3

PROBABILITY DISTRIBUTIONS

Given the definitions of "event," "sample space," and "probability" in the last chapter, we are now ready to take up a major set of concepts in probability and statistics. First of all, the concept of a random variable will be introduced, and then we will discuss probability distributions of random variables. Finally, measures for summarizing probability distributions will be discussed.

3.1 RANDOM VARIABLES

We defined the sample space as the set of all elementary events that are possible outcomes of some simple experiment. These outcomes may be expressed in terms of numbers (for instance, the number coming up on one throw of a die; the temperature in degrees centigrade at a given place and time; the price of a certain stock on a given day) or they may be expressed in nonnumerical terms (for instance, the color of a ball drawn from an urn; the sex of a respondent to a questionnaire; the occurrence or nonoccurrence of rain on a particular day).

Imagine that a single number could be assigned to each and every possible elementary event in a sample space Ω. If we throw a die, the elementary events may correspond to the six possible faces that can come up. If we wished, we could assign to each elementary event the number of spots that appears on the face that comes up; or we could assign the square of the number that comes up; or we might even assign the number 1 if a six comes up, 2 if a five comes up, and so on, up to 6 if a one comes up. In all of these cases, each elementary event is associated with exactly one number. The most natural choice in the example, of course, is to assign the number of spots on the face which comes up. The other examples point out the fact that this is not the only choice.

It is important to note that although each elementary event can be associated with only one number, the same number may be associated with more than one elementary event. In the die example, we could assign the number 1 to the events "one comes up," "three comes up," and "five comes up" and assign the number 0 to the remaining elementary events. Thus the number 1 corresponds to the event "odd number comes up" and the number 0 corresponds to the event "even number comes up." This is an example of how more than one elementary event can be associated with a given number. It is also an example of the assignment of numbers to events which are not originally expressed in terms of numbers. The events "odd number comes up" and "even number comes up" are expressed in nonnumerical terms. In a similar manner, in the rain example, we could assign the number 1 to the event "rain" and the number 0 to the event "no rain."

The above examples show how a number can be assigned to each elementary event in a sample space. We will use the following terminology.

If the symbol X represents a function which associates a real number with each and every elementary event in a sample space Ω, then X is called a random variable which is defined on the sample space Ω. In other words, a random variable X is a real-valued function defined on a sample space.

This notion is illustrated in Figure 3.1.1. The dot in Ω represents an elementary event, and the function X associates a real number with that elementary event. Recalling the discussion of functions in Chapter 1, we note that a value x may be associated with a number of distinct elementary events, but each elementary event is associated with only one value.

We will use upper case letters, such as X, Y, and Z, to denote random variables, and lower case letters, such as x, y, z, a, b, c, and so on, to denote particular values of random variables. The expression $P(X = x)$ symbolizes the probability that the random variable X takes on the particular value x. When there is no chance for confusion, $P(X = x)$ will be abbreviated

![Figure 3.1.1 A Random Variable as a Function from the Sample Space to the Real Line](image-url)
3.2 PROBABILITY DISTRIBUTIONS

Consider the set of events \( \{A_1, A_2, \ldots, A_K\} \), and suppose that they form a partition of the sample space \( S \). That is, they are mutually exclusive and exhaustive. The corresponding set of probabilities, \( \{P(A_1), P(A_2), \ldots, P(A_K)\} \), is a probability distribution. In general,

any statement of a probability function having a set of mutually exclusive and exhaustive events for its domain is a probability distribution.

As a simple example of a probability distribution, imagine a sample space of all employees of a particular firm. An employee is selected at random and classified as “right-handed,” “left-handed,” or “ambidextrous.” The probability distribution might be as given in Table 3.2.1.

Table 3.2.1 Probability Distribution for Hand Preference of Employees

<table>
<thead>
<tr>
<th>Classification</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Right</td>
<td>.60</td>
</tr>
<tr>
<td>Left</td>
<td>.30</td>
</tr>
<tr>
<td>Ambidextrous</td>
<td>.10</td>
</tr>
<tr>
<td></td>
<td>1.00</td>
</tr>
</tbody>
</table>

Or, perhaps, the height of an employee drawn at random is measured. Some seven class intervals are used to record the height of employees, and the probability distribution might be as given in Table 3.2.2.

Table 3.2.2 Probability Distribution for Height of Employees

<table>
<thead>
<tr>
<th>Height in inches</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>78-82</td>
<td>.002</td>
</tr>
<tr>
<td>73-77</td>
<td>.021</td>
</tr>
<tr>
<td>68-72</td>
<td>.136</td>
</tr>
<tr>
<td>63-67</td>
<td>.082</td>
</tr>
<tr>
<td>58-62</td>
<td>.136</td>
</tr>
<tr>
<td>53-57</td>
<td>.021</td>
</tr>
<tr>
<td>48-52</td>
<td>.002</td>
</tr>
<tr>
<td></td>
<td>1.000</td>
</tr>
</tbody>
</table>

The set of all possible elementary events forms a partition of \( S \). As a result, the set consisting of the probabilities of all of the elementary events in a sample space is a probability distribution. If we throw a fair die, the distribution would be as given in Table 3.2.3.
Table 3.2.3 Probability Distribution of Number Showing on Die

