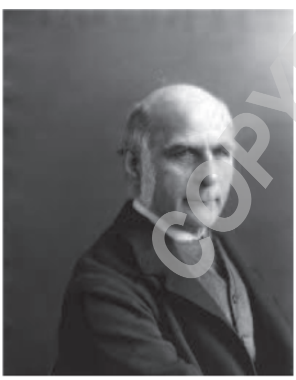
CHAPTER 1

TWO-DIMENSIONAL DISCRETE RANDOM VARIABLES AND DISTRIBUTIONS

Francis Galton (Birmingham 1822 – Haslemere, England 1911)



Sir Francis Galton, a half-cousin of Charles Darwin, was born on 16 February 1822, in Birmingham, England, and died on 17 January 1911, at the age of 88, in Haslemere, England. He was a renowned statistician, sociologist, psychologist, and anthropologist, in addition to being a tropical explorer and geographer. Galton published numerous papers and books. He was very much interested in the joint behavior of quantities, created the statistical concept of correlation, and promoted regression toward the mean, a method which is currently of widespread use in all areas of statistical applications.

Introduction to Probability: Multivariate Models and Applications, Volume 2, First Edition.
N. Balakrishnan, Markos V. Koutras, and Konstadinos G. Politis.

1.1 INTRODUCTION

Quite often, in order to study a random experiment, it is not enough to observe a single characteristic (i.e. a random variable); we might be interested in the study of two or more (usually numerical) characteristics. For example:

- when rolling two dice, we focus on both the indication of the first and the second die;
- when studying the operation of a gas station it makes sense to look at the number of cars waiting to be served in each of the gas pumps of the station.

In these cases, apart from the study of each random variable separately, the determination of the behavior of one in relation to the behavior of the others might be of particular interest. In this chapter, we will deal systematically with the case where we have two discrete random variables (as in the first example above), while the case of two continuous random variables will be considered in Chapter 2. The joint behavior of more than two random variables will be covered later on in Chapter 3.

When one studies a discrete variable, two functions which play a prominent role are the probability function and the (cumulative) distribution function. In this chapter, we study the simultaneous behavior, from a probabilistic viewpoint, of two discrete random variables and we extend the definition of these functions to cover the bivariate case, that is the case of two variables. In such a case, one speaks of a *joint probability function* and a *joint distribution function*, respectively. We shall also consider the expectation for a function of these two random variables (which is a random variable itself) and the conditional expectation of one variable given the other.

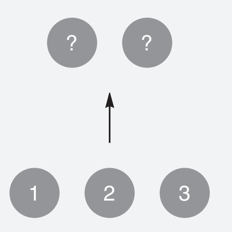
1.2 JOINT PROBABILITY FUNCTION

We begin our discussion with the simple case wherein, on the same random experiment, we are interested simultaneously in two discrete random variables. In the next chapter, we will discuss the case of two continuous random variables, while in Chapter 3, we will consider the case of more than two variables.

We shall start with an introductory example to illustrate some of the ideas involved.

Example 1.2.1 A box contains three balls numbered 1 to 3. We select two balls at random, without replacement, and let us define

X: the number on the first ball drawn,
Y: the largest number on the two balls selected.



As in the case of a single variable, it is convenient to define a sample space Ω whose elements are equally likely. Because the experiment involves selecting two balls, we consider the elementary events (i, j) associated with the experiment, in which i denotes the number on the first ball and j the number on the second ball. There are clearly three choices for the number on the first ball, so that i can take any value in the set $\{1, 2, 3\}$. However, as the sampling is without replacement, j cannot be the same as i, which means that for each value of $i = 1, 2, 3$, the number j on the second ball can take a value from the set $\{1, 2, 3\} - \{i\}$. We thus arrive at the following sample space

$$\Omega = \{(1, 2), (1, 3), (2, 1), (2, 3), (3, 1), (3, 2)\}.$$

This contains six elementary events which are clearly equally likely, i.e. $P(\omega) = 1/6$ for any $\omega \in \Omega$ (in familiar terminology, the elementary events in our sample space are equiprobable). Suppose, in a realization of this experiment, we get the outcome $\omega = (1, 2)$. Then, form the definition of the variables X and Y, we obtain $X = X(\omega) = 1$ and $Y = Y(\omega) = 2$. Instead, if we get the outcome $\omega = (3, 2)$, then the variables X and Y take the values $X = X(\omega) = 3$ and $Y = Y(\omega) = 3$. Table 1.1 presents all possible outcomes from this experiment, along with the values of X and Y in each case:

Looking at the last row of this table, it is apparent that the outcome $X = 1, Y = 2$ may occur only with the outcome $\omega = (1, 2)$ and thus corresponds to the elementary event

$$A = \{\omega \in \Omega : X(\omega) = 1, Y(\omega) = 2\} = \{(1, 2)\}.$$

The probability of the event A is $1/6$, and so we see that the probability the pair of variables (X, Y) takes the value $(1, 2)$ is $1/6$. We denote this probability by $P(X = 1, Y = 2)$, so that we have

$$P(X = 1, Y = 2) = \frac{1}{6}.$$

Consider now the case when both X and Y take the value 3. From Table 1.1, we observe that this occurs if **either** of the outcomes $\omega = (3, 1)$, $\omega = (3, 2)$ occur. Hence, the result $X = 3, Y = 3$ corresponds to the event

$$B = \{\omega \in \Omega : X(\omega) = 3, Y(\omega) = 3\} = \{(3, 1), (3, 2)\},$$

and so the probability of event B is $2/6$, i.e.

$$P(X = 3, Y = 3) = \frac{2}{6}.$$

$\omega \in \Omega$	$(1, 2)$	$(1, 3)$	$(2, 1)$	$(2, 3)$	$(3, 1)$	$(3, 2)$
$P(\omega)$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$
$X = X(\omega)$	1	1	2	2	3	3
$Y = Y(\omega)$	2	3	2	3	3	3
(X, Y)	$(1, 2)$	$(1, 3)$	$(2, 2)$	$(2, 3)$	$(3, 3)$	$(3, 3)$

Table 1.1 Values of the pair (X, Y) with their respective probabilities.

The quantity $P(X = x, Y = y)$, in the above example, expresses the joint allocation of probabilities for values that X, Y may take on together. This generalizes the concept of a probability function for a single variable, $P(X = x)$, and is therefore called the **joint probability function** of the pair (X, Y).

Definition 1.2.1 Let Ω be the sample space for a chance experiment, and X, Y be two discrete random variables defined on it. Denote by $R_{X,Y}$ the set of all possible values for the pair (X, Y). Then, the function defined by

$$f(x, y) = P(X = x, Y = y) = P(\{\omega \in \Omega : X(\omega) = x \text{ and } Y(\omega) = y\}),$$

for any $(x, y) \in R_{X,Y}$, is called the joint probability function of the variables X, Y or, equivalently, the probability function of the two-dimensional random variable (or, the random pair) (X, Y). The distribution of (X, Y) is referred to as a bivariate (or two-dimensional) discrete distribution.

The set $R_{X,Y}$ is referred to as **the range of values** for the pair (X, Y). It is obvious that if (x, y) belongs to $R_{X,Y}$, then x must belong to the range of values of X, say R_X, and y to the range of values of Y, say R_Y. Thus, we have

$$R_{X,Y} \subseteq R_X \times R_Y \subseteq \mathbb{R}^2,$$

where $\times$ denotes the Cartesian product between two sets; moreover,

$$R_X = \{x \in \mathbb{R} : (x, y) \in R_{X,Y}\} \subseteq \mathbb{R}, \quad R_Y = \{y \in \mathbb{R} : (x, y) \in R_{X,Y}\} \subseteq \mathbb{R}.$$

In the univariate case, two main properties of the probability function for a variable X are that it takes nonnegative values, and the sum of its values over the entire range of values of X is equal to one. These properties also hold in the bivariate case, when we consider the probability function $f(x, y)$ associated with the pair (X, Y), as we explain below.

First, in Definition 1.2.1, we see that the function $f(x, y)$ is defined as a probability, and so its values cannot be negative; i.e. we have $0 \leq f(x, y) \leq 1$ for any x, y. Next, if we consider all pairs (x, y) in the set $R_{X,Y}$, we see that the events

$$\{\omega \in \Omega : X(\omega) = x, Y(\omega) = y\}$$

form a *partition* of the sample space Ω (try to explain why!). This, in particular, means that if we add all the values of $f(x, y)$ for all pairs (x, y), we obtain the probability of the entire sample space, i.e.

$$\sum_{(x,y)\in R_{X,Y}} f(x, y) = 1.$$

We state these properties in the following proposition, along with another property of the function f which is evident.

Proposition 1.2.1 *The joint probability function $f(x, y)$ of a random pair (X, Y), with range $R_{X,Y}$, satisfies the following properties:*

1. $f(x, y) = 0$ *for any* $(x, y) \notin R_{X,Y}$,

2. $f(x, y) \geq 0$ *for any* $(x, y) \in R_{X,Y}$,

3. $\displaystyle\sum_{(x,y)\in R_{X,Y}} f(x, y) = 1.$

We should mention at this point that the converse of Proposition 1.2.1 is also true. To be specific, if a real function $f(x, y)$ of two variables satisfies the three properties given in Proposition 1.2.1, then there exists a pair (X, Y) of random variables such that f is the joint probability function of (X, Y).

Example 1.2.1 (Continued)
From Table 1.1, we can present the joint probability function of the pair (X, Y) in Example 1.2.1 in a concise from, as in Table 1.2:

$x \backslash y$	2	3
1	1/6	1/6
2	1/6	1/6
3	–	2/6

Table 1.2 The joint probability function $f(x, y) = P(X = x, Y = y)$ for the pair (X, Y).

In this example, we have the range of values of (X, Y) as

$$R_{X,Y} = \{(1, 2), (1, 3), (2, 2), (2, 3), (3, 3)\},$$

from which we observe

$$R_{X,Y} \subseteq R_X \times R_Y = \{1, 2, 3\} \times \{2, 3\}.$$

Moreover, it is easy to verify that

$$\sum_{(x,y)\in R_{X,Y}} f(x, y) = f(1, 2) + f(1, 3) + f(2, 2) + f(2, 3) + f(3, 3)$$

$$= \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{2}{6} = 1.$$

Example 1.2.2 Find the value of the constant c in each of the two cases below, so that $f(x, y)$ is the joint probability function of a random pair (X, Y):

(i) $f(x, y) = cxy$, where $x = 1, 2, 3$ and $y = 1, 2, 3$
(ii) $f(x, y) = \dfrac{c}{2^x 4^y}$, where $x, y = 0, 1, 2, \ldots$.

SOLUTION In each case, we find the value of c such that the function f satisfies

$$f(x, y) \geq 0 \quad \text{for any } (x, y) \in R_{X,Y}$$

and

$$\sum_{(x,y)\in R_{X,Y}} f(x, y) = 1$$

(as in the case of one variable, we do not have to check the condition $f(x, y) = 0$ for $(x, y) \notin R_{X,Y}$, as this is implicit by the statement of the example).

(i) The first condition above is satisfied whenever $c \geq 0$. The second condition gives

$$\sum_{(x,y)\in R_{X,Y}} f(x, y) = \sum_{x=1}^{3} \sum_{y=1}^{3} f(x, y) = 1.$$

The joint probability function of the pair (X, Y) is as follows:

Using the formula for $f(x, y)$ given in the statement, we get

$$1 = \sum_{x=1}^{3} \sum_{y=1}^{3} f(x, y) = c + 2c + 3c + 2c + 4c + 6c + 3c + 6c + 9c = 36c,$$

which implies that the required value of the constant c, so that $f(x, y)$ is a proper joint probability function, is $c = 1/36$ (Table 1.3).

$x \backslash y$	1	2	3
1	c	$2c$	$3c$
2	$2c$	$4c$	$6c$
3	$3c$	$6c$	$9c$

Table 1.3 The joint probability function $f(x, y)$ of the pair (X, Y) in Part (i) of Example 1.2.2.

(ii) The condition $f(x, y) \geq 0$ leads to $c \geq 0$ again. In order to calculate the (double) sum over all x, y for $x = 0, 1, 2 \ldots$ and $y = 0, 1, 2, \ldots$, we observe first that

$$\sum_{x=0}^{\infty} \sum_{y=0}^{\infty} f(x, y) = \sum_{x=0}^{\infty} \sum_{y=0}^{\infty} \frac{c}{2^x 4^y} = c \sum_{x=0}^{\infty} \left(\frac{1}{2}\right)^x \left\{ \sum_{y=0}^{\infty} \left(\frac{1}{4}\right)^y \right\}.$$

Now, upon recalling the (infinite) sum of a geometric series

$$\sum_{k=0}^{\infty} a^k = \frac{1}{1-a}, \quad \text{for } |a| < 1,$$

and applying this twice in the previous formula, we obtain

$$\sum_{x=0}^{\infty}\sum_{y=0}^{\infty} f(x,y) = c\sum_{x=0}^{\infty}\left(\frac{1}{2}\right)^x \cdot \frac{1}{1-\frac{1}{4}} = \frac{4c}{3}\sum_{x=0}^{\infty}\left(\frac{1}{2}\right)^x = \frac{4c}{3}\cdot\frac{1}{1-\frac{1}{2}} = \frac{8c}{3}.$$

Thus, for f to be a valid joint probability function in this case, we must have $8c/3 = 1$, which yields immediately that $c = 3/8$.

For a two-dimensional random variable (X, Y), if we know its joint probability function $f(x, y)$, we can, at least in principle, calculate any probability associated with these variables. Specifically, let A be a subset of the range $R_{X,Y}$ for the pair (X, Y). Then, the probability of the event that (X, Y) takes some value in the set A can be found as

$$P\left((X,Y) \in A\right) = \sum_{(x,y)\in A} P(X = x, Y = y) = \sum_{(x,y)\in A} f(x,y).$$

Using this formula, for the set

$$A = \{(s,t) \in R_{X,Y} \,:\, s \le x \text{ and } t \le y\},$$

where x, y are any given real numbers we obtain

$$P(X \le x, Y \le y) = \sum_{s\le x}\sum_{t\le y} P(X = s, Y = t) = \sum f(s,t), \tag{1.1}$$

where the last summation extends over all pairs (s, t) in the set $R_{X,Y}$ such that $s \le x$ and $t \le y$.

The function

$$F(x,y) = P(X \le x, Y \le y), \quad x, y \in \mathbb{R}, \tag{1.2}$$

is called the **joint (cumulative) distribution function** of the two-dimensional random variable (X, Y).

Example 1.2.3 A shop has two cashiers. Let X denote the number of customers waiting for service by the first cashier at 11 : 00am on a given day, and Y denote the number of customers waiting for service by the second cashier, at the same time instant. Suppose the shop manager has available data from the previous five months and (interpreting a probability as the limit of the corresponding relative frequency, as is done in the case of single variable), she has determined the joint probabilities $P(X = x, Y = y)$, for X and Y, as in Table 1.4.

