APPENDIX A

REVIEW OF STATISTICS: PROBABILITY AND PROBABILITY DISTRIBUTIONS

The purpose of this and the following three appendixes is to review some fundamental statistical concepts that are needed to understand *Essentials of Econometrics.* These four appendixes will serve as a refresher course for those students who have had a basic course in statistics and will provide a unified framework for following discussions of the material in the main parts of this book for those whose knowledge of statistics has become somewhat rusty. Students who have had very little statistics should supplement these four appendixes with a good statistics book. (Some references are given at the end of this appendix.) Note that the discussion in Appendixes A through D is nonrigorous and is by no means a substitute for a basic course in statistics. It is simply an overview that is intended as a bridge to econometrics.

A.1 SOME NOTATION

In this appendix we come across several mathematical expressions that often can be expressed more conveniently in shorthand forms.

The Summation Notation

The Greek capital letter Σ (sigma) is used to indicate summation or addition. Thus,

$$\sum_{i=1}^{i=n} X_i = X_1 + X_2 + \cdots + X_n$$

where i is the index of summation and the expression on the left-hand side is the shorthand for "take the sum of the variable X from the first value ($i = 1$) to the

nth value ($i = n$)"; X_i stands for the ith value of the X variable.

$$\sum_{i=1}^{i=n} X_i \text{ (or } \sum_{i=1}^{n} X_i)$$

is often abbreviated as

$$\sum X_i$$

where the upper and lower limits of the sum are known or can be easily determined or also expressed as

$$\sum_X X$$

which simply means take the sum of all the relevant values of X. We will use all these notations interchangeably.

Properties of the Summation Operator

Some important properties of Σ are as follows:

1. Where k is a constant

$$\sum_{i=1}^{n} k = nk$$

That is, a constant summed n times is n times that constant. Thus,

$$\sum_{i=1}^{4} 3 = 4 \times 3 = 12$$

In this example $n = 4$ and $k = 3$.

2. Where k is a constant

$$\sum k X_i = k \sum X_i$$

That is, a constant can be pulled out of the summation sign and put in front of it.

3. $$\sum (X_i + Y_i) = \sum X_i + \sum Y_i$$

That is, the summation of the sum of two variables is the sum of their individual summations.

4. $$\sum (a + bX_i) = na + b \sum X_i$$

where a and b are constants and where use is made of properties 1, 2, and 3.

We will make extensive use of the summation notation in the remainder of this appendix and in the main parts of the book.

We now discuss several important concepts from probability theory.

A.2 EXPERIMENT, SAMPLE SPACE, SAMPLE POINT, AND EVENTS

Experiment

The first important concept is that of a **statistical** or **random experiment.** In statistics this term generally refers to any process of observation or measurement that has more than one possible outcome and for which there is uncertainty about which outcome will actually materialize.

Example A.1.

Tossing a coin, throwing a pair of dice, and drawing a card from a deck of cards are all experiments. Although it may seem completely different, the sales of Coca-Cola in a future quarter can also be considered an experiment since we don't know the outcome. Also, there are several possible values that could occur. It is implicitly assumed that in performing these experiments certain conditions are fulfilled, for example, that the coin or the dice are fair (not loaded). The outcomes of such experiments could be a head or a tail if a coin is tossed or any one of the numbers 1, 2, 3, 4, 5, or 6 if a die is thrown. The Coca-Cola sales figure could be any one of a seemingly infinite number of possibilities, depending on many factors. Note that the outcomes are unknown before the experiment is performed. The objectives of such experiments may be to establish a law (e.g., How many heads are you likely to obtain in a toss of, say, 1000 coins?) or to test the proposition that the coin is loaded (e.g., Would you regard a coin as being loaded if you obtained 70 heads in 100 tosses of a coin?).

Sample Space or Population

The set of all possible outcomes of an experiment is called the **population** or **sample space.** The concept of sample space was first introduced by von Mises, an Austrian mathematician and engineer, in 1931.

Example A.2.

Consider the experiment of tossing two fair coins. Let H denote a head and T a tail. Then we have these outcomes: HH, HT, TH, TT, where HH means a head on the first toss and a head on the second toss, HT means a head on the first toss and a tail on the second toss, etc.

In this example the totality of the outcomes, or sample space or population, is 4—no other outcomes are logically possible. (Don't worry about the coin landing on its edge.)

Example A.3.

The New York Mets are scheduled to play a doubleheader. Let O_1 indicate the outcome that they win both games, O_2 that they win the first game but lose the second, O_3 that they lose the first game but win the second, and O_4 that they lose both games. Here the sample space consists of four outcomes: O_1, O_2, O_3, and O_4.

Sample Point

Each member, or outcome, of the sample space or population is called a **sample point.** In Example A.2 each outcome, *HH, HT, TH,* and *TT*, is a sample point. In Example A.3 each outcome, O_1, O_2, O_3, and O_4, is a sample point.

Events

An **event** is a particular collection of outcomes and is thus a *subset* of the sample space.

Example A.4.

Let event *A* be the occurrence of one head and one tail in the coin-tossing experiment. From Example A.2 we see that only outcomes *HT* and *TH* belong to event *A.* (*Note: HT* and *TH* are a subset of the sample space *HH, HT, TH,* and *TT*). Let *B* be the event that two heads occur in a toss of two coins. Then, obviously, only the outcome *HH* belongs to event *B.* (Again, note that *HH* is a subset of the sample space *HH, HT, TH,* and *TT.*)

Events are said to be **mutually exclusive** if the occurrence of one event prevents the occurrence of another event at the same time. In Example A.3, if O_1 occurs, that is, the Mets win both the games, it rules out the occurrence of any of the other three outcomes. Two events are said to be **equally likely** if we are confident that one event is as likely to occur as the other event. In a single toss of a coin a head is as likely to appear as a tail. Events are said to be **collectively exhaustive** if they exhaust all possible outcomes of an experiment. In our coin-tossing example, since *HH, HT, TH,* and *TT* are the only possible outcomes, they are (collectively) exhaustive events. Likewise, in the Mets example, O_1, O_2, O_3, and O_4 are the only possible outcomes, barring, of course, rain or natural calamities such as the earthquake that occurred during the 1989 World Series in San Francisco.

Venn Diagrams

A simple graphic device, called the **Venn diagram,** originally introduced by Venn in his book, *Symbolic Logic,* published in 1881, can be used to depict sample point, sample space, events, and related concepts, as shown in Figure A-1. In this figure each *rectangle represents the sample space S* and the two circles represent two events *A* and *B.* If there are more events, we can draw more circles to represent all those events. The various subfigures in this diagram depict various situations.

Figure A-1(*a*) shows outcomes that belong to *A* and the outcomes that do not belong to *A*, which are denoted by the symbol A', which is called the *complement* of A.

Figure A-1(*b*) shows the *union* (i.e., sum) of *A* and *B*, that is the event whose outcomes belong to set *A* or set *B*. Using set theory notation, it is often denoted as $A \cup B$ (read as *A* union *B*), which is the equivalent of $A + B$.

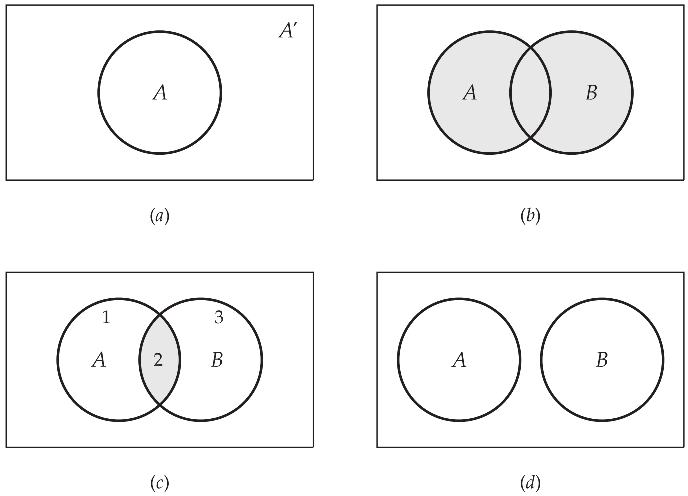

FIGURE A-1 Venn diagram

The shaded area in Figure A-1(*c*) denotes events whose outcomes belong to both set A and set B, which is represented as $A \cap B$ (read as A intersects B), and is the equivalent of the product AB.

