Contents

Chapter 6

Probability Distributions

6.1 Introduction

Statistics and statistical reasoning is based on probability and probabilistic reasoning. Statistical conclusions or proofs are never absolutely certain, but are always qualified, in that a certain probability or likelihood is attached to the conclusion or proof. For example, in opinion polls, summary data might be reported as being accurate to within plus or minus 4 percentage points, nineteen times in twenty, or with probability 0.95. This type of reasoning stands in contrast to the method of reasoning carried out in formal logic, where, given the premises, the conclusion can be shown to be either true or false. In statistics, a conclusion may be reported as being true, but attached to this conclusion is a probability that the conclusion is false. Hopefully this latter probability is very small. For example, a pollster may conclude that a particular political party will win an election but, if this conclusion is based on a poll, this conclusion might be made with only 95% certainty. If there is a stronger basis for the conclusion, the conclusion may be 99% sure. Statistical proofs are never 100% certain.

This approach should not be taken to mean that statistical proofs and results are not well founded. In general, statistical arguments are soundly based and well constructed, but they do rest on a different foundation than do arguments in formal logic. When statistical arguments and methods are used, there is always some uncertainty concerning the truth or falsehood of the conclusion. But this uncertainty does not mean that there is no knowledge concerning the truth or falsehood of a particular conclusion. If a particular problem can be analyzed using probabilistic reasoning, then probabilities or degrees of belief can be attached to particular conclusions. That is, principles of probability allow uncertainty to be quantified, and limits can be placed on this uncertainty. Where this can be done, it is often possible to devise methods which can be used to increase the probability of, or degree of belief in, the conclusion. For example, under some conditions, if the size of a sample is increased, it is possible to make conclusions that hold with very high levels of probability. This increases the certainty with which the conclusion is held.

Since probabilistic reasoning is such a basic part of statistics, it is necessary to become somewhat familiar with probabilistic reasoning, the circumstances in which it can be applied, and how it can be interpreted. One of the main reasons that probability is used in the social sciences is that many conclusions are based either on sampling or on controlled experiments. If a sample is drawn on a probabilistic basis, the principles of probability can be used to draw inferences concerning the nature of the whole population, based on the sample. Similarly, probability can be used to examine the results of certain types of experiments, extending some of the conclusions to a broader situation than that encountered in the experiment itself.

Another reason for using probabilistic reasoning is that there are many probabilistic mathematical models which are useful in the social sciences. These are models such as the binomial or normal distribution. In the social sciences, it is difficult to investigate all possible variables, and there is an unpredictability to human behaviour and actions. These are some of the reasons that social scientists often use models that include both factors which can be investigated and measured, and other less known or less well understood factors. These latter factors are often assumed to behave in a random or probabilistic manner.

Chapter Outline. In applying probability to the social sciences, it is important to understand the principles of probability, so that these applications can be properly carried out. This chapter begins by discussing these principles, and uses these to develop some of the basic rules of probability. Following this, the two most commonly used probability distributions, the binomial and the normal, are discussed. The manner in which the normal distribution can be used to approximate the binomial probabilities follows. The different sections of this chapter may seem disjointed, and they may seem to have little connection with the earlier chapters. However, the developments of the earlier chapters, along with those in this chapter, are pulled together in Chapters 7-9. In particular, this chapter sets the stage for the inferential statistics of Chapters 7-9.

This chapter is not primarily concerned with a study of probability theory as a subject of study and investigation in and of itself. Rather, the aim of the chapter is outline the basic principles and applications of probability so that these can be used in statistical reasoning. Understanding some probability theory is essential in this, but the main concern of the chapter is to show how probability can be applied in statistical reasoning.

6.2 Principles of Probability

6.2.1 Introduction

The term **probability** is more or less synonymous with terms such as **likelihood**, **chances** or **odds**. In mathematics, the term **stochastic** is often used as a more technical and precise term to refer to experiments or games where there is considered to be a **random** or stochastic basis for the outcomes. These are games or **experiments** where the specific outcome is not predictable in an exact sense. In these games, however, it is possible to make statements about the likely or probable outcome in a larger sense. In doing this, **games of chance** are used as examples because they clearly illustrate the basis on which probability rests.

Games of chance are familiar to most people, and have an historical connection with the development of probability theory. Some writers suggest that the origin of probability theory came when gamblers in seventeenth century France consulted mathematicians in order to have the mathematicians improve the basis for placing winning bets. While this story may be more myth than fact, the principles of probability are clearly illustrated by games of chance such as rolling dice or flipping coins. These games have always been used as examples or problems to demonstate how probabilities are determined. Some of the basic principles of these games, and some examples of these games are used in the following sections of the chapter. After introducing the principles of these games of chance, some social science applications are given.

6.2.2 Conditions for Probability

While there are many social science applications for probability, the principles of probability are most clearly illustrated with the traditional games of chance such as rolling a pair of honest dice, flipping a coin or coins, spinning a roulette wheel, or drawing cards from a well shuffled deck of cards. Another common type of example, and one which is closer to the idea of random sampling, is drawing balls from an urn. It is useful to examine the conditions under which these games are conducted in order to obtain a clear idea of the principles of probability. Then it is possible to see how these conditions may be properly applied in social science applications, and where they might be misapplied. In examining these principles, it will be assumed here that the games of chance are fair or **honest**. That is, it will be assumed that whoever flips the coin will flip it well, that the coin itself has no rough edges or chips, and is equally weighted throughout. Similar assumptions are made concerning the other games.

Games of chance, as just described, have several common characteristics. These are:

1. **The exact outcome cannot be predicted**. It is not possible to predict the exact outcome of a particular trial of the game. There is a certain randomness or uncertainty regarding the possible outcomes and this is part of what makes this a game of chance. If it were possible to predict the exact outcome, then it would no longer be considered a game of chance.

2. **All possible outcomes are known**. All the possible outcomes of the game are known or could be listed. For example, for a coin, the outcomes are head and tail; for a deck of cards, there are 52 possible outcomes, or cards, if a card is drawn from the deck. This means that it is not possible to produce some other outcome, say a 53rd card or, for a coin, some outcome other than a head or tail. As will be seen, this ensures that the degree of certainty or uncertainty can be quantified and calculated. It may not be necessary to actually list all the possibilities. For example, it would take too much time to list all possible poker hands that can be drawn from a deck of 52 cards. However, in cases such as this, it is possible to conceive of listing all possible outcomes, and imagine what these are.

   When there is some situation where the outcomes are so uncertain that no one can determine what all of these might be, then the principles of probability can not be applied. For example, if the earth is subject to considerable global warming, some of the possible outcomes might be known, but not all of the possibilities can be imagined at present.

This makes it difficult, if not impossible, to attach probabilities to the various possible outcomes that might occur if global warming continues.

3. **Equally likely outcomes**. There is a certain **symmetry** in the possible outcomes in the sense that all of the outcomes are equally likely. In a game of chance, there is nothing in the structure of the game that makes one outcome any more likely than any other outcome, or nothing that favours one outcome over another. This is equivalent to the notion of an honest coin, one which is equally dense throughout so that when the coin is flipped, each possible side, head or tail, is equally likely. This condition can be modified later by considering games where some outcomes are more likely to occur than others. However, these more likely outcomes can usually be considered to be combinations of individual, equally likely outcomes.

4. **Repeatable under uniform conditions**. The game must be repeatable under uniform conditions. That is, to build up the theory of probability, it must be possible to conduct, or at least be able to imagine conducting, the game over and over again under exactly the same conditions each time. In the case of a coin, the coin is made of a durable metal which does not chip or wear out, even when flipped a large number of times. The process of flipping the coin also must be done in the same manner each time, that is, by giving the coin a good flip each time, so that it lands in such a way that neither head nor tail is favoured by the process of flipping the coin.

5. **Regularity**. Together, the above conditions produce a certain **regularity** in the sense that **patterns** for the results or outcomes can be observed. Individual outcome of a particular game of chance cannot be predicted any better than before, but it becomes possible to begin observing a pattern in terms of the proportion of times a particular outcome occurs. For example, if a coin is flipped 1000 times, about one half of these flips are likely to be heads, and one half tails. That is, it is very likely that there will be approximately 500 heads, perhaps a few more, or perhaps a few less. However, if there were 900 heads, we would probably be willing to say that the coin was not an honest one because this deviates so far from the view of the symmetry of the situation. That is, neither heads nor tails is favoured in the structure of the experiment, so that approximately half the time there should

be heads and half the time tails.

The above conditions lay out the basis for simple games of chance, and on the basis of those principles, probability theory can be constructed. This is begun in the following sections. It is also possible to see how the above principles apply in some social science situations. A short discussion of this follows.

Social Science Applications. A **random sample** can be seen to conform to the above principles. Such a sample is one where each member of a population has an equal chance of being selected in the sample. In this type of sampling, it is not possible to predict who will be selected in the sample. But all possible people in the population are considered for selection, and each person is as likely to be selected as any other person. Many samples could be selected, so that the selection process is repeatable under uniform conditions. It is also possible to notice the ways that these conditions might be violated. If the list used to select people is incomplete, then those people not on the list have no chance of being selected, while each of those people on the list have some chance of being selected. This violates principle (3), that each member of the population has an equal likelihood of being selected. Further, if the results of the sample are published, or made known to the population from which the sample is drawn, the views of people may be changed. This would then make it difficult to satisfy condition (4) above, and conduct another survey under exactly the same conditions. Worse yet, the sampling process itself could contaminate or destroy parts of the population being sampled, meaning that the experiment cannot be repeated.

Many other types of social science problems do not conform to the above principles. The effect of political statements on a population, or the effects of social action or social movements are not likely to conform to these principles. In the social sciences, one of the major problems in applying probability derives from the non-repeatability of actions in society. When dealing with people, actions are usually not repeatable under uniform conditions, as they may be in many natural science experiments. Further, all possible outcomes may not be known for many activities related to human action. There seems to be a basic element of unpredictability to individual human actions.

However, there are many areas in the social sciences where these conditions either hold, or come very close to describing what is happening. If this is the case, then probabilistic methods and models may be safely used.

Approaches to Probability. In examining probability, there are three possible interpretations of probability. Each approach is a legitimate one and each approach has its own particular uses. These interpretations are the classical approach, the frequency interpretation and subjective probabilities. These are outlined as follows.

6.2.3 Classical Approach to Probability

The **classical** or **theoretical** approach to probability depends only on theoretical considerations. In this interpretation, various assumptions or postulates are used and these form the basis for the theory of probability. Building on these assumptions, probability distributions can be constructed. Games of chance are used to illustrate this approach. However, these need not be real games of chance that are actually conducted. Rather, it is only necessary to **imagine** the possibility of the game of chance. For example, imagine a completely fair or honest coin, one that has little or no thickness, so it cannot land on its edge, and one which is uniformly dense throughout. Imagine that the coin is flipped well, and then it is also possible to imagine that it could land on a surface and show either heads or tails.