$x \backslash y$	0	1	2
0	0.10	0.04	0.02
1	0.08	0.20	0.06
2	0.06	0.14	0.30

Table 1.4 The joint probability function $f(x, y) = P(X = x, Y = y)$ for the pair (X, Y) in Example 1.2.3.

To check that the function $f(x, y)$ given in Table 1.4 is indeed a valid joint probability function, we note first that

$$f(x, y) \geq 0 \quad \text{for any } (x, y) \in R_{X,Y} = \{(0, 0), (0, 1), \ldots, (2, 2)\}.$$

Also, adding all probabilities in the table, we find that

$$\sum_{(x,y)\in R_{X,Y}} f(x, y) = \sum_{x=0}^{2} \sum_{y=0}^{2} f(x, y) = f(0, 0) + f(0, 1) + \cdots + f(2, 2)$$
$$= 0.10 + 0.04 + \cdots + 0.30 = 1.$$

Thus, for example, the probability that at 11 : 00am there is one person waiting for service by each cashier is

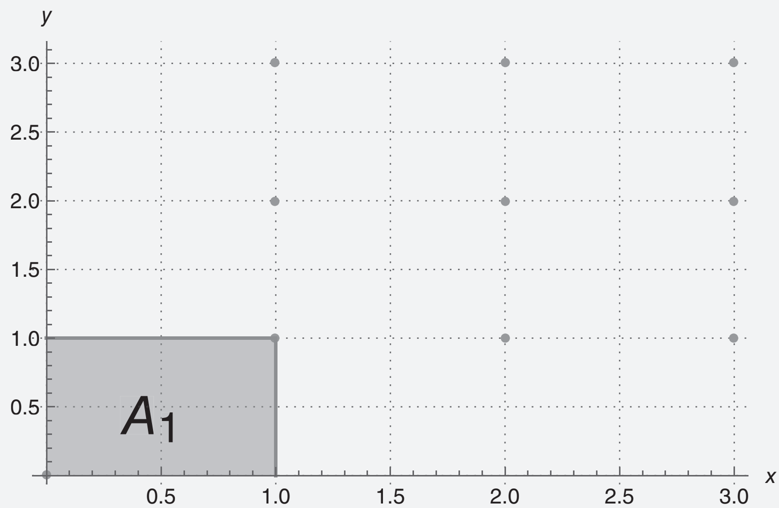
$$P(X = 1, Y = 1) = f(1, 1) = 0.20,$$

while the probability that there are two persons waiting for service by the first cashier and none waiting to be served by the second cashier is

$$P(X = 2, Y = 0) = f(2, 0) = 0.06.$$

Next, the probability that there is at most one person waiting for service by each cashier is the probability that (X, Y) takes values in the shaded area A_1 in the graph below and this can happen if the value of (X, Y) belongs to the set $\{(0, 0), (0, 1), (1, 0), (1, 1)\}$. Thus,

$$P(X \leq 1, Y \leq 1) = F(1, 1) = P((X, Y) \in A_1) = \sum_{x=0}^{1} \sum_{y=0}^{1} f(x, y)$$
$$= f(0, 0) + f(0, 1) + f(1, 0) + f(1, 1) = 0.42.$$



If we want the probability that the number of persons waiting for service is the same for both cashiers, this is simply

$$\begin{aligned} P(X = Y) = P\left((X, Y) \in \{(0, 0), (1, 1), (2, 2)\}\right) &= f(0, 0) + f(1, 1) + f(2, 2) \\ &= 0.60, \end{aligned}$$

while the probability that the *total number* of persons waiting for service by the two cashiers is equal to 3 is

$$\begin{aligned} P(X + Y = 3) = P\left((X, Y) \in \{(1, 2), (2, 1)\}\right) &= f(1, 2) + f(2, 1) \\ &= 0.06 + 0.14 = 0.20. \end{aligned}$$

Finally, the probability that there is one more person waiting for service by the first cashier than by the second one is equal to

$$\begin{aligned} P(X = Y + 1) = P\left((X, Y) \in \{(1, 0), (2, 1)\}\right) &= f(1, 0) + f(2, 1) \\ &= 0.08 + 0.14 = 0.22. \end{aligned}$$

We now list some properties of a bivariate distribution function in the discrete case, most of which parallel and generalize in a straightforward way the properties of a distribution function in the univariate case. If we have a single variable X having distribution F, then we saw in Chapter 4 of Volume I that

$$\lim_{t \to -\infty} F(t) = 0, \quad \lim_{t \to \infty} F(t) = 1.$$

The analogous property in the bivariate case is that

$$\lim_{\substack{x \to -\infty \\ y \to -\infty}} F(x, y) = 0$$

and

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

from the last relation, we see that $F(x, y)$ tends to one when *both* x and y tend to infinity.

Next, when there is only one variable X, its distribution function, $F(t)$, has a single argument t and is nondecreasing in that argument. For the case when two variables X, Y have a joint distribution given in (1.2), we see that for any real x_1, x_2 with $x_1 \leq x_2$ and $y \in \mathbb{R}$,

$$F(x_1, y) = P(X \leq x_1, Y \leq y) \leq P(X \leq x_2, Y \leq y) = F(x_2, y).$$

Similarly, if $y_1 \leq y_2$, we then obtain for any real x that

$$F(x, y_1) = P(X \leq x, Y \leq y_1) \leq P(X \leq x, Y \leq y_2) = F(x, y_2).$$

We thus see that, in the bivariate case, $F(x, y)$ is nondecreasing in *each of its arguments* x and y.

Finally, recall that a univariate discrete distribution function $F(t)$ changes value (more precisely, it has an upward jump) only at the points belonging to the range of the corresponding random variable. The same is true in the bivariate case; in particular, $F(x, y)$ defined in (1.2) changes values *only* at the points (x, y) belonging to the range $R_{X,Y}$.

The following example illustrates all these aspects.

Example 1.2.4 Assume that the random variables X and Y can take on only the values 0 and 1, and that

$$P(X = 0, Y = 0) = 1/3, \quad P(X = 0, Y = 1) = 1/6, \quad P(X = 1, Y = 0) = 1/9.$$

Obtain the joint cumulative distribution function of X and Y.

SOLUTION First, we observe that the joint probability function is not completely specified in the statement. If we add the probabilities of the three events $\{X = 0, Y = 0\}$, $\{X = 0, Y = 1\}$ and $\{X = 1, Y = 0\}$ we do not get 1, but rather

$$\frac{1}{3} + \frac{1}{6} + \frac{1}{9} = \frac{11}{18}.$$

This is due to the fact that in the statement we are only given the probabilities for three values of the pair (X, Y), while (since both X and Y assume two values each) it is clear that there are four possibilities. The probability which is *not given* is that of the event

$\{X = 1, Y = 1\}$ and since we must have the sum of values of a joint probability function over its range to be 1, we readily find

$$\begin{aligned} P(X = 1, Y = 1) & \\ &= 1 - [P(X = 0, Y = 0) + P(X = 0, Y = 1) + P(X = 1, Y = 0)] \\ &= 1 - \frac{11}{18} = \frac{7}{18}. \end{aligned}$$

Now, in order to find the joint distribution function of X and Y, we first note that when either $x < 0$ or $y < 0$, then $F(x, y) = 0$. To see what happens in the remaining cases, we need to consider only the four pairs

$$(0, 0),\ (0, 1),\ (1, 0),\ (1, 1)$$

in the range of (X, Y). Specifically, we find the following:

- For $0 \le x < 1, 0 \le y < 1$, we have (see area Π_1 in the graph)

$$F(x, y) = P(X \le x, Y \le y) = \sum_{(s,t)\in R_{X,Y} : s\le x, t\le y} f(s, t) = f(0, 0) = \frac{1}{3};$$

- For $0 \le x < 1, y \ge 1$, we get (area Π_2 in the graph)

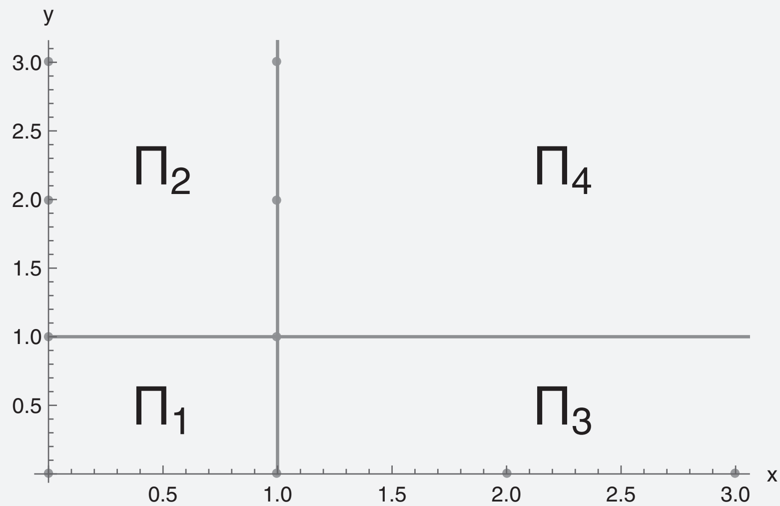
$$\begin{aligned} F(x, y) = P(X \le x, Y \le y) &= \sum_{(s,t)\in R_{X,Y} : s\le x, t\le y} f(s, t) \\ &= f(0, 0) + f(0, 1) = \frac{1}{3} + \frac{1}{6} = \frac{1}{2}; \end{aligned}$$

- For $x \ge 1, 0 \le y < 1$, which corresponds to area Π_3 in the graph, we have that

$$\begin{aligned} F(x, y) = P(X \le x, Y \le y) &= \sum_{(s,t)\in R_{X,Y} : s\le x, t\le y} f(s, t) \\ &= f(0, 0) + f(1, 0) = \frac{1}{3} + \frac{1}{9} = \frac{4}{9}; \end{aligned}$$

- Finally, for $x \ge 1$ and $y \ge 1$ (area Π_4 in the graph), we obtain similarly that

$$\begin{aligned} F(x, y) = P(X \le x, Y \le y) &= \sum_{(s,t)\in R_{X,Y} : s\le x, t\le y} f(s, t) \\ &= f(0, 0) + f(0, 1) + f(1, 0) + f(1, 1) = \frac{1}{3} + \frac{1}{6} + \frac{1}{9} + \frac{7}{18} = 1. \end{aligned}$$



Summarizing the above, we can express the joint distribution function of the random vector (X, Y) as

$$F(x,y) = P(X \leq x, Y \leq y) = \begin{cases} 0, & \text{if } \; x < 0 \quad \text{or} \quad y < 0 \\ 1/3, & \text{if } \; 0 \leq x < 1 \quad \text{and} \quad 0 \leq y < 1 \\ 4/9, & \text{if } \; x \geq 1 \quad \text{and} \quad 0 \leq y < 1 \\ 1/2, & \text{if } \; 0 \leq x < 1 \quad \text{and} \quad y \geq 1 \\ 1, & \text{if } \; x \geq 1 \quad \text{and} \quad y \geq 1. \end{cases}$$

EXERCISES

Group A

1. At a gas station, there are four dispensing pumps, two of which are for gas and the other two are for diesel. Let X be the number of gas pumps which are in use at a particular time during the day, and Y be the number of diesel pumps in use at the same time. The joint probability function, $f(x, y) = P(X = x, Y = y)$, of the variables X and Y is as given in the following table:

$x \backslash y$	0	1	2
0	0.13	0.07	0.02
1	0.15	0.19	0.21
2	0.04	0.12	0.07

(i) Calculate the probabilities $P(X = 1, Y = 1)$ and $P(X \leq 1, Y \leq 1)$.

(ii) Explain in words what the event $\{X \neq 0, Y \neq 0\}$ represents, and find the probability of that event.

(iii) Obtain the probability for each of the following events:

(a) the number of gas pumps in use is the same as the number of diesel pumps in use;

(b) the number of gas pumps in use is smaller than the number of diesel pumps in use;

(c) the total number of pumps in use is 3;

(d) the total number of pumps in use is at least 3.

2. A mini market has three cashier points. The first of these (Cashier Point I) serves only customers who have purchased at most five different products from the mini market, while the other two serve customers who have purchased more than five different products. Let X be the number of persons for service at Cashier Point I and Y be the total number of customers who have purchased more than five products and are waiting at the other two cashier points. We assume that the joint probability function, $f(x, y) = P(X = x, Y = y)$, of X and Y is known to be as follows:

$x \backslash y$	0	1	2	3
0	0.11	0.06	0.02	0.00
1	0.13	0.09	0.04	0.02
2	0.06	0.10	0.05	0.02
3	0.06	0.05	0.04	0.02
4	0.01	0.02	0.04	0.06

(i) Write down the range, $R_{X,Y}$, of the joint distribution of X and Y. Do we have $R_{X,Y} = R_X \times R_Y$?

(ii) Calculate the probabilities $P(X = 1, Y = 2)$ and $P(X = 1, Y \geq 2)$.

(iii) Explain in words what each of the following events means and find the associated probability:

$$\{X = Y\},\ \{X = Y + 2\},\ \{X \geq Y + 2\},\ \{X \geq 1 \text{ and } 2X \leq Y\}.$$

3. Nick has a red and a green dice and he rolls both of them. Let X be the outcome of the green die and Y be the sum of the two outcomes.

(i) What is the range of values for each of the variables X and Y?

(ii) Write down the range of the two-dimensional random variable (X, Y).

(iii) Find the joint probability function of X and Y, and check that it satisfies the conditions of Proposition 1.2.1.

4. The joint probability function of the variables X and Y is

$$f(x, y) = P(X = x, Y = y) = \begin{cases} \dfrac{cx}{y}, & \text{if } x = 1, 2 \text{ and } y = 1, 2, \\ 0, & \text{otherwise.} \end{cases}$$

(i) Find the value of c.

(ii) Write down the joint distribution function of X and Y. Thence, calculate $P(X \neq Y)$.

5. Obtain the joint distribution function of the variables X and Y in Example 1.2.1.

6. Find the value of $c \in \mathbb{R}$ such that, in each of the cases below, the function f will be a valid joint probability function of a two-dimensional random variable:

   (i) $f(x,y) = cxy$, for $x = 1, 2$ and $y = 2, 4, 6$;

   (ii) $f(x,y) = c$, for $x = 1, 2, \ldots, m$ and $y = 1, 2, \ldots, n$, where m, n are two given positive integers;

   (iii) $f(x,y) = c(x + y)$, for $x = 0, 1, 2$ and $y = 0, 1, 2$;

   (iv) $f(x,y) = c(x^2 + y^2)$, for $x = -1, 0, 1, 2$ and $y = -1, 2, 3$;

   (v) $f(x,y) = \dfrac{c}{3^x 7^y}$, for $x = 0, 1, 2, \ldots$ and $y = 0, 1, 2, \ldots$.