<table>
<thead>
<tr>
<th>Face coming up on die</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1/6</td>
</tr>
<tr>
<td>2</td>
<td>1/6</td>
</tr>
<tr>
<td>3</td>
<td>1/6</td>
</tr>
<tr>
<td>4</td>
<td>1/6</td>
</tr>
<tr>
<td>5</td>
<td>1/6</td>
</tr>
<tr>
<td>6</td>
<td>1/6</td>
</tr>
<tr>
<td>-</td>
<td>1.0</td>
</tr>
</tbody>
</table>

However, suppose that we are interested in the square of the number of dots showing on the die. If we define the random variable X to be the square of the number of dots showing, then the probability distribution of X is as given in Table 3.2.4. All values of X other than the values listed in the table have probabilities of zero. Recall that a random variable defined on a sample space associates a real number with each and every elementary event in the sample space. In the example, the probability distribution on the elementary events determines the probability distribution of the random variable X. This exemplifies the fact that, in general, a set of probabilities, or a probability distribution, for the elementary events in 5 determines a set of probabilities, or a probability distribution, for any random variable defined on 5.

Table 3.2.4 Probability Distribution of Square of Number Showing on Die

<table>
<thead>
<tr>
<th>x</th>
<th>P(X = x)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1/6</td>
</tr>
<tr>
<td>4</td>
<td>1/6</td>
</tr>
<tr>
<td>9</td>
<td>1/6</td>
</tr>
<tr>
<td>16</td>
<td>1/6</td>
</tr>
<tr>
<td>25</td>
<td>1/6</td>
</tr>
<tr>
<td>36</td>
<td>1/6</td>
</tr>
<tr>
<td>-</td>
<td>1.0</td>
</tr>
</tbody>
</table>

This introduces the important concept of a probability distribution, or simply a distribution, of a random variable. In the above example, each value of the random variable has the same probability. This need not always be true. Consider the random variable Y, which is defined as the number of heads occurring in two flips of a fair coin. The probability distribution of Y is given in Table 3.2.5. In general, the various values of a random variable are not necessarily equally probable.

Table 3.2.5 Probability Distribution of Number of Heads in Two Tosses of a Fair Coin

<table>
<thead>
<tr>
<th>y</th>
<th>P(Y = y)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1/4</td>
</tr>
<tr>
<td>1</td>
<td>1/2</td>
</tr>
<tr>
<td>2</td>
<td>1/4</td>
</tr>
</tbody>
</table>

In these examples we have specified probability distributions of random variables by simply listing the possible values of the random variables and the corresponding probabilities. If the possible values are few in number, this is an easy way to specify a probability distribution. On the other hand, if there is a large number of possible values, a listing may become cumbersome. In the extreme case, where we have an infinite number of possible values (for example, all real numbers between zero and one), it is clearly impossible to make a listing. Fortunately, there are other ways to specify a probability distribution; in some instances we may be able to specify a mathematical function from which the probability of any value or interval of values can be computed, or we may be able to express the distribution in graphical form. These methods will be discussed in the following sections. First, it is necessary to distinguish between two types of random variables, discrete and continuous.

3.3 DISCRETE RANDOM VARIABLES

There are many situations where the random variable X can assume only a particular finite or countably infinite set of values; recall from Chapter 1 that this means that the possible values of X are finite in number or they are infinite in number but can be put in a one-to-one correspondence with the positive integers. For example, consider a simple experiment consisting of selecting an individual at random and determining the number of months
in the past year during which the individual made at least one trip of 50 miles or more from his or her home. Here, \( X \) can assume only the values 0, 1, 2, \ldots, 12. The numbers 3.68, -1.7, or \( x \) simply cannot be associated appropriately with any elementary event in the sample space. The random variable \( X \) can assume only a finite set of values. As another example, suppose that the simple experiment consists of tossing a fair coin repeatedly until a head appears. Let \( X \) be the number of tosses it takes to get the first head. The first head could appear on the first toss, in which case \( X \) would be 1; it might not appear until the one-hundredth toss, in which case \( X \) would be 100; in general, it might not appear until the \( K \)th toss, where \( K \) is any positive integer. Here we have a countably infinite set of values for \( X \).

If the random variable \( X \) can assume only a particular finite or countably infinite set of values, it is said to be a discrete random variable.

Not all random variables are discrete. Consider the simple experiment which consists of recording the temperature at a given place and time. Suppose that we can measure the temperature exactly; that is, our measuring device allows us to record the temperature to any number of decimal places. If \( X \) is the temperature reading, it is not possible for us to specify a finite or countably infinite set of values as the only possible values for \( X \). If we specify a finite set of values, it is a simple matter to find more values that can be assumed by \( X \), just by taking one of the finite set of values and adding more decimal places. For example, if one of the finite set of values is 75.965, we can determine values 75.9651, 75.9652, and so on, which are also possible values of \( X \). What is being pointed out here is that the possible values of \( X \) consist of the set of real numbers, a set which contains an infinite (and uncountable) number of values. In this case, it is said that \( X \) is a continuous random variable. We will give a more formal definition of a continuous random variable in Section 3.5 after we finish discussing discrete random variables in Sections 3.3 and 3.4.