Finally, Figure A-1(*d*) shows that the two events are *mutually exclusive* because they have no outcomes in common. In set notation, this means $A \cap B = 0$ (or that $AB = 0$).

A.3 RANDOM VARIABLES

Although the outcome(s) of an experiment can be described verbally, such as a head or a tail, or the ace of spades, it would be much simpler if the results of all experiments could be described numerically, that is, in terms of numbers. As we will see later, for statistical purposes such representation is very useful.

Example A.5.

Reconsider Example A.2. Instead of describing the outcomes of the experiment by *HH*, *HT*, *TH*, and *TT*, consider the "variable" number of heads in a toss of two coins. We have the following situation:

First coin	Second coin	Number of heads
T	*T*	0
T	*H*	1
H	*T*	1
H	*H*	2

We call the variable "number of heads" a **stochastic** or **random variable** (r.v., for short). *More generally, a variable whose (numerical) value is determined by the outcome of an experiment is called a random variable*. In the preceding example the r.v., number of heads, takes three different values, 0, 1, or 2, depending on whether no heads, one head, or two heads were obtained in a toss of two coins. In the Mets example the r.v., the number of wins, likewise takes three different values, 0, 1, or 2.

By convention, random variables are denoted by capital letters, X, Y, Z, etc., and the values taken by these variables are often denoted by small letters. Thus, if X is an r.v., x denotes a particular value taken by X.

An r.v. may be either **discrete** or **continuous**. A **discrete random variable** takes on only a finite number of values or countably infinite number of values (i.e., as many values as there are whole numbers). Thus, the number of heads in a toss of two coins can take on only three values, 0, 1, or 2. Hence, it is a discrete r.v. Similarly, the number of wins in a doubleheader is also a discrete r.v. since it can take only three values, 0, 1, or 2 wins. A **continuous random variable,** on the other hand, is an r.v. that can take on any value in some interval of values. Thus, the height of an individual is a continuous variable—in the range of, say, 60 to 72 inches it can take any value, depending on the precision of measurement. Similar factors such as weight, rainfall, or temperature also can be regarded as continuous random variables.

A.4 PROBABILITY

Having defined experiment, sample space, sample points, events, and random variables, we now consider the important concept of probability. First, we define the concept of probability of an event and then extend it to random variables.

Probability of an Event: The Classical or A Priori Definition

If an experiment can result in n *mutually exclusive* and *equally likely* outcomes, and if m of these outcomes are favorable to event A, then $P(A)$, the **probability** that A occurs, is the ratio m/n. That is,

$$P(A) = \frac{\text{number of outcomes favorable to A}}{\text{total number of outocmes}} \qquad \textbf{(A.1)}$$

Note the two features of this definition: The outcomes must be *mutually exclusive* (that is, they cannot occur at the same time), and each outcome must have an *equal chance of occurring* (for example, in a throw of a die, any one of the six numbers has an equal chance of appearing).

Example A.6.

In a throw of a die numbered 1 through 6, there are six possible outcomes: 1, 2, 3, 4, 5, or 6. These outcomes are mutually exclusive since, in a single throw of the die, two or more numbers cannot turn up simultaneously. These six outcomes are also equally likely. Hence, by the classical definition, the probability that any of these six numbers will show up is 1/6—there are six total outcomes and each outcome has an equal chance of occurring. Here $n = 6$ and $m = 1$.

Similarly, the probability of obtaining a head in a single toss of a coin is 1/2 since there are two possible outcomes, *H* and *T*, and each has an equal chance of coming up. Likewise, in a deck of 52 cards, the probability of drawing any single card is 1/52. (Why?) The probability of drawing a spade, however, is 13/52. (Why?)

The preceding examples show why the classical definition is called an **a priori definition** since the probabilities are derived from purely *deductive* reasoning, or simply by the structure of the event. One doesn't have to throw a coin to state that the probability of obtaining a head or a tail is 1/2, since logically, these are the only possible outcomes.

But the classical definition has some deficiencies. What happens if the outcomes of an experiment are not finite or are not equally likely? What, for example, is the probability that the gross domestic product (GDP) next year will be a certain amount or what is the probability that there will be a recession next year? The classical definition is not equipped to answer these questions. A more widely used definition that can handle such cases is the relative frequency definition of probability, which we will now discuss.

Relative Frequency or Empirical Definition of Probability

To introduce this concept of probability, consider the following example.

Example A.7.

Table A-1 gives the distribution of marks received by 200 students on a microeconomics examination. Table A-1 is an example of a **frequency distribution** showing how the r.v. marks in the present example are distributed. The numbers in column 3 of the table are called **absolute frequencies,** that is, the number of occurrences of a given event. The numbers in column 4 are called **relative frequencies,** that is, the absolute frequencies divided by the total number of occurrences (200 in the present case). Thus, the absolute frequency of marks between 70 and 79 is 45 but the relative frequency is 0.225, which is 45 divided by 200.

Can we treat the relative frequencies as probabilities? Intuitively, it seems reasonable to consider the relative frequencies as probabilities provided the number of observations on which the relative frequencies are based is reasonably large. This is the essence of the *empirical,* or *relative frequency, definition of probability.*

TABLE A-1 THE DISTRIBUTION OF MARKS RECEIVED BY 200 STUDENTS ON A MICROECONOMICS EXAMINATION

Marks (1)	Midpoint of interval (2)	Absolute frequency (3)	Relative frequency (4) = (3)/200
0–9	5	0	0
10–19	15	0	0
20–29	25	0	0
30–39	35	10	0.050
40–49	45	20	0.100
50–59	55	35	0.175
60–69	65	50	0.250
70–79	75	45	0.225
80–89	85	30	0.150
90–99	95	10	0.050
	Total	200	1.000

More formally, if in n trials (or observations), m of them are favorable to event A, then $P(A)$, the probability of event A, is simply the ratio m/n (i.e., the relative frequency) *provided n, the number of trials, is sufficiently large* (*technically, infinite*).[1] Notice that, unlike the classical definition, we do not have to insist that the outcome be mutually exclusive and equally likely.

In short, if the number of trials is sufficiently large, we can treat the relative frequencies as fairly good measures of true probabilities. In Table A-1 we can, therefore, treat the relative frequencies given in column 4 as probabilities.[2]

Properties of Probabilities The probability of an event as defined earlier has the following important properties:

1. The probability of an event always lies between 0 and 1. Thus, the probability of event A, $P(A)$, satisfies this relationship:

$$0 \leq P(A) \leq 1 \qquad \textbf{(A.2)}$$

$P(A) = 0$ means event A will not occur, whereas $P(A) = 1$ means event A will occur with certainty. Typically, the probability will lie somewhere between these numbers, as in the case of the probabilities shown in Table A-1.

[1]What constitutes a large or small number depends on the context of the problem. Sometimes a number as small as 30 can be regarded as reasonably large. In presidential elections in the United States, election polls based on a sample of about 800 people are fairly accurate in predicting the final outcome, although the actual number of voters runs into the millions.