7. Two discrete variables X and Y have a joint probability function as

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

   Calculate the following probabilities:

$$P(X = Y),\ P(X > Y),\ P(X < Y),\ P(X + Y = 2),\ P(X + Y > 2).$$

Group B

8. In the experiment of tossing two dice, let X be the number of times that a six appears and Y be the number of times an even integer appears.

   (i) Derive the joint probability function of X and Y.

   (ii) Calculate the following probabilities:

$$P(X \leq 2, Y = 2),\ P(X > 0, Y < 2),\ P(Y > X),\ P(Y = 2X).$$

9. In a certain city, there are three candidates – Smith, Jones, and Allen – running for Mayor in the forthcoming election. It has been estimated by a recent poll that 45% of eligible voters support Mr. Smith, 30% support Ms. Jones, and the remaining 25% support Mr. Allen. We select at random a person who is eligible to vote, and define the following random variables:

$$X = \begin{cases} 1, & \text{if the person supports Mr. Smith,} \\ 0, & \text{otherwise,} \end{cases}$$

   and

$$Y = \begin{cases} 1, & \text{if the person supports Ms. Jones,} \\ 0, & \text{otherwise.} \end{cases}$$

(i) Find the joint probability function of the random variable (X, Y).

(ii) Find $P(X + Y = 0)$ and $P(X + Y = 1)$.

10. The random vector (X, Y) has its joint probability function as follows:

$x \backslash y$	0	1	2
2	0.20	0.15	0.05
3	0.08	0.12	0.08
4	0.07	0.13	0.12

(i) Write down, in a similar tabular form, the joint probability function of the random pair (Z, W), where

$$Z = X + Y, \quad W = X - Y,$$

after specifying the range of values for (Z, W).

(ii) Derive the joint distribution function of (Z, W), and then use it to calculate $P(Z \leq 4, W \leq 1)$.

11. Let (X, Y) be a random vector with range

$$R_{X,Y} = \{(x, y) : x = 0, \pm 1, \pm 2, \ldots, \pm k \text{ and } y = 0, \pm 1, \pm 2, \ldots, \pm k\}$$

for a given positive integer k. Express each of

$$P(X > Y), \quad P(X > |Y|), \quad P(|X| \geq Y), \quad P(X > Y + 1)$$

in terms of the joint probability function

$$f(x, y) = P(X = x, Y = y)$$

of X and Y.

1.3 MARGINAL DISTRIBUTIONS

For an experiment in which we consider two variables X and Y, even if we know their joint probability function, we may still be interested in the probability function of just X and/or that of just Y. More generally, having studied both one- and two-dimensional discrete random variables, two questions that arise in relation to a given pair of variables X, Y are:

- If we know the joint probability function of X, Y, can we obtain the probability function for each of X and Y separately?
- If we know the probability function for each of X, Y separately, can we derive from them their joint probability function?

The answer to the second question is negative, while for the first question it is affirmative, as we explain below.

In Table 1.1, we have the joint probability function for the pair (X, Y) in Example 1.2.1. Can we deduce from Table 1.1, for example, the probability function of X alone? A look at the columns of Table 1.1 shows that, for the pair (X, Y), there are five possible outcomes, namely

$$(1,2), (2,2), (1,3), (2,3), (3,3),$$

and that the first four of these have a probability of $1/6$, while the event $\{(X, Y) = (3,3)\}$ has a probability of $1/3$. Suppose now we are interested in the probability of the event $\{X = 1\}$. Upon observing that this event is equivalent to the union of the events $\{X = 1, Y = 2\}$ and $\{X = 1, Y = 3\}$, them being disjoint, we have

$$\begin{aligned} f_X(1) = P(X = 1) &= P(\{X = 1, Y = 2\} \text{ or } \{X = 1, Y = 3\}) \\ &= P(X = 1, Y = 2) + P(X = 1, Y = 3) \\ &= f(1,2) + f(1,3) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}. \end{aligned}$$

The same can be done for the events $\{X = 2\}$ and $\{X = 3\}$, so that we obtain

$$P(X = i) = \frac{1}{3}, \quad \text{for } i = 1, 2, 3.$$

We now present the following general result.

Proposition 1.3.1 *Let (X, Y) be a jointly discrete random variable with joint probability function $f(x, y)$, with range $R_{X,Y}$. Then, the probability function for each of X and Y can be found as*

$$f_X(x) = P(X = x) = \sum_{y:(x,y)\in R_{X,Y}} f(x, y) \quad \textit{for any } x \in R_X,$$

and

$$f_Y(y) = P(Y = y) = \sum_{x:(x,y)\in R_{X,Y}} f(x, y) \quad \textit{for any } y \in R_Y.$$

Proof: Since (X, Y) have joint probability function $f(x, y)$, we have, for (x, y) in the range of (X, Y),

$$f(x, y) = P(X = x, Y = y).$$

Now, let $x \in R_X$ be fixed. For all values y in the range of Y, consider the events

$$\{X = x, Y = y\}.$$

Taking the union of these events over all $y \in R_Y$, we see that this is simply

$$\bigcup_{y\in R_Y} \{X = x, Y = y\} = \{X = x\}.$$

Further, for different values of $y \in R_Y$, the events on the left-hand side above are mutually exclusive, which implies that

$$P(X = x) = P\left(\bigcup_{y \in R_Y} \{X = x, Y = y\}\right) = \sum_{y \in R_Y} P(X = x, Y = y) = \sum_{y \in R_Y} f(x, y).$$

We thus see that by adding all probabilities $f(x, y)$ with respect to y, we arrive at the probability function of just X, i.e. when it is considered regardless of Y. In a similar way, we obtain

$$\bigcup_{x \in R_X} \{X = x, Y = y\} = \{Y = y\}$$

which implies, using again that the events on the left-hand side are mutually exclusive, that

$$P(Y = y) = P\left(\bigcup_{x \in R_X} \{X = x, Y = y\}\right) = \sum_{x \in R_X} P(X = x, Y = y) = \sum_{x \in R_X} f(x, y). \qquad \square$$

The key point to remember from the above discussion is that if we want the probability function of just X, then we sum over y, while if we want the probability function of just Y, we sum over all values of x.

The distribution of a random variable X alone, when it arises from the joint distribution of X and some other variable, is usually referred to as the **marginal distribution** of X. Similarly, the probability functions f_X and f_Y in Proposition 1.3.1 are called **marginal probability functions**.

Example 1.3.1 At a coffee shop in New York, customers may order coffee and hot chocolate in each of three sizes: Small (S), Medium (M), and Large (L). The percentages of customers, who order either coffee or hot chocolate, according to the size of the drink, is as given in the table. We assume, as usual, that these percentages (which are in fact relative frequencies) are based on a large amount of data so that each relative frequency can be viewed as a probability.

	Small (%)	Medium (%)	Large (%)
Coffee	14	23	19
Hot chocolate	11	20	13

Let X be the variable denoting the drink that a customer orders. For convenience, we assign numerical values to X, and set $X = 0$ if the customer orders coffee and $X = 1$ if the customer orders hot chocolate. Similarly, let Y be the size of the drink and we then set $Y = 0, 1$, and 2 according to the size being S, M, and L, respectively. Then, we see from the table that the joint probability function of X and Y is given by

$$f(0, 0) = 0.14, \quad f(0, 1) = 0.23, \quad f(0, 2) = 0.19,$$

$$f(1, 0) = 0.11, \quad f(1, 1) = 0.20, \quad f(1, 2) = 0.13.$$

It should be clear that the choice of values for X, Y here is arbitrary, and that we could have used, for example, $X = 5$ and $X = 10$ for the events “choice of coffee” and “choice of hot chocolate”.

Suppose now we are interested in the probability function of X alone. As X takes only two values, we simply want to find (based on the table of percentages) the percentage of customers who select coffee for their drink and the percentage of those who select hot chocolate. We then see that the former equals $14\% + 23\% + 19\% = 56\%$, i.e. a probability of 0.56 that a randomly selected customer orders coffee. The percentage of those who choose hot chocolate is simply the sum of the percentages in the second row of the table, i.e. $11\% + 20\% + 13\% = 44\%$, or a probability of 0.44. Thus, the marginal probability function of X is

$$f_X(0) = 0.56, \quad f_X(1) = 0.44.$$

Working in the same way for the variable Y, we see that we now have to add the values *in each column*, to arrive at

$$f_Y(0) = 0.25, \quad f_Y(1) = 0.43, \quad f_Y(2) = 0.32.$$

The use of the term *marginal probability function* for f_X and f_Y above comes from the fact the values of these functions are frequently inserted at the margins of the table along with the joint probability function, as they correspond to the row and column totals of that table, as shown below.

	Small (%)	Medium (%)	Large (%)	
Coffee	14	23	19	56%
Hot chocolate	11%	20%	13%	44%
	25%	43%	32%	100%

Example 1.3.2 Let us reconsider Example 1.2.3 regarding the number of persons waiting to be served at two cashier points of the shop. For the number of customers, X, waiting to be served by the first cashier, we find

$$f_X(0) = P(X = 0) = \sum_{y=0}^{2} f(0, y) = f(0, 0) + f(0, 1) + f(0, 2) = 0.16,$$
$$f_X(1) = P(X = 1) = \sum_{y=0}^{2} f(1, y) = f(1, 0) + f(1, 1) + f(1, 2) = 0.34,$$
$$f_X(2) = P(X = 2) = \sum_{y=0}^{2} f(2, y) = f(2, 0) + f(2, 1) + f(2, 2) = 0.50.$$

In an analogous manner, we find the (marginal) probability function of the number of customers, Y, awaiting service by the second cashier, to be

$$f_Y(0) = P(Y = 0) = \sum_{x=0}^{2} f(x, 0) = f(0, 0) + f(1, 0) + f(2, 0) = 0.24,$$
$$f_Y(1) = P(Y = 1) = \sum_{x=0}^{2} f(x, 1) = f(0, 1) + f(1, 1) + f(2, 1) = 0.38,$$
$$f_Y(2) = P(Y = 2) = \sum_{x=0}^{2} f(x, 2) = f(0, 2) + f(1, 2) + f(2, 2) = 0.38.$$

Inserting these two probability functions at the margins of Table 1.4, we simply obtain the table below. Note that the value (in bold) in the far right-hand cell of the last row in this table is 1 (or 100%, if expressed as a percentage, as in the previous example), and this is always true since this is the sum of the probabilities $f_X(x)$ (or $f_Y(y)$), while it is evident that

$$\sum_{x\in R_X} f_X(x) = \sum_{y\in R_Y} f_Y(y) = 1.$$

$x \backslash y$	0	1	2	f_X
0	0.10	0.04	0.02	0.16
1	0.08	0.20	0.06	0.34
2	0.06	0.14	0.30	0.50
f_Y	0.24	0.38	0.38	**1.00**

The (cumulative) distribution functions of the random variables X and Y alone when these arise from the joint distribution of the pair (X, Y) are called the **marginal distribution functions**, and are denoted by F_X and F_Y, respectively.

Note that, upon using

$$F_X(x) = P(X \leq x) = P(X \leq x \text{ and } (X, Y) \in R_{X,Y}),$$

we simply obtain

$$F_X(x) = \sum_{s\leq x} \left(\sum_{y:(s,y)\in R_{X,Y}} f(s, y) \right) = \sum_{s\leq x} f_X(s).$$

Similarly, for $F_Y(y)$, we have

$$F_Y(y) = \sum_{t\leq y} f_Y(t).$$

Observe that this is exactly the same relation between the probability function and the distribution function of a discrete variable in the univariate case, as seen in Volume I (see Chapter 4).

Further, we should mention that the marginal distribution functions can be found directly from the joint distribution function as

$$F_X(x) = \lim_{y\to\infty} F(x,y), \quad F_Y(y) = \lim_{x\to\infty} F(x,y). \tag{1.3}$$

To justify these expressions, note first that

$$F_X(x) = P(X \le x) = P\left(\lim_{n\to\infty} \{X \le x, Y \le n\}\right)$$

and by an appeal to the continuity property of probability, we get

$$F_X(x) = \lim_{n\to\infty} P(X \le x, Y \le n) = \lim_{n\to\infty} F(x,n) = \lim_{y\to\infty} F(x,y).$$

The proof of the second expression in (1.3) follows in an analogous manner.

Example 1.3.3 Let (X, Y) be a two-dimensional discrete random variable having the joint probability function as

$$f(x,y) = \frac{6}{3^x 4^y}, \quad x = 1, 2, \dots \text{ and } y = 1, 2, \dots$$

(i) Find the marginal probability functions of the variables $X$and Y.

(ii) Calculate the probability $P(X \ge 3 \text{ and } Y \ge 4)$.

SOLUTION

(i) For the marginal probability function of X, we obtain from Proposition 1.3.1 that

$$f_X(x) = \sum_{y=1}^{\infty} f(x,y) = \sum_{y=1}^{\infty} \frac{6}{3^x 4^y} = \frac{6}{3^x \cdot 4} \sum_{y=1}^{\infty} \left(\frac{1}{4}\right)^{y-1}$$
$$= \frac{6}{3^x \cdot 4} \cdot \frac{1}{1 - \frac{1}{4}} = \frac{2}{3^x}, \quad \text{for } x = 1, 2, \dots$$

Similarly, for the variable Y, we find

$$f_Y(y) = \sum_{x=1}^{\infty} f(x,y) = \sum_{x=1}^{\infty} \frac{6}{3^x 4^y} = \frac{6}{3 \cdot 4^y} \sum_{x=1}^{\infty} \left(\frac{1}{3}\right)^{x-1}$$
$$= \frac{2}{4^y} \cdot \frac{1}{1 - \frac{1}{3}} = \frac{3}{4^y}, \quad \text{for } y = 1, 2, \dots$$

(ii) Let us denote the events $\{X < 3\}$ and $\{Y < 4\}$ by A and B, respectively. Then, we see that

$$P(X \geq 3, Y \geq 4) = 1 - P(A \cup B) = 1 - P(A) - P(B) + P(AB),$$

which readily gives

$$\begin{aligned} P(X \geq 3, Y \geq 4) &= 1 - P(X \leq 2) - P(Y \leq 3) + P(X \leq 2, Y \leq 3) \\ &= 1 - F_X(2) - F_Y(3) + F(2,3), \end{aligned}$$

where $F_X(x)$ and $F_Y(y)$ are the marginal distribution functions of X and Y, and $F(x, y)$ is the joint distribution function of the pair (X, Y). Next, we find

$$F_X(2) = P(X \leq 2) = P(X = 1) + P(X = 2) = f_X(1) + f_X(2) = \frac{2}{3} + \frac{2}{3^2} = \frac{8}{9}$$

and

$$F_Y(3) = P(Y \leq 3) = f_Y(1) + f_Y(2) + f_Y(3) = \frac{3}{4} + \frac{3}{4^2} + \frac{3}{4^3} = \frac{63}{64},$$

using f_X and f_Y found in Part (i). In addition, we find

$$\begin{aligned} F(2,3) &= P(X \leq 2, Y \leq 3) \\ &= f(1,1) + f(1,2) + f(1,3) + f(2,1) + f(2,2) + f(2,3) \\ &= \frac{6}{3 \cdot 4} + \frac{6}{3 \cdot 4^2} + \frac{6}{3 \cdot 4^3} + \frac{6}{3^2 \cdot 4} + \frac{6}{3^2 \cdot 4^2} + \frac{6}{3^2 \cdot 4^2} = \frac{7}{8}, \end{aligned}$$

using the expression of $f(x, y)$ given in the statement. Thus, we arrive at

$$P(X \geq 3, Y \geq 4) = 1 - \frac{8}{9} - \frac{63}{64} + \frac{7}{8} = \frac{1}{576}.$$

EXERCISES

Group A

1. The joint probability function of a random vector (X, Y) is

$$f(x, y) = c(x^2 + y^2), \quad \text{for } x = 0, 1, 2 \text{ and } y = 1, 2, 3,$$

for some real constant c.