Given a discrete random variable taking on only the \( K \) different values \( x_1, x_2, \ldots, x_K \), the following two conditions must be satisfied.

1. \( P(X = x_i) \geq 0, \quad i = 1, 2, \ldots, K. \) \quad (3.3.1*)

2. \( \sum_{i=1}^{K} P(X = x_i) = 1. \) \quad (3.3.2*)

That is, the probability of each value that \( X \) can assume must be non-negative and the sum of the probabilities over all of the different values must be 1.00. These two conditions follow directly from the axioms of probability presented in Section 2.4.

For example, let \( X \) represent the number of new houses that will be completed by a particular contractor during the coming month. Because of uncertainties concerning the weather and delivery of some appliances, the contractor is unsure as to how many houses will be completed, and the distribution of \( X \) is presented in Table 3.3.1. Note that the probabilities in Table 3.3.1 satisfy Equations (3.3.1) and (3.3.2); they are non-negative, and they sum to one.

<table>
<thead>
<tr>
<th>( x )</th>
<th>( P(X = x) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>.02</td>
</tr>
<tr>
<td>1</td>
<td>.07</td>
</tr>
<tr>
<td>2</td>
<td>.10</td>
</tr>
<tr>
<td>3</td>
<td>.14</td>
</tr>
<tr>
<td>4</td>
<td>.18</td>
</tr>
<tr>
<td>5</td>
<td>.17</td>
</tr>
<tr>
<td>6</td>
<td>.10</td>
</tr>
<tr>
<td>7</td>
<td>.10</td>
</tr>
<tr>
<td>8</td>
<td>.08</td>
</tr>
<tr>
<td>9</td>
<td>.04</td>
</tr>
<tr>
<td></td>
<td>1.00</td>
</tr>
</tbody>
</table>

Quite often we are interested in finding the probability that the obtained value of \( X \) will fall between two particular values, or in some interval. For instance, what is the probability that the contractor will finish at least 3 houses but no more than 5 houses? The various possible values of \( X \) are mutually exclusive, so we simply need to add the appropriate probabilities:

\[
P(3 \leq X \leq 5) = P(X = 3) + P(X = 4) + P(X = 5) = .14 + .18 + .17 = .49.
\]

Similarly,

\[
P(4 \leq X \leq 8) = P(X = 4) + P(X = 5) + P(X = 6) + P(X = 7) + P(X = 8) = .18 + .17 + .10 + .10 + .08 = .63.
\]

In general, the probability that \( X \) falls in the interval between any two numbers \( a \) and \( b \), inclusive, is found by the sum of probabilities for \( X \) over
all possible values between \( a \) and \( b \), inclusive:

\[
P(a \leq X \leq b) = \sum_{z=a}^{b} P(X = z)
\]

for all \( x \) such that \( a \leq x \leq b \). \( \sum_{z=a}^{b} P(X = z) \) \( (3.3.3^*) \)

By the same argument,

\[
P(X \geq a) = \sum_{z=a}^{\infty} P(X = z)
\]

for all \( x \) such that \( x \geq a \). \( (3.3.4) \)

Furthermore,

\[
P(X < a) = 1 - P(a \leq X) = \sum_{z=a}^{\infty} P(X = z).
\]

For the contractor example, the probability of completing 5 or more houses is

\[
P(X \geq 5) = P(X = 5) + P(X = 6) + \cdots + P(X = 9)
\]

\[
= .17 + .10 + .10 + .08 + .04 = .49,
\]

and the probability of completing less than 5 houses is

\[
P(X < 5) = P(X = 0) + P(X = 1) + \cdots + P(X = 4)
\]

\[
= .02 + .07 + .10 + .14 + .18 = .51.
\]

Alternatively, the latter probability could have been computed as follows once the former probability was computed:

\[
P(X < 5) = 1 - P(X \geq 5) = 1 - .49 = .51.
\]

Be careful to note that in probability calculations such as those given above, it makes a difference whether or not the endpoints of intervals are included. For instance, \( P(X < 5) \) was just computed to be .51, but \( P(X \leq 5) \) is .68:

\[
P(X \leq 5) = P(X = 0) + P(X = 1) + \cdots + P(X = 5)
\]

\[
= P(X < 5) + P(X = 5) = .51 + .17 = .68.
\]

Similarly, \( P(3 \leq X < 5) \) = .49, but \( P(3 < X < 5) = .18, \) \( P(3 \leq X < 5) = .32, \) and \( P(3 < X \leq 5) = .35. \) In computing probabilities for intervals of values of a discrete random variable, it is important to note whether neither, one, or both of the endpoints are included in the interval in each instance.

3.4 Probability Distributions of Discrete Random Variables

Thus, once the probability distribution of a discrete random variable is specified, probabilities of intervals can be determined. In this section we specified the probability distribution in the example by a listing. This is just one possible way to specify the distribution, and often it is not the most convenient way, as we shall see in the next section.

