_1, X_2, \ldots, X_n$ be RVs with $E|X_i|^2 < \infty$ for $i = 1, 2, \ldots, n$. Let $a_1, a_2, \ldots, a_n$ be real numbers and write $S = \sum_{i=1}^{n} a_i X_i$. Then the variance of S exists and is given by

$$\operatorname{var}(S) = \sum_{i=1}^{n} a_i^2 \operatorname{var}(X_i) + \sum_{\substack{i=1 \\ i\neq j}}^{n} \sum_{j=1}^{n} a_i a_j \operatorname{cov}(X_i, X_j). \tag{16}$$

If, in particular, $X_1, X_2, \ldots, X_n$ are such that $\operatorname{cov}(X_i, X_j) = 0$ for $i, j = 1, 2, \ldots, n$, $i \neq j$, then

$$\operatorname{var}(S) = \sum_{i=1}^{n} a_i^2 \operatorname{var}(X_i). \tag{17}$$

Proof. We have

$$\begin{aligned}\operatorname{var}(S) &= E\left(\sum_{i=1}^{n} a_i X_i - \sum_{i=1}^{n} a_i EX_i\right)^2 \\ &= E\left[\sum_{i=1}^{n} a_i^2 (X_i - EX_i)^2 + \sum_{i\neq j} a_i a_j (X_i - EX_i)(X_j - EX_j)\right]\end{aligned}$$

$$= \sum_{i=1}^{n} a_i^2 E(X_i - EX_i)^2 + \sum_{i \neq j} a_i a_j E[(X_i - EX_i)(X_j - EX_j)].$$

If the X_i's satisfy

$$\operatorname{cov}(X_i, X_j) = 0 \qquad \text{for } i, j = 1, 2, \ldots, n; \quad i \neq j,$$

the second term on the right side of (16) vanishes, and we have (17).

Corollary 1. Let $X_1, X_2, \ldots, X_n$ be exchangeable RVs with $\operatorname{var}(X_i) = \sigma^2$, $i = 1, 2, \ldots, n$. Then

$$\operatorname{var}\left(\sum_{i=1}^{n} a_i X_i\right) = \sigma^2 \sum_{i=1}^{n} a_i^2 + \rho\sigma^2 \sum_{i \neq j}^{n} a_i a_j,$$

where ρ is the correlation coefficient between X_i and X_j, $i \neq j$. In particular,

$$\operatorname{var}\left(\sum_{i=1}^{n} \frac{X_i}{n}\right) = \frac{\sigma^2}{n} + \frac{n-1}{n}\rho\sigma^2.$$

Corollary 2. If $X_1, X_2, \ldots, X_n$ are exchangeable and uncorrelated, then

$$\operatorname{var}\left(\sum_{i=1}^{n} a_i X_i\right) = \sigma^2 \sum_{i=1}^{n} a_i^2,$$

and

$$\operatorname{var}\left(\sum_{i=1}^{n} \frac{X_i}{n}\right) = \frac{\sigma^2}{n}.$$

Theorem 7. Let $X_1, X_2, \ldots, X_n$ be iid RVs with common variance σ^2. Also, let $a_1, a_2, \ldots, a_n$ be real numbers such that $\sum_1^n a_i = 1$, and let $S = \sum_{i=1}^n a_i X_i$. Then the variance of S is least if we choose $a_i = 1/n$, $i = 1, 2, \ldots, n$.

Proof. We have

$$\operatorname{var}(S) = \sigma^2 \sum_{i=1}^{n} a_i^2,$$

which is least if and only if we choose the a_i's so that $\sum_{i=1}^n a_i^2$ is smallest, subject to the condition $\sum_{i=1}^n a_i = 1$. We have

$$\sum_{i=1}^{n} a_i^2 = \sum_{i=1}^{n} \left(a_i - \frac{1}{n} + \frac{1}{n}\right)^2$$

$$= \sum_{i=1}^{n} \left(a_i - \frac{1}{n}\right)^2 + \frac{2}{n} \sum_{i=1}^{n} \left(a_i - \frac{1}{n}\right) + \frac{1}{n}$$

$$= \sum_{i=1}^{n} \left(a_i - \frac{1}{n}\right)^2 + \frac{1}{n},$$

which is minimized for the choice $a_i = 1/n$, $i = 1, 2, \ldots, n$.

Note that the result holds if we replace independence by the condition that X_i's are exchangeable and uncorrelated.

Example 4. Suppose that r balls are drawn one at a time without replacement from a bag containing n white and m black balls. Let S_r be the number of black balls drawn.

Let us define RVs X_k as follows:

$$X_k = \begin{cases} 1 & \text{if the } k\text{th ball drawn is black} \\ 0 & \text{if the } k\text{th ball drawn is white} \end{cases} \qquad k = 1, 2, \ldots, r.$$

Then

$$S_r = X_1 + X_2 + \cdots + X_r.$$

Also,

$$(18) \qquad P\{X_k = 1\} = \frac{m}{m+n}, \quad \text{and} \quad P\{X_k = 0\} = \frac{n}{m+n}.$$

Thus $EX_k = m/(m+n)$, and

$$\operatorname{var}(X_k) = \frac{m}{m+n} - \frac{m^2}{(m+n)^2} = \frac{mn}{(m+n)^2}.$$

To compute $\operatorname{cov}(X_j, X_k)$, $j \neq k$, note that the RV $X_j X_k = 1$ if the jth and kth balls drawn are black, and $= 0$ otherwise. Thus

$$(19) \qquad E(X_j X_k) = P\{X_j = 1, X_k = 1\} = \frac{m}{m+n} \frac{m-1}{m+n-1}$$

and

$$\operatorname{cov}(X_j, X_k) = -\frac{mn}{(m+n)^2(m+n-1)}.$$

Thus

$$ES_r = \sum_{k=1}^{r} EX_k = \frac{mr}{m+n}$$

and

$$\operatorname{var}(S_r) = r\frac{mn}{(m+n)^2} - r(r-1)\frac{mn}{(m+n)^2(m+n-1)}$$
$$= \frac{mnr}{(m+n)^2(m+n+1)}(m+n-r).$$

Readers are asked to satisfy themselves that (18) and (19) hold.

Example 5. Let $X_1, X_2, \ldots, X_n$ be independent, and $a_1, a_2, \ldots, a_n$ be real numbers such that $\sum a_i = 1$. Assume that $E|X_i^2| < \infty$, $i = 1, 2, \ldots, n$, and let $\operatorname{var}(X_i) = \sigma_i^2$, $i = 1, 2, \ldots, n$. Write $S = \sum_{i=1}^n a_i X_i$. Then $\operatorname{var}(S) = \sum_{i=1}^n a_i^2 \sigma_i^2 = \sigma$, say. To find weights a_i such that σ is minimum, we write

$$\sigma = a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + \cdots + (1 - a_1 - a_2 - \cdots - a_{n-1})^2\sigma_n^2,$$

and differentiate partially with respect to $a_1, a_2, \ldots, a_{n-1}$, respectively. We get

$$\frac{\partial\sigma}{\partial a_1} = 2a_1\sigma_1^2 - 2(1 - a_1 - a_2 - \cdots - a_{n-1})\sigma_n^2 = 0,$$
$$\vdots$$
$$\frac{\partial\sigma}{\partial a_{n-1}} = 2a_{n-1}\sigma_{n-1}^2 - 2(1 - a_1 - a_2 - \cdots - a_{n-1})\sigma_n^2 = 0.$$

It follows that

$$a_j\sigma_j^2 = a_n\sigma_n^2, \qquad j = 1, 2, \ldots, n-1,$$

that is, the weights a_j, $j = 1, 2, \ldots, n$, should be chosen proportional to $1/\sigma_j^2$. The minimum value of σ is then

$$\sigma_{\min} = \sum_{i=1}^n \frac{k^2}{\sigma_i^4}\sigma_i^2 = k^2 \sum_{i=1}^n \frac{1}{\sigma_i^2},$$

where k is given by $\sum_{j=1}^n (k/\sigma_j^2) = 1$. Thus

$$\sigma_{\min} = \frac{1}{\sum_{j=1}^n (1/\sigma_j^2)} = \frac{H}{n},$$

where H is the harmonic mean of the σ_j^2.

