CHAPTER 2

Random Variables and Probability Distributions

Random Variables

Suppose that to each point of a sample space we assign a number. We then have a *function* defined on the sample space. This function is called a *random variable* (or *stochastic variable*) or more precisely a *random function* (*stochastic function*). It is usually denoted by a capital letter such as X or Y. In general, a random variable has some specified physical, geometrical, or other significance.

EXAMPLE 2.1 Suppose that a coin is tossed twice so that the sample space is $S = \{HH, HT, TH, TT\}$. Let X represent the number of heads that can come up. With each sample point we can associate a number for X as shown in Table 2-1. Thus, for example, in the case of HH (i.e., 2 heads), $X = 2$ while for TH (1 head), $X = 1$. It follows that X is a random variable.

Table 2-1

Sample Point	*HH*	*HT*	*TH*	*TT*
X	2	1	1	0

It should be noted that many other random variables could also be defined on this sample space, for example, the square of the number of heads or the number of heads minus the number of tails.

A random variable that takes on a finite or countably infinite number of values (see page 4) is called a *discrete random variable* while one which takes on a noncountably infinite number of values is called a *nondiscrete random variable*.

Discrete Probability Distributions

Let X be a discrete random variable, and suppose that the possible values that it can assume are given by $x_1, x_2, x_3, \ldots$, arranged in some order. Suppose also that these values are assumed with probabilities given by

$$P(X = x_k) = f(x_k) \qquad k = 1, 2, \ldots \tag{1}$$

It is convenient to introduce the *probability function*, also referred to as *probability distribution*, given by

$$P(X = x) = f(x) \tag{2}$$

For $x = x_k$, this reduces to (1) while for other values of x, $f(x) = 0$.

In general, $f(x)$ is a probability function if

1. $f(x) \geq 0$
2. $\sum_x f(x) = 1$

where the sum in 2 is taken over all possible values of x.

EXAMPLE 2.2 Find the probability function corresponding to the random variable X of Example 2.1. Assuming that the coin is fair, we have

$$P(HH) = \frac{1}{4} \qquad P(HT) = \frac{1}{4} \qquad P(TH) = \frac{1}{4} \qquad P(TT) = \frac{1}{4}$$

Then

$$P(X = 0) = P(TT) = \frac{1}{4}$$

$$P(X = 1) = P(HT \cup TH) = P(HT) + P(TH) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$$

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

The probability function is thus given by Table 2-2.

Table 2-2

x	0	1	2
$f(x)$	1/4	1/2	1/4

Distribution Functions for Random Variables

The *cumulative distribution function*, or briefly the *distribution function*, for a random variable X is defined by

$$F(x) = P(X \leq x) \tag{3}$$

where x is any real number, i.e., $-\infty < x < \infty$.

The distribution function $F(x)$ has the following properties:

1. $F(x)$ is nondecreasing [i.e., $F(x) \leq F(y)$ if $x \leq y$].
2. $\lim_{x\to-\infty} F(x) = 0$; $\lim_{x\to\infty} F(x) = 1$.
3. $F(x)$ is continuous from the right [i.e., $\lim_{h\to 0^+} F(x + h) = F(x)$ for all x].

Distribution Functions for Discrete Random Variables

The distribution function for a discrete random variable X can be obtained from its probability function by noting that, for all x in $(-\infty, \infty)$,

$$F(x) = P(X \leq x) = \sum_{u \leq x} f(u) \tag{4}$$

where the sum is taken over all values u taken on by X for which $u \leq x$.

If X takes on only a finite number of values $x_1, x_2, \ldots, x_n$, then the distribution function is given by

$$F(x) = \begin{cases} 0 & -\infty < x < x_1 \\ f(x_1) & x_1 \leq x < x_2 \\ f(x_1) + f(x_2) & x_2 \leq x < x_3 \\ \vdots & \vdots \\ f(x_1) + \cdots + f(x_n) & x_n \leq x < \infty \end{cases} \tag{5}$$

EXAMPLE 2.3 (a) Find the distribution function for the random variable X of Example 2.2. (b) Obtain its graph.

