STATISTICS FOR ECONOMISTS: A BEGINNING

John E. Floyd
University of Toronto

July 2, 2010

PREFACE

The pages that follow contain the material presented in my introductory quantitative methods in economics class at the University of Toronto. They are designed to be used along with any reasonable statistics textbook. The most recent textbook for the course was James T. McClave, P. George Benson and Terry Sincich, *Statistics for Business and Economics*, Eighth Edition, Prentice Hall, 2001. The material draws upon earlier editions of that book as well as upon John Neter, William Wasserman and G. A. Whitmore, *Applied Statistics*, Fourth Edition, Allyn and Bacon, 1993, which was used previously and is now out of print. It is also consistent with Gerald Keller and Brian Warrack, *Statistics for Management and Economics*, Fifth Edition, Duxbury, 2000, which is the textbook used recently on the St. George Campus of the University of Toronto. The problems at the ends of the chapters are questions from mid-term and final exams at both the St. George and Mississauga campuses of the University of Toronto. They were set by Gordon Anderson, Lee Bailey, Greg Jump, Victor Yu and others including myself.

This manuscript should be useful for economics and business students enrolled in basic courses in statistics and, as well, for people who have studied statistics some time ago and need a review of what they are supposed to have learned. Indeed, one could learn statistics from scratch using this material alone, although those trying to do so may find the presentation somewhat compact, requiring slow and careful reading and thought as one goes along.

I would like to thank the above mentioned colleagues and, in addition, Adonis Yatchew, for helpful discussions over the years, and John Maheu for helping me clarify a number of points. I would especially like to thank Gordon Anderson, who I have bothered so frequently with questions that he deserves the status of mentor.

After the original version of this manuscript was completed, I received some detailed comments on Chapter 8 from Peter Westfall of Texas Tech University, enabling me to correct a number of errors. Such comments are much appreciated.

J. E. Floyd
July 2, 2010

Chapter 2

Probability

2.1 Why Probability?

We have seen that statistical inference is a methodology through which we learn about the characteristics of a population by analyzing samples of elements drawn from that population. Suppose that a friend asks you to invest \$10000 in a joint business venture. Although your friend's presentation of the potential for profit is convincing, you investigate and find that he has initiated three previous business ventures, all of which failed. Would you think that the current proposed venture would have more than a 50/50 chance of succeeding? In pondering this question you must wonder about the likelihood of observing three failures in a sample of three elements from the process by which your friend chooses and executes business ventures if, in fact, more than half the population of ventures emanating from that process will be successful. This line of thinking is an essential part of statistical inference because we are constantly asking ourselves, in one way or other, what the likelihood is of observing a particular sample if the population characteristics are what they are purported to be. Much of statistical inference involves making an hypothesis about the characteristics of a population (which we will later call the null hypothesis) and then seeing whether the sample has a low or high chance of occurring if that hypothesis is true.

Let us begin our study of probability by starting with a population whose characteristics are known to us and inquire about the likelihood or chances of observing various samples from that population.

2.2 Sample Spaces and Events

Suppose we toss a single coin and observe whether it comes up heads or tails. The relevant population here is the infinite sequence of tosses of a single coin. With each toss there is uncertainty about whether the result will be a head or a tail. This coin toss is an example of a *random trial* or *experiment*, which can be defined as an activity having two or more possible outcomes with uncertainty in advance as to which outcome will prevail. The different possible outcomes of the random trial are called the *basic outcomes*. The set of all basic outcomes for a random trial is called the *sample space* for the trial. The sample space for a single coin toss, which we denote by S, contains two basic outcomes, denoted as H (head) and T (tail). This represents a sample of one from the infinite population of single coin tosses. The set of basic outcomes can be written

$$S = \{H, T\} \tag{2.1}$$

These basic outcomes are also called *sample points* or *simple events*. They are *mutually exclusive*—that is, only one can occur—and *mutually exhaustive*—that is, at least one of them must occur.