We conclude this section with some important moment inequalities. We begin with the simple inequality

$$|a+b|^r \le c_r(|a|^r + |b|^r), \tag{20}$$

where $c_r = 1$ for $0 \le r \le 1$, and $= 2^{r-1}$ for $r > 1$. For $r = 0$ and $r = 1$, (20) is trivially true.

First note that it is sufficient to prove (20) when $0 < a \le b$. Let $0 < a \le b$, and write $x = a/b$. Then

$$\frac{(a+b)^r}{a^r + b^r} = \frac{(1+x)^r}{1+x^r}.$$

Writing $f(x) = (1+x)^r/(1+x^r)$, we see that

$$f'(x) = \frac{r(1+x)^{r-1}}{(1+x^r)^2}(1 - x^{r-1}),$$

where $0 < x \le 1$. It follows that $f'(x) > 0$ if $r > 1$, $= 0$ if $r = 1$, and < 0 if $r < 1$. Thus

$$\max_{0 \le x \le 1} f(x) = f(0) = 1 \qquad \text{if } r \le 1,$$

while

$$\max_{0 \le x \le 1} f(x) = f(1) = 2^{r-1} \qquad \text{if } r \ge 1.$$

Note that $|a+b|^r \le 2^r(|a|^r + |b|^r)$ is trivially true since

$$|a+b| \le \max(2|a|, 2|b|).$$

An immediate application of (20) is the following result.

Theorem 8. Let X and Y be RVs and $r > 0$ be a fixed number. If $E|X|^r$, $E|Y|^r$ are both finite, so also is $E|X+Y|^r$.

Proof. Let $a = X$ and $b = Y$ in (20). Taking the expectation on both sides, we see that

$$E|X+Y|^r \le c_r(E|X|^r + E|Y|^r),$$

where $c_r = 1$ if $0 < r \le 1$ and $= 2^{r-1}$ if $r > 1$.

Next we establish Hölder's inequality,

$$|xy| \le \frac{|x|^p}{p} + \frac{|y|^q}{q}, \tag{21}$$

where p and q are positive real numbers such that $p > 1$ and $1/p + 1/q = 1$. Note that for $x > 0$ the function $w = \log x$ is concave. It follows that for $x_1, x_2 > 0$,

$$\log[tx_1 + (1-t)x_2] \ge t\log x_1 + (1-t)\log x_2.$$

Taking antilogarithms, we get

$$x_1^t x_2^{1-t} \geq t x_1 + (1-t)x_2.$$

Now we choose $x_1 = |x|^p, x_2 = |y|^q, t = 1/p, 1-t = 1/q$, where $p > 1$ and $1/p + 1/q = 1$, to get (21).

Theorem 9. Let $p > 1, q > 1$, so that $1/p + 1/q = 1$. Then

$$E|XY| \leq (E|X|^p)^{1/p}(E|Y|^q)^{1/q}. \tag{22}$$

Proof. By Hölder's inequality, letting $x = X[E|X|^p]^{-1/p}$, $y = Y[E|Y|^q]^{-1/q}$, we get

$$|XY| \leq p^{-1}|X|^p[E|X|^p]^{1/p-1}[E|Y|^q]^{1/q} + q^{-1}|Y|^q[E|Y|^q]^{1/q-1}[E|X|^p]^{1/p}.$$

Taking the expectation on both sides leads to (22).

Corollary. Taking $p = q = 2$, we obtain the *Cauchy–Schwarz inequality*,

$$E|XY| \leq E^{1/2}|X|^2 E^{1/2}|Y|^2.$$

The final result of this section is an inequality due to Minkowski.

Theorem 10. For $p \geq 1$

$$[E|X+Y|^p]^{1/p} \leq [E|X|^p]^{1/p} + [E|Y|^p]^{1/p}. \tag{23}$$

Proof. We have, for $p > 1$,

$$|X+Y|^p \leq |X|\,|X+Y|^{p-1} + |Y|\,|X+Y|^{p-1}.$$

Taking expectations and using Hölder's inequality with Y replaced by $|X+Y|^{p-1}$ $(p > 1)$, we have

$$\begin{aligned} E|X+Y|^p &\leq [E|X|^p]^{1/p}[E|X+Y|^{(p-1)q}]^{1/q} + [E|Y|^p]^{1/p}[E|X+Y|^{(p-1)q}]^{1/q} \\ &= \{[E|X|^p]^{1/p} + [E|Y|^p]^{1/p}\} \cdot [E|X+Y|^{(p-1)q}]^{1/q}. \end{aligned}$$

Excluding the trivial case in which $E|X+Y|^p = 0$, and noting that $(p-1)q = p$, we have, after dividing both sides of the last inequality by $[E|X+Y|^p]^{1/q}$,

$$[E|X+Y|^p]^{1/p} \leq [E|X|^p]^{1/p} + [E|Y|^p]^{1/p}, \qquad p > 1.$$

The case $p = 1$ being trivial, this establishes (23).

PROBLEMS 4.5

1. Suppose that the RV (X, Y) is uniformly distributed over the region $R = \{(x, y): 0 < x < y < 1\}$. Find the covariance between X and Y.

2. Let (X, Y) have the joint PDF given by

$$f(x, y) = \begin{cases} x^2 + \dfrac{xy}{3} & \text{if } 0 < x < 1,\ 0 < y < 2, \\ 0 & \text{otherwise.} \end{cases}$$

Find all moments of order 2.

3. Let (X, Y) be distributed with joint density

$$f(x, y) = \begin{cases} \frac{1}{4}[1 + xy(x^2 - y^2)] & \text{if } |x| \le 1,\ |y| \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

Find the MGF of (X, Y). Are X, Y independent? If not, find the covariance between X and Y.

4. For a positive RV X with finite first moment, show that (a) $E\sqrt{X} \le \sqrt{EX}$ and (b) $E(1/X) \ge 1/EX$.

5. If X is a nondegenerate RV with finite expectation and such that $X \ge a > 0$, then

$$E\{\sqrt{X^2 - a^2}\} < \sqrt{(EX)^2 - a^2}.$$

(Kruskal [54])

6. Show that for $x > 0$,

$$\left(\int_x^\infty te^{-t^2/2}\,dt\right)^2 \le \int_x^\infty e^{-t^2/2}\,dt \int_x^\infty t^2e^{-t^2/2}\,dt,$$

and hence that

$$\int_x^\infty e^{-t^2/2}\,dt \ge \tfrac{1}{2}[(4 + x^2)^{1/2} - x]e^{-x^2/2}.$$

7. Given a PDF f that is nondecreasing in the interval $a \le x \le b$, show that for any $s > 0$

$$\int_a^b x^{2s} f(x)\,dx \ge \frac{b^{2s+1} - a^{2s+1}}{(2s + 1)(b - a)} \int_a^b f(x)\,dx,$$

with the inequality reversed if f is nonincreasing.