(a) The distribution function is

$$F(x) = \begin{cases} 0 & -\infty < x < 0 \\ \frac{1}{4} & 0 \leq x < 1 \\ \frac{3}{4} & 1 \leq x < 2 \\ 1 & 2 \leq x < \infty \end{cases}$$

(b) The graph of $F(x)$ is shown in Fig. 2-1.

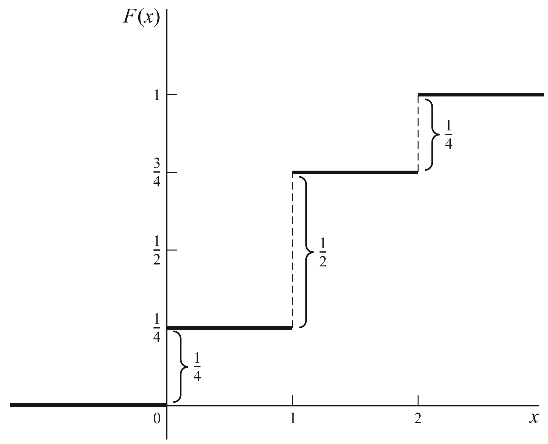

Fig. 2-1

The following things about the above distribution function, which are true in general, should be noted.

1. The magnitudes of the jumps at 0, 1, 2 are $\frac{1}{4}, \frac{1}{2}, \frac{1}{4}$ which are precisely the probabilities in Table 2-2. This fact enables one to obtain the probability function from the distribution function.
2. Because of the appearance of the graph of Fig. 2-1, it is often called a *staircase function* or *step function*. The value of the function at an integer is obtained from the higher step; thus the value at 1 is $\frac{3}{4}$ and not $\frac{1}{4}$. This is expressed mathematically by stating that the distribution function is *continuous from the right* at 0, 1, 2.
3. As we proceed from left to right (i.e. going *upstairs*), the distribution function either remains the same or increases, taking on values from 0 to 1. Because of this, it is said to be a *monotonically increasing function*.

It is clear from the above remarks and the properties of distribution functions that the probability function of a discrete random variable can be obtained from the distribution function by noting that

$$f(x) = F(x) - \lim_{u \to x^-} F(u). \tag{6}$$

Continuous Random Variables

A nondiscrete random variable X is said to be *absolutely continuous*, or simply *continuous*, if its distribution function may be represented as

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(u)\,du \qquad (-\infty < x < \infty) \tag{7}$$

where the function $f(x)$ has the properties

1. $f(x) \geq 0$
2. $\displaystyle\int_{-\infty}^{\infty} f(x)\,dx = 1$

It follows from the above that if X is a continuous random variable, then the probability that X takes on any one particular value is zero, whereas the *interval probability* that X lies *between two different values*, say, a and b, is given by

$$P(a < X < b) = \int_{a}^{b} f(x)\,dx \tag{8}$$

EXAMPLE 2.4 If an individual is selected at random from a large group of adult males, the probability that his height X is precisely 68 inches (i.e., 68.000 . . . inches) would be zero. However, there is a probability greater than zero than X is between 67.000 . . . inches and 68.500 . . . inches, for example.

A function $f(x)$ that satisfies the above requirements is called a *probability function* or *probability distribution* for a continuous random variable, but it is more often called a *probability density function* or simply *density function*. Any function $f(x)$ satisfying Properties 1 and 2 above will automatically be a density function, and required probabilities can then be obtained from (8).

EXAMPLE 2.5 (a) Find the constant c such that the function

$$f(x) = \begin{cases} cx^2 & 0 < x < 3 \\ 0 & \text{otherwise} \end{cases}$$

is a density function, and (b) compute $P(1 < X < 2)$.

(a) Since $f(x)$ satisfies Property 1 if $c \geq 0$, it must satisfy Property 2 in order to be a density function. Now

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_0^3 cx^2\,dx = \left.\frac{cx^3}{3}\right|_0^3 = 9c$$

and since this must equal 1, we have $c = 1/9$.