3.4 Probability Distributions of Discrete Random Variables

The probability distribution of a discrete random variable can be specified by a listing of the possible values of the random variable together with the corresponding probabilities. Such a listing is often represented by a graph. Suppose that the set of possible values of a random variable \( X \) can be denoted by \( \{x_1, x_2, \ldots, x_k\} \). Then the probability distribution can be represented by the set of pairs of values \( \{x_i, P(X = x_i)\} \), where \( i = 1, 2, \ldots, K \). We can graph such pairs in the Cartesian plane, as shown in Figure 3.4.1. The graph can be thought of as placing a mass of probability \( P(X = x_i) \) at the point \( x_i \) on the \( x \)-axis. As a result, such a graph is often referred to as the graph of a probability mass function (PMF), or simply a probability function. The advantage of such a graph over a listing is that it is generally easier to read and gives the reader a better notion of the nature of the probability distribution.

For example, consider the PMF of the random variable \( X \) in the contractor example in the previous section; this PMF is shown in Figure 3.4.2. As another example, consider the PMF of the random variable \( Y \) from Section 3.2: the number of heads in two flips of a fair coin. This PMF is shown in Figure 3.4.3.

![Figure 3.4.1 A Probability Mass Function](image-url)
In some cases it may be possible to express the probability distribution of a discrete random variable in the form of a mathematical function. The simplest example of this procedure concerns the case in which there are exactly $K$ possible values of the random variable $X$, the integers 1 through $K$, and each value has probability $1/K$. This distribution can be expressed as follows:

$$P(X = x) = \begin{cases} 1/K & \text{if } x = 1, 2, \ldots, K, \\ 0 & \text{elsewhere.} \end{cases}$$

It is important to specify that the PMF is equal to zero at all values other than 1, 2, $\ldots$, $K$. Since the PMF is a function on the real line, it must be defined at all points on the real line. For an example involving equal probabilities, let $X$ represent the number that comes up on a single toss of a fair die. The probability distribution of $X$, expressed graphically in Figure 3.4.4, can be expressed in functional form as follows:

$$P(X = x) = \begin{cases} 1/6 & \text{if } x = 1, 2, \ldots, 6, \\ 0 & \text{elsewhere.} \end{cases}$$

Consider once again the random variable $Y$, the number of heads in two flips of a fair coin. The PMF of $Y$ can be represented by

$$P(Y = y) = \begin{cases} \frac{\binom{2}{y}}{4} & \text{if } y = 0, 1, \text{ or } 2, \\ 0 & \text{elsewhere.} \end{cases}$$

There are four possible outcomes (hence the 4 in the denominator), and the number of ways of getting $y$ heads in 2 flips is $\binom{2}{y}$. By calculating the specific values of the PMF from the above formula, you can verify for yourself that the numbers are equal to those in the listing of the distribution of $Y$ given in Section 3.2.

In this section, we have seen how a discrete random variable can be represented in graphical or functional form. Whenever it is possible to use one or both of these two forms, they are usually preferable to a simple listing. In most cases they require less space than a listing and they are easier to understand. When we discuss the most important classes of discrete random variables in Chapter 4, we will rely almost exclusively on functional and graphical representations of probability distributions.

### 3.5 Continuous Random Variables

The previous sections dealt with discrete random variables, which take on only a finite or countably infinite number of values. There are many variables that, conceptually at least, could take on an uncountably infinite number of values. For example, consider the weight of an object or the
temperature at a particular time and place. The underlying sample space in these examples consists not just of integers or decimal expressions given to the nearest one-hundredth; conceptually, variables such as weight and temperature can take on any real number as a value. In practice, limitations of measurement devices imply that the observed weight or temperature is discrete. For instance, a balance may only be accurate to the nearest tenth of an ounce. It is still convenient, however, to use continuous probability models as approximate representations in such situations.

Formally, a continuous random variable may be defined as follows.

A random variable $X$ is said to be continuous if for every pair of values $u$ and $v$ such that $P(X \leq u) < P(X \leq v)$, it is true that $P(u < X < v) > 0$.

Notice that for a random variable to be continuous in an interval limited by the values $u$ and $v$, it is necessary that every nonzero interval in this range have a nonzero probability. For this to be true, there must be an infinite number of possible values that the random variable can assume.

As noted above, continuous random variables are idealizations. The fact that not even in the most precise work known can accuracy be obtained to any number of decimals limits the actual possibility of encountering a continuous random variable in practice. Why, then, do statisticians so often deal with these idealizations? The answer is that, mathematically, distributions of continuous random variables often are far more tractable than distributions of discrete random variables. The function rules for many continuous variables are relatively easy to state and to study using the full power of mathematical analysis. This is not generally true of discrete distributions. Moreover, continuous distributions are very good approximations to many discrete distributions. This fact makes it possible to organize statistical theory about a few such idealized distributions and to find methods that are good approximations to results for the more complicated discrete situations. Nevertheless, the student should realize that these continuous distributions are mathematical abstractions that happen to be quite useful; they do not necessarily describe real situations exactly.