(i) Find the value of c.

(ii) Obtain the marginal probability functions of X and Y.

2. The joint probability function of a pair (X, Y) is

$$f(x, y) = \frac{x+y}{c}, \quad \text{for } x = 0, 1, 2, 3 \text{ and } y = 1, 2, 3,$$

for some $c \in \mathbb{R}$.

(i) What is the value of c?

(ii) Find the marginal probability functions of X and Y.

(iii) Calculate

$$P(X \geq 1),\ P(Y \leq 2),\ P(X \geq 1, Y \leq 2),\ P(X \geq 1 \text{ or } Y \leq 2).$$

3. Let X and Y be two random variables, each of which assumes only the values 0 and 1. It is known that their joint probability function, $f(x, y)$, satisfies

$$f(0,1) + f(1,1) = 2\,\{f(0,0) + f(1,0)\}\,.$$

(i) Obtain the marginal probability function of Y, and thence find $E(Y)$ and $Var(3Y - 2)$.

(ii) Is the information given above sufficient to determine the marginal distribution of X? Explain why or why not!

4. Suppose (X, Y) has the joint probability function as

$$f(x, y) = c\left(\frac{x}{y^2}\right), \quad x = 1, 2, 3 \text{ and } y = 1, 2.$$

After finding the value of the constant $c \in \mathbb{R}$, derive the marginal probability functions of X and Y, and thence find $E(X)$ and $E(Y)$.

5. For the pair (X, Y) of random variables in Example 1.3.3, having a joint probability function $f(x, y)$ as in the statement of the example, find a general expression for

$$P(X \geq x, Y \geq y), \quad \text{for } x = 1, 2, \ldots \quad \text{and } y = 1, 2, \ldots.$$

Then, verify that $P(X \geq 3, Y \geq 4)$ is as found in Part (ii) of Example 1.3.3.

6. The range of values for the two-dimensional random variable (X, Y) is the set

$$R_{X,Y} = \{(2,4), (3,4), (2,5), (3,5)\}.$$

It is known that the following hold for the joint probability function $f(x, y)$:

$$f(3,4) = 3 \cdot f(3,5),\ f(2,4) = f(2,5) + 2 \cdot f(3,5),\ f(3,4) = \frac{f(2,5) + f(3,5)}{2}.$$

(i) Obtain the values $f(x, y)$, for all $(x, y) \in R_{X,Y}$.

(ii) Find the marginal probability functions of X and Y, and thence find their expected values.

7. An airline offers its customers, for a certain trip, travel insurance which costs either US$20 or US$35 (depending on the coverage of the policy). Further, the airline offers car rental for one week, which costs US$75, US$100 or US$130, depending on the type of the car chosen. Let X denote the amount that a traveling passenger pays for car rental and Y be the amount that she pays for insurance (X and Y may also take a zero value if one does not opt for the corresponding services). The joint probability function of X and Y is as given in the following table:

x \ y	0	20	35
0	0.21	0.06	0.05
75	0.11	0.09	0.04
100	0.12	0.10	0.06
130	0.04	0.05	0.07

(i) Find the marginal probability functions and the marginal distribution functions of X and Y.

(ii) Find

$$P(X > 0), \quad P(X > 0 \text{ and } Y > 0), \quad P(X \geq 4Y).$$

(iii) What is the probability that a traveler pays at least

(a) US$80,

(b) US$100

in total for the two services combined (travel insurance and car rental)?

8. In the case of a single variable X, the events $\{X \leq a\}$ and $\{X > a\}$ are complementary for any real a; consequently, their probabilities add up to one. When we have two variables X and Y, the events

$$\{X \leq a, Y \leq b\} \text{ and } \{X > a, Y > b\}$$

are no longer complementary (explain why!).

(i) Show that

$$P(\{X > a, Y > b\}) = 1 - P(\{X \leq a\} \text{ or } \{Y \leq b\})$$

for any real a, b.

(ii) Let F be the joint distribution function of X and Y, and F_X and F_Y be the marginal distribution functions of X and Y, respectively. Use the result in Part (i) to establish that

$$P(\{X > a, Y > b\}) = 1 - F_X(a) - F_Y(b) + F(a, b).$$

(Note that the arguments needed to solve this exercise have already been used in Example 1.3.3, with the only difference being that the "$\geq$" signs in the example have been replaced by strict inequalities above.)

Group B

9. Nick tosses a coin three times. Let X be the number of heads that appear in the first toss of the coin and Y be the total number of heads in the three tosses.
   (i) What is the range, $R_{X,Y}$, of the random vector (X, Y)?
   (ii) Show that the joint probability function of (X, Y) is

$$f(x, y) = \frac{1}{8}\binom{2}{y-x}, \quad \text{for } (x, y) \in R_{X,Y}.$$

   (iii) Find the marginal probability functions of X and Y. What do you observe?

10. The joint probability function of a random pair (X, Y) is

$$f(x, y) = c, \quad \text{for } x = 1, 2, \dots, y \quad \text{and } y = 1, 2, \dots, n,$$

   where n is a given positive integer.
   (i) Show that

$$c = \frac{2}{n(n+1)}.$$

   (ii) Find the marginal probability functions of X and Y, and then find their expected values.

11. A pair (X, Y) of discrete random variables has joint probability function

$$f(x, y) = \binom{n}{x}\binom{n-x}{y} p^x q^y r^{n-x-y}, \quad (x, y) \in R_{X,Y},$$

   where n is a positive integer, $p, q, r \geq 0$ are such that $p + q + r = 1$, and

$$R_{X,Y} = \{(x, y) : x, y = 0, 1, \dots, n \text{ and } x + y \leq n\}.$$

   Establish that the marginal distributions of X and Y are binomial with parameters (n, p) and (n, q), respectively. (The bivariate distribution corresponding to $f(x, y)$ above is called a **trinomial distribution**, and will be studied in detail in Chapter 6.)

12. The joint probability function of the pair (X, Y) is given by

$$f(x, y) = e^{-\lambda}\frac{(\lambda p)^x}{x!} \cdot \frac{(\lambda q)^{y-x}}{(y-x)!}, \quad \text{for } x = 0, 1, 2, \dots, y \text{ and } y = 0, 1, 2, \dots,$$

   where $\lambda > 0$, and $p, q \in (0, 1)$ with $p + q = 1$. Show that the marginal distributions of the variables X and Y are Poisson with parameters λp and λ, respectively.

1.4 EXPECTATION OF A FUNCTION

In many situations when the joint distribution of two random variables X and Y is known, we may be interested in calculating probabilities associated with a function of these variables. For example, when X denotes the number of boys and Y denotes the number of girls in a randomly selected family from a population, we may be interested in

- the distribution, or the expected value, of the number of children in the family,
- the probability that there are more girls than boys in the family,

and so on. In the former case, we want to find the distribution, or the expectation of the random variable $X + Y$, while in the latter we want to calculate the probability $P(X - Y < 0)$.

In general, let (X, Y) be a two-dimensional random variable with joint probability function $f(x, y)$, with $(x, y) \in R_{X,Y}$, and let $h(x, y)$ be a real function of two variables. Then, $Z = h(X, Y)$ is a (one-dimensional) random variable. Finding the distribution of Z, from the joint distribution of (X, Y), may not be a straightforward task and we will take this up in Chapter 4 of the book. However, there is always an explicit expression for $E(Z)$ in terms of the joint probability function f.

To provide another motivating example, let us reconsider Example 1.2.3, wherein X and Y correspond to the numbers of customers awaiting service at the two cashier points. Then:

- the variable $Z = X + Y$ represents the total number of customers awaiting service,
- the variable $Z = \max\{X, Y\}$ corresponds to the number of waiting customers at the cashier point that has the largest queue,
- the variable $Z = \min\{X, Y\}$ corresponds to the number of waiting customers at the cashier point that has the shortest queue.

As $Z = h(X, Y)$ is a univariate discrete random variable, its expectation (provided it exists) can be found readily as

$$E[h(X, Y)] = E(Z) = \sum_{z \in R_Z} zP(Z = z) = \sum_{z \in R_Z} z f_Z(z),$$

where $f_Z(z) = P(Z = z)$ denotes the probability function of Z. However, this expression assumes knowledge of $f_Z(z)$, and so we usually employ the following result, which is a direct extension of the result corresponding to a function of a single random variable.

Proposition 1.4.1 *Let (X, Y) be a pair of discrete random variables with joint probability function $f(x, y)$, and $h(x, y)$ be a real function of two variables. Then, the expected value of the random variable $h(X, Y)$ is given by*

$$E[h(X, Y)] = \sum_{(x,y) \in R_{X,Y}} h(x, y) f(x, y),$$

assuming that the sum on the right-hand side converges absolutely (to a finite value).

Proof: The proof is omitted. □

Example 1.4.1 A pair of discrete random variables (X, Y) has range

$$R_{X,Y} = \{(1, 1), ((1, 2), (2, 1), (2, 2)\},$$

while, for the joint probability function $f(x, y)$, it is known that

$$f(2,2) = 2f(1,2),\ f(1,2) = 2f(2,1),\ f(2,1) = f(1,1).$$

(i) Calculate the values of the function $f(x, y)$, for all $(x, y) \in R_{X,Y}$, and find the marginal probability functions of X and Y.

(ii) Find, using two different ways, each of $E(X + Y)$ and $E(XY)$.

(iii) Examine which, if any, of the following hold:

$$E(X + Y) = E(X) + E(Y),\ E(XY) = E(X)E(Y).$$

SOLUTION

(i) Let us set, for convenience, $f(1, 1) = c$. Then, from the expressions given in the statement, we have

$$f(2,1) = c, \quad f(1,2) = 2c, \quad f(2,2) = 4c,$$

so that the condition

$$\sum_{x=1}^{2}\sum_{y=1}^{2} f(x,y) = 1$$

gives $8c = 1$, which means $c = 1/8$. The joint probability function and the marginal probability functions of X and Y are as given in the following table:

$x \backslash y$	1	2	f_X
1	$1/8$	$2/8$	$3/8$
2	$1/8$	$4/8$	$5/8$
f_Y	$2/8$	$6/8$	1

Table 1.5 The joint probability function of (X, Y) in Example 1.4.1

(ii) From Proposition 1.4.1, we have

$$\begin{aligned} E(X+Y) &= \sum_{(x,y)\in R_{X,Y}} (x+y)f(x,y) \\ &= (1+1)f(1,1) + (1+2)f(1,2) + (2+1)f(2,1) + (2+2)f(2,2) \\ &= 2 \cdot \frac{1}{8} + 3 \cdot \frac{2}{8} + 3 \cdot \frac{1}{8} + 4 \cdot \frac{4}{8} = \frac{27}{8}. \end{aligned}$$

Similarly, we have

$$E(XY) = \sum_{(x,y)\in R_{X,Y}} xyf(x,y)$$

$$= (1 \cdot 1)f(1,1) + (1 \cdot 2)f(1,2) + (2 \cdot 1)f(2,1) + (2 \cdot 2)f(2,2)$$
$$= 1 \cdot \frac{1}{8} + 2 \cdot \frac{2}{8} + 2 \cdot \frac{1}{8} + 4 \cdot \frac{4}{8} = \frac{23}{8}.$$

Another way to find the expectations above is to determine the probability function for each of the random variables $Z = X + Y$ and $W = XY$, and then use them to find the corresponding means using the formula for a single variable. For this purpose, we see that the ranges of values of Z and W are $R_Z = \{2, 3, 4\}$ and $R_W = \{1, 2, 4\}$, respectively. Concerning the probability functions of Z and W, we find from Table 1.5, first for variable Z, that

$$f_Z(2) = P(Z = 2) = P(X + Y = 2) = P(X = 1, Y = 1) = \frac{1}{8},$$
$$f_Z(3) = P(Z = 3) = P(X + Y = 3) = P(X = 1, Y = 2) + P(X = 2, Y = 1) = \frac{3}{8},$$
$$f_Z(4) = P(Z = 4) = P(X + Y = 4) = P(X = 2, Y = 2) = \frac{4}{8}.$$

Similarly, for the variable W we obtain

$$f_W(1) = P(W = 1) = P(XY = 1) = P(X = 1, Y = 1) = \frac{1}{8},$$
$$f_W(2) = P(W = 2) = P(XY = 2) = P(X = 1, Y = 2) + P(X = 2, Y = 1) = \frac{3}{8},$$
$$f_W(4) = P(W = 4) = P(XY = 4) = P(X = 2, Y = 2) = \frac{4}{8}.$$

We then find, using the formula for the expectation of a discrete univariate random variable, that

$$E(Z) = \sum_{z \in R_Z} z f_Z(z) = 2 \cdot \frac{1}{8} + 3 \cdot \frac{3}{8} + 4 \cdot \frac{4}{8} = \frac{27}{8}$$

and

$$E(W) = \sum_{w \in R_W} w f_W(w) = 1 \cdot \frac{1}{8} + 2 \cdot \frac{3}{8} + 4 \cdot \frac{4}{8} = \frac{23}{8}.$$

Clearly, these agree with the values obtained earlier.

(iii) Because the row and column totals of Table 1.5 give the probability functions of X and Y, it is easy to find their expectations. In particular, we have

$$E(X) = \sum_{x \in R_X} x f_X(x) = 1 \cdot f_X(1) + 2 \cdot f_X(2) = 1 \cdot \frac{3}{8} + 2 \cdot \frac{5}{8} = \frac{13}{8},$$
$$E(Y) = \sum_{y \in R_Y} y f_Y(y) = 1 \cdot f_Y(1) + 2 \cdot f_Y(2) = 1 \cdot \frac{2}{8} + 2 \cdot \frac{6}{8} = \frac{14}{8}.$$

It is now apparent that the equality

$$E(X+Y) = E(X) + E(Y)$$

holds; but, we do find

$$E(XY) = \frac{23}{8} \neq \frac{13}{8} \cdot \frac{14}{8} = E(X) \cdot E(Y).$$

Proposition 1.4.2 *Let $h_1(X,Y)$ and $h_2(X,Y)$ be two functions of the two-dimensional random variable (X,Y). Then, we have*

$$E\left[h_1(X,Y) + h_2(X,Y)\right] = E\left[h_1(X,Y)\right] + E\left[h_2(X,Y)\right]$$

provided the expectations on the right-hand side exist.