Now suppose we toss two coins simultaneously and record whether they come up heads or tails. One might think that there would be three basic outcomes in this case—two heads, head and tail, and two tails. Actually, there are four simple events or sample points because the combination head and tail can occur in two ways—head first and then tail, and tail first followed by head. Thus, the sample space for this random trial or experiment will be

$$S = \{HH, HT, TH, TT\} \tag{2.2}$$

A subset of the set of sample points is called an *event*. For example, consider the event 'at least one head'. This would consist of the subspace

$$E_1 = \{HH, HT, TH\} \tag{2.3}$$

containing three of the four sample points. Another event would be 'both faces same'. This event, which we can call E_2, is the subset

$$E_2 = \{HH, TT\}. \tag{2.4}$$

The set of outcomes not contained in an event E_j is called the *complementary event* to the event E_j which we will denote by E_j^c. Thus, the complementary events to E_1 and E_2 are, respectively,

$$E_1^c = \{TT\} \tag{2.5}$$

and

$$E_2^c = \{HT, TH\}. \tag{2.6}$$

The set of sample points that belongs to both event E_i and event E_j is called the *intersection* of E_i and E_j. The intersection of E_1 and E_2^c turns out to be the event E_2^c because both sample points in E_2^c are also in E_1. We can write this as

$$E_1 \cap E_2^c = \{HT, TH\} = E_2^c. \tag{2.7}$$

The intersection of E_1^c and E_2^c contains no elements, that is

$$E_1^c \cap E_2^c = \phi \tag{2.8}$$

where ϕ means *nil* or nothing. An event containing no elements is called the *null set* or *null event*. When the intersection of two events is the null event, those two events are said to be *mutually exclusive*. It should be obvious that the intersection of an event and its complement is the null event.

The set of sample points that belong to at least one of the events E_i and E_j is called the *union* of E_i and E_j. For example, the union of E_1^c and E_2^c is

$$E_3 = E_1^c \cup E_2^c = \{HT, TH, TT\}, \tag{2.9}$$

the event 'no more than one head'. Each sample point is itself an event—one of the elementary events—and the union of all these elementary events is the sample space itself. An event that contains the entire sample space is called the *universal event*.

We can express the intersection and union of several events as, respectively,

$$E_1 \cap E_2 \cap E_3 \cap E_4 \cap \cdots\cdots$$

and

$$E_1 \cup E_2 \cup E_3 \cup E_4 \cup \cdots\cdots.$$

The set of all possible events that can occur in any random trial or experiment, including both the universal event and the null event, is called the *event space*.

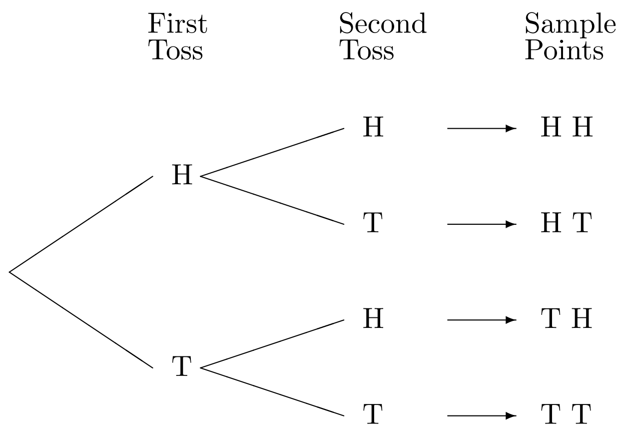
The above examples of random trials and sample spaces resulting therefrom represent perhaps the simplest cases one could imagine. More complex situations arise in experiments such as the daily change in the Dow Jones Industrial Average, the number of students of the College involved in accidents in a given week, the year-over-year rate of inflation in the United Kingdom, and so forth. Sample points, the sample space, events and the event space in these more complicated random trials have the same meanings and are defined in the same way as in the simple examples above.