[2]There is yet another definition of probability, called *subjective probability*, which is the foundation of *Bayesian statistics*, that is a rival to classical statistics. Under the subjective or "degrees of belief" definition of probability we can ask questions such as: What is the probability that Iraq will have a democratic government? What is the probability that the Chicago Cubs will win the World Series next year? Or what is the probability that there will be a stock market crash in the year 2010?

2. If *A*, *B*, *C*, . . . are *mutually exclusive events*, the probability that any one of them will occur is equal to the sum of the probabilities of their individual occurrences. Symbolically,

$$P(A + B + C + \cdots) = P(A) + P(B) + P(C) + \cdots \qquad \textbf{(A.3)}$$

where the expression on the left-hand side of the equality means the probability of *A* or *B* or *C*, etc.[3]

3. If *A*, *B*, *C*, . . . are *mutually exclusive* and *collectively exhaustive* sets of events, the sum of the probabilities of their individual occurrences is 1. Symbolically,

$$P(A + B + C + \cdots) = P(A) + P(B) + P(C) + \cdots = 1 \qquad \textbf{(A.4)}$$

Example A.8.

In Example A.6 we saw that the probability of obtaining any of the six numbers on a die is 1/6 since there are six equally likely outcomes, and each one of them has an equal chance of turning up. Since the numbers 1, 2, 3, 4, 5, and 6 form an exhaustive set of events, $P(1+2+3+4+5+6)=1$, where 1, 2, 3, . . . means the probability of number 1 or number 2 or number 3, etc. And since 1, 2, . . . 6 are mutually exclusive events in that two numbers cannot occur simultaneously in a throw of a single die, $P(1+2+3+4+5+6)=P(1)+P(2)+\cdots+P(6)=1/6+1/6+1/6+1/6+1/6+1/6=1$.

In passing, note the following rules of probability that will come in handy later on.

1. If *A*, *B*, *C*, . . . are any events, they are said to be *statistically independent* if the probability of their occurring together is equal to the product of their individual probabilities. Symbolically,

$$P(ABC\cdots) = P(A)\,P(B)\,P(C)\cdots \qquad \textbf{(A.5)}$$

where $P(ABC\cdots)$ means the probability of events $ABC\cdots$ occurring simultaneously or jointly.[4] Hence, it is called a *joint probability.* In relation to the joint probability $P(ABC\cdots)$, $P(A)$, $P(B)$, etc. are called *unconditional, marginal,* or *individual probabilities,* for reasons that will become clear in Section A.6.

Example A.9.

Suppose we throw two coins simultaneously. What is the probability of obtaining a head on the first coin and a head on the second coin? Let *A* denote the event of obtaining a head on the first coin and *B* on the second coin. We

[3] In set theory notation, this would be written as $P(A \bigcup B \bigcup C. \ldots)$.
[4] In set theory notation, this would be written as $P(A \bigcap B \bigcap C. \ldots)$.

therefore want to find the probability $P(AB)$. Common sense would suggest that the probability of obtaining a head on the first coin is independent of the probability of obtaining a head on the second coin. Hence, $P(AB) = P(A)P(B) = (1/2)\ (1/2) = 1/4$ since the probability of obtaining a head (or a tail) is 1/2.

2. If events $A, B, C, \ldots$ are *not* mutually exclusive, Eq. (A.3) needs to be modified. Thus, if events A and B are not mutually exclusive, we have

$$P(A + B) = P(A) + P(B) - P(AB) \tag{A.6}$$

   where $P(AB)$ is the joint probability that the two events occur together (see Fig. A-1[*c*]).[5] Of course, if A and B are mutually exclusive, $P(AB) = 0$ (Why?), and we are back to Eq. (A.3). Equation (A.6) can be easily generalized to more than two events.
3. For every event A there is an event A', called the **complement** of A, with these properties:
   a. $P(A + A') = 1$, and
   b. $P(AA') = 0$

These properties can be easily verified from Fig. A-1(*a*).

Example A.10.

A card is drawn from a deck of cards. What is the probability that it will be either a heart or a queen? Clearly a heart and a queen are not mutually exclusive events, for one of the four queens is a heart. Hence,

$$\begin{aligned} P\text{ (a heart or a queen)} &= P(\text{heart}) + P(\text{queen}) - P(\text{heart and queen}) \\ &= 13/52 + 4/52 - 1/52 \\ &= 4/13 \end{aligned}$$

Let A and B be two events. Let us suppose we want to find out the probability that the event A occurs knowing that the event B has already occurred. This probability, called the **conditional probability of *A***, *conditional on event B occurring,* and denoted by the symbol $P(A \mid B)$, is computed from the formula

$$P(A \mid B) = \frac{P(AB)}{P(B)}; \quad P(B) > 0 \tag{A.7}$$

That is, the conditional probability of A, given B, is equal to the ratio of their joint probability to the marginal probability of B. In like manner,

$$P(B \mid A) = \frac{P(AB)}{P(A)}; \quad P(A) > 0 \tag{A.8}$$

[5]To avoid the shaded area in Fig. A-1(*c*) being counted twice, we have to subtract $P(AB)$ on the right-hand side of this equation.

To visualize Eq. (A.7), we can resort to a Venn diagram, as shown in Figure A-1(*c*). As you can see from this figure, regions 2 and 3 represent event B and regions 1 and 2 represent event A. Because region 2 is common to both events, and since B has occurred, if we divide the area of region 2 by the sum of the areas of regions 2 and 3, we will get the (conditional) probability that event A has occurred, knowing that B has occurred. Simply put, the conditional probability is the fraction of the time that event A occurs when event B has occurred.

Example A.11.

In an introductory accounting class there are 500 students, of which 300 are males and 200 are females. Of these, 100 males and 60 females plan to major in accounting. A student is selected at random from this class, and it is found that this student plans to be an accounting major. What is the probability that the student is a male?

Let A denote the event that the student is a male and B that the student is an accounting major. Therefore, we want to find out $P(A \mid B)$. From the formula of conditional probability just given, this probability can be obtained as

$$\begin{aligned} P(A \mid B) &= \frac{P(AB)}{P(B)} \\ &= \frac{100/500}{160/500} \\ &= 0.625 \end{aligned}$$

From the data given previously, it can be readily seen that $P(A) = 300/500 = 0.6$; that is, the unconditional probability of selecting a male student is 0.6, which is different from the preceding probability 0.625.

This example brings out an important point, namely, *conditional and unconditional probabilities in general are different*. However, if the two events are independent, then we can see from Eq. (A.7) that

$$P(A \mid B) = \frac{P(AB)}{P(B)} = \frac{P(A)P(B)}{P(B)} = P(A) \quad \textbf{(A.9)}$$

Note that $P(AB) = P(A)P(B)$ when the two events are independent, as noted earlier. In this case, the conditional probability of A given B is the same as the unconditional probability of A. In this case it does not matter if B occurs or not.

An interesting application of conditional probability is contained in the famous **Bayes' Theorem,** which was originally propounded by Thomas Bayes, a nonconformist minister in Turnbridge Wells, England (1701–1761). This theorem, published after Bayes' death, led to the so-called Bayesian School of Statistics, a rival to the school of classical statistics, which still predominates statistics teaching in most universities in the world. The knowledge that an event B has occurred can be used to revise or update the probability that an event A has occurred. This is the essence of Bayes' Theorem.

To explain this theorem, let A and B be two events, each with a positive probability. Bayes' Theorem then states that:[6]

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B \mid A)P(A) + P(B \mid A')P(A')} \tag{A.10}$$

where A', called the *complement of A*, means the event A does not occur.

In words, Bayes' Theorem shows how conditional probabilities of the form $P(B \mid A)$ may be combined with initial probability of A (i.e., P[A]) to obtain the final probability $P(A \mid B)$. Notice how the roles of conditioning event (B) and outcome event (A) have been interchanged. The following example will show how Bayes' Theorem works.