(b)
$$P(1 < X < 2) = \int_1^2 \frac{1}{9}x^2\,dx = \left.\frac{x^3}{27}\right|_1^2 = \frac{8}{27} - \frac{1}{27} = \frac{7}{27}$$

In case $f(x)$ is continuous, which we shall assume unless otherwise stated, the probability that X is equal to any particular value is zero. In such case we can replace either or both of the signs $<$ in (8) by $\leq$. Thus, in Example 2.5,

$$P(1 \leq X \leq 2) = P(1 \leq X < 2) = P(1 < X \leq 2) = P(1 < X < 2) = \frac{7}{27}$$

EXAMPLE 2.6 (a) Find the distribution function for the random variable of Example 2.5. (b) Use the result of (a) to find $P(1 < x \leq 2)$.

(a) We have

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(u)\,du$$

If $x < 0$, then $F(x) = 0$. If $0 \leq x < 3$, then

$$F(x) = \int_0^x f(u)\,du = \int_0^x \frac{1}{9}u^2\,du = \frac{x^3}{27}$$

If $x \geq 3$, then

$$F(x) = \int_0^3 f(u)\,du + \int_3^x f(u)\,du = \int_0^3 \frac{1}{9}u^2\,du + \int_3^x 0\,du = 1$$

Thus the required distribution function is

$$F(x) = \begin{cases} 0 & x < 0 \\ x^3/27 & 0 \leq x < 3 \\ 1 & x \geq 3 \end{cases}$$

Note that $F(x)$ increases monotonically from 0 to 1 as is required for a distribution function. It should also be noted that $F(x)$ in this case is continuous.

(b) We have

$$\begin{aligned} P(1 < X \le 2) &= P(X \le 2) - P(X \le 1) \\ &= F(2) - F(1) \\ &= \frac{2^3}{27} - \frac{1^3}{27} = \frac{7}{27} \end{aligned}$$

as in Example 2.5.

The probability that X is between x and $x + \Delta x$ is given by

$$P(x \le X \le x + \Delta x) = \int_x^{x+\Delta x} f(u)\,du \tag{9}$$

so that if Δx is small, we have approximately

$$P(x \le X \le x + \Delta x) = f(x)\Delta x \tag{10}$$

We also see from (7) on differentiating both sides that

$$\frac{dF(x)}{dx} = f(x) \tag{11}$$

at all points where $f(x)$ is continuous; i.e., the derivative of the distribution function is the density function.

It should be pointed out that random variables exist that are neither discrete nor continuous. It can be shown that the random variable X with the following distribution function is an example.

$$F(x) = \begin{cases} 0 & x < 1 \\ \dfrac{x}{2} & 1 \le x < 2 \\ 1 & x \ge 2 \end{cases}$$

In order to obtain (11), we used the basic property

$$\frac{d}{dx}\int_a^x f(u)\,du = f(x) \tag{12}$$

which is one version of the Fundamental Theorem of Calculus.

Graphical Interpretations

If $f(x)$ is the density function for a random variable X, then we can represent $y = f(x)$ graphically by a curve as in Fig. 2-2. Since $f(x) \ge 0$, the curve cannot fall below the x axis. The entire area bounded by the curve and the x axis must be 1 because of Property 2 on page 36. Geometrically the probability that X is between a and b, i.e., $P(a < X < b)$, is then represented by the area shown shaded, in Fig. 2-2.

The distribution function $F(x) = P(X \le x)$ is a monotonically increasing function which increases from 0 to 1 and is represented by a curve as in Fig. 2-3.

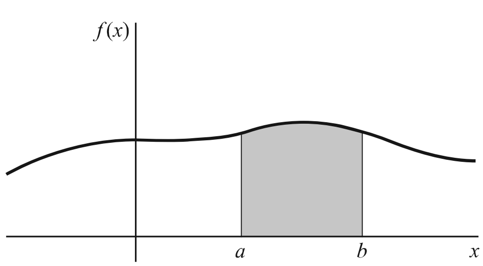

Fig. 2-2

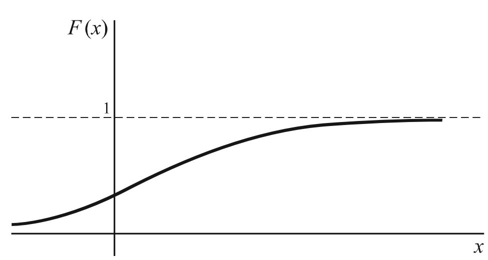

Fig. 2-3

Joint Distributions

The above ideas are easily generalized to two or more random variables. We consider the typical case of two random variables that are either both discrete or both continuous. In cases where one variable is discrete and the other continuous, appropriate modifications are easily made. Generalizations to more than two variables can also be made.