Proof: Let $f(x,y)$, for $(x,y) \in R_{X,Y}$, be the joint probability function of (X,Y). Let us introduce the function

$$h(x,y) = h_1(x,y) + h_2(x,y).$$

In view of Proposition 1.4.1, we can then write

$$\begin{aligned}
E\left[h_1(X,Y) + h_2(X,Y)\right] &= E\left[h(X,Y)\right] \\
&= \sum_{(x,y)\in R_{X,Y}} h(x,y)f(x,y) \\
&= \sum_{(x,y)\in R_{X,Y}} [h_1(x,y) + h_2(x,y)]f(x,y) \\
&= \sum_{(x,y)\in R_{X,Y}} [h_1(x,y)f(x,y) + h_2(x,y)f(x,y)] \\
&= \sum_{(x,y)\in R_{X,Y}} h_1(x,y)f(x,y) + \sum_{(x,y)\in R_{X,Y}} h_2(x,y)f(x,y) \\
&= E\left[h_1(X,Y)\right] + E\left[h_2(X,Y)\right],
\end{aligned}$$

which yields the desired result. □

With reference to the question in Part (iii) of Example 1.4.1, the last proposition shows that $E(X+Y) = E(X) + E(Y)$ is always true (just put $h_1(X,Y) = X, h_2(X,Y) = Y$ in the proposition), meaning that

the expected value of the sum of two variables is equal to the sum of their expected values.

In contrast, and as we demonstrated in Example 1.4.1, this is not true when we consider products instead of sums for random variables.

In addition, by applying Proposition 1.4.2 to the special case when $h_1(X, Y) = aX$ and $h_2(X, Y) = bY$, with $a, b \in \mathbb{R}$, we readily obtain the following corollary.

Corollary 1.4.1 (*Expectation for a linear combination of random variables*) *For any random variables X and Y and any real numbers a and b, we have*

$$E(aX + bY) = aE(X) + bE(Y).$$

EXERCISES

Group A

1. Assume that (X, Y) has a joint probability function

$$f(x, y) = c\ 2^{x-y+2}, \quad x = 1, 2 \text{ and } y = 1, 2.$$

   (i) Find the value of c and then obtain the marginal probability functions of X and Y.

   (ii) Calculate $E(X)$, $E(Y)$ and $E(X/Y)$, and then verify that $E(X/Y) \neq E(X)/E(Y)$.

2. An examination consists of two parts, each of which has three multiple-choice questions. Each question carries 10 points. Let X be the number of points that a student gets in the first part of the exam and Y be the number of points that the same student receives in the second part of the exam.

   Suppose the joint probability function of (X, Y) is given as follows:

$$P(X = x, Y = y)$$

x \ y	0	10	20	30
0	0.02	0.06	0.02	0.10
10	0.02	0.15	0.10	0.06
20	0.01	0.10	0.14	0.01
30	0.02	0.05	0.10	0.04

   (i) What is the average total mark achieved by a student in this exam (i.e. when we add the points from both parts of the exam)?

   (ii) Find the average mark achieved by a student in the exam, if the first part has a weight of 60% and the second part has a weight of 40%.

   (iii) Suppose a student's final mark is formed by the points the student receives for only one of the two parts of the exam, which is the one in which the student performed the best. What is the expected value of the student's final mark in this case?

3. In a small ship that carries vehicles, the cost per trip for a regular car is US\$20, while the cost for a truck is US\$75. Let X be the number of passenger cars that the ship carries during a trip and Y be the number of trucks on the same trip. The joint probability function of X and Y is as follows:

$$P(X = x, Y = y)$$

x \ y	0	1	2	3
0	0.000	0.005	0.015	0.025
1	0.010	0.020	0.030	0.040
2	0.020	0.030	0.050	0.055
3	0.040	0.050	0.065	0.075
4	0.040	0.035	0.045	0.045
5	0.030	0.040	0.045	0.050
6	0.035	0.035	0.040	0.030

Calculate the expected amount of money that the company owning the ship would make during a trip.

4. Calculate the expectations $E(XY), E(X + Y), E(X^2 + 3Y^2 - XY)$ for each of the joint probability functions for (X, Y) below:

(i) $f(x, y) = \dfrac{x + y}{21}, x = 1, 2$ and $y = 1, 2, 3$;

(ii) $f(x, y) = \dfrac{1}{6}, x = 1, 2, \ldots, y$ and $y = 1, 2, 3$;

(iii) $f(x, y) = \dfrac{1}{9}, x = 2, 3, 4$ and $y = -1, 1, 2$;

(iv) $f(x, y) = \dfrac{x^2 + y^2}{32}, (x, y) \in \{(1, 1), (1, 4), (2, 3)\}$.

5. Let X be a random variable which takes the values $-1, 0$, and 1, each with probability $1/3$, i.e.

$$P(X = -1) = P(X = 0) = P(X = 1) = \frac{1}{3},$$

and Y be another variable defined as

$$Y = \begin{cases} 0, & \text{if } X \neq 0, \\ 1, & \text{if } X = 0. \end{cases}$$

Check that $E(XY) = E(X)E(Y)$.

6. Let X and Y be two discrete random variables with joint probability function

$$f(x, y) = \begin{cases} \dfrac{1}{4}, & \text{if } (x, y) \in \{(1, 2), (-1, 2), (1, 3), (-1, 3)\}, \\ 0, & \text{otherwise.} \end{cases}$$

Show that, for these variables, $E(XY) = E(X)E(Y)$.

7. The joint probability function of X and Y is given by

$$f(0,0) = f(1,0) = f(0,1) = \frac{1}{3}.$$

   (i) Find the marginal probability functions of X and Y.

   (ii) Calculate

$$E(X),\ E(Y),\ E(X+Y),\ E(XY),\ E\left(\frac{X+1}{Y+1}\right),\ E\left(\frac{Y+1}{X+1}\right).$$

8. Suppose, for the random variables X and Y, it is known that $E(X) = 1, E(Y) = 3$, $E(XY) = 2$, $Var(X) = 2$ and $Var(Y) = 4$. Obtain the expectation for each of the following random variables:

$$X^2 + Y^2,\ (X - Y)^2,\ 4X^2 - 2Y^2,\ (X+Y)(X-Y),\ (X+Y-2)(X-2Y+1).$$

Group B

9. Nick tosses two dice. Let X be the outcome of the first die and Y be the outcome of the second one.

   (i) Find the distribution for each of

$$Z = \min\{X, Y\} \text{ and } W = \max\{X, Y\}.$$

   (ii) Obtain $E(Z)$ and $E(W)$ using two different methods.

   (iii) Verify that

$$E(Z) + E(W) = E(X) + E(Y).$$

   Can you interpret the last result intuitively?

10. (Continuation of the previous exercise) Find the joint probability function of Z and W. Thence, calculate $E(ZW)$. Is it true in this case that $E(ZW) = E(Z)E(W)$?

11. The joint probability function of X and Y is given by

$$f(x, y) = \frac{3}{2^{x+1}4^{y+1}}$$

   for $x = 0, 1, 2, \ldots$ and $y = 0, 1, 2, \ldots$.

   (i) Find the marginal probability functions of X and Y, and then calculate their expected values.

   (ii) Obtain the expected value of the product XY. What do you observe?

12. Six persons sit at random around a circular table. When person A wants to pass a written message to person B, the transfer of the message is done by the direction in which A and B are closer to one another, and let Z be the number of persons sitting between A and B in that direction. Then, find $E(Z)$.

   [Hint: Let X and Y be the places of the two persons, and then $Z = h(X, Y)$. Record in a two-way table (with respect to X and Y) the values that Z can take on, and then calculate the quantity $E(Z) = E[h(X, Y)]$.]

1.5 CONDITIONAL DISTRIBUTIONS AND EXPECTATIONS

When we consider probabilities of events in a sample space, we have seen in Volume I that the concept of conditional probability arises naturally. There are many occasions wherein we want to find the probability that an event A occurs *knowing* that another event, B, has already occurred. In this case, the conditional probability of A, given B has occurred, is given by

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

provided $P(B) > 0$.

Suppose now we have two random variables, X and Y, with joint probability function $f(x, y)$. If we know that Y has taken the value $Y = y$, we may then be interested in calculating the conditional probabilities

$$P(X = x|Y = y)$$

for any value of X in its range. This is provided in the following definition.

Definition 1.5.1 Let (X, Y) be a pair of random variables with joint probability function $f(x, y)$ and marginal probability functions $f_X(x)$ and $f_Y(y)$. Then, the **conditional probability function** of X, given $Y = y$ (here, y is a fixed value in the range of Y such that $P(Y = y) > 0$), is denoted by $f_{X|Y}(x|y)$ and is given by

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

where x takes any value such that $(x, y) \in R_{X,Y}$. Similarly, the conditional probability function of Y, given $X = x$ (here, x is in the range of X such that $P(X = x) > 0$), is given by

$$f_{Y|X}(y|x) = P(Y = y|X = x) = \frac{P(X = x, Y = y)}{P(X = x)} = \frac{f(x, y)}{f_X(x)},$$

where y takes any value such that $(x, y) \in R_{X,Y}$.

It is easy to check that the function $f_{X|Y}(x|y)$ is a (discrete, univariate) probability function. In fact, it is obvious that it takes only nonnegative values. Moreover, we recall from the definition of a marginal probability function that

$$f_Y(y) = \sum_{x:(x,y)\in R_{X,Y}} f(x, y),$$

which immediately yields

$$\sum_{x:(x,y)\in R_{X,Y}} f_{X|Y}(x|y) = \sum_{x:(x,y)\in R_{X,Y}} \frac{f(x, y)}{f_Y(y)} = 1$$

showing that $f_{X|Y}(x|y)$ is a valid probability function. It is important to keep in mind that we view this as a function of x alone, while the value of y is kept fixed (and is often assumed known, like $y = 3$). In a similar fashion, we can show that for a fixed $x \in R_X$, the function $f_{Y|X}(y|x)$ is a valid probability function.

Example 1.5.1 Let us consider again Table 1.4 in Example 1.2.3, concerning the numbers of customers waiting to be served by two cashiers in a shop. The joint probability function of the variables X and Y, representing the customers awaiting service by the first and second cashiers, respectively, is given below (this is similar to Table 1.4, but also contains the marginal probability functions of X and Y):

x \ y	0	1	2	$f_X(x)$
0	0.10	0.04	0.02	0.16
1	0.08	0.20	0.06	0.34
2	0.06	0.14	0.30	0.50
$f_Y(y)$	0.24	0.38	0.38	1.00

Suppose now we know that $Y = 0$ (no customer is waiting to be served by the second cashier). Then, according to Definition 1.5.1, the conditional probability function of X, given $y = 0$, is given by

$$f_{X|Y}(x|0) = \frac{f(x,0)}{f_Y(0)} = \begin{cases} \dfrac{0.10}{0.24} = 0.42, & x = 0, \\ \dfrac{0.08}{0.24} = 0.33, & x = 1, \\ \dfrac{0.06}{0.24} = 0.25, & x = 2, \end{cases}$$

which can be written concisely as in the following table:

x	$f_{X|Y}(x|0)$
0	0.42
1	0.33
2	0.25

This has the following interpretation. If we know that there is no one waiting to be served by the second cashier, then the probability that the queue at the first cashier is also empty has probability 42%, while there is a probability of 33% that exactly one person is waiting to be served, and a probability of 25% that exactly two persons are waiting for service.

In a similar way, we find the conditional probability function of Y, given $X = 2$, as

$$f_{Y|X}(y|2) = \frac{f(2,y)}{f_X(2)} = \begin{cases} \dfrac{0.06}{0.50} = 0.12, & y = 0, \\ \dfrac{0.14}{0.50} = 0.28, & y = 1, \\ \dfrac{0.30}{0.50} = 0.60, & y = 2, \end{cases}$$

which can be written concisely as in the following table:

y	$f_{Y\|X}(y\|2)$
0	0.12
1	0.28
2	0.60

We have found above the conditional probability function of X, given $Y = 0$. It should be noted that, if we instead know that $Y = 1$, then the conditional probability function of X is

$$f_{X|Y}(x|1) = \frac{f(x,1)}{f_Y(1)} = \begin{cases} \dfrac{0.04}{0.38} = 0.105, & x = 0, \\ \dfrac{0.20}{0.38} = 0.526, & x = 1, \\ \dfrac{0.14}{0.38} = 0.369, & x = 2. \end{cases}$$

We thus observe that the conditional probability function of X is different, depending on whether $Y = 0$ or $Y = 1$. Intuitively, this means that the number of customers that wait to be served by the first cashier depends on the number of persons waiting for service by the second cashier. A formal detailed discussion of the concept of *independence* between two random variables is made in Chapter 3.

Once we have defined conditional probability functions, we may proceed to consider conditional distribution functions. Since the function

$$g(y) = f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)} \tag{1.4}$$

is, for a given $x \in R_X$, a probability function, it corresponds to a (cumulative) distribution function

$$G(y) = \sum_{t \le y} g(t).$$

This is called the **conditional (cumulative) distribution function** of Y, given $X = x$, and is denoted by $F_{Y|X}(y|x)$. Similarly,

$$\sum_{y:(x,y)\in R_{X,Y}} y\, g(y)$$

defines the expectation corresponding to the probability function in (1.4), and is called the **conditional expectation** of Y, given $X = x$. We denote it by $E(Y|X = x)$. Putting together all the notation above, we have the following:

$$F_{Y|X}(y|x) = P(Y \le y|X = x) = \sum_{t \le y} f_{Y|X}(t|x)$$

and

$$E(Y|X = x) = \sum_{y:(x,y)\in R_{X,Y}} y f_{Y|X}(y|x). \tag{1.5}$$

Of course, we can define analogously the cumulative distribution function of X, given $Y = y$ (written as $F_{X|Y}(x|y)$) and the expectation of X, given $Y = y$, denoted by $E(X|Y = y)$.

Finally, as with the usual (i.e. unconditional) expectation, the conditional expectation of a function $h(Y)$, given $X = x$, is simply given by

$$E[h(Y)|X = x] = \sum_{y:(x,y)\in R_{X,Y}} h(y) f_{Y|X}(y|x).$$

Example 1.5.2 In Example 1.5.1 we found, based on the joint probability function in Table 1.4 (people waiting to be served at two cashier points), the conditional distribution of Y, given $X = 2$, as

y	$f_{Y\|X}(y\|2)$
0	0.12
1	0.28
2	0.60
	1.00

In contrast with formula (1.5), which looks rather complicated, when a conditional probability function is available (as in the table above), it is very simple to work out the corresponding conditional expectation. We simply multiply each value of y in the table by the associated probability (in the same row of the table) and then sum up the values. Thus, we have

$$E(Y|X = 2) = 0 \cdot (0.12) + 1 \cdot (0.28) + 2 \cdot (0.60) = 1.48.$$

It may be noted that (in analogy with the unconditional case), even though the variable Y is integer-valued, its conditional expectation need not be so. Here we have found that, if we know that there are 2 persons awaiting service by the first cashier, the expected number of persons waiting to be served by the second cashier is 1.48.