Example A.12. Bayes' Theorem

Suppose a woman has two coins in her handbag. One is a fair coin and one is two-headed. She takes a coin at random from her handbag and tosses it. Suppose a head shows up. What is the probability that the coin she tossed was two-headed?

Let A be the event that the coin is two-headed and A' the event that the coin is fair. The probability of selecting either of these coins is $P(A) = P(A') = 1/2$. Let B be the event that a head turns up. If the coin has two heads, B is certain to occur. Hence, $(B \mid A) = 1$. But if the coin is fair, $P(B \mid A') = 0.5$. Therefore by Bayes' Theorem we obtain

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)} = \frac{(1)(0.5)}{0.75} = \frac{2}{3} \approx 0.66$$

Notice the particular feature of this theorem. Before the coin was tossed, the probability of selecting a regular or two-headed coin was the same, namely, 0.5. But knowing that a head was actually observed, the probability that the coin selected was two-headed is revised upwards to about 0.66.

In Bayesian language, $P(A)$ is called the **prior probability** (i.e., before the fact or evidence) and $P(A \mid B)$ is called the revised or **posterior probability** (after the fact or evidence). The knowledge that B has occurred leads us to reassess or revise the (prior) probability assigned to A.

[6]If the sample space is partitioned into A (event A occurs) and A' (event A does not occur), then for any event B, it is true that $P(B) = P(BA) + P(BA')$; that is, the probability that B occurs is the sum of the common outcomes between B and each partition of A. This result can be generalized if A is partitioned into several segments. This can be seen easily from a Venn diagram.

You might see this finding as somewhat puzzling. Intuitively you will think that this probability should be 1/2. But look at the problem this way. There are three ways in which heads can come up, and in two of these cases, the hidden face will also be heads.

Notice another interesting feature of the theorem. In classical statistics we assume that the coin is fair when we toss it. In Bayesian statistics we question that premise or hypothesis.

Probability of Random Variables

Just as we assigned probabilities to sample outcomes or events of a sample space, we can assign probabilities to random variables, for as we saw, random variables are simply numerical representations of the outcomes of the sample space, as shown in Example A.5. In this textbook we are largely concerned with random variables such as GDP, money supply, prices, and wages, and we should know how to assign probabilities to random variables. Technically, we need to study the *probability distributions* of random variables, a topic we will now discuss.

A.5 RANDOM VARIABLES AND THEIR PROBABILITY DISTRIBUTIONS

By probability distribution of a random variable we mean the possible values taken by that variable and the probabilities of occurrence of those values. To understand this clearly, we first consider the probability distribution of a discrete r.v., and then we consider the probability distribution of a continuous r.v., for there are differences between the two.

Probability Distribution of a Discrete Random Variable

As noted before, a discrete r.v. takes only a finite (or countably infinite) number of values.

Let X be an r.v. with distinct values of $x_1, x_2, \ldots$. The function f defined by

$$\begin{aligned} f(X = x_i) &= P(X = x_i) \quad i = 1, 2, \ldots, \\ &= 0 \text{ if } x \neq x_i \end{aligned} \tag{A.11}$$

is called the **probability mass function** (**PMF**) or simply the **probability function** (**PF**), where $P(X = x_i)$ means that the probability that the discrete r.v. X takes the numerical value of x_i.[7] Note these properties of the PMF given in Eq. (A.11):

$$0 \leq f(x_i) \leq 1 \tag{A.12}$$

$$\sum_x f(x_i) = 1 \tag{A.13}$$

[7]The values taken by a discrete r.v. are often call *mass points* and $f(X = x_i)$ denotes the mass associated with the mass point x_i.

where the summation extends over all the values of X. Notice the similarities of these properties with the properties of probabilities discussed earlier.

To see what this means, consider the following example.

Example A.13.

Let the r.v. X represent the number of heads obtained in two tosses of a coin. Now consider the following table:

Number of heads X	PF $f(X)$
0	1/4
1	1/2
2	1/4
Sum	1.00

In this example the r.v. X (the number of heads) takes three different values—$X = 0$, 1, or 2. The probability that X takes a value of zero (i.e., no heads are obtained in a toss of two coins) is 1/4, for of the four possible outcomes of throwing two coins (i.e., the sample space), only 1 is favorable to the outcome *TT*. Likewise, of the four possible outcomes, only one is favorable to the outcome of two heads; hence, its probability is also 1/4. On the other hand, two outcomes, *HT* and *TH*, are favorable to the outcome of one head; hence, its probability is $2/4 = 1/2$. Notice that in assigning these probabilities we have used the classical definition of probability.

Geometrically, the PMF of this example is as shown in Figure A-2.

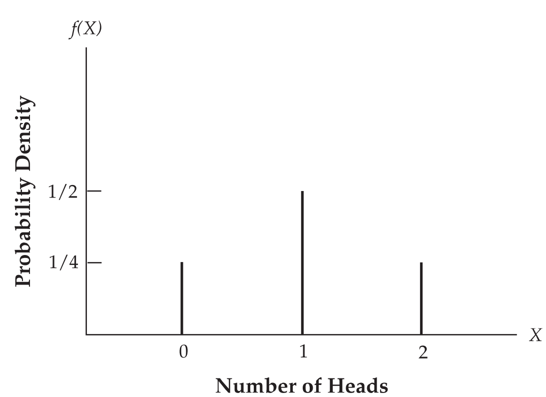

FIGURE A-2 The probability mass function (PMF) of the number of heads in two tosses of a coin (Example A.13)

Probability Distribution of a Continuous Random Variable

The probability distribution of a continuous random variable is conceptually similar to the probability distribution of a discrete r.v. except that we now measure the probability of such an r.v. over a certain range or interval. Instead of calling it a PMF, we call it a **probability density function (PDF).** An example will make the distinction between the two clear.

Let X represent the continuous r.v. height, measured in inches, and suppose we want to find out the probability that the height of an individual lies in the interval, say, 60 to 68 inches. Further, suppose that the r.v. height has the PDF as shown in Figure A-3.

The probability that the height of an individual lies in the interval 60 to 68 inches is given by the shaded area lying between 60 and 68 marked on the curve in Figure A-3. (How this probability is actually measured is shown in Appendix C.) In passing, note that since a continuous r.v. can take an uncountably infinite number of values, the probability that such an r.v. takes a particular numerical value (e.g., 63.00 inches) is always zero; *the probability for a continuous r.v. is always measured over an interval,* say, between 62.5 and 63.5 inches.

More formally, for a continuous r.v. X the probability density function (PDF), $f(X)$, is such that

$$P(x_1 < X < x_2) = \int_{x_1}^{x_2} f(x)\,dx \qquad \textbf{(A.14)}$$

for all $x_1 < x_2$, where $\int$ is the integral symbol of calculus, which is the equivalent of the summation symbol ($\sum$) used for taking the sum of the values of a discrete random variable, and where dx stands for a small interval of x values.

A PDF has the following properties:

1. The total area under the curve $f(x)$ given in Eq. (A.14) is 1,
2. $P(x_1 < X < x_2)$ is the area under the curve between x_1 and x_2, where $x_2 > x_1$,

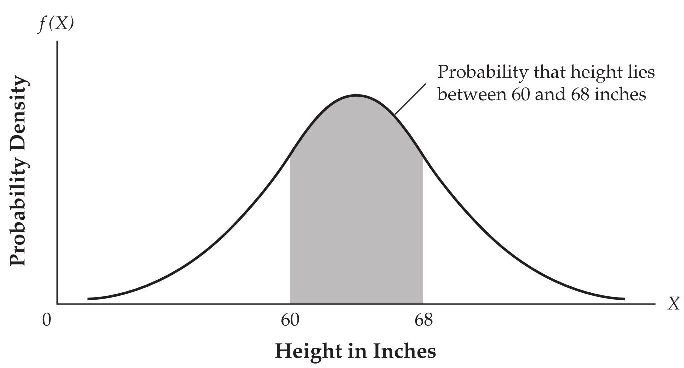

FIGURE A-3 The PDF of a continuous random variable

3. since the probability that a continuous r.v. takes a particular value is zero because probabilities for such a variable are measured over an area or interval, the left-hand side of Eq. (A.14) can be expressed in any of these forms:

$$P(x_1 \le X \le x_2) = P(x_1 < X \le x_2) = P(x_1 \le X < x_2) = P(x_1 < X < x_2)$$

An important example of a continuous PDF is the **normal probability density function,** which is discussed in Appendix C. We will use this function extensively in the later chapters.