8. Derive the Lyapunov inequality (Theorem 3.4.3)

$$[E|X|^r]^{1/r} \le [E|X|^s]^{1/s}, \qquad 1 < r < s < \infty,$$

from Hölder's inequality (22).

9. Let X be an RV with $E|X|^r < \infty$ for $r > 0$. Show that the function $\log E|X|^r$ is a convex function of r.

10. Show with the help of an example that Theorem 9 is not true for $p < 1$.

11. Show that the converse of Theorem 8 also holds for independent RVs; that is, if $E|X+Y|^r < \infty$ for some $r > 0$ and X and Y are independent, then $E|X|^r < \infty$, $E|Y|^r < \infty$. (*Hint:* Without loss of generality, assume that the median of both X and Y is 0. Show that for any $t > 0$, $P\{|X+Y| > t\} > \frac{1}{2}P\{|X| > t\}$. Now use the remarks preceding Lemma 3.2.2 to conclude that $E|X|^r < \infty$.)

12. Let $(\Omega, \mathcal{S}, P)$ be a probability space and $A_1, A_2, \ldots, A_n$ be events in $\mathcal{S}$ such that $P(\cup_{k=1}^n A_k) > 0$. Show that

$$2 \sum_{1 \le j < k < n} P(A_j A_k) \ge \frac{(\sum_{k=1}^n PA_k)^2 - \sum_{k=1}^n PA_k}{P(\cup_{k=1}^n A_k)}.$$

(*Hint:* Let X_k be the indicator function of A_k, $k = 1, 2, \ldots, n$. Use the Cauchy–Schwarz inequality.) (Chung and Erdös [13])

13. Let $(\Omega, \mathcal{S}, P)$ be a probability space and $A, B \in \mathcal{S}$ with $0 < PA < 1$, $0 < PB < 1$. Define $\rho(A, B)$ by $\rho(A, B) =$ correlation coefficient between RVs I_A and I_B, where I_A, I_B, are the indicator functions of A and B, respectively. Express $\rho(A, B)$ in terms of PA, PB, and $P(AB)$, and conclude that $\rho(A, B) = 0$ if and only if A and B are independent. What happens if $A = B$ or if $A = B^c$?

(a) Show that

$$\rho(A, B) > 0 \Leftrightarrow P\{A \mid B\} > P(A) \Leftrightarrow P\{B \mid A\} > P(B)$$

and

$$\rho(A, B) < 0 \Leftrightarrow P\{A \mid B\} < PA \Leftrightarrow P\{B \mid A\} < PB.$$

(b) Show that

$$\rho(A, B) = \frac{P(AB)\,P(A^cB^c) - P(AB^c)\,P(A^cB)}{(PA\,PA^c \cdot PB\,PB^c)^{1/2}}.$$

14. Let $X_1, X_2, \ldots, X_n$ be iid RVs, and define

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n} \quad \text{and} \quad S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}.$$

Suppose that the common distribution is symmetric. Assuming the existence of moments of appropriate order, show that $\operatorname{cov}(\bar{X}, S^2) = 0$.

15. Let X, Y be iid RVs with common standard normal density

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \qquad -\infty < x < \infty.$$

Let $U = X + Y$ and $V = X^2 + Y^2$. Find the MGF of the random variable (U, V). Also, find the correlation coefficient between U and V. Are U and V independent?

16. Let X and Y be two discrete RVs:

$$P\{X = x_1\} = p_1, \qquad P\{X = x_2\} = 1 - p_1,$$

and

$$P\{Y = y_1\} = p_2, \qquad P\{Y = y_2\} = 1 - p_2.$$

Show that X and Y are independent if and only if the correlation coefficient between X and Y is zero.

17. Let X and Y be dependent RVs with common means 0, variances 1, and correlation coefficient ρ. Show that

$$E[\max(X^2, Y^2)] \le 1 + \sqrt{1 - \rho^2}.$$

18. Let X_1, X_2 be independent normal RVs with density functions

$$f_i(x) = \frac{1}{\sigma_i\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x - \mu_i}{\sigma_i}\right)^2\right], \qquad -\infty < x < \infty; \quad i = 1, 2.$$

Also let

$$Z = X_1 \cos\theta + X_2 \sin\theta \quad \text{and} \quad W = X_2 \cos\theta - X_1 \sin\theta.$$

Find the correlation coefficient, ρ, between Z and W, and show that

$$0 \le \rho^2 \le \left(\frac{\sigma_1^2 - \sigma_2^2}{\sigma_1^2 + \sigma_2^2}\right)^2.$$

19. Let $(X_1, X_2, \ldots, X_n)$ be an RV such that the correlation coefficient between each pair $X_i, X_j, i \ne j$, is ρ. Show that $-(n-1)^{-1} \le \rho \le 1$.

20. Let $X_1, X_2, \dots, X_{m+n}$ be iid RVs with finite second moment. Let $S_k = \sum_{j=1}^{k} X_j, k = 1, 2, \dots, m+n$. Find the correlation coefficient between S_n and $S_{m+n} - S_m$, where $n > m$.

21. Let f be the PDF of a positive RV, and write

$$g(x, y) = \begin{cases} \dfrac{f(x+y)}{x+y} & \text{if } x > 0,\ y > 0, \\ 0 & \text{otherwise.} \end{cases}$$

Show that g is a density function in the plane. If the mth moment of f exists for some positive integer m, find EX^m. Compute the means and variances of X and Y and the correlation coefficient between X and Y in terms of moments of f. (Adapted from Feller [23, p. 100].)

22. A die is thrown $n+2$ times. After each throw a $+$ sign is recorded for 4, 5, or 6, and a $-$ sign for 1, 2, or 3, the signs forming an ordered sequence. Each sign, except the first and the last, is attached a characteristic RV that assumes the value 1 if both the neighboring signs differ from the one between them, and 0 otherwise. Let $X_1, X_2, \dots, X_n$ be these characteristic RVs, where X_i corresponds to the $(i+1)$st sign $(i = 1, 2, \dots, n)$ in the sequence. Show that

$$E\left(\sum_1^n X_i\right) = \frac{n}{4} \quad \text{and} \quad \operatorname{var}\left(\sum_1^n X_i\right) = \frac{5n-2}{16}.$$

23. Let (X, Y) be jointly distributed with PDF f defined by $f(x, y) = \frac{1}{2}$ inside the square with corners at the points $(0, 1), (1, 0), (-1, 0), (0, -1)$ in the (x, y)-plane, and $f(x, y) = 0$ otherwise. Are X, Y independent? Are they uncorrelated?

4.6 CONDITIONAL EXPECTATION

In Section 4.2 we defined the conditional distribution of an RV X, given Y. We showed that if (X, Y) is of the discrete type, the conditional PMF of X, given $Y = y_j$, where $P\{Y = y_j\} > 0$, is a PMF when considered as a function of the x_i's (for fixed y_j). Similarly, if (X, Y) is an RV of the continuous type with PDF $f(x, y)$ and marginal densities f_1 and f_2, respectively, then at every point (x, y) at which f is continuous and at which $f_2(y) > 0$ and is continuous, a conditional density function of X, given Y, exists and may be defined by

$$f_{X|Y}(x \mid y) = \frac{f(x, y)}{f_2(y)}.$$

We also showed that $f_{X|Y}(x \mid y)$, for fixed y, when considered as a function of x is a PDF in its own right. Therefore, we can (and do) consider the moments of this conditional distribution.