### 3.6 PROBABILITY DISTRIBUTIONS OF CONTINUOUS RANDOM VARIABLES

In Section 3.3 it was possible for us to discuss discrete random variables in terms of probabilities such as $P(X = x)$, the probability that the random variable $X$ takes on some value exactly. However, for continuous random variables, the occurrence of any exact value of $X$ may be regarded as having zero probability. Instead of considering probabilities for single values, then, it is necessary to consider probabilities for intervals of values, such as $P(a \leq X \leq b)$. Thus, if $X$ represents the maximum official temperature in degrees Fahrenheit at Chicago on December 25, 1990, and if it is assumed that $X$ is continuous, then it is worthwhile to consider probabilities such as $P(20 \leq X \leq 40)$, $P(23 \leq X \leq 35)$, $P(29 \leq X \leq 31)$, $P(29.9 \leq X \leq 30.1)$, $P(29.99 \leq X \leq 30.01)$, and so on, but it is not worthwhile to consider $P(X = 30)$. The smaller the interval, the smaller the probability, and the probability that the high temperature is exactly 30 is, in effect, zero.

In the continuous case, then, we do not consider the probability that $X$ takes on some value $x$. Instead, we deal with the so-called probability density of $X$ at $x$, symbolized by

$$f(x) = \text{probability density of } X \text{ at } x.$$  

A fairly rough definition of a probability density can be given as follows. For any distribution, it is proper to speak of the probability associated with an interval. Imagine an interval with limits $a$ and $b$. Then the probability of that interval is

$$P(a \leq X \leq b) = P(X \leq b) - P(X < a).$$

Now suppose that we let the size of the interval be denoted by

$$\Delta x = b - a,$$

so that

$$b = a + \Delta x.$$  

Then the probability of the interval relative to $\Delta x$ is

$$\frac{P(a \leq X \leq b)}{\Delta x} = \frac{P(X \leq a + \Delta x) - P(X < a)}{\Delta x}.$$  

Now, suppose that we fix $a$, but allow $\Delta x$ to vary; in fact, let $\Delta x$ approach zero in size. What happens to the probability of this interval relative to the interval size? Both numbers will change, and the ratio will change, but this ratio will approach some limiting value as $\Delta x$ comes close to zero. That is, we can speak of the limit of $P(a \leq X \leq b)/\Delta x$ as $\Delta x$ approaches zero. This limit gives the probability density of the variable $X$ at value $a$. Loosely speaking, we can say that the probability density at $a$ is the rate of change in the probability of an interval with lower limit $a$, for minute changes in the size of the interval. This rate of change will depend on two things: the function rule assigning probabilities to intervals such as $(X \leq a)$ and the particular "region" of $X$ values we happen to be talking about. Rather than talk about probabilities of $X$ values per se, for continuous random variables it is
mathematically far more convenient to discuss the probability density and reserve the idea of probability only for the discussion of intervals of $X$ values. This need not trouble us very much, as we will generally be interested only in intervals of values in the first place.

The above development of a probability density function (PDF) becomes clearer when it is presented in graphical form. Consider a histogram (bar graph) in which the probability of an interval is represented by the area of the "bar" over the interval (Figure 3.6.1). This area is, of course, given by the product of the height of the bar and the width of the interval; thus, the height of the bar equals the probability of the interval divided by the width of the interval. If each of the intervals is divided into two subintervals, a new histogram could be drawn using the new subintervals, as in Figure 3.6.2. If this process of dividing intervals is repeated over and over, the width of each interval becomes smaller and the histogram begins to look more like a smooth curve. In the limit, it is a smooth curve, the probability density function (Figure 3.6.3). Since the height of a bar of the histogram equals the probability of the interval divided by the width of the interval, the limit of this ratio is the height of the density function, as indicated in the previous paragraph.

To illustrate this process, suppose that $X$ represents the weight (in pounds) of a person randomly chosen from the student body of a particular university. First, the weights might be grouped into 50-pound intervals, yielding a histogram comparable to Figure 3.6.1. Next, to obtain greater precision, the weights might be grouped into 25-pound intervals, yielding a histogram comparable to Figure 3.6.2. This process of reducing the length of each interval can be repeated, yielding histograms with 12.5-pound intervals, 6.25-pound intervals, and so on. As the interval length decreases (and correspondingly, the number of intervals increases), the histogram looks more and more like the smooth curve illustrated in Figure 3.6.3.

Intervals can always be given probabilities, regardless of whether the random variable is discrete or continuous. For continuous random variables the probability of an interval depends on the probability density associated with each value in the interval. The probability of the interval from $a$ to $b$ is

$$P(a \leq X \leq b) = \int_a^b f(x) \, dx. \quad \text{(3.6.1*)}$$

Since the definite integral of a function over some interval corresponds to the area under the curve of the function in that interval, we can say that the probability of an interval is the same as the area cut off by that interval under the curve for the probability densities when the random variable is continuous, and the total area is equal to 1.00. The expression $f(x) \, dx$ can be thought of as the area of a minute interval with midpoint $X = x$, somewhere between $a$ and $b$. When the number of such intervals approaches an infinite number and their size approaches zero, the sum of all these areas is the entire area cut off by the limits $a$ and $b$. Since there is an infinite number of such intervals, this sum is expressed by the definite integral sign

$$\int_a^b.$$
Readers who are uncomfortable with calculus, then, can think of a definite integral as the limiting form of a sum or as the area under a curve.