Example A.14.

The PDF of a (continuous) r.v. X is given by

$$f(x) = \frac{x^2}{9} \qquad 0 \le x \le 3$$

What is the probability that $0 < x < 1$?

To get the answer, we have to evaluate the integral of the preceding PDF over the stated range. That is,

$$\int_0^1 \frac{x^2}{9}\,dx = \frac{1}{9}\int_0^1 x^2\,dx = \frac{1}{9}\left[\frac{x^3}{3}\right]_0^1 = \frac{1}{27}$$

That is, the probability that x lies between 0 and 1 is 1/27. (*Note:* The integral of x^2 is $(x^3/3)$, which can be checked easily by taking the derivative of the latter, which is $\frac{d}{dx}\left(\frac{x^3}{3}\right) = x^2$.)

Incidentally, if you evaluate the given PDF over the entire range of 0 to 3, you will see that $\int_0^3 \frac{x^2}{9}\,dx = 1$, as it should be. (Why?) Of course, $f(x) \ge 0$ for all x values in the range 0 to 3.

Cumulative Distribution Function (CDF)

Associated with the PMF or PDF of an r.v. X is its **cumulative distribution function (CDF),** $F(X)$, which is defined as follows:

$$F(X) = P(X \le x) \qquad \textbf{(A.15)}$$

where $P(X \le x)$ means the probability that the r.v. X takes a value less than or equal to x, where x is given. (Of course, for a continuous r.v., the probability that such an r.v. takes the exact value of x is zero.) Thus, $P(X \le 2)$ means the probability that the r.v. X takes a value less than or equal to 2. The following properties of CDF should be noted:

1. $F(-\infty) = 0$ and $F(\infty) = 1$, where $F(-\infty)$ and $F(\infty)$ are the limits of $F(x)$ as x tends to $-\infty$ and ∞, respectively.
2. $F(x)$ is a nondecreasing function such that if $x_2 > x_1$ then $F(x_2) \ge F(x_1)$.

3. $P(X \geq k) = 1 - F(k)$; that is, the probability that X assumes a value equal to or greater than k is 1 minus the probability that X takes a value below k.
4. $P(x_1 \leq X \leq x_2) = F(x_2) - F(x_1)$; that is, the probability that X lies between values x_1 and x_2 is the probability that X lies below x_2 *minus* the probability that X lies below x_1, a property that will help us in computing probabilities in practice.

Example A.15.

What are the PDF and CDF of the r.v. number of heads obtained in four tosses of a fair coin? These functions are as follows:

	PDF		CDF	
Number of heads (X)	Value of X	PDF $f(X)$	Value of X	CDF $f(X)$
0	$0 \leq X < 1$	1/16	$X \leq 0$	1/16
1	$1 \leq X < 2$	4/16	$X \leq 1$	5/16
2	$2 \leq X < 3$	6/16	$X \leq 2$	11/16
3	$3 \leq X < 4$	4/16	$X \leq 3$	15/16
4	$4 \leq X$	1/16	$X \leq 4$	1

As this example and the definition of CDF suggest, a CDF is merely an "accumulation" or simply the sum of the PDF for the values of X less than or equal to a given x. That is,

$$F(X) = \sum^{x} f(x) \tag{A.16}$$

where $\sum^{x} f(X)$ means the sum of the PDF for values of X less than or equal to the specified x, as shown in the preceding table. Thus, in this example the probability that X takes the value of less than 2 (heads) is 5/16, but the probability that it takes a value of less than 3 is 11/16. Of course, the probability that it takes a value of 4 or less than four heads is 1. (Why?)

Geometrically, the CDF of Example A.15 looks like Figure A-4. Since we are dealing with a discrete r.v. in Example A.15, its CDF is a discontinuous function, known as a **step function.** If we were dealing with the CDF of a continuous r.v., its CDF would be a continuous curve, as shown in Figure A-5.[8]

Example A.16.

Referring to Example A.15, what is the probability that X lies between 2 and 3? Here we want to find $F(X=3) - F(X=2)$. From the table given in Example A.15

[8]If x is continuous with a PDF of $f(x)$, then $F(x) = \int_{-\infty}^{x} f(x)\,dx$ and $f(x) = F'(x)$, where $F'(x)$ is the derivative of $F(x) = \frac{d}{dx} F(X)$.

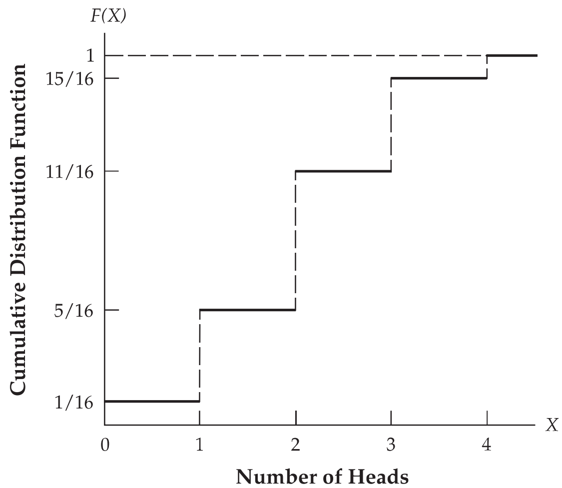

FIGURE A-4 The cumulative distribution function (CDF) of a discrete random variable (Example A.15)

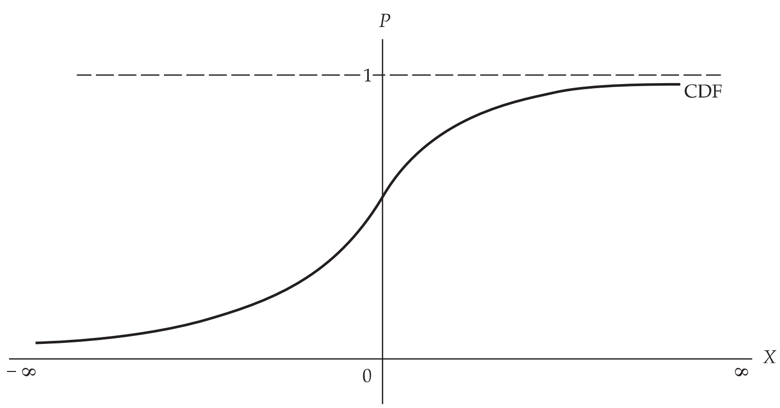

FIGURE A-5 The CDF of a continuous random variable

we find that $F(X \leq 3) = 15/16$ and that $F(X \leq 2) = 11/16$. Therefore, the probability that X lies between 2 and 3 is $4/16 = 1/4$.