Generalizing this idea, consider a game of chance which conforms to the principles discussed in the last section. Sometimes such a game is called an **experiment**, so conducting the game of chance once would be referred to as a single trial of the experiment. Flipping the imaginary coin once would be one trial of that experiment and the outcome could be either a head or a tail. The probability of a particular outcome on any trial is defined as follows.

Definition 6.2.1 If a probability experiment has N possible outcomes, each of which is equally likely, then the **probability** of each of the N outcomes is $1/N$.

This definition of probability follows directly from the principles of probability discussed in the last section. Any one of the N outcomes could occur and there is nothing that favours any one of these outcomes over any other outcome. As a result, each outcome has one chance in N of occurring. This can be generalized to define the probability of an event which contains several individual outcomes.

Definition 6.2.2 If a probability experiment has N possible outcomes, each of which is equally likely, and if **event** E has N_E outcomes in it, then the **probability of event** E is defined as

$$P(E) = \frac{N_E}{N}.$$

That is, each of the N_E outcomes has probability $1/N$ of occurring, and there are N_E of these outcomes, so that the chance of one of these outcomes ocurring is N_E/N.

Example 6.2.1 Some Simple Probabilities

1. *Suppose you imagine flipping a two sided coin. The coin has only* $N = 2$ *possible outcomes, head and tail. Each of these is equally likely so that the probability of each of these outcomes is* $1/2$. *If head is denoted by* H *and tail by* T, *these probabilities are*

$$P(\text{head}) = P(H) = \frac{1}{2}$$

$$P(\text{tail}) = P(T) = \frac{1}{2}$$

2. *If we imagine a single die with 6 sides then* $N = 6$. *If the die is rolled, each side occurs with equal likelihood. Thus the probability of any particular side turning up is* $1/6$, *by Definition 6.2.1. The probability of getting an outcome of 2 on a single roll of the die could be written*

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

If E *is the event of obtaining an even number, then this can be obtained by rolling a 2, 4, or 6. Event* E *has* $N_E = 3$ *outcomes in it, so that*

$$P(E) = \frac{N_E}{N} = \frac{3}{6} = 0.5$$

3. *A deck of cards has* $N = 52$ *cards in it. If the deck is well shuffled and one of the cards is randomly picked from the deck, then the probability of selecting any particular card is* $1/52$. *For example, the probability of selecting the queen of spades is* $1/52$. *The probability of selecting a queen is* $4/52$ *since, if event* E *is the event of selecting one of the four queens, then* $N_E = 4$, *so that* $P(E) = 4/52$.

Rules for Probability. Some further rules for classical probabilities state that probabilities are always numbers between 0 and 1, and that the sum of the probabilities of all the outcomes of a probability experiment totals 1. These rules are:

Rule 1. If E is any event defined over the set of possible outcomes of a probability experiment, then

$$0 \leq P(E) \leq 1.$$

Rule 2. If $e_1, e_2, \cdots, e_N$ is the set of all N outcomes of a probability experiment, and each of these is equally likely to occur, then

$$\sum_{i=1}^{N} P(e_i) = 1.$$

The first rule states that probabilities are always numbers between 0 and 1. Probabilities could be defined on a scale from 0 to 100, and treated as percentages. This is sometimes done, but in this textbook, probabilities are always defined as being between 0 and 1. If an event cannot occur, then its probability is 0, and when an event is absolutely certain to occur, then it has a probability equal to 1.

The second rule states that probabilities are additive. If all the possible outcomes are considered, the sum of their probabilities is equal to 1.

The above definitions and rules can be used to build up the whole theory of probability. These definitions and rules are used in later sections to examine the probabilities of various combinations of events. Before doing this, some other approaches to probability are discussed.

6.2.4 Frequency Interpretation of Probability

An alternative view of probability begins by examining the frequency of occurrence of particular outcomes in a large series of trials of a probability experiment. The number of times that a particular outcome occurs in these trials forms the basis for defining probabilities in this manner. This can be done as follows.

Definition 6.2.3 If a probability experiment is performed n times under uniform conditions, and there are n_E occurrences of event E, then an **estimate** of the probability of event E is n_E/n. The **probability of event E**

is defined as

$$P(E) = \lim_{n \to \infty} \frac{n_E}{n}.$$

For those unfamiliar with the mathematical notation of this definition, the following description and example should show how this works. Suppose the probability experiment is conducted n times. At this stage, the best estimate of the probability of event E is the proportion of times that event E occurs in the n trials, that is, n_E/n. However, this is only an estimate and the experiment could be conducted more and more times. It is possible to imagine conducting the experiment an indefinitely large number of times. In this case, the probability of event E is the proportion of times that event E occurs when the probability experiment is carried out an indefinitely large number of times. The symbol

$$\lim_{n \to \infty}$$

means the limit as n approaches infinity. The limit is the number that is approached as the number of trials of the experiment becomes extremely large or infinite. In the above definition, the ratio n_E/n approaches a particular number as n becomes larger and larger. This number is $P(E)$ and is defined as the probability of event E. This number is between 0 and 1, and the rules of probability apply to this.

Example 6.2.2 Flipping a Coin Many Times

This hypothetical example illustrates what might happen if an honest coin is flipped a very large number of times. In Table 6.1, let n be the number of times the coin is flipped and let H be the event of a head occurring. Then n_H is the number of heads that occur in n flips of the coin. In this hypothetical example, after 10 flips of the coin, 7 heads occur and the estimate of the probability of a head occurring is $7/10 = 0.7$. After 100 flips of the coin, 58 heads have occurred and the estimate of the probability of obtaining a head is $58/100 = 0.58$. As can be seen in the table, as n becomes larger and larger, the proportion of times that a head occurs gets closer and closer to one half. If the coin is really an honest one, then it seems likely that the probability of a head would be

$$P(\text{Head}) = P(H) = \lim_{n \to \infty} \frac{n_H}{n} = 0.5000$$

Suppose the coin is flipped an extremely large number of times. The larger the number of flips of the coin, the closer the ratio of the number of heads to the number of flips of the coin gets to 0.5. In the limit, when n *becomes indefinitely large, the ratio of the number of heads to number of flips of the coin will approach some specific number,* $P(E)$*. If this is an honest coin, it is to be expected that this limit will be* $P(E) = 0.5$.

n	n_H	n_H/n
10	7	0.7000
100	58	0.5800
1000	520	0.5200
10,000	5150	0.5150
100,000	50,793	0.5079
∞		0.5000

Table 6.1: Estimates of Probability of Heads in Flips of a Coin

When probabilities based on the frequency interpretation can be compared with those derived from the classical approach, it might be expected that the probabilities obtained from the two approaches would be more or less equal to each other. That is, if a real coin is more or less honest, then about half the time that this coin is flipped, heads would be expected. This conforms to the probability of heads, based on the classical interpretation. As shown in Example 6.2.1, for a coin there are $N = 2$ outcomes. Obtaining a head is one of these two possible, and equally likely, outcomes. Based on the classical approach the probability of a head is $1/2$, and this represents our view of what an honest coin is like. The frequency approach is then expected to yield the same probability for a head.

In cases where the probabilities based on the two interpretations do not conform to each other, this is a sign that one of the conditions listed in Section 6.2.2 does not hold. For example, if a coin is flipped a very large number of times and the proportion of heads which occurs is 0.75, then the conclusion would be that the coin was not an honest one. This may mean than principle (3) of Section 6.2.2, that of equally likely outcomes, could be the principle violated. Alternatively, the conclusion might be that the coin was not flipped under uniform conditions, as given in principle (4).

In many cases, there is no alternative but to use the frequency interpretation. Suppose we want to know the probability of contracting lung cancer among those people who smoke at least one package of cigarettes a day for 20 years. There is no theoretical way to calculate this, at least given the present state of knowledge on this issue. What can be done in situations like this is to collect data from as large a sample as possible, and from as many samples as possible. Then the data from these samples can be used to make an estimate of the required probability, using the frequency interpretation.

It should also be noted that once the probabilities of individual outcomes are defined on the basis of the frequency interpretation, the other rules of classical probability, to be discussed later, can be applied to these probabilities.

6.2.5 Subjective Interpretation of Probability

A third possible way of defining probabilities is to take an event and subjectively attach a number to this event. This number then becomes an estimate of the probability of the occurrence of the event. This type of probability has also been called **judgment** or **personal** probability, because a personal judgment concerning the likelihood or probability of a particular event occurring has been made. In doing this, the number representing the subjective probability of the event should be between 0 and 1. Numbers in this range are used in order to obtain probabilities which are consistent with the other interpretations of probability, where probabilities are always between 0 and 1.

If an event is judged to be impossible, then it is assigned a subjective probability of 0. If some other event is judged to be completely certain to occur, it is assigned a subjective probability of 1. In between, there are no general rules concerning how to assign subjective probabilities, and each person assigning these would use his or her own judgment. Presumably, if a number close to 1 is assigned as a subjective probability of an event, this means that the event is more probable than an event that has been assigned a number much less than 1. For example,if event A has probability 0.8 of occurring, it is being judged as more likely to occur than an event B which has been assigned a subjective probability of only 0.6. In this way, the various events can be ranked or ordered according to the probabilities assigned on the basis of personal judgment. For example, if it was necessary to determine the probability of an earthquake, neither the frequency nor classical interpretation could be used, at least not given the present state of knowledge concerning earthquakes. But a geologist could be asked, and the geologist would presumably be able to provide a reasonably good judgment of the probability of an earthquake occurring. An ordinary person could also make a judgment of the probabiity of the same event, but this judgment would probably not be any better than a guess.

In this approach to probability, it can be seen that the ordering of events is not usually done on any objective basis, but is merely based on someone's judgments. As a result, these probabilities are only as good as the judgment of the person who attached the numbers to the events. Someone who is knowledgable in the particular subject area may make a very good judgment based on his or her knowledge of this subject matter. Others who are less knowledgable are likely to make poorer estimates.

While this type of judgment is called a probability, instead of a probability in the strict sense of the term, it is really more properly considered to be a **degree of belief** that the event will occur. In order to have probability make sense, the experiment must be repeatable under reasonably close to uniform conditions. Many events to which subjective probabilities are attached are not repeatable in this sense. Then the subjective probability becomes a degree of belief that the event will occur. This type of probability thus becomes a means of attaching numbers to events, in much the same way one attaches numbers to attitudes using an attitude scale.

6.2.6 Comparison and Uses of Interpretations

Each of the above interpretations of probability has circumstances in which it is best used. Where it is possible to use more than one of the interpretations to compute the probability of an event, then these probabilities are hopefully more or less the same, regardless of interpretation.

Subjective probability is the interpretation that is least solidly based. The number attached to a particular event will differ, depending on who makes the judgment. As a consequence, there could be several different estimates of the probability of a particular event. While it may be quite useful to have these different estimates when discussing the chance that an uncertain event will occur, it becomes more difficult to use subjective probabilities in further developments of probability and statistics. If there is no other method available, then a subjective probability can certainly be used. However, if it is possible to make the estimate on the basis of a classical or frequency approach, these latter methods are generally considered preferable.