In Example 1.5.1, we also found the conditional probability functions of X, given $Y = 0$ and $Y = 1$. From these functions, we find immediately

$$E(X|Y = 0) = 0 \cdot (0.42) + 1 \cdot (0.33) + 2 \cdot (0.25) = 0.83,$$

$$E(X|Y = 1) = 0 \cdot (0.10) + 1 \cdot (0.53) + 2 \cdot (0.37) = 1.27.$$

We observe that these two expectations are unequal, that is,

$$E(X|Y = 0) \neq E(X|Y = 1).$$

In the examples we have considered so far, we have assumed that the range of values for a random pair (X, Y) is known in advance, that is, before the random experiment is carried out. But, in several instances, this may not be the case, in particular when the range of values of one variable depends on the value that the other variable takes. The following example illustrates this issue, as well as the use of conditional expectations for making decisions.

Example 1.5.3 Suppose Karen, who is participating in a TV quiz show, has reached the final stage of the show. At this stage, she has the option to choose whether she wants to answer one, two, or three questions. If she answers all questions correctly, she wins

- US$500 if she has chosen to answer one question,
- US$1000 if she has chosen to answer two questions, or
- US$2000 if she has chosen to answer three questions.

If she gives at least one wrong answer (or does not give an answer to a question at all), she receives nothing.

The probability that Karen answers a given question correctly is 0.80. What is the best strategy for Karen to maximize her expected profit from the show?

SOLUTION Let us set X to take the values 1, 2, and 3 if Karen decides to choose one, two, and three questions, respectively. As we are interested in the amount of money she receives from the show, let us also define a variable Y representing her earnings. The ranges of values for X and Y are

$$R_X = \{1, 2, 3\}, \quad R_Y = \{0, 500, 1000, 2000\};$$

but, it is clear that for (X, Y), some combinations are not possible (for example, the pairs of values $(1, 1000)$ and (3,500) cannot occur).

The best strategy for Karen is the one that maximizes

$$E(Y|X = i)$$

for $i = 1, 2, 3$. So, we need to calculate all three conditional expectations.

For $i = 1$, the probability function of Y is given by

$$P(Y = 500|X = 1) = 0.80, \quad P(Y = 0|X = 1) = 0.20,$$

and so

$$\begin{aligned} E(Y|X = 1) &= 0 \cdot P(Y = 0|X = 1) + 500 \cdot P(Y = 500|X = 1) \\ &= 0 + 500 \cdot (0.80) = 400. \end{aligned}$$

For $i = 2$, the values that Y can take are $Y = 0$ (if Karen does not answer both questions correctly) and $Y = 1000$. The probability that she wins US$1000 in this case is $(0.8)^2$ (assuming implicitly that the events corresponding to her answering correctly the two questions are independent, which seems to be a reasonable assumption here). Consequently,

$$\begin{aligned} E(Y|X = 2) &= 0 \cdot P(Y = 0|X = 2) + 1000 \cdot P(Y = 1000|X = 2) \\ &= 0 + 1000 \cdot (0.80)^2 = 640. \end{aligned}$$

Finally, for $i = 3$, the possibilities are $Y = 2000$ or $Y = 0$ depending on whether Karen answers all three questions correctly or not. The former event has probability $(0.80)^3$, assuming again that the events corresponding to her answering correctly successive questions are independent. This gives

$$\begin{aligned} E(Y|X = 3) &= 0 \cdot P(Y = 0|X = 3) + 2000 \cdot P(Y = 2000|X = 3) \\ &= 0 + 2000 \cdot (0.80)^3 = 1024. \end{aligned}$$

We thus see that the best strategy, to maximize her expected profit, is to choose three questions to answer.

An important point to note here is that, given $X = x$, the *conditional distribution* of the variable W, denoting the number of correct answers that Karen gives, is binomial with parameters $n = x$ and $p = 0.80$. See Exercises 8 and 11 at the end of this section for variants of the present problem, where this remark could prove useful!

EXERCISES

Group A

1. The basketball teams of two Universities, A and B, have reached the Final Four of the NCAA tournament. The two teams are drawn to play in different semifinals, and so if they both win their first games, they will play against each other in the final. Suppose each team has a 60% chance of winning their semifinal game, and further suppose if one or both teams reach the final, there is an even chance for the finalists to win the trophy.

   Let X and Y be the number of wins that the teams from Universities A and B, respectively, achieve in the Final Four.

   (i) Write down the range of values for the pair (X, Y) and its joint probability function.

   (ii) Derive the marginal probability functions of X and Y.

   (iii) Obtain the conditional probability function and the conditional expectation of X, given that the second team does not qualify for the final.

2. The joint probability function of (X, Y) is

   $$f(x, y) = c(x + y), \quad x = 1, 2, 3 \text{ and } y = 5, 10.$$

   Find the value of c, and then calculate the conditional expectations $E(X|Y = 5)$ and $E(X|Y = 10)$. Are these two expectations equal?

3. A box contains two balls marked “0” and two balls marked “1”. Suppose we select two balls from the box without replacement, and let X and Y be the numbers on the first and second balls drawn, respectively.

   (i) Give the probability function of X.

(ii) Calculate the conditional probability functions

$$f_{Y|X}(y|0) \quad \text{and}\, f_{Y|X}(y|1),\ y = 0, 1,$$

and write down the joint probability function of X and Y in the form of a two-way table.

(iii) Find the conditional probability functions of X, given $Y = 0$ and $Y = 1$. Thence, calculate $E(X|Y = 0)$ and $E(X|Y = 1)$ and verify that

$$E(X|Y = 0) = 2E(X|Y = 1).$$

4. Suppose four paintings by a famous painter are on sale at a forthcoming auction. The company organizing the auction is particularly interested in the number of paintings that will be sold at a price which is at least three times as much as the asking price for a painting. Let X be the total number of paintings that will be sold at the auction and Y be the number of those that will fetch at least three times the asking price. Suppose the joint probability function of X and Y is as given in the following table (the organizers are certain that at least one painting will be sold, and so the event $\{X = 0\}$ has probability zero):

x \ y	0	1	2	3	4
1	0.03	0.02	0	0	0
2	0.05	0.05	0	0	0
3	0.10	0.15	0.17	0.10	0
4	0.07	0.08	0.10	0.05	0.03

(i) If exactly one painting is sold at the auction, what is the probability that it is sold at a price at least three times the asking price?

(ii) Let W be the number of paintings that will be sold, but will not fetch at least three times the asking price. Then, find the expected value of W.

(iii) Obtain the conditional probability function of Y, given $X = 3$, and thence find $E(Y|X = 3)$.

5. From an urn containing n tickets numbered $1, 2, \ldots, n$, suppose we select a ticket at random, and denote by X the number on that ticket. If $X = x$, we select a number Y randomly from the set $\{1, 2, \ldots, x\}$.

(i) Write down the probability function of X and the conditional probability function of Y, given $X = x$.

(ii) Obtain the joint probability function of (X, Y).

(iii) Give an expression, in terms of a sum, for the marginal distribution function of Y.

6. Consider the joint probability function of (X, Y) (see Example 1.3.3) given by

$$f(x, y) = \frac{6}{3^x 4^y}, \quad x = 1, 2, \ldots \text{ and } y = 1, 2, \ldots$$

In Example 1.3.3, we found the marginal probability functions f_X and f_Y. Using these:

(i) Calculate $E(X)$, $E(Y)$ and $E(3X - 2Y)$.

(ii) For a given $y \in \{1, 2, \dots\}$, obtain the conditional probability function $f_{X|Y}(x|y)$. Does this depend on the value of y chosen?

(iii) For a given $x \in \{1, 2, \dots\}$, obtain the conditional probability function $f_{Y|X}(y|x)$. Does this depend on the value of x chosen?

(iv) Obtain $E(X|Y = y)$ and $E(Y|X = x)$. What do you observe?

(Hint: You can answer most parts of this question without any calculations, by simply identifying the marginal distributions of X and Y to a well-known univariate distribution. The same also applies to the conditional probability functions you will find in Parts (ii) and (iii), hence all expectations can be found immediately by appealing to known properties of that distribution.)

7. Suppose the joint probability function of (X, Y) is as given in the following table:

$x \backslash y$	2	3	4
1	0.06		0.10
2	0.04	0.07	
3		0.15	0.25

(i) Fill the missing entries in this table if it is known that

$$f_{X|Y}(2|2) = 0.2 \quad \text{and} \quad f_{Y|X}(2|1) = 0.2.$$

(ii) Obtain $E(Y|X = x)$ for $x = 1, 2, 3$.

8. Consider the following modification of Example 1.5.3:

- If Karen decides to answer one question, she receives US\$500 if she answers correctly, otherwise she receives nothing.
- If she decides to answer two questions, she receives US\$300 for each correct answer she gives (that is, she gets US\$0, US\$300 or US600 when she gives 0, 1 and 2 correct answers, respectively).
- If she decides to answer three questions, she receives US\$200 for each correct answer she gives.

Find the number of questions that Karen should choose under this scenario in order to maximize her expected earnings from the show.

9. The joint probability function of (X, Y) is given by

$x \backslash y$	1	2
5	c	$2c$
10	$3c$	$1/3$

(i) After finding the value of c, derive the marginal probability functions of X and Y.

(ii) Calculate the conditional probabilities $f_{X|Y}(5|y)$ for $y = 1, 2$.

(iii) Obtain $E(X|Y = 1)$, $E(X|Y = 2)$.

10. The joint probability function of (X, Y) is given by

$x \backslash y$	1	2
0	a	b
1	$3a$	$1/8$
2	$2b$	$1/3$

for some suitable constants a and b.

Find the values of a and b if it is known that

$$E(X|Y = 1) = E(Y|X = 2).$$

Group B

11. Consider Example 1.5.3 again, with X denoting the number of questions that Karen chooses to answer and W denoting the number of correct answers she gives. Obtain the joint probability function of (X, W), and thence find the marginal probability function of W.

12. The number of fish caught by a fisherman in an hour is a random variable X taking values $0, 1, 2, 3, 4$ with probabilities 0.1, 0.2, 0.3, 0.25, 0.15, respectively. A percentage of 20% among the fish caught have a weight exceeding 5 lb.

Let Y be the number of fish whose weight exceeds 5 lb caught by the fisherman.

(i) Verify that the conditional distribution of Y, given $X = 4$, is the binomial distribution with parameters $(n = 4, p = 0.2)$. Thence, calculate $P(X = 4, Y = 3)$.

(ii) Using arguments similar to those in Part (i), obtain the joint probability function of (X, Y).

(iii) Derive the marginal probability function of Y.

13. The random variables X and Y have joint probability function as

$$f(x, y) = \frac{1}{e^2 y!(x - y)!}, \quad x = 0, 1, 2, \dots, \quad y = 0, 1, 2, \dots, x.$$

(i) Find the marginal probability function of X and the conditional probability function of Y, given $X = x$, for $x = 0, 1, 2, \dots$;

(ii) Thence, obtain

$$E(Y|X = x), \; E(Y^2|X = x), \quad x = 0, 1, 2, \dots .$$

1.6 BASIC CONCEPTS AND FORMULAS

Joint probability function for a pair of discrete random variables	$f(x,y) = P(X = x, Y = y)$ $= P(\{\omega \in \Omega : X(\omega) = x \text{ and } Y(\omega) = y\})$
Properties of the joint probability function	• $f(x,y) = 0$ if $(x,y) \notin R_{X,Y}$; • $f(x,y) \geq 0$ for any $(x,y) \in R_{X,Y}$; • $\sum_{(x,y)\in R_{X,Y}} f(x,y) = 1$
Joint (cumulative) distribution function	$F(x,y) = \sum_{s\leq x, t\leq y} f(s,t)$
Calculation of probabilities using joint probability function	$P((X,Y) \in A) = \sum_{(x,y)\in A} f(x,y)$
Marginal probability functions	$f_X(x) = \sum_{y:(x,y)\in R_{X,Y}} f(x,y)$, $f_Y(y) = \sum_{x:(x,y)\in R_{X,Y}} f(x,y)$
Expectation of function of two random variables	$E[h(X,Y)] = \sum_{(x,y)\in R_{X,Y}} h(x,y)f(x,y)$
Expectation of a sum and linear combination	$E\left[h_1(X,Y) + h_2(X,Y)\right]$ $= E\left[h_1(X,Y)\right] + E\left[h_2(X,Y)\right]$ $E(aX + bY) = aE(X) + bE(Y)$
Conditional probability function of X, given $Y = y$	$f_{X\|Y}(x\|y) = \dfrac{f(x,y)}{f_Y(y)}$, for $f_Y(y) > 0$
Conditional probability function of Y, given $X = x$	$f_{Y\|X}(y\|x) = \dfrac{f(x,y)}{f_X(x)}$, for $f_X(x) > 0$
Conditional (cumulative) distribution functions	$F_{X\|Y}(x\|y) = P(X \leq x\|Y = y) = \sum_{t\leq x} f_{X\|Y}(t\|y)$, $F_{Y\|X}(y\|x) = P(Y \leq y\|X = x) = \sum_{s\leq y} f_{Y\|X}(s\|x)$
Conditional expectations	$E(X\|Y = y) = \sum_{x:(x,y)\in R_{X,Y}} x f_{X\|Y}(x\|y)$; $E(h(X)\|Y = y) = \sum_{x:(x,y)\in R_{X,Y}} h(x) f_{X\|Y}(x\|y)$; $E(Y\|X = x) = \sum_{y:(x,y)\in R_{X,Y}} y f_{Y\|X}(y\|x)$; $E(h(Y)\|X = x) = \sum_{y:(x,y)\in R_{X,Y}} h(y) f_{Y\|X}(y\|x)$

1.7 COMPUTATIONAL EXERCISES

1. Let us assume we have a random vector (X, Y) with joint probability function

$$f(x, y) = c(x^2 + y^2), \quad x = -5, -3, -1, 1, 3, 5, 7, \quad y = -1, 0, 1, 2, 3,$$

for a suitable real constant c.