A.6 MULTIVARIATE PROBABILITY DENSITY FUNCTIONS

So far we have been concerned with *single variable,* or *univariate,* probability distribution functions. Thus, the PMFs of Examples A.5 and A.13 are univariate PDFs, for we considered single random variables, such as the number of heads in a toss of two coins or the number of heads in a toss of four coins. However, we do not need to be so restrictive, for the outcomes of an experiment could be

TABLE A-2 THE FREQUENCY DISTRIBUTION OF TWO RANDOM VARIABLES: NUMBER OF PCS SOLD (X) AND NUMBER OF PRINTERS SOLD (Y)

	Number of PCs Sold (X)					
	0	1	2	3	4	Total
Number of Printers Sold (Y)						
0	6	6	4	4	2	22
1	4	10	12	4	2	32
2	2	4	20	10	10	46
3	2	2	10	20	20	54
4	2	2	2	10	30	46
Total	16	24	48	48	64	200

described by more than one r.v., in which case we would like to find their probability distributions. Such probability distributions are called **multivariate probability distributions.** The simplest of these is the *bivariate,* or a two-variable PMF or PDF. Let us illustrate with an example.[9]

Example A.17.

A retail computer store sells personal computers (PCs) as well as printers. The number of computers and printers sold varies from day to day, but the store manager obtained the sales history over the past 200 days in the form of Table A-2.

In this example we have two random variables, X (number of PCs sold) and Y (number of printers sold). The table shows that of the 200 days, there were 30 days when the store sold 4 PCs and 4 printers, but on 2 days, although it sold 4 PCs, it sold no printers. Other entries in the table are to be interpreted similarly. Table A-2 provides an example of what is known as a **joint frequency distribution;** it gives the number of times a combination of values is taken by the two variables. Thus, in our example the number of times 4 PCs and 4 printers were sold together is 30. Such a number is called an *absolute frequency*. All the numbers shown in Table A-2 are thus absolute frequencies.

Dividing the absolute frequencies given in the preceding table by 200, we obtain the relative frequencies, which are shown in Table A-3.

Since our sample is reasonably large, we can treat these (joint) relative frequencies as measures of joint probabilities, as per the frequency interpretation of probabilities. Thus, the probability that the store sells three PCs and three printers is 0.10, or about 10 percent. Other entries in the table are to be interpreted similarly.

[9]This example is adapted from Ron C. Mittelhammer, *Mathematical Statistics for Economics and Business*, Springer, New York, 1995, p. 107.

TABLE A-3 THE BIVARIATE PROBABILITY DISTRIBUTION OF NUMBER OF PCS SOLD (X) AND NUMBER OF PRINTERS SOLD (Y)

	Number of PCs Sold (x)					
	0	1	2	3	4	Total $f(Y)$
Number of Printers Sold (y)						
0	0.03	0.03	0.02	0.02	0.01	0.11
1	0.02	0.05	0.06	0.02	0.01	0.16
2	0.01	0.02	0.10	0.05	0.05	0.23
3	0.01	0.01	0.05	0.10	0.10	0.27
4	0.01	0.01	0.01	0.05	0.15	0.23
Total $f(X)$	0.08	0.12	0.24	0.24	0.32	1.00

Because the two variables are discrete, Table A-3 provides an example of what is known as **bivariate** or **joint probability mass function (PMF).**

More formally, let X and Y be two discrete random variables. Then the function

$$\begin{aligned} f(X, Y) &= P(X = x \text{ and } Y = y) \\ &= 0 \text{ when } X \neq x \text{ and } Y \neq y \end{aligned} \tag{A.17}$$

is known as the joint PMF. This gives the joint probability that X takes the value of x and Y takes the value of y simultaneously, where x and y are some specific values of the two variables. Notice the following properties of the joint PMF:

1. $f(X, Y) \geq 0$ for all pairs of X and Y. This is so because all probabilities are nonnegative.
2. $\sum_x \sum_y f(X, Y) = 1$. This follows from the fact that the sum of the probabilities associated with all joint outcomes must equal 1.

Note that we have used the double summation sign because we are now dealing with two variables. If we were to deal with a three-variable joint PMF, we would be using the triple summation sign, and so on.

The joint probability of two continuous random variables (i.e., joint PDF) can be defined analogously, although the mathematical expressions are somewhat involved and are given by way of exercises for the benefit of the more mathematically inclined reader.

The discussion of joint PMF or joint PDF leads to a discussion of some related concepts, which we will now discuss.

Marginal Probability Functions

We have studied the univariate PFs, such as $f(X)$ or $f(Y)$, and the bivariate, or joint, PF $f(X, Y)$. Is there any relationship between the two? Yes, there is.

In relation to $f(X, Y)$, $f(X)$ and $f(Y)$ are called **univariate, unconditional, individual,** or **marginal PMFs or PDFs.** More technically, the probability that X

TABLE A-4 MARGINAL PROBABILITY DISTRIBUTIONS OF *X* (NUMBER OF PCS SOLD) AND *Y* (NUMBER OF PRINTERS SOLD)

Value of *X*	$f(X)$	Value of *Y*	$f(Y)$
0	0.08	0	0.11
1	0.12	1	0.16
2	0.24	2	0.23
3	0.24	3	0.27
4	0.32	4	0.23
Sum	1.00		1.00

assumes a given value (e.g., 2) *regardless of the values taken by* Y is called the marginal probability of X, and the distribution of these probabilities is called the marginal PMF of X. How do we compute these marginal PMFs or PDFs? That is easy. In Table A-3 we see from the column totals that the probability that X takes the value of 1 regardless of the values taken by Y is 0.12; the probability that it takes the value of 2 regardless of Y's value is 0.24, and so on. Therefore, the marginal PMF of X is as shown in Table A-4. Table A-4 also shows the marginal PMF of Y, which can be derived similarly. Note that the sum of each of the PFs, $f(X)$ and $f(Y)$, is 1. (Why?)

You will easily note that to obtain the marginal probabilities of X, we sum the joint probabilities corresponding to the given value of X regardless of the values taken by Y. *That is, we sum down the columns*. Likewise, to obtain the marginal probabilities of Y, we sum the joint probabilities corresponding to the given value of Y regardless of the values taken by X. *That is, we sum across the rows*. Once such marginal probabilities are computed, finding the marginal PMFs is straightforward, as we just showed. More formally, if $f(X, Y)$ is the joint PMF of random variables X and Y, then the marginal PFs of X and Y are obtained as follows:

$$f(X) = \sum_{y} f(X, Y) \text{ for all } X \qquad \textbf{(A.18)}$$

and

$$f(Y) = \sum_{x} f(X, Y) \text{ for all } Y \qquad \textbf{(A.19)}$$

If the two variables are continuous, we will replace the summation symbol with the integral symbol. For example, if $f(X, Y)$ represents a joint PDF, to find the marginal PDF of X, we will integrate the joint PDF with respect to Y values, and to find to marginal PDF of Y, we will integrate it with respect to the X values (see Problem A.20).

Conditional Probability Functions

Continuing with Example A.17, let us now suppose we want to find the probability that 4 printers were sold, knowing that 4 PCs were sold. In other words, what is the probability that $Y = 4$, conditional upon the fact that $X = 4$? This is known as **conditional probability** (recall our earlier discussion of conditional probability of an event). This probability can be obtained from the **conditional probability mass function** defined as

$$f(Y \mid X) = P(Y = y \mid X = x) \tag{A.20}$$

where $f(Y \mid X)$ stands for the *conditional PMF* of Y; it gives the probability that Y takes on the value of y (number of printers sold) conditional on the knowledge that X has assumed the value of x (number of PCs sold). Similarly,

$$f(X \mid Y) = P(X = x \mid Y = y) \tag{A.21}$$

gives the conditional PMF of X.

Note that the preceding two conditional PFs are for two discrete random variables, Y and X. Hence, they may be called *discrete conditional PMFs.* Conditional PDFs for continuous random variables can be defined analogously, although the mathematical formulas are slightly involved (see Problem A.20).