Definition 1. Let X and Y be RVs defined on a probability space $(\Omega, \mathcal{S}, P)$, and let h be a Borel-measurable function. Then the *conditional expectation* of $h(X)$, given Y, written as $E\{h(X) \mid Y\}$, is an RV that takes the value $E\{h(X) \mid y\}$, defined by

$$(1)\quad E\{h(X) \mid y\} = \begin{cases} \sum_x h(x) P\{X = x \mid Y = y\} & \text{if } (X, Y) \text{ is of the discrete type and } P\{Y = y\} > 0, \\ \int_{-\infty}^{\infty} h(x) f_{X|Y}(x \mid y)\, dx & \text{if } (X, Y) \text{ is of the continuous type and } f_2(y) > 0, \end{cases}$$

when the RV Y assumes the value y.

Needless to say, a similar definition may be given for the conditional expectation $E\{h(Y) \mid X\}$.

It is immediate that $E\{h(X) \mid Y\}$ satisfies the usual properties of an expectation provided we remember that $E\{h(X) \mid Y\}$ is not a constant but an RV. The following results are easy to prove. We assume the existence of indicated expectations.

$$(2)\qquad E\{c \mid Y\} = c \qquad \text{for any constant } c$$

and

$$(3)\qquad E\{[a_1 g_1(X) + a_2 g_2(X)] \mid Y\} = a_1 E\{g_1(X) \mid Y\} + a_2 E\{g_2(X) \mid Y\},$$

for any Borel functions g_1, g_2.

$$(4)\qquad P(X \geq 0) = 1 \Longrightarrow E\{X \mid Y\} \geq 0$$

and

$$(5)\qquad P(X_1 \geq X_2) = 1 \Longrightarrow E\{X_1 \mid Y\} \geq E\{X_2 \mid Y\}.$$

The statements in (3), (4), and (5) should be understood to hold with probability 1.

$$(6)\qquad E\{X \mid Y\} = E(X), \qquad E\{Y \mid X\} = E(Y)$$

for independent RVs X and Y.

If $\phi(X, Y)$ is a function of X and Y, then

$$(7)\qquad E\{\phi(X, Y) \mid y\} = E\{\phi(X, y) \mid y\},$$

and

$$E\{\psi(X)\phi(X,Y) \mid X\} = \psi(X)E\{\phi(X,Y) \mid X\} \tag{8}$$

for any Borel function ψ.

Again, (8) should be understood as holding with probability 1. Relation (7) is useful as a computational device. See Example 3 below.

The moments of a conditional distribution are defined in the usual manner. Thus, for $r \geq 0$, $E\{X^r \mid Y\}$ defines the rth moment of the conditional distribution. We can define the central moments of the conditional distribution and, in particular, the variance. There is no difficulty in generalizing these concepts for n-dimensional distributions when $n > 2$. We leave the reader to furnish the details.

Example 1. An urn contains three red and two green balls. A random sample of two balls is drawn (a) with replacement, and (b) without replacement. Let $X = 0$ if the first ball drawn is green, $= 1$ if the first ball drawn is red, and let $Y = 0$ if the second ball drawn is green, $= 1$ if the second ball drawn is red.

The joint PMF of (X, Y) is given in the following tables:

(a) With replacement

$Y \backslash X$	0	1	
0	$\frac{4}{25}$	$\frac{6}{25}$	$\frac{2}{5}$
1	$\frac{6}{25}$	$\frac{9}{25}$	$\frac{3}{5}$
	$\frac{2}{5}$	$\frac{3}{5}$	1

(b) Without replacement

$Y \backslash X$	0	1	
0	$\frac{2}{20}$	$\frac{6}{20}$	$\frac{2}{5}$
1	$\frac{6}{20}$	$\frac{6}{20}$	$\frac{3}{5}$
	$\frac{2}{5}$	$\frac{3}{5}$	1

The conditional PMFs and the conditional expectations are as follows:

(a) $$P\{X = x \mid 0\} = \begin{cases} \frac{2}{5}, & x = 0, \\ \frac{3}{5}, & x = 1, \end{cases} \qquad P\{Y = y \mid 0\} = \begin{cases} \frac{2}{5}, & y = 0, \\ \frac{3}{5}, & y = 1, \end{cases}$$

$$P\{X = x \mid 1\} = \begin{cases} \frac{2}{5}, & x = 0, \\ \frac{3}{5}, & x = 1, \end{cases} \qquad P\{Y = y \mid 1\} = \begin{cases} \frac{2}{5}, & y = 1, \\ \frac{3}{5}, & y = 1, \end{cases}$$

$$E\{X \mid Y\} = \begin{cases} \frac{3}{5}, & y = 0 \\ \frac{3}{5}, & y = 1, \end{cases} \qquad E\{Y \mid X\} = \begin{cases} \frac{3}{5}, & x = 0, \\ \frac{3}{5}, & x = 1; \end{cases}$$

(b) $$P\{X = x \mid 0\} = \begin{cases} \frac{1}{4}, & x = 0, \\ \frac{3}{4}, & x = 1, \end{cases} \qquad P\{Y = y \mid 0\} = \begin{cases} \frac{1}{4}, & y = 0, \\ \frac{3}{4}, & y = 1, \end{cases}$$

$$P\{X = x \mid 1\} = \begin{cases} \frac{1}{2}, & x = 0, \\ \frac{1}{2}, & x = 1, \end{cases} \qquad P\{Y = y \mid 1\} = \begin{cases} \frac{1}{2}, & y = 0, \\ \frac{1}{2}, & y = 1, \end{cases}$$

$$E\{X \mid Y\} = \begin{cases} \frac{3}{4}, & y = 0, \\ \frac{1}{2}, & y = 1, \end{cases} \qquad E\{Y \mid X\} = \begin{cases} \frac{3}{4}, & x = 0, \\ \frac{1}{2}, & x = 1. \end{cases}$$

Example 2. For the RV (X, Y) considered in Examples 4.2.5 and 4.2.7,

$$E\{Y \mid x\} = \int_x^1 y f_{Y|X}(y \mid x)\, dy = \frac{1}{2}\frac{1-x^2}{1-x} = \frac{1+x}{2}, \qquad 0 < x < 1,$$

and

$$E\{X \mid y\} = \int_0^y x f_{X|Y}(x \mid y)\, dx = \frac{y}{2}, \qquad 0 < y < 1.$$

Also,

$$E\{X^2 \mid y\} = \int_0^y x^2 \frac{1}{y}\, dx = \frac{y^2}{3}, \qquad 0 < y < 1$$

and

$$\begin{aligned}\operatorname{var}\{X \mid y\} &= E\{X^2 \mid y\} - [E\{X \mid y\}]^2 \\ &= \frac{y^2}{3} - \frac{y^2}{4} = \frac{y^2}{12}, \qquad 0 < y < 1.\end{aligned}$$

Theorem 1. Let $Eh(X)$ exist. Then

$$Eh(X) = E\{E\{h(X) \mid Y\}\}. \tag{9}$$

Proof. Let (X, Y) be of the discrete type. Then

$$\begin{aligned}E\{E\{h(X) \mid Y\}\} &= \sum_y \left[\sum_x h(x) P\{X = x \mid Y = y\}\right] P\{Y = y\} \\ &= \sum_y \left[\sum_x h(x) P\{X = x, Y = y\}\right] \\ &= \sum_x h(x) \sum_y P\{X = x, Y = y\} \\ &= Eh(X).\end{aligned}$$

The proof in the continuous case is similar.