One area of statistics based on the classical interpretation is sampling theory. If samples are selected on the basis of probability, then the classical approach can be used to determine the probability of the various possible events and also to determine what can be inferred about a whole population on the basis of a sample drawn from the population. Since such statistical inference is a major part of statistics, the classical approach to probability forms the basis for much of the rest of the course. The binomial and normal distributions are both based on the classical interpretation of probability. These distributions, in turn, are extensively used in statistical inference later in the textbook.

When estimating probabilities of particular events, the classical interpretation may be difficult to use in real world situations. Once the real world situation begins to differ from the conditions in which ideal games of chance are conducted, then there may be little basis for estimating probabilities, using the classical approach. In this case, the frequency interpretation is likely to be used. For example, suppose the probability of being unemployed, for people of different levels of education, is to be determined. In general, studies of the labour force find that among those with higher levels of schooling completed, there is a smaller proportion of unemployed. This conclusion is based on data collected concerning education levels and labour force status of labour force members. There is no mathematical reason why the results need come out the way they do. Based on labour force studies though, the frequency interpretation of probability could be used to provide estimates showing that the probability of being unemployed is lower, the higher the education level of the person. This estimate is entirely based on the frequency interpretation. In doing this, data which is as accurate as possible should be used. This includes using as representative a sample as possible, and as large a sample size as possible. However, it should be noted that under these circumstances, a frequency estimate of probability will likely change somewhat depending on which population is being surveyed and exactly how the variables are defined. In addition, the estimate of the probability will likely change as the size of sample increases.

All the approaches to probability have their distinct and legitimate uses. You should be familiar with each approach and how it can be properly applied. Section 6.2.8 builds on the classical approach to probability. More rules, definitions and examples are given there to show how the theoretical approach to probability is constructed and used. First, some problems concerning interpretations of probability are given, and Section 6.2.7 shows how probability can be applied to a sampling problem.

Problems Concerning Interpretations of Probability. For each of the following, which interpretation of probability would seem most appropriate? In each case, briefly explain your choice. Answers to these problems are given on page 349.

1. The probability that it will snow tomorrow.
2. The probability that it will snow on June 3.
3. The probability that the ozone layer of the earth will be damaged beyond repair in the next 5 years.
4. The probability that in a random sample of 100 people in Regina, there will be at least two supporters of the Rhinocerous Party.
5. The probability that a random sample of 5 Saskatchewan MLA's will yield at least two NDP MLAs.
6. The probability that the Tories will win the next provincial election.
7. The probability that the Roughriders will win the Grey Cup this year.

6.2.7 An Application of Probability to Sampling

Some of the implications of the classical approach to probabilitiy can be seen by considering a simple case of random sampling from a very large population. Suppose this large population is composed of exactly one half males and one half females. If a person is randomly chosen from this population, then that person might be either a male or a female. Since the population is exactly half male and half female, the 2 possible outcomes of male (M) or female (F) are equally likely. Thus the conditions of Section 6.2.2 are satisfied and the rules of Section 6.2.3 can be applied. In this case, $N = 2$, $P(M) = 1/2$ and $P(F) = 1/2$. The distribution of outcomes and associated probabilities is given in Table 6.2.

Now suppose a random sample of size 2 is taken from this same population. This sample can be obtained either by picking the two cases simultaneously, or by first picking one person and then by picking another. Using the latter method, after selecting one person, the population is still very close to being composed of half males and half females. This is because the population is large, and selecting only one person from this population does not change the sex composition of the population. For the second draw, the

Outcome	Probability
M	1/2
F	1/2

Table 6.2: Probabilities for Sample of Size 1, $N = 2$

chance of picking a male is thus still regarded as being exactly equal to the chance of picking a female.

When a sample of size two is selected, it may seem that there are only 3 outcomes: 2 males, a male and a female, and 2 females. While there are only these three situations which can occur, these situations are not equally likely. The event of obtaining a male and a female can occur in two ways, the male could be picked first, followed by selection of a female (MF), or the female could be drawn first and this could be followed by selection of a male (FM). In contrast, there is only one way to select 2 males, and only one way to select 2 females. What this means is that the **order** of occurrence of the various outcomes must be considered. This must be done to ensure equal likelihood of selection for the different outcomes.

Based on these considerations, for a sample of size 2, there are four equally likely outcomes: MM, MF, FM and FF. The middle two outcomes, MF and FM, both produce a male and a female, and by listing the order of the occurrences, it can be seen that there are two ways to produce the combination of a male and a female. The list of $N = 4$ outcomes is given in Table 6.3. Each of these outcomes is equally likely, and thus the probability of each outcome is 1/4.

Outcome	Probability
MM	1/4
MF	1/4
FM	1/4
FF	1/4

Table 6.3: Probabilities for Sample of Size 2, $N = 4$

Now extend this same problem to randomly picking a sample of 3 people from the very large population which is half male and half female. Again, since the population is very large, the population has one half males and one half females, even after selection of the first and second person. In Table 6.4 it can be seen that there are 8 equally likely possible outcomes, each of which will have probability 1/8 of occurring. This set of 8 outcome exhausts all the possible ways of selecting males and females in a sample of size 3.

Outcome	Probability
MMM	1/8
MMF	1/8
MFM	1/8
FMM	1/8
FFM	1/8
MFF	1/8
FMF	1/8
FFF	1/8

Table 6.4: Probabilities for Sample of Size 3, $N = 8$

Also note that the probabilities for males or females in these examples are identical with the probabilities of heads or tails derived from flipping a coin. If M is redefined as H, for head, and F is redefined as T, for tail, then these tables give the probabilities of obtaining different numbers of heads and tails for 1, 2, or 3 flips of a coin. In terms of the nature of the probability experiment, the game of flipping coins is identical with the structure of the experiment of randomly drawing samples from a large population, where this population is equally divided into two groups.

Events. The set of all outcomes, as listed in Tables 6.2 to 6.4, can be used to determine the probabilities of various **events**. As noted in Definition 6.2.2, an event is any set of outcomes defined according to some criterion. For example, suppose a sample of size 3 is selected, as in Table 6.4. Let event E be defined as the event of selecting exactly 1 male in a random sample of size 3 from a large population which is half male and half female. As can be seen in Table 6.4, event E is composed of 3 outcomes: FFM, MFF and FMF, and $N_E = 3$. These are the only outcomes for which there is exactly one male. In a sample of size 3, the number of possible outcomes is $N = 8$. The number of these outcomes for which event E occurs is $N_E = 3$, and based on Definitions 6.2.1 and 6.2.2,

$$P(1 \text{ male}) = \frac{N_E}{N} = \frac{3}{8}.$$

Similarly, the probabilities of various other events could be determined as follows:

$$P(\text{at least one male}) = \frac{7}{8} = 0.875$$

$$P(\text{at least one of each sex}) = \frac{6}{8} = 0.75$$

$$P(2 \text{ or more females}) = \frac{4}{8} = 0.5$$

6.2.8 Combinations of Events

Once there is a list of all the possible outcomes for a probability experiment, various events can be defined over this set of outcomes. In addition, various **combinations of events** can be defined by using the words *and* and *or*. This section examines the manner in which two or more events, defined over the set of all the outcomes, can be joined together using the connectors *and* and *or*.

Using the connector *and*. In probability, *and* is used to denote that both events A and B occur together. The event (A *and* B) denotes that set of outcomes of a probability experiment that is common to both events A and B.

Definition 6.2.4 If there are two events A and B, defined over a set of N outcomes of a probability experiment, and if $N(A \textit{ and } B)$ is the number of outcomes common to A and B, then the probability of the event (A *and* B) is defined as

$$P(A \textit{ and } B) = \frac{N(A \textit{ and } B)}{N}$$

Example 6.2.3 Events in a Sample of Size 3

Consider the sampling example of Section 6.2.7, where a sample of size 3 is randomly selected from a large population which is equally divided between males and females. Suppose event A is defined as the event of selecting at least one person of each sex and event B is defined as the event of selecting at least two males. In this case, the events A and B can be pictured as in Table 6.5.

Outcomes

	MMM
	MMF
B	MFM
	FMM
	FFM
A	FMF
	MFF
	FFF

Table 6.5: Outcomes and Events for Sample of Size 3, N=8

As can be seen in Table 6.5, $P(A) = 6/8$ and $P(B) = 4/8$. The number of outcomes common to the two events is 3 out of the 8 possible outcomes so that $N(A\ and\ B) = 3$ and

$$P(A\ and\ B) = \frac{N(A\ and\ B)}{N} = \frac{3}{8}.$$

Mutually Exclusive Events. In some cases, two events have no outcomes in common. When this is the case, the two events are said to be **mutually exclusive** of each other.

Definition 6.2.5 Events A and B are said to be **mutually exclusive** if $N(A\ and\ B) = 0$. If $N(A\ and\ B) = 0$, then it follows that $P(A\ and\ B) = 0$.

In Table 6.5 of Example 6.2.3, let C be defined as the event of all males being selected. This is the outcome MMM, the first outcome listed. Now let D be the event of at least two females being selected. This event is composed of the last four outcomes, FFM, FMF, MFF and FFF. By examining these events, or from Table 6.5, it can be seen that these two events have no outcomes in common. As a result, there is a zero probability of the two events occurring together and the two events are mutually exclusive.

In any probability experiment, each of the individual outcomes is mutually exclusive of each of the other individual outcomes. For example, in Table 6.5, outcome MMM is a different outcome than is MMF and these are, in turn, both different from MFM, and so on. Thus, each outcome is mutually exclusive of each of the other individual outcomes.

Using the Connector *or*. The use of *or* in probability is an all inclusive or. That is, it denotes inclusion of any of the events listed in an *or* statement. In ordinary language, this is equivalent to using "either or, or possibly both."

Definition 6.2.6 If two events A and B are defined over a set of outcomes of a probability experiment, then $N(A \textit{ or } B)$ is the set of outcomes that is contained in either event A or in event B, or in both events A and B. The probability of $(A \textit{ or } B)$ is

$$P(A \textit{ or } B) = \frac{N(A \textit{ or } B)}{N}.$$

In Example 6.2.3, where A was the event of selecting at least one person of each sex and B was the event of selecting 2 or more males, the outcomes of Table 6.5 can be counted to see that $N(A \textit{ or } B) = 7$. Thus

$$P(A \textit{ or } B) = \frac{7}{8}.$$

In this case it can be seen that the two events overlap and have three outcomes in common. In total though, there are 7 of the 8 outcomes which fall into either A or B or possibly both. Another way of obtaining $P(A \textit{ or } B)$ is to add the probability of event A to the probability of event B, and subtract from this, the probability that events A and B occurr together. This leads to the following rule.