One simple method of computing the conditional PF is as follows:

$$\begin{aligned} f(Y \mid X) &= \frac{f(X, Y)}{f(X)} \\ &= \frac{\text{joint probability of } X \text{ and } Y}{\text{marginal probability of } X} \end{aligned} \tag{A.22}$$

$$\begin{aligned} f(X \mid Y) &= \frac{f(X, Y)}{f(Y)} \\ &= \frac{\text{joint probability of } X \text{ and } Y}{\text{marginal probability of } Y} \end{aligned} \tag{A.23}$$

In words, *the conditional PMF of one variable, given the value of the other variable, is simply the ratio of the joint probability of the two variables divided by the marginal or unconditional PF of the other (i.e., the conditioning) variable.* (Compare this with the conditional probability of an event A, given that event B has happened, i.e., $P[A \mid B]$.)

Returning to our example, we want to find out $f(Y = 4 \mid X = 4)$, which is

$$\begin{aligned} f(Y = 4 \mid X = 4) &= \frac{f(Y = 4 \text{ and } X = 4)}{f(X = 4)} \\ &= \frac{0.15}{0.32} \text{ (from Table A-3)} \\ &= \approx 0.47 \end{aligned} \tag{A.24}$$

From Table A-3 we observe that the marginal, or unconditional, probability that Y takes a value of 4 is 0.23, but knowing that 4 PCs were sold, the probability that 4 printers will be sold increases to ≈ 0.47. Notice how the knowledge about the other event, the conditioning event, changes our assessment of the probabilities. This is in the spirit of Bayesian statistics.

In regression analysis, as we show in Chapter 2, we are interested in studying the behavior of one variable, say, stock prices, conditional upon the knowledge of another variable, say, interest rates. Or, we may be interested in studying the female fertility rate, knowing a woman's level of education. Therefore, the knowledge of conditional PMFs or PDFs is very important for the development of regression analysis.

Statistical Independence

Another concept that is vital for the study of regression analysis is the concept of **independent random variables,** which is related to the concept of independence of events discussed earlier. We explain this with an example.

Example A.18.

A bag contains three balls numbered 1, 2, and 3, respectively. Two balls are drawn at random, with replacement, from the bag (i.e., every time a ball is drawn it is put back before another is drawn). Let the variable X denote the number on the first ball drawn and Y the number on the second ball. Table A-5 gives the joint as well as the marginal PMFs of the two variables.

Now consider the probabilities $f(X=1, Y=1)$, $f(X=1)$, and $f(Y=1)$. As Table A-5 shows, these probabilities are 1/9, 1/3, and 1/3, respectively. Now the first of these is a joint probability, whereas the last two are marginal probabilities. However, the joint probability in this case is equal to the product of the two marginal probabilities. When this happens, we say that the two variables are **statistically independent,** that is, the value taken by one variable has no effect on the value taken by the other variable. More formally, *two variables X and Y are statistically independent if and only if their joint PMF or PDF can be expressed as the product of their individual, or marginal, PMFs or PDFs for all combinations of X and Y values.* Symbolically,

$$f(X, Y) = f(X)\, f(Y) \tag{A.25}$$

TABLE A-5 STATISTICAL INDEPENDENCE OF TWO RANDOM VARIABLES

		X			
		1	2	3	$f(Y)$
	1	1/9	1/9	1/9	3/9
Y	2	1/9	1/9	1/9	3/9
	3	1/9	1/9	1/9	3/9
$f(X)$		3/9	3/9	3/9	1

You can easily verify that for any other combination of X and Y values given in Table A-5 the joint PF is the product of the respective marginal PFs; that is, the two variables are statistically independent. *Bear in mind that Equation (A.25) must be true for all combinations of X and Y values.*

Example A.19.

Are the number of PCs sold and the number of printers sold in Example A.17 independent random variables? To determine this, let us apply the definition of independence given in Eq. (A.25). Let $X = 3$ (3 computers sold) and $Y = 2$ (2 printers sold). From Table A-3 we see that $f(X = 3, Y = 2) = 0.05$; $f(X = 3) = 0.24$ and $f(Y = 2) = 0.23$. Obviously, in this case $0.05 \neq (0.24)(0.23)$. Hence, in the present case, the number of PCs sold and the number of printers sold are not independent variables. This may not be surprising, especially for those who buy a computer for the first time. Sometimes a store may offer a special discount if a customer buys both.

A.7 SUMMARY AND CONCLUSIONS

In econometrics, mathematical statistics plays a key role. And the foundation of mathematical statistics is based on probability theory. Therefore, without some background in probability, we will not be able to appreciate the theory behind several econometric techniques that we discuss in the main chapters of the book.

That is why in this appendix we introduced some fundamental concepts of probability, such as sample space, sample points, events, random variables, and probability distributions of random variables. Since in econometrics we deal with relationships between (economic) variables, we have to consider the joint probability distributions of such variables. This led to a discussion of concepts, such as joint events and joint variables and their probability distributions, conditional probability distributions, unconditional probability distributions, and statistical independence. An interesting application of the conditional probability distribution is Bayes' Theorem, which shows how experimental knowledge can be used to revise probabilities.

All the concepts discussed in this appendix are illustrated with several examples. You may want to refer to this appendix when these concepts are addressed in the econometric techniques explained in the main text of the book.

KEY TERMS AND CONCEPTS

The key terms and concepts introduced in this appendix are

Statistical or random experiment
Population or sample space
  a) sample point
  b) events—mutually exclusive; equally likely; collectively exhaustive

Venn diagram
Stochastic or random variable
  a) discrete random variable
  b) continuous random variable
Probability and features of probability

A priori definition (classical definition) of probability
Frequency distribution; absolute frequency; relative frequency
Complement
Conditional probability of *A*
Bayes' Theorem
 Prior probability
 Posterior probability
Probability mass function (PMF) or probability function (PF), and probability density function (PDF)
Normal probability density function
PMFs of discrete and PDFs of continuous random variable
Cumulative distribution function (CDF)
Step function
Multivariate PDF
 a) bivariate or joint PMF and PDF; joint frequency distribution
 b) marginal (or univariate, unconditional, or individual) PMF and PDF
 c) conditional probability; conditional PMF
Statistical independence; independent random variables

REFERENCES

As noted in the introduction to this appendix, the discussion presented here is, of necessity, brief and intuitive and not meant as a substitute for a basic course in statistics. Therefore, the reader is advised to keep on hand one or two of the many good books on statistics. The following short list of such references is only suggestive.

1. Newbold, Paul. *Statistics for Business and Economics* (latest ed.). Prentice-Hall, Englewood Cliffs, N.J. This is a comprehensive nonmathematical introduction to statistics with lots of worked-out examples.
2. Hoel, Paul G. *Introduction to Mathematical Statistics* (latest ed.). Wiley, New York. This book provides a fairly simple introduction to various aspects of mathematical statistics.
3. Mood, Alexander M., Franklin A. Graybill, and Duane C. Boes. *Introduction to the Theory of Statistics.* McGraw-Hill, New York, 1974. This is a standard but mathematically advanced book.
4. Mosteller, F., R. Rourke, and G. Thomas. *Probability with Statistical Applications* (latest ed.). Addison-Wesley, Reading, Mass.
5. DeGroot, Morris H. *Probability and Statistics* (3rd ed.). Addison-Wesley, Reading, Mass.
6. Ron C. Mittelhammer. *Mathematical Statistics for Economics and Business.* Springer, New York, 1999.

QUESTIONS

A.1. What is the meaning of
 a. sample space
 b. sample point
 c. events
 d. mutually exclusive events
 e. PMF and PDF
 f. joint PDF
 g. marginal PDF
 h. conditional PDF
 i. statistical independence

A.2. A and B are two events. Can they be mutually exclusive and independent simultaneously?