Note that this agrees with the definition of the probability of an interval for a discrete variable. In Section 3.3 it was stated that the probability that $X$ lies between two numbers $a$ and $b$, inclusive, is a sum of probabilities. Since a definite integral is analogous to a sum, the probability of an interval for a continuous random variable is analogous to the probability of an interval for a discrete random variable.

As noted above, when $X$ is continuous, $P(X = a) = 0$ and $P(X = b) = 0$. As a result, the probability of an interval for a continuous random variable is the same whether or not the endpoints of the interval are included. Obviously, inclusion or exclusion of the endpoints will not affect the area given by

$$\int_a^b f(x) \, dx.$$ 

In the discrete case, on the other hand, the points do make a difference. If $X$ is the number coming up on one throw of a fair die, for instance,

$$P(2 < X < 4) = P(X = 3) = 1/6,$$

while

$$P(2 \leq X \leq 4) = P(X = 2) + P(X = 3) + P(X = 4) = 3/6 = 1/2.$$ 

It should be strongly emphasized that the height of the density function does not represent probability; indeed, the probability of any single point is 0. It is the area under the curve that represents probability. To satisfy the basic axioms of probability theory, a probability density function $f(x)$ must satisfy two properties:

1. $f(x) \geq 0$ for all $x$. \hfill (3.6.2*)

2. $\int_{-\infty}^{\infty} f(x) \, dx = 1$. \hfill (3.6.3*)

That is, $f(x)$ cannot be negative, and the total area under the curve must equal 1. These conditions are analogous to the requirements imposed on a discrete PMF: $P(X = a) \geq 0$ and $\sum_a P(X = a) = 1$. In order to clarify the concept of a continuous random variable and to emphasize the analogy between discrete and continuous random variables, a graphical analysis is quite valuable.

A PDF for some hypothetical distribution is shown in Figure 3.6.4. The two points $a$ and $b$ marked off on the horizontal axis represent limits of some interval. The shaded portion between $a$ and $b$ shows

$$P(a \leq X \leq b) = \int_a^b f(x) \, dx,$$

the probability that $X$ takes on a value between the limits $a$ and $b$.

For an example of a specific PDF, consider $X$, the length of life (in hours) of a light bulb. Suppose that $X$ is uniformly distributed over the interval from 80 to 100. That is, the light bulb is sure to last at least 80 hours, but it will not last longer than 100 hours. Between 80 hours and 100 hours, the probability is uniformly distributed. This means that if we divide the interval from 80 to 100 into subintervals of equal width $i$, then these subintervals have the same probabilities and we can do this for any value of $i$. The probability density function of $X$ can be represented in functional form:

$$f(x) = \begin{cases} 
1/20 & \text{if } 80 \leq x \leq 100, \\
0 & \text{elsewhere}.
\end{cases}$$

Notice that $f(x)$ satisfies the two requirements of a PDF, since $f(x)$ is nonnegative and

$$\int_{-\infty}^{\infty} f(x) \, dx = \int_{-\infty}^{80} 0 \, dx + \int_{80}^{100} \frac{1}{20} \, dx + \int_{100}^{\infty} 0 \, dx$$

$$= \int_{80}^{100} \frac{1}{20} \, dx = \left[ \frac{x}{20} \right]_{80}^{100} = \frac{100}{20} - \frac{80}{20} = 1.$$
The graph of $f(x)$ is shown in Figure 3.6.5. Note that the total area under $f(x)$ is simply the area of a rectangle with length 20 and height 1/20. In this example, it is easy to calculate the probability of any interval. Suppose that $80 \leq a < b \leq 100$. Then $P(a \leq X \leq b)$ is the area of a rectangle with length $(b - a)$ and height 1/20; this area is simply $(b - a)/20$. Formally,

$$P(a \leq X \leq b) = \int_{a}^{b} \left( \frac{1}{20} \right) dx = \left[ \frac{x}{20} \right]_{a}^{b} = \frac{(b - a)}{20}.$$

It should be pointed out that it is possible for a random variable to be discrete over part of its range and continuous over another part of its range. Let $X$ represent the amount of rain (in inches) that will fall on a particular summer day in Paris, France. Suppose that $P(X = 0) = .6$ and that the remaining .4 of probability is distributed according to the density function

$$f(x) = \begin{cases} \frac{4 - x}{20} & \text{if } 0 < x < 4, \\ 0 & \text{elsewhere.} \end{cases}$$

The graph of the distribution of $X$ is a combination of a PMF and a PDF, as shown in Figure 3.6.6, and a random variable such as this is often called a mixed random variable. Mixed random variables arise in practice because some random variables are continuous except at a few points with positive probability. In the rain example, the quantity of rain is continuous provided that it rains, but there is a positive probability (.6) that it will not rain at all. Such random variables can be handled easily with a combination of techniques for working with discrete random variables and techniques for working with continuous random variables.

### 3.7 CUMULATIVE DISTRIBUTION FUNCTIONS

We have seen that the probability distribution of a random variable $X$ can be represented by a probability mass function (PMF) in the discrete case and by a probability density function (PDF) in the continuous case. There is a useful alternative method for specifying a probability distribution for both discrete and continuous random variables. This involves the concept of cumulative probabilities.