Rule 3. The probability of A *or* B is

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B).$$

The derivation of this rule comes from the basic definitions of *and* and *or* and recognizing that

$$N(A \text{ or } B) = N(A) + N(B) - N(A \text{ and } B).$$

Dividing each of these parts of the expression by N gives

$$\frac{N(A \text{ or } B)}{N} = \frac{N(A)}{N} + \frac{N(B)}{N} - \frac{N(A \text{ and } B)}{N}$$

and these are the probabilities of Rule 3. In the Example 6.2.3, $P(A) = 6/8$, $P(B) = 4/8$ and $P(A \text{ and } B) = 3/8$ so that

$$P(A \text{ or } B) = \frac{6}{8} + \frac{4}{8} - \frac{3}{8} = \frac{7}{8}.$$

Where two events A and B are **mutually exclusive**, then

$$N(A \text{ and } B) = 0.$$

so that

$$P(A \text{ or } B) = P(A) + P(B) - 0 = P(A) + P(B).$$

That is, if the two events are mutually exclusive, then the probability of one or other of these events is the simple sum of their individual probabilities.

6.2.9 A Survey Sampling Example

Another type of example, again involving sampling, is a two way cross classification of respondents from a sample. As an example of this, consider the following table based on respondents in the Social Studies 203 Labour Force Survey. This example is set up a little differently than Example 6.2.3, but illustrates a useful way to examine survey data.

In this example there are 610 respondents in the Survey who answered questions concerning their sex and social class. In this Survey, the variable measuring social class is self identification of social class. Respondents were asked which of 5 social classes they would say they were in, going from upper to lower class. None said upper class and the distribution of responses for the other choices are shown in Table 6.6

This table can be turned into a problem in probability by considering the 610 people as a population or set of possible outcomes. Suppose one person is randomly picked out of the 610 people in this group. There is an equal chance of selecting a male as a female, because there are 305 out of 610 respondents who are male and 305 of the 610 who are female. Going back to the definitions of probability, there are $N = 610$ possible outcomes. If event M is defined as the event of picking a male, then $N(M) = 305$, since event M has 305 outcomes in it. Similarly, event F, the event of the random choice being a female, has 305 outcomes, so $N(F) = 305$ and $P(F) = 305/610 = 1/2$.

	Sex		
Social Class	Male (M)	Female (F)	Total
Upper Middle (A)	33	29	62
Middle (B)	153	181	334
Working (C)	103	81	184
Lower (D)	16	14	30
Total	305	305	610

Table 6.6: Social Class Cross Classified by Sex of Respondents

Let the event of picking a person of each of the different social classes be identified as in the table. The probability of picking a person who is middle class is defined as the probability of event B and since there are 334 of the 610 respondents who have this characteristic,

$$P(B) = \frac{N(B)}{N} = \frac{334}{610} = 0.548.$$

Other probabilities could also be calculated, for example,

$$P(\text{lower class}) = P(D) = \frac{30}{610} = 0.049.$$

$$P(\text{working class}) = P(C) = \frac{184}{610} = 0.302.$$

The probabilities of various combinations of events can also be determined. For example, if one of the 610 respondents is randomly selected, what is the probability that this person would be a middle class female? In this case, the probability is

$$P(B \text{ and } F) = \frac{N(B \text{ and } F)}{N} = \frac{181}{610} = 0.297.$$

The probability of randomly selecting a working class male is

$$P(C \text{ and } M) = \frac{N(C \text{ and } M)}{N} = \frac{103}{610} = 0.169.$$

The probability of the above *and* combinations are relatively straightforward to calculate. The *or* combinations may be a little more difficult to envisage and Rule 3 is very useful here. For example, if one respondent is randomly selected, what is the probability that this person is either female or middle class? This probability is

$$\begin{aligned} P(F \text{ or } B) &= P(F) + P(B) - P(F \text{ and } B) \\ &= \frac{305}{610} + \frac{334}{610} - \frac{181}{610} \\ &= \frac{305 + 334 - 181}{610} \\ &= \frac{458}{610} = 0.751. \end{aligned}$$

Alternatively, it would be possible go through the table and pick out all those respondents for whom one of either conditions F or B were true. This involves counting $N(F \text{ or } B)$. All the respondents for whom event F, being female, holds are those in the female column and going down cell by cell, this is $29 + 181 + 81 + 14$ for a total of 305. In addition, there are another 153 for whom condition B holds, but which are not female. So $N(F \text{ or } B) = 29 + 181 + 81 + 14 + 153 = 458$. In doing this, make sure not to count twice the 181 respondents who are both female and middle class. If these are included twice, then this would be double counting and this would violate the principle of equally likely outcomes for each one of the 610 respondents.

Similarly, other *or* probabilities can be calculated. The probability of selecting either a male *or* a working class respondent is given by:

$$\begin{aligned} P(M \text{ or } C) &= P(M) + P(C) - P(M \text{ and } C) \\ &= \frac{305}{610} + \frac{184}{610} - \frac{103}{610} \end{aligned}$$

$$\begin{aligned} &= \frac{305 + 184 - 103}{610} \\ &= \frac{386}{610} = 0.633. \end{aligned}$$

The probability of selecting a middle or working class person is somewhat easier because these two events are mutually exclusive so that:

$$\begin{aligned} P(B \textit{ or } C) &= P(B) + P(C) - P(B \textit{ and } C) \\ &= \frac{334}{610} + \frac{184}{610} - 0 \\ &= \frac{334 + 184}{610} \\ &= \frac{518}{610} = 0.849. \end{aligned}$$

Application to the Whole Population. Under certain conditions, the above probabilities can be more broadly applied so that they refer to more than just the above table. If the sample on which the table is based is a representative sample of a population, then the above probabilities can be considered to be probabilities of events that are defined over the whole population. Under certain conditions, a random sample can be regarded as being quite representative of a population. In this case, if one person is randomly selected from the whole population, the probability that this person is middle class might be said to be 0.548. In addition, the probability of randomly selecting a male or working class person from the population is 0.633, as calculated earlier.

These examples begin to move into the area of statistical inference, where conclusions concerning a whole population are being inferred from a sample. This also shows that these probabilities have considerable application beyond this table. If the table really is **exactly representative** of the whole population, then the probabilities above do apply to random selection from the whole population. However, it is very unlikely that the sample is exactly representative of the whole population, even where the sampling is truly random. In addition, many nonsampling errors may occur, so there are likely to be additional problems of applying this approach beyond this specific table and random selection from this table. The discussion of statistical inference later in the textbook will deal with this problem and the extent to which generalizations beyond a sample can be made. You should be aware of such a possibility, but at this stage, the discussion will be left at the level of the specific probabilities in the tables.

Problem - Social Class and Education. Table 6.7 is based on the Social Studies 203 Labour Force Survey. For this table, H represents high school or less education, P some post secondary education, and U having a University degree or degrees. The events of being in the different social classes are as labelled in Table 6.7 and are defined in the same way as in Table 6.6. Answers to this problem are given on page 351.

	Education Level			
Social Class	H	P	U	Total
Upper Middle (A)	16	21	24	61
Middle (B)	152	100	81	333
Working (C)	128	40	15	183
Lower (D)	24	4	2	30
Total	320	165	122	607

Table 6.7: Social Class Cross Classified by Education of Respondents

If a person is randomly selected from the 607 respondents for whom information is available in Table 6.6, what is the probability of obtaining a person who

1. is working class?
2. has a degree?
3. is working class and has high school or less education?
4. is lower class and has more than a high school education?
5. has high school or less and is middle or upper middle class?
6. has a degree or is working class?
7. does not have a degree or is working or lower class?

6.2.10 Conditional Probabilities

Frequently a situation arises where the probability of some event happening, given that some other event has already occurred, is desired. In this case, it is necessary to determine the **conditional probabilities**. For example, in the above cross classification example of Table 6.6, suppose that a female has been selected, but the social class of this female is not known. In this circumstance, suppose that someone wishes to determine the probability that the person chosen is working class, given that we know the person is female. Based on Table 6.6, this probability can be determined by considering only the 305 outcomes that are female. Thus $N(F) = 305$ and, of these outcomes, the number which are also working class, $N(F \textit{ and } C)$, is 81. Thus the probability of selecting a working class respondent, given that the respondent is female, is 81/305 or 0.266. This probability is referred to as the conditional probability of being working class, given that the respondent is female.

Definition 6.2.7 If there are two events A and B, defined on a set of outcomes of a probability experiment, then the **conditional probability of A given B**, is

$$P(A/B) = \frac{N(A \textit{ and } B)}{N(B)},$$

where $N(A \textit{ and } B)$ is the number of outcomes common to A and B and $N(B)$ is the number of outcomes in event B. The conditional probability of B given A can similarly be defined as

$$P(B/A) = \frac{N(A \textit{ and } B)}{N(A)}.$$

Based on these definitions, the following rules can be derived.

Rule 5. Using conditional probabilities, the probability of $(A \textit{ and } B)$ can be written as

$$P(A \textit{ and } B) = P(A)P(B/A)$$

or

$$P(A \textit{ and } B) = P(B)P(A/B).$$

These rules are derived as follows:

$$\begin{aligned}
P(A \text{ and } B) &= \frac{N(A \text{ and } B)}{N} \\
&= \frac{N(A \text{ and } B)}{N(A)} \frac{N(A)}{N} \\
&= P(B/A)P(A) \\
&= P(A)P(B/A).
\end{aligned}$$

$$\begin{aligned}
P(A \text{ and } B) &= \frac{N(A \text{ and } B)}{N} \\
&= \frac{N(A \text{ and } B)}{N(B)} \frac{N(B)}{N} \\
&= P(A/B)P(B) \\
&= P(B)P(A/B).
\end{aligned}$$

Example 6.2.4 Conditional Probabilities

Based on random selection of a single case from Table 6.6, some conditional probabilities are:

1. *The probability of selecting a working class person, given that this person is male is*
$$P(C/M) = \frac{N(C \text{ and } M)}{N(M)} = \frac{103}{305} = 0.338$$
2. *The probability of selecting an upper class person, given that this person is male is*
$$P(A/M) = \frac{N(A \text{ and } M)}{N(M)} = \frac{33}{305} = 0.108$$
3. *The probability of selecting a male, given that this person is lower class is*
$$P(M/D) = \frac{N(M \text{ and } D)}{N(D)} = \frac{16}{30} = 0.533$$
4. *The probability of selecting a female, given that this person is working class is*
$$P(F/C) = \frac{N(F \text{ and } C)}{N(C)} = \frac{81}{184} = 0.440$$

Application to the Whole Population. Under certain conditions, these conditional probabilities can be applied to the whole population from which the sample is drawn. Again, assume that Table 6.7 represents the distribution of class and education across the whole population. If you were to randomly meet a member of the population, and if this person is female, the probability that she will be middle class is $P(B/F) = 181/305 = 0.593$. If you happened to meet a male, again considering this to be a random process, the probability that the male is middle class is $P(B/M) = 153/305 = 0.502$. In contrast, if you did not know the sex of the person, the best estimate of the probability that the person is middle class would be $P(B) = 334/610 = 0.548$. Beginning from this last situation where there is no information concerning the person, before meeting her or him, the best estimate that can be made of the probability of that person being middle class is 0.548. It can be seen that the estimate of the probability of the person being middle class can be improved if the sex of the person is known. This happens because the conditional probabilities are conditional on already knowing something about the situation. Given this partial knowledge, estimates of the probabilities concerning the remaining uncertain aspects of the situation can often be improved.