A.3. For every event A, there is the *complement of A*, denoted by A', which means that A does not occur. Are the following statements true or false?

a. $P(A + A')$ or $P(A \cup A') = 1$

b. $P(AA')$ or $P(A \cap A') = 0$

A.4. Four economists have predicted the following rates of growth of GDP (%) for the next quarter:

E_1 = below 2%, E_2 = 2 or greater than 2% but below 4%, E_3 = 4 or greater than 4% but less than 6%, and E_4 = 6% or more.

Let A_i be the actual rate of % GDP growth rate according to the same four classifications as E_i (e.g., A_1 = GDP growth rate of less than 2%).

a. Are the events E_1 through E_4 mutually exclusive? Are they collectively exhaustive?

b. What is the meaning of the events (1) E_1A_2 (or $E_1 \cap A_2$), (2) $E_3 + A_3$ (or $E_3 \cup A_3$), (3) $E_i + A_i$ (or $E_i \cup A_i$) where $I = 1$ through 4, and (4) E_iA_j (or $E_i \cap A_j$) where $i > j$?

A.5. What is the difference between a PDF and a PMF?

A.6. What is the difference between the CDFs of continuous and discrete random variables?

A.7. By the conditional probability formula, we have

1. $P(A \mid B) = \dfrac{P(AB)}{P(B)}$ and

2. $P(B \mid A) = \dfrac{P(AB)}{P(A)} \rightarrow P(AB) = P(B \mid A)P(A)$

where $\rightarrow$ means "implies." If you substitute for $P(AB)$ from the right-hand side of (2) into the numerator of (1), what do you get? How do you interpret this result?

PROBLEMS

A.8. What do the following stand for?

a. $\sum_{i=1}^{4} x^{i-1}$

b. $\sum_{i=2}^{6} ay_i$, a is a constant

c. $\sum_{i=1}^{2} (2x_i + 3y_i)$

d. $\sum_{i=1}^{3} \sum_{j=1}^{2} x_i y_j$

e. $\sum_{i=1}^{4} (i + 4)$

f. $\sum_{i=1}^{3} 3^i$

g. $\sum_{i=1}^{10} 2$

h. $\sum_{i=1}^{3} (4x^2 - 3)$

A.9. Express the following in the Σ notation:

a. $x_1 + x_2 + x_3 + x_4 + x_5$

b. $x_1 + 2x_2 + 3x_3 + 4x_4 + 5x_5$

c. $\left(x_1^2 + y_1^2\right) + \left(x_2^2 + y_2^2\right) + \cdots + \left(x_k^2 + y_k^2\right)$

A.10. It can be shown that the sum of the first n positive numbers is:

$$\sum_{k=1}^{n} k = \frac{n(n+1)}{2}$$

Use the preceding formula to evaluate

a. $\sum_{k=1}^{500} k$ **b.** $\sum_{k=10}^{100} k$ **c.** $\sum_{k=10}^{100} 3k$

A.11. It can be proved that the sum of squares of the first n positive numbers is:

$$\sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6}$$

Using this formula, obtain

a. $\sum_{k=1}^{10} k^2$ **b.** $\sum_{k=10}^{20} k^2$ **c.** $\sum_{k=11}^{19} k^2$ **d.** $\sum_{k=1}^{10} 4k^2$

A.12. An r.v. X has the following PDF:

TABLE A-6

X	$f(X)$
0	b
1	$2b$
2	$3b$
3	$4b$
4	$5b$

a. What is the value of b? Why?
b. Find the $P(X \leq 2)$; prob$(X \leq 3)$; prob$(2 \leq X \leq 3)$.

A.13. The following table gives the joint probability distribution, $f(X, Y)$, of two random variables X and Y.

TABLE A-7

	X		
Y	1	2	3
1	0.03	0.06	0.06
2	0.02	0.04	0.04
3	0.09	0.18	0.18
4	0.06	0.12	0.12

a. Find the marginal (i.e., unconditional) distributions of X and Y, namely, $f(X)$ and $f(Y)$.
b. Find the conditional PDF, $f(X \mid Y)$ and $f(Y \mid X)$.

A.14. Of 100 people, 50 are Democrats, 40 are Republicans, and 10 are Independents. The percentages of the people in these three categories who read *The Wall Street Journal* are known to be 30, 60, and 40 percent, respectively. If one of these people is observed reading the *Journal*, what is the probability that he or she is a Republican?

A.15. Let A denote the event that a person lives in New York City. Let $P(A) = 0.5$. Let B denote the event that the person does not live in New York City but works in the city. Let $P(B) = 0.4$. What is the probability that the person either lives in the city or does not live in the city but works there?

A.16. Based on a random sample of 500 married women, the following table gives the joint PMF of their work status in relation to the presence or absence of children in the household.[10]

TABLE A-8 CHILDREN AND WORK STATUS OF WOMEN IN THE UNITED STATES

	Works outside home	Does not work outside home	Total
Has children	0.2	0.3	0.5
Does not have children	0.4	0.1	0.5
Total	0.6	0.4	1.0

a. Are children and working outside of the home mutually exclusive?

b. Are working outside of the home and presence of children independent events?

A.17. The following table gives the joint probability of X and Y, where X represents a person's poverty status (below or above the poverty line as defined by the U.S. government), and Y represents the person's race (white, blacks only, and all Hispanic).

TABLE A-9 POVERTY IN THE UNITED STATES, 2007

	X	
Y	Below poverty line	Above poverty line
White	0.0546	0.6153
Black	0.0315	0.0969
Hispanic	0.0337	0.1228
Asian	0.0046	0.0406

Source: These data are derived from the U.S. Census Bureau, Current Population Reports, *Poverty in the United States: 2007*, September 2008, Table 1. Although the poverty line varies by several socioeconomic characteristics, for a family of four in 2007, the dividing line was about $21,302. Families below this income level can be classified as poor.

a. Compute $f(X \mid Y = white)$; $f(X \mid Y = black)$, $f(X \mid Y = Hispanic)$, and $f(X \mid Y = Asian)$, where X represents below the poverty line. What general conclusions can you draw from these computations?

b. Are race and poverty status independent variables? How do you know?

[10]Adapted from Barry R. Chiswick and Stephen J. Chiswick, *Statistics and Econometrics: A Problem Solving Approach*, University Park Press, Baltimore, 1975.

A.18. The following table gives joint probabilities relating cell phone usage to stopping properly at intersections.

a. Compute the probability of failing to stop at an intersection, given the driver was on the cell phone.
b. Compute the probability of failing to stop at an intersection, given the driver was not using a cell phone.
c. Compute the probability of stopping properly at an intersection, given the driver was on the cell phone.
d. Are cell phone usage and failing to stop at intersections independent of each other? Why or why not?

TABLE A-10

	Failed to stop at intersection	Stopped at intersection
On cell phone	0.047	0.016
Not using cell phone	0.201	0.736

Source: David L. Strayer and Frank A. Drews, "Multitasking in the Automobile," Chapter 9.

***A.19.** The PDF of a continuous random variable X is as follows:

$$\begin{aligned} f(X) &= c(4x - 2x^2) \quad 0 \le x \le 2 \\ &= 0 \textit{ otherwise} \end{aligned}$$

a. For this to be a proper density function, what must be the value of c?
b. Find $P(1 < x < 2)$
c. Find $P(x > 2)$

***A.20.** Consider the following joint PDF:

$$\begin{aligned} f(x, y) &= \frac{12}{5}x(2 - x - y);\ 0 < x < 1;\ 0 < y < 1 \\ &= 0 \textit{ otherwise} \end{aligned}$$

a. Find $P(x > 0.5)$ and $P(y < 0.5)$
b. What is the conditional density of X given that $Y = y$, where $0 < y < 1$?

*Optional.