The following set of commands in Mathematica enables us to find the value of c, to write the joint probability function in a tabular form, and then to calculate the expectation of $U = h(X, Y)$, where $h(x, y) = 3x^2y^3$.

```
In[1]:= s1:= Sum[x^2 + y^2, {x, −5, 7, 2}, {y, −1, 3}];
c = 1/s1;
f[x_, y_]:= c*(x^2 + y^2);
t = Table[f[x, y], {x, −5, 7, 2}, {y, −1, 3}];
Print["c= ", c];
Print[MatrixForm[N[t]]]; h[x_, y_]:= 3 x^2 y^3;
m:= 0; m2:= 0;
Do[m = m + f[x, y]*h[x, y]; m2 = m2 + f[x, y]*h[x, y]^2,
{x, −5, 7, 2}, {y, −1, 3}];
Print["Mean= ", N[m]];
Print["Variance= ", N[m2 − m^2]]

Out[1]= c= 1/700
(0.0371429  0.0357143   0.0371429   0.0414286   0.0485714
 0.0142857  0.0128571   0.0142857   0.0185714   0.0257143
 0.0028571  0.0014285   0.0028571   0.0071428   0.0142857
 0.0028571  0.0014285   0.0028571   0.0071428   0.0142857
 0.0142857  0.0128571   0.0142857   0.0185714   0.0257143
 0.0371429  0.0357143   0.0371429   0.0414286   0.0485714
 0.0714286  0.07        0.0714286   0.0757143   0.0828571)
Mean= 712.5
Variance= 1.3637*10^6
```

By an appropriate modification of the above, in each of the following cases find the value of c, then write the joint probability function in the form of a table, and finally find the expectation of XY:

(i) $f(x, y) = cxy,\ x = 1, 2, 3, 5, 7, \quad y = 4, 8, 12, 16, 20;$

(ii) $f(x, y) = \dfrac{cx^2}{y^3},\ x = 1, 2, 3, \quad y = 1, 2, 3;$

(iii) $f(x, y) = \dfrac{x + y}{c},\ x = 1, 2, 3, \quad y = 1, 2, 3;$

(iv) $f(x, y) = c(x + y + 1)^2,\ x = -3, -2, -1, 0, 1, 2, 3,\ y = 0, 1, 2, 3, 4.$

2. In the following cases, $f(x, y)$ is the joint probability function of X and Y taking values in the specified range for each of the two variables. Use Mathematica to calculate the value of c appearing in $f(x, y)$ below:

(i) $f(x,y) = \dfrac{c}{4^{x+1}7^y}$, $x = 0,1,2,\ldots,\ y = 0,1,2,\ldots$;

(ii) $f(x,y) = \dfrac{c}{5^{x+1}9^{y+1}}$, $x,y = 0,1,2,\ldots$;

(iii) $f(x,y) = \dfrac{c}{5^{x+1}9^{y+4}}$, $x,y = 0,1,2,\ldots$;

(iv) $f(x,y) = c\,\dfrac{5^x 2^y}{x!y!}$, $x,y = 0,1,2,\ldots$;

(v) $f(x,y) = \dfrac{c}{x!(x-y)!}$, $x = 0,1,2,\ldots,\ y = 0,1,2,\ldots,x$.

3. A pair of discrete random variables (X, Y) has joint probability function as

$$f(x,y) = \frac{x+y+xy}{c}, \quad x = 1,2,\ldots,10, \quad y = 5,6,\ldots,25,$$

for a suitable real constant c. Identify the value of c, and then calculate

$$E(X),\ E(Y),\ E(XY),\ E\left(\frac{X}{Y}\right).$$

Compare the last two quantities with the values of $E(X)E(Y)$ and $E(X)/E(Y)$, respectively. What do you observe?

4. The joint probability function of X and Y is given by

$$f(x,y) = c(x+y), \quad x = 1,2,\ldots,12,\ y = 1,2,\ldots,12,$$

for some suitable constant c. After finding the value of c, use Mathematica to calculate

(i) the marginal distributions of X and Y;

(ii) the expected values of X and Y;

(iii) the value of $E(X+Y)$ and verify that $E(X+Y) = E(X) + E(Y)$.

5. The joint probability function of X and Y in Example 1.2.3 is given in Table 1.4. Use Mathematica to

(i) present the values of this function in a table format, as in Table 1.4;

(ii) obtain the marginal probability functions of X and Y;

(iii) find the conditional probability functions $f_{X|Y}(x|0)$ and $f_{Y|X}(y|2)$, and verify that the results you obtain agree with those in Examples 1.2.3 and 1.5.1.

6. Let $f(x,y)$ be the probability function of (X,Y), wherein X takes values in the set $\{1,2,\cdots,n\}$ and for a given $X = x$, the range of Y is the set $\{1,2,\cdots,x\}$. Write a program to calculate $E\left[g(X,Y)\right]$ for a function g of the two variables. As an application,

(i) solve Exercise 5 of Section 1.5;

(ii) calculate

$$E(3X^2 + 2Y^4),\ E(X/Y),\ E(X - \sqrt{Y}),\ E(\ln(X + Y))$$

when the joint probability function of (X, Y) is given by

$$f(x, y) = c, \quad x = 1, 2, \ldots, 10,\ y = 1, 2, \ldots, x$$

(you need to find the value of c first).

7. We toss a coin 10 times, and denote by X the number of times the sequence HT occurs and by Y the number of times the sequence THH occurs (here, H stands for "Heads" and T stands for "Tails", as usual).
   (i) In Mathematica, create a set having all 2^{10} outcomes for this experiment (it might be convenient to code H and T as 0 and 1, respectively).
   (ii) After identifying the range of values, $R_{X,Y}$, for (X, Y), find their joint probability function.
   (iii) Thence, calculate

$$P(X = 1|Y = 2) \text{ and } P(Y = 0|X = 3).$$

8. The table below gives the joint distribution (that is, the joint probability function) of the pair (X, Y).

x \ y	0	1	2
0	2/21	3/21	4/21
1	3/21	4/21	5/21

With the following commands, we obtain the marginal probability functions of X and Y as well as all conditional distributions of X given Y and those of Y given X.

```
In[1]:= n = 2; m = 3;
t = {{2/21, 3/21, 4/21}, {3/21, 4/21, 5/21}};
Print[MatrixForm[t]]
Print["Marginal probability function of X"]
Do[Xmarg[i] = Sum[t[[i, j]], {j, 1, m}]; Print[Xmarg[i]], {i, 1, n}]
Print["Marginal probability function of Y"]
Do[Ymarg[j] = Sum[t[[i, j]], {i, 1, n}]; Print[Ymarg[j]], {j, 1, m}]
Print["Marginal probability functions of X given Y"]
Do[Print ["------"]
Do[Print[t[[i, j]]/Ymarg[j]], {i, 1, n}], {j, 1, m}]
Print["Marginal probability functions of Y given X"]
Do[Print ["------"]
Do[Print[t[[i, j]]/Xmarg[i]], {j, 1, m}], {i, 1, n}]
```

```
Out[1]=
(2/21   1/7     4/21
 1/7    4/21    5/21)
Marginal probability function of X
3/7
4/7
Marginal probability function of Y
5/21
1/3
3/7
Marginal probability functions of X given Y
-------
2/5
3/5
-------
3/7
4/7
-------
4/9
5/9
Marginal probability functions of Y given X
-------
2/9
1/3
4/9
-------
1/4
1/3
5/12
```

By a suitable modification of this program and adding some new commands as appropriate, calculate the following for the pair (X, Y) with joint probability function given in the table below:

(i) the marginal probability functions as well as the expectation and variance for each of X and Y;

(ii) all the conditional distributions of X given Y and those of Y given X;

(iii) all the conditional expectations and variances of X given Y and those of Y given X.

$x \backslash y$	0	1	2	3
0	7/370	6/185	17/370	11/185
1	9/370	7/185	19/370	12/185
2	11/370	8/185	21/370	13/185
3	13/370	9/185	23/370	14/185
4	3/74	2/37	5/74	3/37

1.8 SELF-ASSESSMENT EXERCISES

1.8.1 True–False Questions

1. When we toss a coin twice, let X be the number of heads and Y be the number of tails observed. Then, $P(X + Y = 2) = 1$.

2. If (X, Y) has joint probability function $f(x, y)$, then

$$\sum_{x:(x,y)\in R_{X,Y}} f(x, y) = 1.$$

3. Let (X, Y) be a discrete, two-dimensional random variable with range $R_{X,Y}$ and $F_X(x)$ be the marginal distribution function of X. Then,

$$F_X(x) = \sum_{y:(x,y)\in R_{X,Y}} P(X \le x, Y = y).$$

Questions 4–12 below refer to a pair of random variables (X, Y) with joint probability function

$$f(0,0) = f(0,1) = \frac{1}{4}, \quad f(1,0) = \frac{1}{8}, \quad f(1,1) = \frac{3}{8}.$$

4. $P(X = 1) = \frac{1}{2}$.

5. $P(X = Y) = \frac{5}{8}$.

6. $f_{X|Y}(1|0)$ cannot be defined.

7. $P(X = 1|Y = 1) = \frac{3}{5}$.

8. X and Y have the same expectation.

9. The conditional expectation of X, given $Y = 1$, is $\frac{1}{3}$.

10. $E(X^2|Y = 0) = \frac{1}{3}$.

11. $E(XY) = \frac{3}{8}$.

12. Let $F_Y(y)$ be the marginal distribution function of Y. Then, $F_Y(1/2) = \frac{1}{2}$.

13. Let (X, Y) have joint probability function $f(x, y)$ and denote the marginal probability function of X by $f_X(x)$. Then,

$$\sum_{y:(x,y)\in R_{X,Y}} \frac{f(x, y)}{f_X(x)} = 1.$$

14. If X and Y are discrete random variables, then

$$E(2X^2 + 3Y^2) = 2[E(X)]^2 + 3[E(Y)]^2.$$

15. For any pair of discrete variables (X, Y),

$$E(2X - Y) + E(X + Y) = 3E(X).$$

16. An urn contains 10 black balls and 8 red balls. We select from it two balls without replacement, and let X and Y denote the number of black and red balls selected, respectively. Then,

$$E(X|Y = 1) = 1.$$

17. For a random pair (X, Y), let $F_{X|Y}(x|y)$ be the conditional distribution function of X, given $Y = y$ (assuming that $P(Y = y) > 0$). Then, the following holds for $(x, y) \in R_{X,Y}$:

$$P(X > x|Y = y) = 1 - F_{X|Y}(x|y).$$

1.8.2 Multiple Choice Questions

1. Assume that the joint probability function of (X, Y) is

$$f(x, y) = \frac{c(x + y)}{9}, \quad x = 1, 2,\ y = 2, 3, 4.$$

Then, the value of c is

(a) 3 (b) $\frac{1}{3}$ (c) $\frac{1}{27}$ (d) $\frac{5}{9}$ (e) $\frac{9}{5}$

2. Assume that the pair (X, Y) has joint probability function

$$f(1, 1) = f(1, 2) = \frac{1}{6},\ f(2, 1) = \frac{1}{4},\ f(2, 2) = \frac{5}{12}.$$

Then, the marginal probability function of Y is

(a) $f_Y(1) = \frac{2}{7}, \quad f_Y(2) = \frac{5}{7}$

(b) $f_Y(1) = \frac{5}{12}, \quad f_Y(2) = \frac{7}{12}$

(c) $f_Y(1) = \frac{1}{3}, \quad f_Y(2) = \frac{2}{3}$

(d) $f_Y(1) = \frac{2}{3}, \quad f_Y(2) = \frac{1}{3}$

(e) $f_Y(1) = \frac{2}{5}, \quad f_Y(2) = \frac{3}{5}$

3. The joint probability function of (X, Y) is

$$f(10, 20) = f(10, 40) = \frac{1}{5},\ f(20, 20) = \frac{1}{2},\ f(20, 40) = \frac{1}{10}.$$

Then, $E(X+Y)$ is equal to

(a) 22.5 (b) 36 (c) 40 (d) 42 (e) 45

4. Nick tosses a die twice, and let X and Y denote the outcomes of the first and second tosses, respectively. Then, $P(X \le 2|Y = 4)$ is

(a) $\frac{1}{18}$ (b) $\frac{1}{3}$ (c) $\frac{1}{6}$ (d) $\frac{1}{2}$ (e) $\frac{1}{9}$

5. Assume that each of the discrete random variables X and Y takes only the values 1 and 2, while their joint probability function is

$$f(1,1) = \frac{4}{15},\ f(1,2) = \frac{1}{6},\ f(2,1) = \frac{1}{5},\ f(2,2) = \frac{11}{30}.$$

Then, $P(X = 1|Y = 2)$ is equal to

(a) $\frac{5}{11}$ (b) $\frac{5}{17}$ (c) $\frac{1}{6}$ (d) $\frac{5}{16}$ (e) $\frac{11}{17}$

6. Assume that the joint probability function of (X, Y) is

$$f(x,y) = \frac{x^2+y^2}{c}, \quad x = 2,3,\ y = 2,3,$$

for a suitable constant c. Then, $P(Y = 3)$ equals

(a) $\frac{9}{31}$ (b) $\frac{3}{8}$ (c) $\frac{9}{13}$ (d) $\frac{31}{48}$ (e) $\frac{31}{52}$

7. Wendy has a green and a blue dice and she tosses them once. Let X be the outcome of the blue die and Y be the sum of the two outcomes. Then, $P(X = 3|Y = 8)$ is

(a) $\frac{1}{5}$ (b) $\frac{1}{6}$ (c) $\frac{1}{4}$ (d) $\frac{5}{36}$ (e) $\frac{1}{9}$

8. A box contains 5 blue chips and 4 red ones. We select two chips at random without replacement. Let X be the number of blue chips and Y be the number of red chips selected. The marginal probability function of Y is

(a) $f_Y(0) = \frac{25}{81}$, $f_Y(1) = \frac{40}{81}$, $f_Y(2) = \frac{16}{81}$

(b) $f_Y(0) = \frac{1}{6}$, $f_Y(1) = \frac{5}{9}$, $f_Y(2) = \frac{5}{18}$

(c) $f_Y(0) = \frac{5}{18}$, $f_Y(1) = \frac{5}{9}$, $f_Y(2) = \frac{1}{6}$

(d) $f_Y(0) = \frac{1}{3}$, $f_Y(1) = \frac{5}{9}$, $f_Y(2) = \frac{1}{9}$

(e) $f_Y(0) = \frac{5}{9}$, $f_Y(1) = \frac{1}{3}$, $f_Y(2) = \frac{1}{9}$

9. The joint probability function of (X, Y) is

$$f(x, y) = \frac{1}{5^x 2^{y-2}}, \quad x = 1, 2, \ldots, \quad y = 1, 2, \ldots .$$

The marginal distribution of X is (for $x = 1, 2, \ldots$)

(a) $f_X(x) = \dfrac{4}{5^x}$

(b) $f_X(x) = \dfrac{2}{5^x}$

(c) $f_X(x) = \dfrac{1}{5^x}$

(d) $f_X(x) = \dfrac{2}{5^{x+1}}$

(e) $f_X(x) = \dfrac{4}{5^{x+1}}$

Questions 10–15 refer to a random pair (X, Y) whose joint probability function is as follows:

$x \backslash y$	1	2	3
1	0.10	0.15	0.15
2	0.12	0.08	0.05
3	0.06	0.14	0.15

10. $f_{X|Y}(3|2)$ is

(a) $\frac{1}{5}$ (b) $\frac{35}{37}$ (c) $\frac{14}{37}$ (d) $\frac{2}{5}$ (e) $\frac{3}{7}$

11. $P(X > 1, Y \le 2)$ equals

(a) 0.32 (b) 0.43 (c) 0.45 (d) 0.40 (e) 0.60

12. The value of the conditional (cumulative) distribution function $F_{X|Y}(x|y)$ for $x = y = 2$ is

(a) $\frac{13}{20}$ (b) $\frac{8}{37}$ (c) $\frac{23}{37}$ (d) $\frac{9}{20}$ (e) $\frac{9}{13}$

13. $E(X + Y)$ equals

(a) 2.00 (b) 2.07 (c) 1.70 (d) 4.02 (e) 3.90

14. $E(Y|X = 1)$ equals

(a) 0.357 (b) 1.857 (c) 2.125 (d) 2.050 (e) 2.750

15. $E(X^2 - Y^2)$ equals

(a) 0.36 (b) 1.30 (c) -0.0824 (d) 9.46 (e) -0.36

1.9 REVIEW PROBLEMS

1. We toss a coin n times and denote by X the number of heads in the first toss and by Y the total number of heads in the n tosses.
   (i) Find the joint probability function of (X, Y).
   (ii) Derive the marginal probability functions of X and Y.
   (iii) Show that the conditional probability function of Y, given $X = x$, is a (univariate) binomial distribution, while the conditional distribution of X, given $Y = y$, is a (univariate) hypergeometric distribution.