Problems on Conditional Probability. Suppose a case is randomly selected from the set of people in Table 6.7. What are the following conditional probabilities? Answers are given on page 352.

1. What is the conditional probability of selecting a middle class person, given that a person with a University degree has been selected?

2. What is the conditional probability of selecting a person who is working class, given that the person has some post secondary education but does not have a University degree?

3. What is the conditional probability of selecting an upper middle class person, given that this person does not have a University degree?

4. How and why does the probability in 3 differ from the probability of selecting a lower class person, given that the person does not have a University degree?

5. Compare the conditional probability of finding a person with a high school education or less (i) given the person is middle class with (ii) given the person is working class.

6.2.11 Independence and Dependence

The example of conditional probabilities of the last section shows that the probability of an event may change depending on what knowledge is available concerning which other events have or have not occurred. If the probabilities do change in the manner shown in the last section, then the two events are said to be **dependent**. In contrast, if the probability of an event does not change, even after we know which other event has occurred, then these events are said to be **independent** of each other. The formal definitions are as follows.

Definition 6.2.8 Events A and B, defined over the set of outcomes of a probability experiment, are said to be **independent** of each other if

$$P(A/B) = P(A)$$

and

$$P(B/A) = P(B).$$

If events A and B are independent of each other, then it follows that

$$P(A \text{ and } B) = P(A)P(B),$$

so that the probability of the event (A *and* B) can be determined by multiplying together the individual probabilities. That is, if A and B are independent, so that $P(A/B) = P(A)$, then

$$P(A \text{ and } B) = P(A/B)P(B) = P(A)P(B).$$

When two events are independent of each other, the probability of one of the events occurring does not depend on whether or not the other event occurs. If A is independent of B, then the probability of A occurring does not in any way depend on whether or not event B has occurred. In this case, the conditonal probabilities equal ordinary probabilities, and only the latter need to be considered.

The formal definition of independence corresponds to the ordinary usage of the term as well. If two events are independent of each other, then we consider these as events which have no connection with each other or are unlikely to affect each other in any way. A game of chance is structured so that different trials of the game are independent of each other. Flipping the coin, rolling the die, or shuffling the deck of cards, when each of these is done well, assures independence of succcessive trials in the game of chance. In these ideal games, no trial of the game is supposed to affect the probabilities of outcomes on other trials.

In the social sciences, various events may be more likely to be dependent on each other. For example, the amount of wheat produced in Saskatchewan may have an effect on world wheat production and prices and these may indirectly have a small effect on the size of workers' wage settlements in Italy. However, the connection is so indirect that we would likely imagine the probability of different levels of Italian workers' wage settlements to be independent of levels of wheat production in Saskatchewan. On the other hand, the price of wheat may have a lot to do with election results in Saskatchewan. The probability of the government being reelected may be strongly dependent on the event of higher wheat prices.

Also note that the idea of independence can be extended to consider the possibility of the independence of several events. Suppose there are 4 events A, B, C and D, each of which is independent of each of the other events, and there is no form of joint dependence among these events. Then it can be shown that

$$P(A \text{ } and \text{ } B \text{ } and \text{ } C \text{ } and \text{ } D) = P(A) \times P(B) \times P(C) \times P(D).$$

This result is not proven here, but the proof is a relatively straightforward extension of the earlier results. It will be seen in Section **??** that this result provides one of the bases for determination of the binomial probabilities.

Example 6.2.5 Sample of Size 3

The sampling example in Table 6.5 can be used to illustrate independence. There event A was defined as the event of selecting at least one person of each sex in a random sample of size 3 from a large population that is divided equally between males and females. Event B was defined as the event of at least 2 females in the same type of sample.

Based on these definitions, the probabilities of A and of B were $P(A) = 6/8 = 3/4$ and $P(B) = 4/8 = 1/2$. The conditional probabilities can be seen to equal these probabilities.

$$P(A/B) = \frac{N(A \text{ and } B)}{N(B)} = \frac{3}{4} = P(A).$$

$$P(B/A) = \frac{N(A \text{ and } B)}{N(A)} = \frac{3}{6} = P(B).$$

As a result, for this sampling problem, events A and B can be considered to be independent of each other.

Example 6.2.6 Sample of Size 150

Tables 6.8 to 6.10 contain three hypothetical sets of data. In each case a group of 150 people is composed of 100 males and 50 females. The distribution of these people according to whether they agree or disagree with some opinion question is also given, and this differs in each table. A quick glance at Table 6.8 shows that, in this table, the opinions are the same for each sex, with 60 per cent of each sex agreeing with the opinion. This means that opinion is independent of sex and opinion does not depend on sex. In this table, knowing the sex of the person does not help in predicting that person's opinion.

In Table 6.9 the situation changes so that while females are evenly split on the issue, males tend to agree rather than disagree with the opinion. In Table 6.10, the situation is more extreme and the difference between males and females there is even greater than in Table 6.9. Based on these considerations, independence would be expected in Table 6.8, but dependence in Table 6.9, with even stronger dependence in Table 6.10.

Opinion	Male (M)	Female (F)	Total
Agree (A)	60	30	90
Disagree (D)	40	20	60
Total	100	50	150

Table 6.8: No Relation Between Opinion and Sex

In order to analyze these tables in terms of the definition of independence, let the events be labelled as in the tables, that is, let event A be the event of agreeing, event D the event of disagreeing, and events M and F as the events of being male and female, respectively. Then consider the *probability of event A, that of agreeing. Again, as in the previous survey sampling examples, imagine that one person is randomly selected from the total group of 150 people.*

Examining Table 6.8 first, events A and M can be seen to be independent and events A and F are also independent. From this table, the following calculations can be made and these show the independence of the various events.

$$P(A) = \frac{90}{150} = 0.6,$$

$$P(A/M) = \frac{60}{100} = 0.6 = P(A),$$

$$P(A/F) = \frac{30}{50} = 0.6 = P(A).$$

As a result, the event of agreeing is independent of being male and the event of agreeing is also independent of being female. Also note that

$$P(A \text{ and } M) = \frac{60}{150} = 0.4$$

$$P(A)P(M) = \frac{90}{150}\frac{100}{150} = 0.4$$

and this illustrates that when two events A and M are independent of each other, $P(A \text{ and } B)$ is the product of the probabilities of A and of B.

One consequence of independence in this table is that it is not possible to improve the estimate of whether a person is likely to agree or disagree by knowing the sex of the person. The characteristic of being male or being female seems to have nothing to do with opinion because both sexes have the same views. In this case, knowing the sex of the person chosen does not provide any basis for predicting whether the person is likely to agree or disagree.

In Table 6.9, the situation changes considerably. Here the corresponding probabilities are:

$$P(A) = \frac{90}{150} = 0.60,$$

$$P(A/M) = \frac{65}{100} = 0.65 > P(A),$$

$$P(A/F) = \frac{25}{50} = 0.50 < P(A).$$

As a result, A and M are not quite independent since the probability of A differs from the conditional probability of A, given that event M has *occurred. In the case of A and F, these events are even more dependent in the sense that the probability of A differs quite considerably from the conditional probability, $P(A/F)$. This means that for this table, knowledge of the sex of the person chosen does considerably improve the estimate of the probability that the person agrees or disagrees, compared to the situation where the sex of the person is not known.*

In order to predict opinion, if the person is male, the conditional probability of agreeing (0.65) is somewhat greater than the overall probability of agreeing (0.60). However, the probability that the person agrees, given that the person is female, is considerably lower (0.5) than the probability of agreeing where the sex of the person is not known (0.60). The event of agreeing is more dependent on the event of being female than on the event of being male.

Opinion	Male (M)	Female (F)	Total
Agree (A)	65	25	90
Disagree (D)	35	25	60
Total	100	50	150

Table 6.9: Weak Relation Between Opinion and Sex

Opinion	Male (M)	Female (F)	Total
Agree (A)	75	25	100
Disagree (D)	25	25	50
Total	100	50	150

Table 6.10: Strong Relation Between Opinion and Sex

In Table 6.10 the situation is even more marked. Based on Table 6.10, probabilities are:

$$P(A) = \frac{100}{150} = 0.667,$$

$$P(A/M) = \frac{75}{100} = 0.75 > P(A),$$
$$P(A/F) = \frac{25}{50} = 0.50 < P(A).$$

Here event A is quite dependent on both M and F. In this table, knowing the sex of the person chosen considerably helps in improving the estimate of the probability of that person agreeing or disagreeing. If the sex of the person was not known, then the probability of agreeing is 0.667. If it is known that the person is male, then the probability of agreeing is much greater, 0.75 or 3 chances in 4. In constrast, if the person is female, the probability of agreeing is much less, being just 0.5, or one chance in two.

The relevance of these considerations, when these ideas can be applied to larger populations should also be clear. If Table 6.10 were to represent the whole population, then the probability that females agree is much lower than the probability that males agree. This may be useful is examining the structure of attitudes in the population as a whole.

Example 6.2.7 Relationship Between Class and Sex

The survey sampling example of Table 6.6 can be used to look at independence and dependence as well. There, M was male and F female with events A through D being the various social classes from upper middle (A) through to lower class (D). Based on this, the possible dependence or independence of various events can be determined.

Class	P(class)	P(class/M)	P(class/F)
Upper Middle (A)	0.102	0.108	0.095
Middle (B)	0.548	0.502	0.593
Working (C)	0.302	0.338	0.267
Lower (D)	0.049	0.052	0.046

Table 6.11: Independence and Dependence of Various Events

Table 6.11 contains the probabilities and conditional probabilities of the various events, based on the data in Table 6.6 In Table 6.11, the column headed P(class) gives the probabilities of randomly selecting a person of each social class from the total group of 610 people. The column to the right *of this, headed P(class/M), gives the conditional probabilities of selecting a person of each social class, given that the person is male. These two columns can be compared to see whether the event of being in any of the social classes is independent of the event of being male. It can be seen that for none of the social classes is the overall probability exactly equal to the corresponding conditional probability. Thus none of the social classes is independent of being male.*

Comparison of the probabilities also shows that some of the events are more dependent on each other than others. This happens because the gap between the overall probability and the corresponding conditional probability is greater in some rows than in others. Take the first two rows. There the probability of the event of being upper middle class (0.102), is almost equal to the conditional probability of being upper middle class, given that the person is male (0.108). This means that the event of being male is practically independent of being upper middle class. In the second row, the probability of being middle class is somewhat greater for the group as a whole (0.548), than it is for males (0.502), meaning that the characteristics of being middle class and being male are more dependent than are the characteristics of being upper middle class and being male.

Strength of Dependence or Independence. The implication of the last examples is that there are **degrees of dependence** or **degrees of independence**. If the overall probability is very close to the corresponding conditional probability, then the two events can be considered to be close to being independent or **relatively independent**. In contrast, when the conditional probability differs considerably from the corresponding overall probability, then the two events can be considered to be **relatively dependent**. The greater the gap between the overall and the conditional probability, the greater the dependence of the two events.