Theorem 1 is quite useful in computation of $Eh(X)$ in many applications.

Example 3. Let X and Y be independent continuous RVs with respective PDF f and g and DFs F and G. Then $P\{X < Y\}$ is of interest in many statistical applications. In view of Theorem 1,

$$P\{X < Y\} = EI_{\{X<Y\}} = E\{E\{I_{\{X<Y\}}|Y\}\}$$

where I_A is the indicator function of event A. Now

$$\begin{aligned}E\{I_{\{X<Y\}}|Y=y\} &= E\{I_{[X<y]} \mid y\}\\ &= E(I_{[X<y]}) = F(y)\end{aligned}$$

and it follows that

$$P\{X<Y\} = E\{F(Y)\} = \int_{-\infty}^{\infty} F(y)g(y)\,dy.$$

If, in particular, $X \stackrel{d}{=} Y$, then

$$P\{X<Y\} = \int_{-\infty}^{\infty} F(y)f(y)\,dy = \tfrac{1}{2}.$$

More generally,

$$\begin{aligned}P\{X-Y\le z\} &= E\{E\{I_{\{X-Y\le z\}} \mid Y\}\} = E[F(Y+z)]\\ &= \int_{-\infty}^{\infty} F(y+z)g(y)\,dy\end{aligned}$$

gives the DF of $Z = X - Y$ as computed in corollary to Theorem 4.4.3.

Example 4. Consider the joint PDF

$$f(x,y) = xe^{-x(1+y)}, \qquad x \ge 0, \quad y \ge 0, \quad \text{and zero otherwise}$$

of (X, Y). Then

$$f_X(x) = e^{-x}, \qquad x \ge 0, \quad \text{and zero otherwise}$$

and

$$f_Y(y) = \frac{1}{(1+y)^2}, \qquad y \ge 0, \quad \text{and zero otherwise.}$$

Clearly, EY does not exist but

$$E\{Y \mid x\} = \int_0^{\infty} yxe^{-xy}\,dy = \frac{1}{x}.$$

Theorem 2. If $EX^2 < \infty$, then

(10) $$\operatorname{var}(X) = \operatorname{var}(E\{X \mid Y\}) + E(\operatorname{var}\{X \mid Y\}).$$

Proof. The right-hand side of (10) equals, by definition,

$$\begin{aligned}\{E(E\{X \mid Y\})^2 &- [E(E\{X \mid Y\})]^2\} + E(E\{X^2 \mid Y\} - (E\{X \mid Y\})^2)\\ &= \{E(E\{X \mid Y\})^2 - (EX)^2\} + EX^2 - E(E\{X \mid Y\})^2\\ &= \text{var}(X).\end{aligned}$$

Corollary. If $EX^2 < \infty$, then

$$\text{var}(X) \geq \text{var}(E\{X \mid Y\}) \tag{11}$$

with equality if and only if X is a function of Y.

Equation (11) follows immediately from (10). The equality in (11) holds if and only if

$$E(\text{var}\{X \mid Y\}) = E(X - E\{X \mid Y\})^2 = 0,$$

which holds if and only if with probability 1

$$X = E\{X \mid Y\}. \tag{12}$$

Example 5. Let $X_1, X_2, \ldots$ be iid RVs and let N be a positive integer-valued RV. Let $S_N = \sum_{k=1}^{N} X_k$ and suppose that the X's and N are independent. Then

$$E(S_N) = E\{E\{S_N \mid N\}\}.$$

Now

$$E\{S_N \mid N = n\} = E\{S_n \mid N = n\} = nEX_1$$

so that

$$E(S_N) = E(NEX_1) = (EN)(EX_1).$$

Again, we have assumed above and below that all indicated expectations exist. Also,

$$\text{var}(S_N) = \text{var}(E\{S_N \mid N\}) + E(\text{var}\{S_N \mid N\}).$$

First,

$$\text{var}(E\{S_N \mid N\}) = \text{var}(NEX_1) = (EX_1)^2 \,\text{var}(N).$$

Second,

$$\text{var}\{S_N \mid N = n\} = n \,\text{var}(X_1),$$

so

$$E(\operatorname{var}\{S_N \mid N\}) = (EN)\operatorname{var}(X_1).$$

It follows that

$$\operatorname{var}(S_N) = (EX_1)^2 \operatorname{var}(N) + (EN)\operatorname{var}(X_1).$$

PROBLEMS 4.6

1. Let X be an RV with PDF given by

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

Find $E\{X \mid a < X < b\}$, where a and b are constants.

2. (a) Let (X, Y) be jointly distributed with density

$$f(x, y) = \begin{cases} y(1+x)^{-4} e^{-y(1+x)^{-1}}, & x, y \geq 0, \\ 0, & \text{otherwise.} \end{cases}$$

Find $E\{Y \mid X\}$.

(b) Do the same for the joint density

$$f(x, y) = \begin{cases} \frac{4}{5}(x + 3y)e^{-x-2y}, & x, y \geq 0, \\ 0, & \text{otherwise.} \end{cases}$$

3. Let (X, Y) be jointly distributed with bivariate normal density

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

Find $E(X \mid y)$ and $E\{Y \mid x\}$. (Here, $\mu_1, \mu_2 \in \mathcal{R}, \sigma_1, \sigma_2 > 0$, and $|\rho| < 1$.)

4. Find $E(Y - E\{Y \mid X\})^2$.

5. Show that $E(Y - \phi(X))^2$ is minimized by choosing $\phi(X) = E\{Y \mid X\}$.

6. Let X have PMF

$$P_\lambda\{X = x\} = \frac{\lambda^x e^{-\lambda}}{x!}, \qquad x = 0, 1, 2, \ldots$$

and suppose that λ is a realization of a RV Λ with PDF

$$f(\lambda) = e^{-\lambda}, \qquad \lambda > 0.$$

Find $E\{e^{-\Lambda} \mid X = 1\}$.

7. Find $E(XY)$ by conditioning on X or Y for the following cases:
   (a) $f(x, y) = xe^{-x(1+y)}$, $x > 0$, $y > 0$, and zero otherwise.
   (b) $f(x, y) = 2$, $0 \le y \le x \le 1$, and zero otherwise.

8. Suppose that X has uniform PDF $f(x) = 1, 0 \le x \le 1$ and zero otherwise. Let Y be chosen from interval $(0, X]$ according to the PDF

$$g(y \mid x) = \frac{1}{x}, \qquad 0 < y \le x, \qquad \text{and zero otherwise}$$

Find $E\{Y^k \mid X\}$ and EY^k for any fixed constant $k > 0$.

4.7 ORDER STATISTICS AND THEIR DISTRIBUTIONS

Let $(X_1, X_2, \ldots, X_n)$ be an n-dimensional random variable, and $(x_1, x_2, \ldots, x_n)$ be an n-tuple assumed by $(X_1, X_2, \ldots, X_n)$. Arrange $(x_1, x_2, \ldots, x_n)$ in increasing order of magnitude so that

$$x_{(1)} \le x_{(2)} \le \cdots \le x_{(n)},$$

where $x_{(1)} = \min(x_1, x_2, \ldots, x_n)$, $x_{(2)}$ is the second smallest value in $x_1, x_2, \ldots, x_n$, and so on, $x_{(n)} = \max(x_1, x_2, \ldots, x_n)$. If any two x_i, x_j are equal, their order does not matter.