1. DISCRETE CASE. If X and Y are two discrete random variables, we define the *joint probability function* of X and Y by

$$P(X = x, Y = y) = f(x, y) \tag{13}$$

where 1. $f(x, y) \geq 0$

2. $\sum_x \sum_y f(x, y) = 1$

i.e., the sum over all values of x and y is 1.

Suppose that X can assume any one of m values $x_1, x_2, \ldots, x_m$ and Y can assume any one of n values $y_1, y_2, \ldots, y_n$. Then the probability of the event that $X = x_j$ and $Y = y_k$ is given by

$$P(X = x_j, Y = y_k) = f(x_j, y_k) \tag{14}$$

A joint probability function for X and Y can be represented by a *joint probability table* as in Table 2-3. The probability that $X = x_j$ is obtained by adding all entries in the row corresponding to x_i and is given by

$$P(X = x_j) = f_1(x_j) = \sum_{k=1}^{n} f(x_j, y_k) \tag{15}$$

Table 2-3

X \ Y	y_1	y_2	$\ldots$	y_n	Totals ↓	
x_1	$f(x_1, y_1)$	$f(x_1, y_2)$	$\ldots$	$f(x_1, y_n)$	$f_1(x_1)$	
x_2	$f(x_2, y_1)$	$f(x_2, y_2)$	$\ldots$	$f(x_2, y_n)$	$f_1(x_2)$	
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$	
x_m	$f(x_m, y_1)$	$f(x_m, y_2)$	$\ldots$	$f(x_m, y_n)$	$f_1(x_m)$	
Totals →	$f_2(y_1)$	$f_2(y_2)$	$\ldots$	$f_2(y_n)$	1	← Grand Total

For $j = 1, 2, \ldots, m$, these are indicated by the entry totals in the extreme right-hand column or margin of Table 2-3. Similarly the probability that $Y = y_k$ is obtained by adding all entries in the column corresponding to y_k and is given by

$$P(Y = y_k) = f_2(y_k) = \sum_{j=1}^{m} f(x_j, y_k) \tag{16}$$

For $k = 1, 2, \ldots, n$, these are indicated by the entry totals in the bottom row or margin of Table 2-3.

Because the probabilities (15) and (16) are obtained from the margins of the table, we often refer to $f_1(x_j)$ and $f_2(y_k)$ [or simply $f_1(x)$ and $f_2(y)$] as the *marginal probability functions* of X and Y, respectively.

It should also be noted that

$$\sum_{j=1}^{m} f_1(x_j) = 1 \quad \sum_{k=1}^{n} f_2(y_k) = 1 \tag{17}$$

which can be written

$$\sum_{j=1}^{m}\sum_{k=1}^{n} f(x_j, y_k) = 1 \tag{18}$$

This is simply the statement that the total probability of all entries is 1. The *grand total* of 1 is indicated in the lower right-hand corner of the table.

The *joint distribution function* of X and Y is defined by

$$F(x, y) = P(X \leq x, Y \leq y) = \sum_{u \leq x}\sum_{v \leq y} f(u, v) \tag{19}$$

In Table 2-3, $F(x, y)$ is the sum of all entries for which $x_j \leq x$ and $y_k \leq y$.

2. CONTINUOUS CASE. The case where both variables are continuous is obtained easily by analogy with the discrete case on replacing sums by integrals. Thus the *joint probability function* for the random variables X and Y (or, as it is more commonly called, the *joint density function* of X and Y) is defined by

1. $f(x, y) \geq 0$
2. $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)\,dx\,dy = 1$

Graphically $z = f(x, y)$ represents a surface, called the *probability surface*, as indicated in Fig. 2-4. The total volume bounded by this surface and the xy plane is equal to 1 in accordance with Property 2 above. The probability that X lies between a and b while Y lies between c and d is given graphically by the shaded volume of Fig. 2-4 and mathematically by

$$P(a < X < b, c < Y < d) = \int_{x=a}^{b}\int_{y=c}^{d} f(x, y)\,dx\,dy \tag{20}$$