These ideas of relative dependence and relative independence are rather vaguely defined here and what is a large difference between the corresponding probabilities is difficult to say. This depends on the particular data set being examined and the type of problem involved. However, the concept is useful when we consider that in social science problems and social science data, it is relatively uncommon to find complete independence of different events. Rather, some events are found to have very little effect on other events, and can be considered to be almost independent of other events. Other pairs or groups of events are more likely to have an effect on each other, or be connected together in some way. In these latter cases, the events are considered to be relatively dependent.

In addition to the question of whether two social phenomena affect each other or not is the question of how the sampling process affects the results. In general, the sample is not exactly representative of the population and there are a considerable number of both sampling and nonsampling errors. Even if the events are theoretically independent of each other, there is unlikely to be exact independence of these same events in an actual sample. As a result of these considerations, if the events are considered to be relatively independent on the basis of an actual sample, then this may be taken as some evidence that the events are really independent of each other in the whole population.

When events A and B are relatively independent of each other, knowledge of A is of little use in predicting outcomes for event B. In this case, we may ignore A when investigating B. Returning to the case of Italian workers' wages, suppose these are to be predicted. In attempting to predict these, production levels of Saskatchewan wheat could be safely ignored, because these two events are independent of each other. In the case of events which are relatively dependent, knowledge of both events and the way in which they are dependent on each other becomes important for further investigation. To return to the same problem, the results of the next Saskatchewan election could depend, in some way, on Saskatchewan wheat production. In analyzing the possibility of these situations, cross classifications of the sort examined here help to analyze the possible relationships.

To conclude the discussion of Table 6.11, based on data from Table 6.6, it seems that none of the social classes are strongly dependent on the sex of the person. This might have been expected because we expect the people of each social class to be more or less equally divided between males and females. If there is any dependence here, it does not appear to be in the upper middle or the lower class, because the overall probabilities and the conditional probabilities are practically identical for both sexes for each of these social classes. If there is any dependence it appears in the working and middle classes. Comparing those probabilities, males are a little more likely than are females to be working class, and in turn, females are a little more likely than males, or than the whole group, to be middle class. However, the dependence is not large, and not too much should be made of these differences without further investigation in this and other samples.

The following examples illustrate some further applications of the principles of independence and dependence.

Problems Concerning Independence and Dependence. Suggested answers to the following problems are given on page 353.

1. Using the data in Table 6.7, compute the following probabilities and comment on the dependence or independence of the various events. In doing this, let H be the event of having a high school or less education, P the event of some post secondary education and U the event of having a degree. Events A through D are as in Tables 6.7.

   (a) $P(B/H)$, $P(B/P)$ and $P(B/U)$.

   (b) $P(H/C)$, $P(P/C)$ and $P(U/C)$.

   (c) $P(D/H)$, $P(D/P)$ and $P(D/U)$.

2. For each of the following quotes, explain how the concepts of independence and dependence are involved. Attempt to identify the events that the authors are referring to and then state whether they are being considered as independent or dependent by the author.

   (a) "A little bit of alcohol apparently does a lot of good for a woman's heart: moderate drinkers – two drinks or less a day – cut their heart attack **risk** by more than a third, a new study finds. (**USA Today**, October 3, 1990).

   (b) "Lower dollar doesn't stop shopping in U.S." Regina **Leader-Post**, February 8, 1992.

   (c) "An interesting finding of the current study was the lack of differences between the male and female children with regard to post-divorce school performance, peer relations, and relations with both the custodial and non-custodial parents. This finding seems to contradict earlier findings ... that show significantly more negative effects of divorce on boys." A. Tuzlak and D. W. Hillock, "Single mothers and their children after divorce: A study of those 'who make it'," in J. E. Veever, **Continuity and Change in Marriage and Family**," page 310.

   (d) A study of 503 inmates of Canadian prisons was recently conducted by the Research and Statistics Branch of the Correctional Service in Canada. In **Forum on Corrections Research**, Volume 3, No. 3, 1991, it was reported that:

> Total number of convictions were compared between offenders who reported no drug use, those who reported irregular drug use (less than once a week) and those who used drugs regularly as teenagers. The findings were surprising. All three groups had a similar average number of convictions: 18.4 for those who reported no drug use as teenagers, 17.7 for irregular drug users, and 18.6 for those who regularly used drugs before they were 18.
> The exact opposite result had been expected – that offenders who frequently took drugs would be involved in a greater number of crimes than those who did not.

3. The data in Table 6.12 comes from Julian Tanner, "Reluctant Rebels: A case study of Edmonton high school drop-outs," **Canadian Review of Sociology and Anthropology**, Volume 27, Number 1, 1990. Tanner took a sample of 152 young people who had left school before completing grade 12.

   (a) If a person is randomly selected from this group, what is the probability that this person:

      i. Has no desire to go back to school?
      ii. Yes, wishes to go back to school?
      iii. Wishes to go to high school and is male?
      iv. Wishes to go to high school or is male?
      v. Is female, given that maybe they have a desire to go back to school?
      vi. Responds 'Maybe' given male?
      vii. Is the sex of the respondent independent of the event of 'No' no desire to go back to school?

   (b) Concerning this data, Tanner comments:

   > Male respondents, especially, were clearly more hesitant about returning to high school than they were about acquiring more education.

   Obtain two conditional probabilities and compare them in order to support Tanner's statement.

Desire to Go Back to School	Sex: Female	Male	Total
No	8	10	18
Maybe	17	12	29
Yes, to High School	28	20	48
Yes, but not to High School	24	33	57
Total	77	75	152

Table 6.12: Desire to Go Back to School by Sex

6.2.12 Sampling With and Without Replacement

One final topic that was glossed over earlier was the discussion of sample selection from a large population which is half male and half female (Section 6.2.7). There a large population was specified so that even when a case, or several cases, were selected from the population, the population remained about half male and half female. This happened because the population was so large that taking out a few males or a few females did not alter the basically equal split between males and females. However, if the population is not all that large, the selection process itself alters the composition of the population, at least if the members of the population are not replaced before further selections from the population are made. In these sampling situations, where the population is not all that large, one distinguishes between **sampling with replacement** and **sampling without replacement**.

In the case of **sampling with replacement**, the member of the population that was selected is replaced in the population, before making the second draw. In this case, the sampling process is the same at each stage and each draw is like starting all over again. In these cases of sampling with replacement, successive draws are independent of each other. This is because each draw is done anew with the whole or original population. The outcome of any draw should in no way affect the probabilities of the outcomes on the next draw. This situation is exactly equivalent to the drawing of a card from a deck of cards, putting the card back in the deck and shuffling the deck well. Then on the next draw, all 52 cards again stand an equal chance of being drawn. One problem with this type of sampling is that the same case may be drawn more than once, especially if the sample size amounts to a reasonably large proportion of the total population, say 10 per cent or more. What one should do with people who are selected twice in surveys is not exactly clear, so that it is more common to sample without replacing the person selected before making the next selection.

Sampling without replacement is the sampling process where those members of the population selected in the sample are not replaced before the next draw or draws are made. Begin with a population of 6 people, 3 females and 3 males. If the first draw is a female, then there are only 5 people left in the population, 3 of whom are male and 2 of whom are female. The probability of drawing a female on the next draw is thus 2/5. In contrast, if a male had been selected on the first draw, there would be 5 people in the population, with 2 being male and 3 being female. In this case, the probability of a female on the second draw would be 3/5. These can be written:

$$P(\text{female on 2nd draw/female on 1st draw}) = \frac{2}{5},$$

$$P(\text{female on 2nd draw/male on 1st draw}) = \frac{3}{5}.$$

In both cases, these conditional probabilities differ from the probability of drawing a female on the first draw (1/2). Thus it can be seen that sampling without replacement leads to dependence of successive draws. When conducting samples, this may affect the probabilities and attention should be paid to how serious this violation of independence is.

6.2.13 Random Variables and Probability Distributions

When determining probabilities from first principles, as in the previous sections, all of the possible outcomes for a probability experiment have been listed. Once this has been done, it is often useful to combine outcomes in various ways. Once these outcomes have been combined, it is then possible to attach probabilities to the reduced set of outcomes and produce a **probability distribution**. An example of this follows, with some definitions and further examples then given.

In Section 6.2.7 the list of all outcomes for a random sample of size 2 in a large population composed of half males and half females, was given in

Table 6.3. Along with the outcomes is the associated set of probabilities. Suppose though that we were not so interested in the set of all possible outcomes, but rather in the probabilities associated with drawing 0, 1 or 2 females in this sample. In this latter case, the original set of all outcomes in Table 6.3 can be combined to show the probability of selecting 0, 1 or 2 females in this sample. This is illustrated in Tables 6.13 and 6.14.

Outcome	Number of Females	Probability
MM	0	1/4
MF	1	1/4
FM	1	1/4
FF	2	1/4

Table 6.13: Probabilities for Number of Females Selected, Sample of Size 2

Table 6.13 shows that the probability of selecting 0 females in the sample is 1/4 and the proability of selecting 2 females in the sample is also 1/4. As can be seen, there are two ways of selecting exactly one female in the sample, outcomes MF or FM. Since these two outcomes are mutually exclusive, the probability of MF or FM is the sum of the probabilities of each of these outcomes. The probability of selecting exactly one female in the sample is thus $1/4 + 1/4 = 1/2$. All this can be combined and presented more compactly by defining a variable X which is defined as the number of females selected in a random sample of size 2 from a large population with half males and half females. Along with the variable X, the associated probabilities can now be given. This probability distribution is given in Table 6.14.

Based on the above considerations, the following definitions provide a means of presenting the outcomes of probability experiments.

Definition 6.2.9 A **random variable** is a function which takes on numerical values, where these values are defined over the set of all outcomes of a probability experiment.

In the above example, X is a random variable, defined as the number of females selected in a sample of size 2, drawn randomly from a large population composed of half males and half females. In general, any rule

X	Probability
0	1/4
1	1/2
2	1/4

Table 6.14: Probability Distribution of X, Number of Females Selected, Sample of Size 2

that assigns numbers to the set of all outcomes of a probability experiment, also defines a random variable. In this way, once the set of outcomes has been determined, then these outcomes can be grouped together according to some criterion. If this grouping is associated with a particular set of numbers, then these values define the values of the random variable. Once the set of possible values for the random variable have been determined, it should then be possible to determine the probability of each of these values. This can be done by combining the probabilities for the individual outcomes.

Definition 6.2.10 A **probability distribution** is a random variable along with the probabilities associated with the values of that random variable.

A probability distribution is much like a frequency distribution. A variable taking on different values, and these values occurr with different probabilities, rather than with different frequencies. The probabilities are usually classical or theoretical probabilities, being derived on the basis of mathematical reasoning. Example 6.2.8 illustrates a simple probability distribution, and Example 6.2.9 compares a frequency and a probability distribution.