Definition 1. The function $X_{(k)}$ of $(X_1, X_2, \ldots, X_n)$ that takes on the value $x_{(k)}$ in each possible sequence $(x_1, x_2, \ldots, x_n)$ of values assumed by $(X_1, X_2, \ldots, X_n)$ is known as the kth-*order statistic* or *statistic of order* k. $\{X_{(1)}, X_{(2)}, \ldots, X_{(n)}\}$ is called the set of order statistics for $(X_1, X_2, \ldots, X_n)$.

Example 1. Let X_1, X_2, X_3 be three RVs of the discrete type. Also, let X_1, X_3 take on values 0, 1, and X_2 take on values 1, 2, 3. Then the RV (X_1, X_2, X_3) assumes these triplets of values: (0, 1, 0), (0, 2, 0), (0, 3, 0), (0, 1, 1), (0, 2, 1), (0, 3, 1), (1, 1, 0), (1, 2, 0), (1, 3, 0), (1, 1, 1), (1, 2, 1), (1, 3, 1); $X_{(1)}$ takes on values 0, 1; $X_{(2)}$ takes on values 0, 1; and $X_{(3)}$ takes on values 1, 2, 3.

Theorem 1. Let $(X_1, X_2, \ldots, X_n)$ be an n-dimensional RV. Let $X_{(k)}, 1 \le k \le n$, be the statistic of order k. Then $X_{(k)}$ is also an RV.

Statistical considerations such as sufficiency, completeness, invariance, and ancillarity (Chapter 8) lead to the consideration of order statistics in problems of statistical inference. Order statistics are particularly useful in nonparametric statistics (Chapter 13), where, for example, many test procedures are based on ranks of observations. Many of these methods require the distribution of the ordered observations, which we now study.

In the following we assume that $X_1, X_2, \ldots, X_n$ are iid RVs. In the discrete case there is no magic formula to compute the distribution of any $X_{(j)}$ or any of the joint distributions. A direct computation is the best course of action.

Example 2. Suppose that X_n's are iid with geometric PMF

$$p_k = P(X = k) = pq^{k-1}, \qquad k = 1, 2, \ldots, 0 < p < 1, \quad q = 1 - p.$$

Then for any integers $x \ge 1$ and $r \ge 1$,

$$P\{X_{(r)} = x\} = P\{X_{(r)} \le x\} - P\{X_{(r)} \le x - 1\}.$$

Now

$$\begin{aligned} P\{X_{(r)} \le x\} &= P\{\text{at least } r \text{ of } X\text{'s are } \le x\} \\ &= \sum_{i=1}^{r} \binom{n}{i} [P(X_1 \le x)]^i [P(X_1 > x)]^{n-i} \end{aligned}$$

and

$$P(X_1 \ge x) = \sum_{k=x}^{\infty} pq^{k-1} = (1 - p)^{x-1}.$$

It follows that

$$P\{X_{(r)} = x\} = \sum_{i=r}^{n} \binom{n}{i} q^{(x-1)(n-i)} \left\{ q^{n-i}[1 - q^x]^i - [1 - q^{x-1}]^i \right\},$$

$x = 1, 2, \ldots$. In particular, let $n = r = 2$. Then

$$P\{X_{(2)} = x\} = pq^{x-1}(pq^{x-1} + 2 - 2q^{x-1}), \qquad x \ge 1.$$

Also, for integers $x, y \ge 1$ we have

$$\begin{aligned} P\{X_{(1)} = x, X_{(2)} - X_{(1)} = y\} &= P\{X_{(1)} = x, X_{(2)} = x + y\} \\ &= P\{X_1 = x, X_2 = x + y\} + P\{X_1 = x + y, X_2 = x\} \end{aligned}$$

$$= 2pq^{x-1} \cdot pq^{x+y-1}$$
$$= 2pq^{2x-2} \cdot pq^{y} = P\{X_{(1)} = x\}P\{X_{(2)} = y\}$$

and

$$P\{X_{(1)} = 1, X_{(2)} - X_{(1)} = 0\} = P\{X_{(1)} = X_{(2)} = 1\} = p^2.$$

It follows that $X_{(1)}$ and $X_{(2)} - X_{(1)}$ are independent RVs and, moreover, that $X_{(2)} - X_{(1)}$ has a geometric distribution.

In the following we assume that $X_1, X_2, \ldots, X_n$ are iid RVs of the continuous type with PDF f. Let $\{X_{(1)}, X_{(2)}, \ldots, X_{(n)}\}$ be the set of order statistics for $X_1, X_2, \ldots, X_n$. Since the X_i are all continuous type RVs, it follows with probability 1 that

$$X_{(1)} < X_{(2)} < \cdots < X_{(n)}.$$

Theorem 2. The joint PDF of $(X_{(1)}, X_{(2)}, \ldots, X_{(n)})$ is given by

$$(1)\quad g(x_{(1)}, x_{(2)}, \ldots, x_{(n)}) = \begin{cases} n! \prod_{i=1}^{n} f(x_{(i)}), & x_{(1)} < x_{(2)} < \cdots < x_{(n)}, \\ 0, & \text{otherwise.} \end{cases}$$

Proof. The transformation from $(X_1, X_2, \ldots, X_n)$ to $(X_{(1)}, X_{(2)}, \ldots, X_{(n)})$ is not one-to-one. In fact, there are $n!$ possible arrangements of $x_1, x_2, \ldots, x_n$ in increasing order of magnitude. Thus there are $n!$ inverses to the transformation. For example, one of the $n!$ permutations might be

$$x_4 < x_1 < x_{n-1} < x_3 < \cdots < x_n < x_2.$$

Then the corresponding inverse is

$$x_4 = x_{(1)}, \quad x_1 = x_{(2)}, \quad x_{n-1} = x_{(3)}, \quad x_3 = x_{(4)}, \quad \ldots, \quad x_n = x_{(n-1)}, \quad x_2 = x_{(n)}.$$

The Jacobian of this transformation is the determinant of an $n \times n$ identity matrix with rows rearranged, since each $x_{(i)}$ equals one and only one of $x_1, x_2, \ldots, x_n$. Therefore, $J = \pm 1$, and

$$g(x_{(2)}, x_{(n)}, x_{(4)}, x_{(1)}, \ldots, x_{(3)}, x_{(n-1)}) = |J| \prod_{i=1}^{n} f(x_{(i)}), \quad x_{(1)} < x_{(2)} < \cdots < x_{(n)}.$$

The same expression holds for each of the $n!$ arrangements.

It follows (see Remark 4.4.2) that

$$g(x_{(1)}, x_{(2)}, \ldots, x_{(n)}) = \sum_{\substack{\text{all } n! \\ \text{inverses}}} \prod_{i=1}^{n} f(x_{(i)})$$

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

Example 3. Let X_1, X_2, X_3, X_4 be iid RVs with PDF f. The joint PDF of $X_{(1)}, X_{(2)}, X_{(3)}, X_{(4)}$ is

$$g(y_1, y_2, y_3, y_4) = \begin{cases} 4!\, f(y_1) f(y_2) f(y_3) f(y_4), & y_1 < y_2 < y_3 < y_4, \\ 0, & \text{otherwise.} \end{cases}$$

Let us compute the marginal PDF of $X_{(2)}$. We have

$$\begin{aligned} g_2(y_2) &= 4! \iiint f(y_1) f(y_2) f(y_3) f(y_4)\, dy_1\, dy_3\, dy_4 \\ &= 4!\, f(y_2) \int_{-\infty}^{y_2} \int_{y_2}^{\infty} \left[\int_{y_3}^{\infty} f(y_4)\, dy_4 \right] f(y_3) f(y_1)\, dy_3\, dy_1 \\ &= 4!\, f(y_2) \int_{-\infty}^{y_2} \left\{ \int_{y_2}^{\infty} [1 - F(y_3)] f(y_3)\, dy_3 \right\} f(y_1)\, dy_1 \\ &= 4!\, f(y_2) \int_{-\infty}^{y_2} \frac{[1 - F(y_2)]^2}{2} f(y_1)\, dy_1 \\ &= 4!\, f(y_2) \frac{[1 - F(y_2)]^2}{2!} F(y_2), \qquad y_2 \in \mathcal{R}. \end{aligned}$$

The procedure for computing the marginal PDF of $X_{(r)}$, the rth-order statistic of $X_1, X_2, \ldots, X_n$, is similar. The following theorem summarizes the result.