   [Hint: For Parts (i) and (ii), generalize the arguments used in Exercise 9 of Section 1.3.]

2. Let X be a random variable having the Bernoulli distribution with success probability p, so that
$$P(X = 1) = p = 1 - P(X = 0).$$
   Define the variables $Y = X^2$ and $W = 1 - X$.
   (i) Obtain the joint probability functions of (X, Y) and (X, W).
   (ii) Calculate the means of XY and X^3 and verify that
$$E(XY) = E(X^3)$$
   (a) by using the joint distribution of (X, Y) found in Part (i), and Proposition 1.4.1 for the expectation of a function of two variables;
   (b) by finding the third moment around zero of a Bernoulli distribution.

3. Let X be the outcome in a single throw of a die. If $X = x$, we toss a coin x times, and denote the number of heads we observe in them by Y.
   (i) Obtain the joint probability function of (X, Y).
   (ii) Derive the marginal probability function of Y.
   (iii) If we know that the number of heads observed in this experiment is 2, what is the expected outcome of the die?

   [Hint: Use the distribution of X and the fact that given $X = x$, the conditional distribution of Y is binomial with parameters x and $1/2$.]

4. **(Bivariate hypergeometric distribution)** A box contains a white, b black and c red balls and we take successively, without replacement, n balls out of the box. Let X be the number of white balls that were selected and Y be the number of black balls selected. Find the joint probability function of (X, Y). The distribution of (X, Y) is a generalization of the familiar (univariate) hypergeometric distribution; this distribution will be further discussed in Chapter 6.

Application: A medical drawer contains 4 boxes with aspirins, 2 boxes with sleeping pills, and 3 boxes with multivitamins. We select two boxes at random (without replacement), and let X and Y be the numbers of boxes with aspirins and with sleeping pills, respectively, that we selected. Then, find

$$P(X + Y = 1), \quad P(X \le 1), \quad P(X \le 1, Y \le 1).$$

5. Let (X, Y) be a pair of discrete random variables whose range is

$$R_{X,Y} = \{(x, y) : x \in \{x_0, x_1, \dots\}, \text{ and } y \in \{y_0, y_1, \dots\}\}.$$

Show that the joint probability function of X and Y, denoted by $f(x, y)$, is related to the joint distribution function, $F(x, y)$, of X and Y through the formula

$$f(x_i, y_j) = \begin{cases} F(x_0, y_0), & i = j = 0, \\ F(x_0, y_j) - F(x_0, y_{j-1}), & i = 0, j \ge 1, \\ F(x_i, y_0) - F(x_{i-1}, y_0), & i \ge 1, j = 0, \\ F(x_i, y_j) - F(x_{i-1}, y_j) & \\ \quad -F(x_i, y_{j-1}) + F(x_{i-1}, y_{j-1}), & i \ge 1, j \ge 1. \end{cases}$$

Application: Assume that the joint distribution function of (X, Y) is

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

Obtain the joint probability function of (X, Y), and thence calculate

$$P(X = 1, Y = 1), \; P(X = 1), \; P(X + Y = 1), \; P(X^2 - Y^2 = -1).$$

6. Let X and Y be two discrete random variables with ranges R_X and R_Y, respectively. If (X, Y) has joint probability function $f(x, y)$ and range $R_{X,Y}$, verify that the marginal distribution of X satisfies

$$f_X(x) = \sum_{y : (x,y) \in R_{X,Y}} f_{X|Y}(x|y) f_Y(y), \; (x, y) \in R_{X,Y}.$$

Application 1: A random variable N follows the Poisson distribution with parameter $\lambda > 0$, while for another variable X it is known that

(i) when the variable N takes the value n, then X can only take one of the values $0, 1, 2, \dots, n$,

(ii) $P(X = x|N = n) = \binom{n}{x} p^x (1 - p)^{n-x}, \; x = 0, 1, \dots, n,$

for some $p \in (0, 1)$, that is, the conditional distribution of X, given $N = n$, is binomial with parameters (n, p).

Show that the distribution of X is Poisson with parameter λp.

Application 2: The number of persons who are treated at an outpatient ward in a hospital has the Poisson distribution with parameter $\lambda = 15$. It has been estimated that 90% of the patients who are treated in the ward leave within three hours of their admittance. Find

(i) the probability that, on a given day, there are at least three patients who stay for more than three hours in the ward;

(ii) the expected number of persons treated daily and *not staying* in the ward for more than three hours.

7. Suppose in a certain town, 15% of married couples have no children, 25% have one child, 35% have two children, and 25% have three children. We assume further that male and female children are equally likely in each family and, after choosing at random a family from that population, let X and Y be the numbers of boys and girls in this family, respectively.

   (i) Obtain the joint probability function of (X, Y).

   (ii) What is the probability that the family does not have the same number of boys and girls?

   (iii) Calculate the expected numbers of boys, girls and children in the family.

8. The joint probability function of (X, Y) is

$$f(x, y) = \frac{x + y}{21}, \quad x = 1, 2, 3, \ y = 1, 2.$$

   (i) Calculate $P(X \leq 2, Y = 1)$, $P(X = Y)$, $P(X + 2Y = 5)$.

   (ii) Write down the marginal probability functions of X and Y.

   (iii) Find the conditional probability functions $f_{X|Y}(x|y)$ and $f_{Y|X}(y|x)$, and thence the values of $P(X = 2|Y = 2)$, $P(Y = 1|X = 1)$.

   (iv) Calculate $E(X|Y = 1)$ and $E(Y|X = 3)$.

9. Nigel and Garry, who like playing chess, agree to play a series of four chess games. Let X be the number of games that Nigel wins and Y be the number of games that Garry wins. Assume that, in each game, each player has a probability of 30% to win and that there is a probability of 40% of the game ending in a draw.

   Calculate the probability that

   (i) there are exactly two draws in the series of four games;

   (ii) Gary wins two games and Nigel wins one;

   (iii) each player wins at least one game;

   (iv) the series of four games ends in a draw (that is, X and Y take the same value).

   Can you write down the marginal distribution of X without doing any calculations?

10. In a city there are two stores selling cars of a certain make. Let X and Y denote the numbers of cars sold weekly in the first and second stores, respectively. The joint probability function of X and Y is as follows:

x \ y	0	1	2	3
0	0.07	0.06	0.05	0.03
1	0.05	0.06	0.05	0.04
2	0.07	0.08	0.10	0.05
3	0.03	0.05	0.07	0.05
4	0	0.02	0.04	0.03

(i) What is the probability that, in a given week, the number of cars sold by the two stores is the same?

(ii) Find the probability that the first store sells exactly two cars more than the second one.

(iii) Calculate the probability that, in a given week, the total number of cars sold together by the two stores is (a) exactly 5, and (b) at least 3.

(iv) Obtain the marginal probability functions of X and Y, and thence find $E(X), E(Y)$ and $E(X+Y)$.

(v) What is the expected number of cars sold by the second store in a week when the first store has sold two cars?

11. The joint probability function of (X, Y) is

$$f(1,2) = f(1,3) = \frac{1}{7}, \quad f(2,2) = \frac{2}{7}, \quad f(2,3) = \frac{3}{7}.$$

(i) Find the marginal probability functions of X and Y.

(ii) Calculate

$$E(X),\ E(Y),\ E(X+Y),\ E(3X-2Y),\ E(XY),\ E(X/Y).$$

(iii) Obtain the conditional probability functions of X, given Y, and of Y, given X. Thence, calculate

$$E(X|Y=2), \quad E(X|Y=3), \quad E(Y|X=1), \quad E(Y|X=2).$$

12. The random variables X and Y have joint probability function as

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

(i) Calculate $P(X \leq 2, Y > 3)$.

(ii) Show that

$$E(Y|X=x) = \frac{3}{2}$$

for all $x = 1, 2, \ldots$.

13. The joint probability function of (X, Y) is given by

$x \backslash y$	10	20
2	a	$2/9$
3	$a - b$	a
4	$2a$	$2b$

for some constants a and b.

(i) Find the values of a and b if it is known that

$$2f(3, 10) + 3f(4, 20) = 2f(2, 20).$$

(ii) Calculate $E(X|Y = 10)$ and $E(X|Y = 20)$. Are these two expectations equal?

(iii) Find the value of $E(X^2 + Y)$.

1.10 APPLICATIONS

1.10.1 Mixture Distributions and Reinsurance

In certain insurance portfolios, especially those associated with catastrophic risks, e.g. due to national disasters, the insurance company may not afford to undertake the accumulated risk of potentially huge losses; in these cases, it is quite common practice the primary insurance company to transfer part of this risk to another company by *reinsurance*. One of the most common types of reinsurance treaties is *excess of loss reinsurance*; under this agreement, the (primary) insurance company pays up to a certain amount, say z_0, to the policyholder who made the claim. If a claim exceeds that amount, the excess over z_0 is paid by the reinsurer. Apparently, if a claim is less than or equal to z_0 monetary units, it does not reach the reinsurer.

Suppose now we examine the following problem, from the reinsurer's point of view: we want to know the probability that within a fixed time period, say a month, there will be at least k claims that the reinsurance company has to pay for. Let N denote the number of claims that will be made to the primary insurer and X be the number, among them, that will be above z_0 so that they reach the reinsurer. We use the generic symbol Z for the size of a claim and assume that the distribution function for the sizes of these claims is F, so that an arbitrary claim reaches the reinsurer with probability

$$p = P(Z > z_0) = 1 - F(z_0).$$

The distribution F may be either discrete or continuous, but we focus first on the variables N, X which are both discrete. The problem may then be formulated as follows. We wish to calculate the probability

$$P(X \geq k)$$

so that the reinsurance company makes provisions that this probability does not exceed a fixed level, say α, for some small value of $\alpha \in (0, 1)$.

A common assumption in this context, which is supported by empirical evidence, is that the distribution of N is Poisson (recall that the Poisson is called the distribution for *rare events*). Write λ for the parameter of this distribution. Can we use this fact to obtain the distribution of X? In order to accomplish this we first note that, given that $N = n$, the distribution of X is binomial with parameters n and p (each claim exceeds the value z_0, independently of the other claims, with the same probability p). Using the result of Exercise 6 in Section 1.9, we then obtain that the distribution of X is Poisson with parameter λp. A distribution which arises in this form is called a **mixture distribution**, while the distribution of N and the conditional distribution of X, given N, are the *components* of that mixture.

Suppose now the reinsurer poses the condition $P(X \geq k) < \alpha$, or equivalently,

$$P(X < k) \geq 1 - \alpha.$$

Since the distribution of X is Poisson with parameter λp we get

$$\sum_{x=0}^{k-1} e^{-\lambda p} \frac{(\lambda p)^x}{x!} \geq 1 - \alpha. \tag{1.6}$$

For catastrophic risks, it is typical that the distribution of individual claims is heavy-tailed. A distribution with such behavior which is used often as a model for insurance claims is the Pareto distribution, with density

$$f_Z(z) = \frac{cb^c}{(z+b)^{c+1}}, \qquad z \geq 0.$$

Here $b > 0, c > 0$ are the parameters of the distribution. Is is easy to see that the distribution function associated with this density is

$$F_Z(z) = 1 - \left(\frac{b}{z+b}\right)^c, \qquad z \geq 0.$$

For a portfolio with Pareto claims, as above, the probability that a claim reaches the reinsurer is simply

$$p = 1 - F_Z(z) = \left(\frac{b}{z+b}\right)^c,$$

and now substituting the value of p in (1.6) we arrive at the formula

$$\sum_{x=0}^{k-1} \exp\left[-\lambda\left(\frac{b}{z+b}\right)^c\right] \frac{\lambda^x \left((b/(z_0+b))\right)^{cx}}{x!} \geq 1 - \alpha.$$

We write for simplicity $g(z_0)$ for the quantity on the left-hand side above. It is easy to check that g is a increasing function of z_0, so that the condition

$$g(z_0) \geq 1 - \alpha$$

is equivalent to the condition that z_0 exceeds a certain value, say z^*. Figure 1.1 shows the function $g(z_0)$ for $\lambda = 40, b = 2, c = 4, k = 4$.

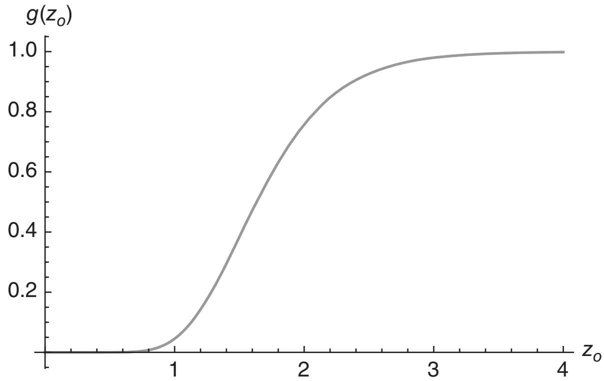


Figure 1.1 The function $g(z_0)$.

Finally, the following table gives the value z^*, which is the minimum value for z_0 so that (1.6) is satisfied, for $b = 2, c = 4$ and for different values of λ, k, and α.

α	λ	k	z^*
0.001	40	2	8.9
0.001	40	4	4.3
0.001	80	2	11.0
0.001	80	4	5.4
0.005	40	2	6.9
0.005	40	4	3.6
0.005	80	2	8.6
0.005	80	4	4.7
0.01	40	2	6.2
0.01	40	4	3.3
0.01	80	2	7.7
0.01	80	4	4.3

KEY TERMS

bivariate (or two-dimensional) distribution
conditional cumulative distribution function
conditional distribution
conditional expectation
conditional probability function
joint cumulative distribution function
joint distribution
joint probability function
marginal distribution
marginal probability function
mixture distribution
random pair (of discrete variables); or two-dimensional (discrete) random variable