Example 6.2.8 Probability Distribution for a Sample of Size 3

If a sample of size three is selected from a large population composed of half males and half females, then the probability distribution for the set of all outcomes was given in Table 6.4. If X is defined as the number of females in the sample, then the probability distribution for X is as shown in Table 6.15. For this table, note that the probability of 0 females is the probability of the event MMM. The probability of exactly 1 female in the sample is the sum of the probabilities of the three mutually exclusive events MMF, MFM and FMM. Each of these has probability $1/8$, *so that the probability of selecting* *exactly one female is* 3/8. *The remainder of the probabilities are derived in a similar manner.*

X	Probability
0	1/8
1	3/8
2	3/8
3	1/8

Table 6.15: Probability Distribution of X, Number of Females Selected, Sample of Size 3

Problem - The Probability Distribution for a Pair of Dice. If a pair of 6 sided dice is rolled, and if random variable X is defined as the sum of the face values that show upon rolling the dice, determine the probability distribution of X. (For the answer, see page 357).

6.2.14 Characteristics of Probability Distributions

The probability distributions shown above are quite similar to frequency distributions for discrete variables, presented earlier in this textbook. In both cases there is a variable X which takes on different numerical values. Instead of the frequencies of occurrence of these values in a frequency distribution, the probability distribution has **probabilities** for the different values of X. In a frequency distribution, if the frequencies are divided by the total number of cases in the sample and the proportional distribution for X is presented, then the proportional and probability distributions look very similar. The only real difference between the two is in the nature of the variable X. In the case of the frequency or proportional distribution, these frequencies and their corresponding proportions are based on observed results of real world experiments or data collection. In the case of the probability distribution, the variable X is a random variable which varies in some manner based on the principles of probability. In this case of a random variable, the proportions represent probabilities of occurrence for the set of possible values of X.

Example 6.2.9 Sample of Size 3

Suppose that 100 random samples, each of size three, are taken from a large population which has approximately half males and half females in it. Suppose that in 10 of these samples all persons selected are males and in 12 of the samples, all persons selected are females. Further suppose that 38 of the samples have 2 males and 1 female, and that 40 of the samples have 1 male and 2 females. Then the frequency distribution, the corresponding proportional distribution and the probability distribution (from Table 6.15, but converted into decimal form) are all presented in Table 6.16.

X	Frequency	Proportion	Probability
0	10	0.10	0.125
1	38	0.38	0.375
2	40	0.40	0.375
3	12	0.12	0.125
Total	100	1.00	1.000

Table 6.16: Frequency, Proportional and Probability Distributions of X, Number of Females Selected, Samples of Size 3

As can be seen in Table 6.16, the frequency and proportional distributions are based on a specific set of data that has been collected. In contrast, the probability distribution is based on theoretical or classical principles of probability. No actual drawing of real samples need be done in order to determine the probabilities for X when it is regarded as a random variable. Rather, the probabilities are based on the principles described earlier.

It should also be noted that frequency and proportional distributions can be regarded as examples of the frequency interpretation of probability. If the experiment of drawing samples were repeated even more times than the 100 shown in Table 6.16, then perhaps the proportional and probability distributions would have even closer values. As it stands, because of the variability of random selection, if only 100 samples are drawn, the proportions based on these 100 samples will not conform exactly with the theoretical probabilities. However, the two distributions can be seen to be fairly similar. In addition, if the two distributions are attempting to determine the same variable, X, then the probabilities form a basis for determining how good the sample is. In this case, the frequencies and proportions do not differ too much from what the theoretical probabilies are, so that this set of 100 samples provides a fairly close approximation to the probabilities. Of course, if there is no theoretical basis for deriving the probabilities, then the proportions would be used as reasonable estimates of what the probabilities might be.

Parameters for the Probability Distributions. Just as the mean and standard deviation, and other statistics, were derived for frequency distributions, so corresponding summary measures can be derived for probability distributions. These measures are referred to as **parameters** when they denote summary measures for theoretical or probability distributions. The corresponding summary measures such as $\bar{X}$ or s are referred to as **statistics** if they are based on actual data that has been obtained and summarized.

Parameters are obtained in much the same manner, and with similar formulae, as are statistics. In this section, the formulae for the mean and standard deviation, as parameters, are given. When working with a theoretical or probability distribution, the parameter for the mean is given the symbol μ, the Greek symbol for the letter mu. For the theoretical or probability distribution, this is defined as

$$\mu = \sum XP(X)$$

The standard deviation is given the symbol σ, the small Greek sigma, when referring to the standard deviation of a theoretical or probability distribution. The variance of such a distribution is then σ^2. These two measures are defined as follows:

$$\sigma^2 = \sum (X - \mu)^2 P(X)$$

$$\sigma = \sqrt{\sum (X - \mu)^2 P(X)}$$

Note that these formulae are analogous to those for $\bar{X}$, s^2 and s, given in Chapter 5. The major difference is that the formulae for σ and σ^2 do not have an n or $n - 1$ in the denominator. In the case of s and s^2, the summation has to be divided by the sum of the frequencies. This is just n, the sample size or $n - 1$, the sample size minus 1, so that a sort of mean of the squares of the deviations about the mean is obtained. In the case of the probability distribution and the associated parameters here, the sum of the probabilities is 1, so in essence, the summation is being divided by 1. But dividing the summation by 1 leaves the value of the summation unchanged. As a result, this 1 does not appear in the formulae. Also note that other parameters could be defined for these theoretical or probability distributions in a manner analogous to the definitions of the statistics presented earlier.

Example 6.2.10 Parameters for Sample of Size 3

For Example 6.2.9, random selection of a sample of size three from a large population composed of half males and half females, the calculations for the mean and standard deviation are shown in Table 6.17.

X	$P(X)$	$XP(X)$	$(X-\mu)$	$(X-\mu)^2P(X)$
0	0.125	0	-1.5	0.28125
1	0.375	0.375	-0.5	0.09375
2	0.375	0.750	0.5	0.09375
3	0.125	0.375	1.5	0.28125
Total	1.000	1.500		0.75000

Table 6.17: Calculations for Parameters of Probability Distribution

Based on Table 6.17, it can be seen that

$$\mu = \sum XP(X) = 1.500$$

$$\sigma^2 = \sum (X-\mu)^2 P(X) = 0.75$$

$$\sigma = \sqrt{\sum (X-\mu)^2 P(X)} = \sqrt{0.75} = 0.866$$

The mean μ can be interpreted as the mean value of X obtained if this sampling experiment is performed an extremely large number of times. This interpretation is very similar to the frequency interpretation of probability. In this case, in the limit, the mean number of females in random samples of size 3, drawn from a large population of half males and half females, is

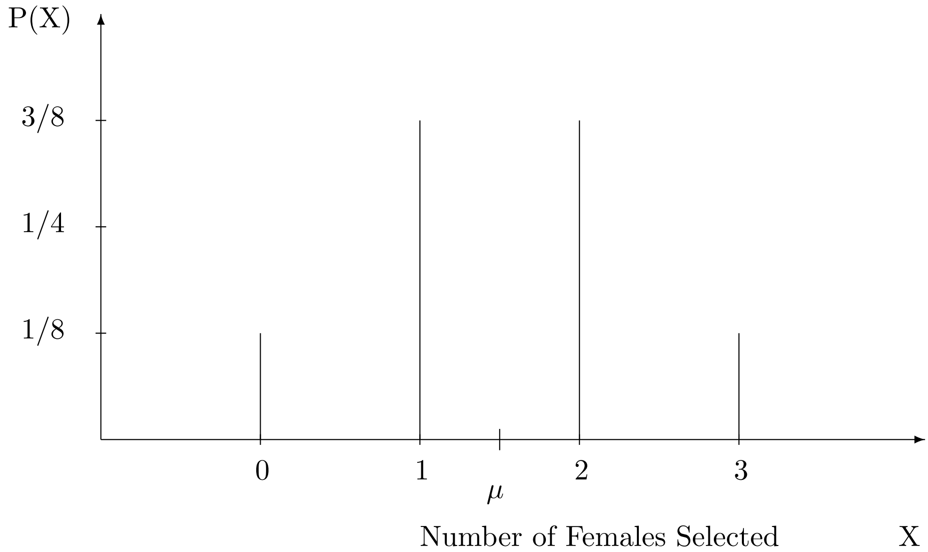

Figure 6.1: Line Chart of Probability Distribution of X, Number of Females Selected in a Sample of Size 3

1.5 females. This results makes intuitive sense in that one would expect the average number of females in this type of sample to be 1.5. Similarly, the standard deviation of 0.866 females can be interpreted as the variablility in the number of females per sample of size 3, drawn from this population. These then become the mean and standard deviation for the theoretical probability distribution of X.

Diagrammatic Representation of Probability Distributions. Examination of Figures 6.1 and 6.2 may help in interpreting the meaning of parameters. Each of these figures present the probability distribution of Table 6.15 diagramatically. Figure 6.1 gives a line chart of the probability distribution for the number of females selected in a random sample of size 3, from a large population half male and half female. The height of each line

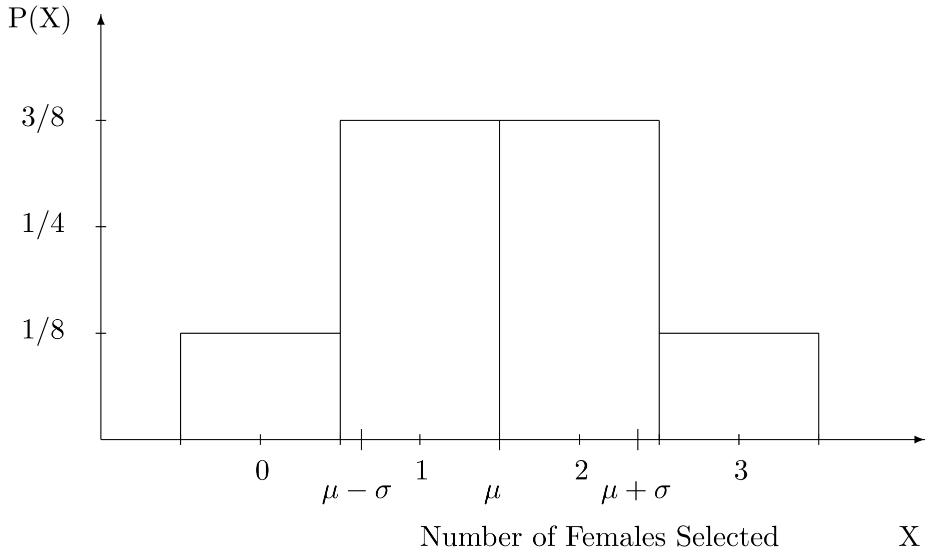

Figure 6.2: Histogram of Probability Distribution of X, Number of Females Selected in a Sample of Size 3

represents the probability of occurrence of each of the number of females in this sample. The mean μ can be seen to be at the centre of the distribution.