Theorem 3. The marginal PDF of $X_{(r)}$ is given by

$$(2) \qquad g_r(y_r) = \frac{n!}{(r-1)!\,(n-r)!} [F(y_r)]^{r-1} [1 - F(y_r)]^{n-r} f(y_r),$$

where F is the common DF of $X_1, X_2, \ldots, X_n$.

Proof.

$$g_r(y_r) = n! f(y_r) \int_{-\infty}^{y_r} \int_{-\infty}^{y_{r-1}} \cdots \int_{-\infty}^{y_2} \int_{y_r}^{\infty} \int_{y_{r+1}}^{\infty} \cdots \int_{y_{n-1}}^{\infty} \prod_{i \neq r}^{n} f(y_i)\, dy_n \cdots dy_{r+1}$$

$$\cdot\, dy_1 \cdots dy_{r-1}$$

$$= n!\, f(y_r) \frac{[1 - F(y_r)]^{n-r}}{(n-r)!} \int_{-\infty}^{y_2} \cdots \int_{-\infty}^{y_r} \prod_{i=1}^{r-1} [f(y_i)\, dy_i]$$

$$= n!\, f(y_r) \frac{[1 - F(y_r)]^{n-r}}{(n-r)!} \frac{[F(y_r)]^{r-1}}{(r-1)!},$$

as asserted.

We now compute the joint PDF of $X_{(j)}$ and $X_{(k)}$, $1 \le j < k \le n$.

Theorem 4. The joint PDF of $X_{(j)}$ and $X_{(k)}$ is given by

$$g_{jk}(y_j, y_k) = \begin{cases} \dfrac{n!}{(j-1)!\,(k-j-1)!\,(n-k)!} F^{j-1}(y_j)[F(y_k) \\ \quad -F(y_j)]^{k-j-1}[1 - F(y_k)]^{n-k} f(y_j) f(y_k) & \text{if } y_j < y_k, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

Proof.

$$g_{jk}(y_j, y_k) = \int_{-\infty}^{y_j} \cdots \int_{-\infty}^{y_2} \int_{y_j}^{y_k} \cdots \int_{y_{k-2}}^{y_k} \int_{y_k}^{\infty} \cdots \int_{y_{n-1}}^{\infty} n!\, f(y_1) \cdots f(y_n)$$

$$\cdot\, dy_n \cdots dy_{k+1}\, dy_{k-1} \cdots dy_{j+1}\, dy_1 \cdots dy_{j-1}$$

$$= n! \int_{-\infty}^{y_j} \cdots \int_{-\infty}^{y_2} \int_{y_j}^{y_k} \cdots \int_{y_{k-2}}^{y_k} \frac{[1 - F(y_k)]^{n-k}}{(n-k)!} f(y_1) f(y_2) \cdots f(y_k)$$

$$\cdot\, dy_{k-1} \cdots dy_{j+1}\, dy_1 \cdots dy_{j-1}$$

$$= n! \frac{[1 - F(y_k)]^{n-k}}{(n-k)!} f(y_k) \int_{-\infty}^{y_j} \cdots \int_{-\infty}^{y_2} \frac{[F(y_k) - F(y_j)]^{k-j-1}}{(k-j-1)!}$$

$$\cdot\, f(y_1) f(y_2) \cdots f(y_j)\, dy_1 \cdots dy_{j-1}$$

$$= \frac{n!}{(n-k)!\,(k-j-1)!} [1 - F(y_k)]^{n-k} [F(y_k) - F(y_j)]^{k-j-1}$$

$$\cdot\, f(y_k) f(y_j) \frac{[F(y_j)]^{j-1}}{(j-1)!}, \qquad y_j < y_k,$$

as asserted.

In a similar manner we can show that the joint PDF of $X_{(j_1)}, \ldots, X_{(j_k)}$, $1 \le j_1 < j_2 < \cdots < j_k \le n$, $1 \le k \le n$, is given by

$$g_{j_1, j_2, \ldots, j_k}(y_1, y_2 \ldots, y_k) = \frac{n!}{(j_1 - 1)!\,(j_2 - j_1 - 1)! \cdots (n - j_k)!} \cdot F^{j_1 - 1}(y_1) f(y_1)[F(y_2) - F(y_1)]^{j_2 - j_1 - 1} f(y_2) \cdots [1 - F(y_k)]^{n - j_k} f(y_k)$$

for $y_1 < y_2 < \cdots < y_k$, and $= 0$ otherwise.

Example 4. Let $X_1, X_2, \ldots, X_n$ be iid RVs with common PDF

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

Then

$$g_r(y_r) = \begin{cases} \dfrac{n!}{(r-1)!\,(n-r)!} y_r^{r-1}(1 - y_r)^{n-r}, & 0 < y_r < 1, (1 \le r \le n), \\ 0 & \text{otherwise.} \end{cases}$$

The joint distribution of $X_{(j)}$ and $X_{(k)}$ is given by

$$g_{jk}(y_j, y_k) = \begin{cases} \dfrac{n!}{(j-1)!\,(k-j-1)!\,(n-k)!} y_j^{j-1}(y_k - y_j)^{k-j-1}(1 - y_k)^{n-k}, \\ \qquad\qquad 0 < y_j < y_k < 1, \\ 0 \qquad\qquad \text{otherwise,} \end{cases}$$

where $1 \le j < k \le n$.

The joint PDF of $X_{(1)}$ and $X_{(n)}$ is given by

$$g_{1n}(y_1, y_n) = n(n-1)(y_n - y_1)^{n-2}, \qquad 0 < y_1 < y_n < 1$$

and that of the range $R_n = X_{(n)} - X_{(1)}$ by

$$g_{R_n}(w) = \begin{cases} n(n-1)w^{n-2}(1 - w), & 0 < w < 1, \\ 0, & \text{otherwise.} \end{cases}$$

Example 5. Let $X_{(1)}, X_{(2)}, X_{(3)}$ be the order statistics of iid RVs X_1, X_2, X_3 with common PDF

$$f(x) = \begin{cases} \beta e^{-x\beta}, & x > 0 \\ 0, & \text{otherwise} \end{cases} \qquad (\beta > 0).$$

Let $Y_1 = X_{(3)} - X_{(2)}$ and $Y_2 = X_{(2)}$. We show that Y_1 and Y_2 are independent. The joint PDF of $X_{(2)}$ and $X_{(3)}$ is given by

$$g_{23}(x, y) = \begin{cases} \dfrac{3!}{1!\,0!\,0!}(1 - e^{-\beta x})\beta e^{-\beta x}\beta e^{-\beta y}, & x < y, \\ 0, & \text{otherwise.} \end{cases}$$

The PDF of (Y_1, Y_2) is

$$\begin{aligned} f(y_1, y_2) &= 3!\,\beta^2(1 - e^{-\beta y_2})e^{-\beta y_2}e^{-(y_1+y_2)\beta} \\ &= \begin{cases} [3!\,\beta e^{-2\beta y_2}(1 - e^{-\beta y_2})](\beta e^{-\beta y_1}), & 0 < y_1 < \infty, 0 < y_2 < \infty, \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

It follows that Y_1 and Y_2 are independent.