2.3 Univariate, Bivariate and Multivariate Sample Spaces

The sample space resulting from a single coin toss is a univariate sample space—there is only one dimension to the random trial. When we toss two coins simultaneously, the sample space has two dimensions—the result of the first toss and the result of the second toss. It is often useful to portray bivariate sample spaces like this one in tabular form as follows:

	One	
Two	H	T
H	HH	TH
T	HT	TT

Each of the four cells of the table gives an outcome of the first toss followed by an outcome of the second toss. This sample space can also be laid out in tree form:

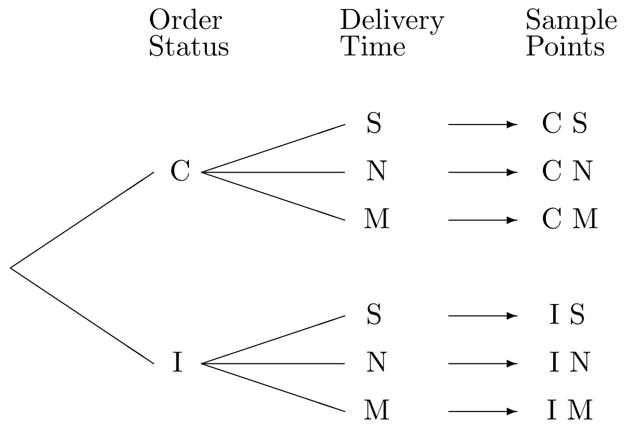


A more interesting example might be the parts delivery operation of a firm supplying parts for oil drilling rigs operating world wide. The relevant random trial is the delivery of a part. Two characteristics of the experiment are of interest—first, whether the correct part was delivered and second, the number of days it took to get the part to the drilling site. This is also a bivariate random trial the essence of which can be captured in the following table:

		Time of Delivery		
		S	N	M
Order	C	C S	C N	C M
Status	I	I S	I N	I M

The status of the order has two categories: 'correct part' (C) and 'incorrect part' (I). The time of delivery has three categories: 'same day' (S), 'next day' (N) and 'more than one day' (M). There are six sample points or basic outcomes. The top row in the table gives the event 'correct part' and the bottom row gives the event 'incorrect part'. Each of these events contain three sample points. The first column on the left in the main body of the table gives the event 'same day delivery', the middle column the event 'next day delivery' and the third column the event 'more than one day delivery'. These three events each contain two sample points or basic outcomes. The event 'correct part delivered in less than two days' would be the left-most two sample points in the first row, (C S) and (C N). The complement of that event, 'wrong part or more than one day delivery' would be the remaining outcomes (C M), (I S), (I N) and (I M).

Notice also that the basic outcome in each cell of the above table is the intersection of two events—(C S) is the intersection of the event C or 'correct part' and the event S 'same day delivery' and (I N) is the intersection of the event I, 'incorrect part', and the event N 'next day delivery'. The event 'correct part' is the union of three simple events, (C S) $\cup$ (C N) $\cup$ (C M). The parts delivery sample space can also be expressed in tree form as follows:



2.4 The Meaning of Probability

Although probability is a term that most of us used before we began to study statistics, a formal definition is essential. As we noted above, a random trial is an experiment whose outcome must be one of a set of sample points with uncertainty as to which one it will be. And events are collections of sample points, with an event occurring when one of the sample points or basic outcomes it contains occurs. *Probability* is a value attached to a sample point or event denoting the likelihood that it will be realized. These probability assignments to events in the sample space must follow certain rules.

1. The probability of any basic outcome or event consisting of a set of basic outcomes must be between zero and one. That is, for any outcome o_i or event E_i containing a set of outcomes we have

$$\begin{aligned} 0 \leq P(o_i) \leq 1 \\ 0 \leq P(E_i) \leq 1. \end{aligned} \tag{2.10}$$

If $P(o_i) = 1$ or $P(E_i) = 1$ the respective outcome or event is certain to occur; if $P(o_i) = 0$ or $P(E_i) = 0$ the outcome or event is certain not to occur. It follows that probability cannot be negative.

2. For any set of events of the sample space S (and of the event space E),

$$P(E_j) = \sum_{i=1}^{J} P(o_i). \tag{2.11}$$

where J is the number of basic events or sample points contained in the event E_j. In other words, the probability that an event will occur is the sum of the probabilities that the basic outcomes contained in that event will occur. This follows from the fact that an event is said to occur when one of the basic outcomes or sample points it contains occurs.

3. Since it is certain that at least one of the sample points or elementary events in the sample space will occur, $P(S) = 1$. And the null event cannot occur, so $P(\phi) = 0$ where ϕ is the null event. These results follow from the fact that $P(S)$ is the sum of the probabilities of all the simple or basic events.