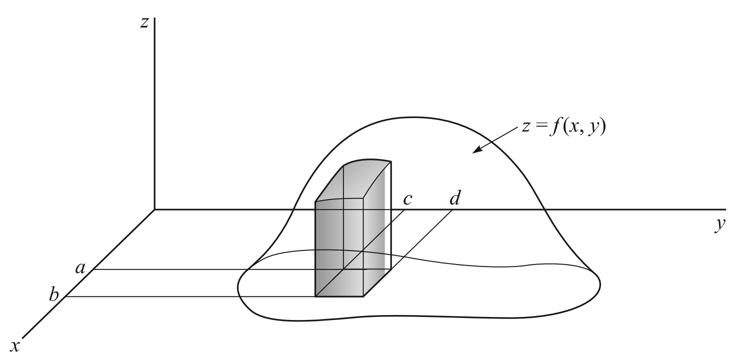

Fig. 2-4

More generally, if A represents any event, there will be a region $\mathcal{R}_A$ of the xy plane that corresponds to it. In such case we can find the probability of A by performing the integration over $\mathcal{R}_A$, i.e.,

$$P(A) = \iint_{\mathcal{R}_A} f(x, y)\,dx\,dy \tag{21}$$

The *joint distribution function* of X and Y in this case is defined by

$$F(x, y) = P(X \leq x, Y \leq y) = \int_{u=-\infty}^{x}\int_{v=-\infty}^{y} f(u, v)\,du\,dv \tag{22}$$

It follows in analogy with (11), page 38, that

$$\frac{\partial^2 F}{\partial x\, \partial y} = f(x, y) \tag{23}$$

i.e., the density function is obtained by differentiating the distribution function with respect to x and y.

From (22) we obtain

$$P(X \leq x) = F_1(x) = \int_{u=-\infty}^{x} \int_{v=-\infty}^{\infty} f(u, v)\, du\, dv \tag{24}$$

$$P(Y \leq y) = F_2(y) = \int_{u=-\infty}^{\infty} \int_{v=-\infty}^{y} f(u, v)\, du\, dv \tag{25}$$

We call (24) and (25) the *marginal distribution functions*, or simply the *distribution functions*, of X and Y, respectively. The derivatives of (24) and (25) with respect to x and y are then called the *marginal density functions,* or simply the *density functions*, of X and Y and are given by

$$f_1(x) = \int_{v=-\infty}^{\infty} f(x, v)\, dv \qquad f_2(y) = \int_{u=-\infty}^{\infty} f(u, y)\, du \tag{26}$$

Independent Random Variables

Suppose that X and Y are discrete random variables. If the events $X = x$ and $Y = y$ are independent events for all x and y, then we say that X and Y are *independent random variables*. In such case,

$$P(X = x, Y = y) = P(X = x)P(Y = y) \tag{27}$$

or equivalently

$$f(x, y) = f_1(x)f_2(y) \tag{28}$$

Conversely, if for all x and y the joint probability function $f(x, y)$ can be expressed as the product of a function of x alone and a function of y alone (which are then the marginal probability functions of X and Y), X and Y are independent. If, however, $f(x, y)$ cannot be so expressed, then X and Y are *dependent*.

If X and Y are continuous random variables, we say that they are *independent random variables* if the events $X \leq x$ and $Y \leq y$ are independent events for all x and y. In such case we can write

$$P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y) \tag{29}$$

or equivalently

$$F(x, y) = F_1(x)F_2(y) \tag{30}$$

where $F_1(z)$ and $F_2(y)$ are the (marginal) distribution functions of X and Y, respectively. Conversely, X and Y are independent random variables if for all x and y, their joint distribution function $F(x, y)$ can be expressed as a product of a function of x alone and a function of y alone (which are the marginal distributions of X and Y, respectively). If, however, $F(x, y)$ cannot be so expressed, then X and Y are dependent.

For continuous independent random variables, it is also true that the joint density function $f(x, y)$ is the product of a function of x alone, $f_1(x)$, and a function of y alone, $f_2(y)$, and these are the (marginal) density functions of X and Y, respectively.

Change of Variables

Given the probability distributions of one or more random variables, we are often interested in finding distributions of other random variables that depend on them in some specified manner. Procedures for obtaining these distributions are presented in the following theorems for the case of discrete and continuous variables.