Finally, we consider the moments: namely, the means, variances, and covariances of order statistics. Suppose that $X_1, X_2, \ldots, X_n$ are iid RVs with common DF F. Let g be a Borel function on $\mathcal{R}$ such that $E|g(X)| < \infty$, where X has DF F. Then for $1 \le r \le n$,

$$\begin{aligned} &\left| \int_{-\infty}^{\infty} g(x) \frac{n!}{(n-r)!(r-1)!}[F(x)]^{r-1}[1 - F(x)]^{n-r} f(x)\,dx \right| \\ &\qquad \le n\binom{n-1}{r-1} \int_{-\infty}^{\infty} |g(x)| f(x)\,dx \qquad (0 \le F \le 1) \\ &\qquad < \infty \end{aligned}$$

and we write

$$Eg(X_{(r)}) = \int_{-\infty}^{\infty} g(y) g_r(y)\,dy$$

for $r = 1, 2, \ldots, n$. The converse also holds. Suppose that $E|g(X_{(r)})| < \infty$ for $r = 1, 2, \ldots, n$. Then

$$n\binom{n-1}{r-1} \int_{-\infty}^{\infty} |g(x)| F^{r-1}(x)[1 - F(x)]^{n-r} f(x)\,dx < \infty$$

for $r = 1, 2, \ldots, n$ and hence

$$\begin{aligned} &n \int_{-\infty}^{\infty} \left\{ \sum_{r=1}^{n} \binom{n-1}{r-1} F^{r-1}(x)[1 - F(x)]^{n-r} \right\} |g(x)| f(x)\,dx \\ &\qquad = n \int_{-\infty}^{\infty} |g(x)| f(x)\,dx < \infty. \end{aligned}$$

Moreover, it also follows that

$$\sum_{r=1}^{n} Eg(X_{(r)}) = nEg(X).$$

As a consequence of the remarks above, we note that if $E|g(X_{(r)})| = \infty$ for some r, $1 \le r \le n$, then $E|g(X)| = \infty$, and conversely, if $E|g(X)| = \infty$, then $E|g|X_{(r)})| = \infty$ for some r, $1 \le r \le n$.

Example 6. Let $X_1, X_2, \dots, X_n$ be iid with Pareto PDF $f(x) = 1/x^2$, if $x \ge 1$, and $= 0$ otherwise.

Then $EX = \infty$. Now for $1 \le r \le n$,

$$\begin{aligned} EX_{(r)} &= n\binom{n-1}{r-1}\int_1^\infty x\left(1-\frac{1}{x}\right)^{r-1}\frac{1}{x^{n-r}}\frac{dx}{x^2} \\ &= n\binom{n-1}{r-1}\int_0^1 y^{r-1}(1-y)^{n-r-1}\,dy. \end{aligned}$$

Since the integral on the right side converges for $1 \le r \le n-1$ and diverges for $r > n-1$, we see that $EX_{(r)} = \infty$ for $r = n$.

PROBLEMS 4.7

1. Let $X_{(1)}, X_{(2)}, \dots X_{(n)}$ be the set of order statistics of independent RVs $X_1, X_2, \dots, X_n$ with common PDF

$$f(x) = \begin{cases} \beta e^{-x\beta} & \text{if } x \ge 0, \\ 0 & \text{otherwise.} \end{cases}$$

   (a) Show that $X_{(r)}$ and $X_{(s)} - X_{(r)}$ are independent for any $s > r$.

   (b) Find the PDF of $X_{(r+1)} - X_{(r)}$.

   (c) Let $Z_1 = nX_{(1)}, Z_2 = (n-1)(X_{(2)} - X_{(1)}), Z_3 = (n-2)(X_{(3)} - X_{(2)}), \dots, Z_n = (X_{(n)} - X_{(n-1)})$. Show that $(Z_1, Z_2, \dots, Z_n)$ and $(X_1, X_2, \dots, X_n)$ are identically distributed.

2. Let $X_1, X_2, \dots, X_n$ be iid from PMF

$$p_k = \frac{1}{N}, \qquad k = 1, 2, \dots, N.$$

   Find the marginal distributions of $X_{(1)}, X_{(n)}$, and their joint PMF.

3. Let $X_1, X_2, \dots, X_n$ be iid with a DF

$$f(y) = \begin{cases} y^\alpha & \text{if } 0 < y < 1, \\ 0 & \text{otherwise}, \qquad \alpha > 0. \end{cases}$$

Show that $X_{(i)}/X_{(n)}$, $i = 1, 2, \dots, n-1$, and $X_{(n)}$ are independent.

4. Let $X_1, X_2, \dots, X_n$ be iid RVs with common Pareto DF $f(x) = \alpha\sigma^\alpha/x^{\alpha+1}$, $x > \sigma$ where $\alpha > 0$, $\sigma > 0$. Show that:

(a) $X_{(1)}$ and $(X_{(2)}/X_{(1)}, \dots, X_{(n)}/X_{(1)})$ are independent.

(b) $X_{(1)}$ has Pareto $(\sigma, n\alpha)$ distribution.

(c) $\sum_{j=1}^{n} \ln(X_{(j)}/X_{(1)})$ has PDF

$$f(x) = \frac{x^{n-2}e^{-\alpha x}}{(n-2)!}, \qquad x > 0.$$

5. Let $X_1, X_2, \dots, X_n$ be iid nonnegative RVs of the continuous type. If $E|X| < \infty$, show that $E|X_{(r)}| < \infty$. Write $M_n = X_{(n)} = \max(X_1, X_2, \dots, X_n)$. Show that

$$EM_n = EM_{n-1} + \int_0^\infty F^{n-1}(x)[1 - F(x)]\,dx, \qquad n = 2, 3, \dots.$$

Find EM_n in each of the following cases:

(a) X_i have the common DF

$$F(x) = 1 - e^{-\beta x}, \qquad x \geq 0.$$

(b) X_i have the common DF

$$F(x) = x, \qquad 0 < x < 1.$$

6. Let $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ be the order statistics of n independent RVs $X_1, X_2, \dots, X_n$ with common PDF $f(x) = 1$ if $0 < x < 1$, and $= 0$ otherwise. Show that $Y_1 = X_{(1)}/X_{(2)}$, $Y_2 = X_{(2)}/X_{(3)}, \dots, Y_{(n-1)} = X_{(n-1)}/X_{(n)}$, and $Y_n = X_{(n)}$ are independent. Find the PDFs of $Y_1, Y_2, \dots, Y_n$.

7. For the PDF in Problem 4, find $EX_{(r)}$.

8. An urn contains N identical marbles numbered 1 through N. From the urn n marbles are drawn, and let $X_{(n)}$ be the largest number drawn. Show that $P(X_{(n)} = k) = \binom{k-1}{n-1} \Big/ \binom{N}{n}$, $k = n, n+1, \dots, N$, and $EX_{(n)} = n(N+1)/(n+1)$.