The probability that a random variable $X$ takes on a value at or below a given number $x$ is often written as

$$F(x) = P(X \leq x).$$ (3.7.1*)

The symbol $F(x)$ represents the function relating the various values of $X$ to the corresponding cumulative probabilities. This function is called the cumulative distribution function (CDF).

A CDF must satisfy certain mathematical properties, the most important of which are:

1. $0 \leq F(x) \leq 1$; (3.7.2*)
2. if $a < b$, $F(a) \leq F(b)$; (3.7.3*)
3. $F(\infty) = 1$ and $F(-\infty) = 0$. (3.7.4*)

The first property reflects the fact that $F(x)$ is a probability, $P(X \leq x)$, and properties 2 and 3 follow directly from the definition of a CDF.
Given the PMF of a discrete random variable or the PDF of a continuous random variable, the cumulative distribution function is uniquely determined. If \( X \) is discrete,

\[
F(x) = P(X \leq x) = \sum_{a \leq x} P(X = a). 
\] (3.7.5*)

If \( X \) is continuous,

\[
F(x) = P(X \leq x) = \int_{-\infty}^{x} f(z) \, dz. 
\] (3.7.8*)

Consider the example in which \( X \) represented the number of heads occurring in two flips of a fair coin. The PMF was

\[
P(X = z) = \begin{cases} 
1/4 & \text{for } z = 0 \text{ or } z = 2, \\
1/2 & \text{for } z = 1, \\
0 & \text{elsewhere.}
\end{cases}
\]

To compute the CDF, note that \( F(x) \) will be 1 for all values of \( x \) greater than or equal to 2, since \( X \) is certain to be less than or equal to all of these values. Also, \( F(x) \) will be zero for all negative values of \( x \), since \( X \) cannot possibly be smaller than a negative number (it only takes on the values 0, 1, and 2). Similarly, we can determine the values of the CDF between 0 and 2, with the result shown in Figure 3.7.1.

![Figure 3.7.1 CDF for Number of Heads in Two Tosses of a Fair Coin](image)

The CDF, like the PMF, is a function defined on the real line, and in the example it is assigned exactly one value for each real number.

For an example of the CDF of a continuous random variable, recall the situation in which \( X \) was uniformly distributed over the interval from 80 to 100, where \( X \) represented the life (in hours) of a light bulb. In this case, \( F(x) \) represents the probability that the life of the bulb is less than or equal to \( x \) hours. Clearly, \( F(x) \) is equal to 1 for values of \( x \) greater than 100, since the light bulb cannot last longer than 100 hours according to our assumptions in Section 3.6. Similarly, \( F(x) \) is equal to 0 for values of \( x \) smaller than 80, for we stated that the bulb is sure to last at least 80 hours. For values between 80 and 100,

\[
F(x) = \int_{-\infty}^{x} f(x) \, dx = \int_{80}^{x} 0 \, dx + \int_{80}^{x} \left( \frac{1}{20} \right) \, dx
\]

\[
= \frac{z}{20} - \frac{80}{20} = \frac{x - 80}{20}.
\]

Thus, the CDF for this example, which is illustrated in Figure 3.7.2, can
be stated in functional form as follows:

\[ F(z) = \begin{cases} 
1 & \text{if } z > 100, \\
\frac{(z - 80)}{20} & \text{if } 80 \leq z \leq 100, \\
0 & \text{if } z < 80.
\end{cases} \]

Observe that the CDF in this example is a continuous function rather than a step function. This is true for all continuous random variables. In fact, the definition of a continuous random variable can be given in terms of the CDF.

A random variable \( X \) is said to be continuous if its cumulative distribution function \( F(z) \) is a continuous function. On the other hand, the CDF of a discrete random variable is always a step function rather than a continuous function.

If we know the CDF of a continuous random variable, it is possible to find the density function \( f(z) \). The relationship is given by

\[ f(z) = \frac{d}{dz} F(z) = F'(z). \tag{3.7.7} \]

In general, the derivative of the CDF is equal to the PDF. This follows from a fundamental theorem of calculus which requires that \( F(z) \) be a continuous function. The reader not familiar with this theorem can safely take the above statement for granted and not worry about its derivation. A graphical interpretation of Equation (3.7.7) is that the value of the density function at a point \( z \) is equal to the slope of the CDF at that point. The derivative of a function at a point \( z \) corresponds to the slope of the line tangent to the function at that point.

In the previous example, it follows that

\[ f(z) = \frac{d}{dz} F(z) = \begin{cases} 
\frac{d}{dz} (1) = 0 & \text{if } z > 100, \\
\frac{d}{dz} \left( \frac{(z - 80)}{20} \right) = \frac{1}{20} & \text{if } 80 \leq z \leq 100, \\
\frac{d}{dz} (0) = 0 & \text{if } z < 80.
\end{cases} \]

This is identical to \( f(z) \) as given in Section 3.6. In geometric terms, the value of the density function at any point \( z \) is equal to the slope of the line tangent to the CDF at the point \( z \). In the example, \( F(z) \) is a straight line for \( 80 \leq z \leq 100 \). At any point \( z \), the tangent line is identical to the line representing \( F(z) \). Then, since this line has a given slope which is obviously constant for all \( z \) between 80 and 100, the density function must simply be constant for all \( z \) between 80 and 100. Compare the graphs of the PDF and CDF (Figures 3.6.5 and 3.7.2).