Figure 6.2 gives the same distribution as a bar chart or histogram. While the probability distribution is discrete, if bars are drawn around each value of X, centred on the exact value of X, the size of these bars represent the probabilities. Each bar has width one unit, and height equal to the probability for each value of X. For example, for $X = 1$, the bar is one unit wide and 3/8 high, for an area of 3/8. Thus the area of each bar represents the probability of each X, and the sum of the areas of the four bars is 1. Again, the mean μ can be seen to be in the centre of the distribution. The standard deviation σ is a little more difficult to picture. However, in Example 6.2.10, it will be seen that the standard deviation in this distribution is 0.866. The interval of one standard deviation on each side of the mean is given in Figure 6.2, and it can be seen that this interval contains a considerable proportion of the area under the curve.

6.2.15 Suggested Answers for Problems on Probability

Problems Concerning the Interpretation of Probability, page 315

1. The probability that it will snow tomorrow is most likely a **subjective** probability for most of us, although it is likely to be informed by some use of the **frequency** interpretation, based on our knowledge of today's weather and how that is likely to influence tomorrow's weather. Also, we use our memories of weather in the past, at this particular time of year, to improve our subjective estimates. A meteorologist more systematically collects, records and analyzes data on weather. These extensive records provide frequencies of occurrences of different types of weather and these can be used in the frequency interpretation to improve estimates of the probability of various types of weather occurring.

2. The probability that it will rain or snow on June 3 could best be determined using the **frequency** interpretation of probability. In order to do this, though, it would be necessary to obtain weather records over many years. Then the probability of snow could be estimated as the proportion of times that there was snow on previous June 3rds. If one did not have access to any such records and had no knowledge of what weather conditions were like on most June 3rds, then the subjective approach to probability would have to be used. There would seem to be not basis here for using the classical approach.

3. For most of us, this would be no more than a guess, and thus would be a subjective estimate of probability. This is not an experiment that is repeateable, the estimate of the probability of this happening depends on many factors. Those natural scientists who have studied the ozone layer would be in a better position to attach subjective probabilities

to this. But given the impossibility of using either the classical or frequency approach, this probability is in the end a subjective probability.

4. Since this is a random sample, and any random sample is based on theoretical methods, this could be a **classical** interpretation of probability. The only catch is that the proportion of adults in Regina who support the Rhinocerous Party is unknon so that a survey might be required in order to determine the proportion of Rhinocerous Party supporters in Regina. This would provide a **frequency interpretation** of probability to estimate the proportion of Regina adults who support the Rhinocerous party. Then the classical approach could be used to estimate the probability requested.

5. This is a **classical** interpretation. The number of NDP MLAs is known and there are 66 seats so this probability could be calculated using binomial probabilities, without needing to collect any further data.

6. The election is not a repeatable experiment so the principles cannot be easily applied. However, in the same way as with the weather, it is possible to conduct surveys and compare the results with past elections. Using large samples, and the **frequency** interpretation of probability, estimates of the probability of each particular party winning an election can be obtained.

7. It is not possible to repeat this experiment under uniform conditions so, strictly speaking, probability cannot be applied. However, the **subjective** view of probability could be used. It would also be possible to use some aspects of the frequency interpretation, by simulating the results of football games using computer programs. In the end though, the results is highly subjective.

Problems on Probabilities on page 325

1. $P(\text{working class}) = P(C) = 183/607 = 0.301$.
2. $P(\text{has a degree}) = P(U) = 122/607 = 0.201$.
3. $$P(\text{working class and high school or less})$$
$$P(C\ and\ H) = \frac{128}{607} = 0.211.$$
4. $$P(\text{lower class and more than high school})$$
The event of having more than a high school education is the event $P\ or\ U$. The number of outcomes this has in common with being lower class is $4 + 2 = 6$ so that
$$P(D\ and\ (P\ or\ U)) = \frac{4+2}{607} = 0.010.$$
5. $$P(\text{high school and middle or upper middle class})$$
Being middle or upper middle class is being in outcome A or B. These have in common with high school or less, event H, $16 + 152 = 168$ outcomes so that $N(H\ and\ (A\ or\ B)) = 168$.
$$P(H\ and\ (A\ or\ B)) = \frac{168}{607} = 0.277.$$
6. $$P(\text{degree or working class}) = P(U\ or\ C)$$
$$P(U\ or\ C) = P(U) + P(C) - P(UandC) = \frac{122 + 183 - 15}{607}$$
$$= \frac{290}{607} = 0.478.$$

7. This can be broken into various parts. The number with a degree is all outcomes in H or in P and this is $320+165 = 485$. The number who are working or lower class is all those in C or D and this is $183+30 = 213$. We do not want to count any of these outcomes twice and the number of outcomes common to all these is $128 + 40 + 24 + 4 = 196$. The required probability is

$$\frac{485}{607} + \frac{213}{607} - \frac{196}{607} = \frac{502}{607} = 0.827.$$

Problems on Conditional Probabilities, page 328

1.

$$P(B/U) = \frac{81}{122} = 0.664.$$

2.

$$P(C/P) = \frac{N(C \text{ and } P)}{N(P)} = \frac{40}{165} = 0.242.$$

3.

$$P(A/(H \text{ or } P)) = \frac{N(A \text{ and } (H \text{ or } P))}{N(H \text{ or } P)} = \frac{16+21}{320+165} = \frac{37}{485} = 0.076.$$

4. What is requested here is $P(D/(H \text{ or } P))$ and this is $(24 + 4)/485 = 0.058$. This probability and the probability in the previous part are fairly close to each other, in spite of the fact that there are twice as many people who say they are upper middle class as say they are lower class. These probabilities are close because a considerable number of upper middle class people have at least some post-secondary education, while most lower class respondents have completed no more than high school.

5.

$$P(H/B) = \frac{152}{333} = 0.456$$

$$P(H/C) = \frac{128}{183} = 0.699$$

These probabilities provide an idea of how the level of education differs for respondents of different class backgrounds. The probability of having no more than a high school education, given that the respondent claims to be working class is 0.699. Since more middle class respondents have completed higher levels of education, the probability of finding a respondent who has completed no more than high school, given that the respondent is middle class, is only 0.456. These results are consistent with the view that persons higher on the social stratification scale are more likely to have higher levels of education than those lower on the scale.

Problems on Independence and Dependence, page 337

1. (a)

$$P(B/H) = \frac{152}{320} = 0.475$$

$$P(B/P) = \frac{100}{165} = 0.606$$

$$P(B/U) = \frac{81}{122} = 0.664$$

Also note that $P(B) = 333/607 = 0.549$ so that none of events H, P or U is independent of event B, being middle class. However, event P is closest to being independent, whereas both events H and U are more dependent. This means that as education level increases, the probability of being middle class increases, with having some post-secondary education being around the average education level for middle class respondents.

(b)

$$P(H/C) = \frac{128}{183} = 0.699$$

$$P(H) = \frac{320}{607} = 0.527$$

$$P(P/C) = \frac{40}{183} = 0.219$$

$$P(P) = \frac{165}{607} = 0.272$$

$$P(U/C) = \frac{15}{183} = 0.082$$

$$P(U) = \frac{122}{607} = 0.201$$

Based on these, it can be seen that the events of being working class and having some post-secondary education are closer to being independent than are being working class and having either no more than a high school education or, alternatively, having some university. Also, the conditional probabilities decline in a regular fashion showing that the probability of a working class person completing higher levels of education is lower, with each successively higher level of education.

(c)

$$P(D/H) = \frac{24}{320} = 0.075$$

$$P(D/P) = \frac{4}{165} = 0.024$$

$$P(D/U) = \frac{2}{122} = 0.016$$

$$P(D) = \frac{30}{607} = 0.049$$

Event D, being lower class, is dependent on the event of being at each education level. However, the probabilities move in a regular fashion, with the probability of being lower class being smaller for each successively higher level of education. This is more or less as expected since most people who say they are lower class have the lowest level of education.

2. (a) If A is the event of being a woman who is a moderate drinker, B is the event of a woman having a heart attack, and C the event of being a women who is not a moderate drinker, then $P(B/A) < P(B/C)$. This also means $P(B/A) < P(B)$, so that a woman's chances of having a heart attack are reduced from the overall probability if she is a moderate drinker. The event of being a moderate drinker and having a heart attack are **dependent** on each other.

   (b) The two events are **independent** of each other. That is, the lower value (of the Canadian dollar) has had no effect on the amount of shopping in the United States. According to this headline, the two events of the value of the dollar and shopping in the United States have nothing to do with each other.

   (c) The events being referred to here are the post-divorce performance and behaviour of male children, the post-divorce performance and behaviour of female children, and the event of a divorce in the family of the child. The quote says that the post-divorce performance is independent of whether or not the child is a boy or girl. Previous studies appear to have argued that the performance would be worse if the child is a boy rather than a girl. That is, previous studies argued that post-divorce performance is **dependent** on the sex of the child, whereas the current study says that the two are **independent**.

   (d) The event of being convicted appears to be **independent** of the event of whether drugs had been used regularly, irregularly, or not at all. That is, the probability of conviction (at least as measured by number of convictions) appears to be the same for each group.

It was expected by the researchers that the chance or number of convictions would be **dependent** on the type or existence of drug use. While the results are not completely independent, they appear so close to each other that one would argue that they are independent.

3. (a) i. P(no desire to go back to school) $= 18/152 = 0.118$.

ii.

$$P(\text{Yes, wishes to go back to school})$$

$$\text{Prob} = (48 + 57)/152 = 105/152 = 0.691$$

iii.

$$P(\text{high school and male}) = 20/152 = 0.132$$

iv.

$$P(\text{high school or male}) =$$

$$P(\text{high school}) + P(\text{male}) - P(\text{high school and male}) =$$

$$\frac{48}{152} + \frac{75}{152} - \frac{(20)}{152} = \frac{103}{152} = 0.678$$

v.

$$P(\text{female given maybe}) = 17/29 = 0.586$$

vi.

$$P(\text{maybe given male}) = 12/75 = 0.160$$

vii.

$$P(\text{no given female}) = 8/77 = 0.104$$

$$P(\text{no given male}) = 10/75 = 0.133$$

$$P(\text{no}) = 18/152 = 0.118$$

These three probabilities are different than each other, with the probability of no given female being less than the probability of no given male, and the overall probability between these two. The sex of the respondent is dependent on the even of 'no' there is no desire to go back to school. However, the differences are not all that large between these probabilities, so the events are not highly, but only weakly, dependent on the lack of desire to return to school.

(b) For this quote, the author is referring only to males, so these are conditional probabilities given the event of a male respondent. Of the 75 males, 20 wished to return to high school while 33 wished to return to school but not to high school. The conditional probability of wishing to return to high school given the event of being male is $20/75 = 0.267$. This is less than the conditional probability of wishing to return to another type of education given male, that is $33/75 = 0.440$. The lower probability for the former than for the latter would back Tanner's statement.

Probability Distribution for a Pair of Dice. The probability distribution for rolling a pair of dice is given in Table 6.18.

X	$P(X)$
2	1/36
3	2/36
4	3/36
5	4/36
6	5/36
7	6/36
8	5/36
9	4/36
10	3/36
11	2/36
12	1/36

Table 6.18: Probability Distribution of X, Sum of Face Values for 2 Dice