In general (that is, for distributions other than uniform distributions), density functions are not constant (graphically, they are not horizontal lines). This means that, in general, cumulative distribution functions are not straight lines, as in the above example. Cumulative distribution functions for continuous random variables often have more or less the characteristic S-shape shown in Figure 3.7.3.

The probability that a random variable takes on any value between limits \( a \) and \( b \) (including \( b \) but not \( a \)) can be found from

\[ P(a < X \leq b) = F(b) - F(a). \tag{3.7.8} \]

This is seen easily if it is recalled that \( F(b) \) is the probability that \( X \) takes on the value \( b \) or below, \( F(a) \) is the probability that \( X \) takes on the value \( a \) or below; their difference must be \( P(a < X \leq b) \). For continuous random variables, this probability is unchanged if the sign is changed to \( \leq \) or vice versa, since the probability that \( X \) exactly equals any particular number such as \( a \) or \( b \) is zero. However, for discrete variables, \( < \) and \( \leq \) signs may lead to different probabilities.

The symbol \( F(z) \) can be used to represent the cumulative probability that \( X \) is less than or equal to \( z \) for either a continuous or a discrete random variable. All random variables have cumulative distribution functions. However, remember that only continuous variables have density functions and that the value of a density function at a particular value of a random variable is not a probability. To avoid possible confusion, the density notation \( f(z) \) will be reserved for continuous variables in all of the following, and \( P(z) \) will be used both for discrete variables and for intervals of values of continuous variables, as in \( P(a < X \leq b) \). Occasionally in mathematical statistics the terms "distribution function" or

![Figure 3.7.3 One Type of CDF for a Continuous Random Variable](image-url)
3.8 SUMMARY MEASURES OF PROBABILITY DISTRIBUTIONS

As we have seen, the probability distribution of a random variable, whether discrete or continuous, can be represented in several alternative ways. If a graphical analysis is desired, a distribution can be represented by a probability mass function (PMF) in the discrete case and by a probability density function (PDF) in the continuous case; or it can be represented by a cumulative distribution function (CDF) in either case. Sometimes it is also possible to represent a probability distribution in functional notation, provided there is some relatively simple mathematical function that can describe the PMF, PDF, or CDF. In the discrete case, a probability distribution can be represented by a simple listing of all possible values of the random variable and the probability corresponding to each value (provided that the random variable takes on only a finite number of values). This last form of representation could also be used as an approximation in the continuous case if the possible values of the random variable are grouped into a finite set of mutually exclusive and exhaustive intervals on the real line.

Thus, there are several ways to represent the probability distribution of a random variable. Each of these methods has some disadvantages. If there are numerous possible values for a discrete random variable, a listing or a graph might be inconvenient; if the mathematical function corresponding to a distribution is quite complicated or if no function can be found to represent the distribution, the use of functional notation might be infeasible. Even if these disadvantages do not apply in a particular case, it may be quite time-consuming to represent the entire probability distribution of interest. Consequently, it is often easier and more efficient to look only at certain characteristics of a distribution than to attempt to specify the distribution as a whole. Such characteristics can be summarized into one or more numerical values, conveying some, though not all, of the information in the entire distribution.

Two such general characteristics of any distribution are its measures of central tendency (or location) and dispersion (or variability). Indices of central tendency are ways of describing the "typical" or the "average" value of the random variable. Indices of dispersion, on the other hand, describe the "spread" or the extent of the difference among the possible values of the random variable.

In the following sections we will discuss several measures of central tendency and dispersion. In many cases, it is possible to adequately describe a probability distribution with a few such measures. It should be remembered, however, that these measures serve only to summarize some important features of the probability distribution; in general, they do not completely describe the entire distribution. In some situations in statistical analysis the entire distribution is of genuine interest, and the summary measures are not sufficient for the problem being attacked. On the other hand, as we will see in Chapter 4, if we know that the distribution is one of a number of frequently encountered types of distributions, it may be possible to get a complete picture of the distribution from a few summary measures.

One of the most common and most useful summary measures of a probability distribution is a measure known as the expectation of a random variable. The concept of expectation plays an important role not only as a useful summary measure, but also as a central concept in the theory of probability and statistics.

3.9 THE EXPECTATION OF A RANDOM VARIABLE

A very prominent place in theoretical statistics is occupied by the concept of the mathematical expectation of a random variable $X$.

If the distribution of $X$ is discrete, then the expectation (or expected value) of $X$ is defined to be

$$E(X) = \sum_z zP(X = z) = \sum_z zp(x),$$

(3.9.1*)

where the sum is taken over all of the different values that the variable $X$ can assume. For a continuous random variable $X$, the expectation is defined to be

$$E(X) = \int_{-\infty}^{\infty} xp(x) \ dx.$$

(3.9.2*)

Note the similarity between the discrete and continuous cases. In the discrete case, $E(X)$ is a sum of products of values of $X$ and their probabilities; in the continuous case, $E(X)$ is an integral (which is, after all, a sum in the limiting case) of products of values of $X$ and the corresponding