A number of results follow from these postulates

- $P(E_i) \leq P(E_j)$ when E_i is a subset of (contained in) E_j.
- If E_i and E_j are mutually exclusive events of a sample space, then $P(E_i \cap E_j) = 0$. That is, both events cannot occur at the same time.

Probability can be expressed as an *odds ratio*. If the probability of an event E_j is a, then the odds of that event occurring are a to $(1 - a)$. If the probability that you will get into an accident on your way home from work tonight is .2, then the odds of you getting into an accident are .2 to .8 or 1 to 4. If the odds in favour of an event are a to b then the probability of the event occurring is

$$\frac{a}{a+b}$$

If the odds that your car will break down on the way home from work are 1 to 10, then the probability it will break down is $1/(10 + 1) = 1/11$.

2.5 Probability Assignment

How are probabilities established in any particular case? The short answer is that we have to assign them. The probability associated with a random trial or experiment can be thought of as a mass or "gob" of unit weight. We have to distribute that mass across the sample points or basic elements in the sample space. In the case of a single coin toss, this is pretty easy to do. Since a fair coin will come up heads half the time and tails half the time we will assign half of the unit weight to H and half to T, so that the probability of a head on any toss is .5 and the probability of a tail is .5. In the case where we flip two coins simultaneously our intuition tells us that each of the four sample points HH, HT, TH, and TT are equally likely, so we would assign a quarter of the mass, a probability of .25, to each of them. When it comes to determining the probability that I will be hit by a car on my way home from work tonight, I have to make a wild guess on the basis of information I might have on how frequently that type of accident occurs between 5 o'clock and 6 o'clock on weekday afternoons in my neighbourhood and how frequently I jay-walk. My subjective guess might be that there is about one chance in a thousand that the elementary event 'get hit by a car on my way home from work' will occur and nine-hundred and ninety-nine chances in a thousand that the mutually exclusive elementary event 'do not get hit by a car on my way home from work' will occur. So I assign a probability of .001 to the event 'get hit' and a probability of .999 to the event 'not get hit'. Note that the implied odds of me getting hit are 1 to 999.

As you might have guessed from the above discussion the procedures for assigning probabilities fall into two categories—objective and subjective. In the case of coin tosses we have what amounts to a mathematical model of a fair coin that will come up heads fifty percent of the time. If the coin is known to be fair this leads to a purely objective assignment of probabilities—no personal judgement or guesswork is involved. Of course, the proposition that the coin is fair is an assumption, albeit a seemingly reasonable one. Before assigning the probabilities in a coin toss, we could toss the coin a million times and record the number of times it comes up heads. If it is a fair coin we would expect to count 500,000 heads. In fact, we will get a few more or less than 500,000 heads because the one million tosses is still only a sample, albeit a large one, of an infinite population. If we got only 200,000 heads in the 1,000,000 tosses we would doubt that the coin was a fair one. A theoretically correct assignment of probabilities would be one based on the frequencies in which the basic outcomes occur in an infinite sequence of experiments where the conditions of the experiment do not change. This uses a basic axiom of probability theory called the *law of large numbers*. The law states essentially that the relative frequency of occurrence of a sample point approaches the theoretical probability of the outcome as the experiment is repeated a larger and larger number of times and the frequencies are cumulated over the repeated experiments. An example is shown in Figure 2.1 where a computer generated single-coin toss is performed 1500 times. The fraction of tosses turning up heads is plotted against the cumulative number of tosses measured in hundreds.

In practice, the only purely objective method of assigning probabilities occurs when we know the mathematics of the data generating process—for example, the exact degree of 'fairness' of the coin in a coin toss. Any non-objective method of assigning probabilities is a subjective method, but subjective assignments can be based on greater or lesser amounts of information, according to the sample sizes used to estimate the frequency of occurrence of particular characteristics in a population. When relative frequencies are used to assign probabilities the only subjective component is the choice of the data set from which the relative frequencies are obtained. For this reason, the assignment of probabilities based on relative frequencies is often also regarded as objective. In fact, inferential statistics essentially involves the use of sample data to try to infer, as objectively as possible, the proximate probabilities of events in future repeated experiments or random draws from a population. Purely subjective assignments of probabilities are those that use neither a model of the data-generating process nor data on relative frequencies.

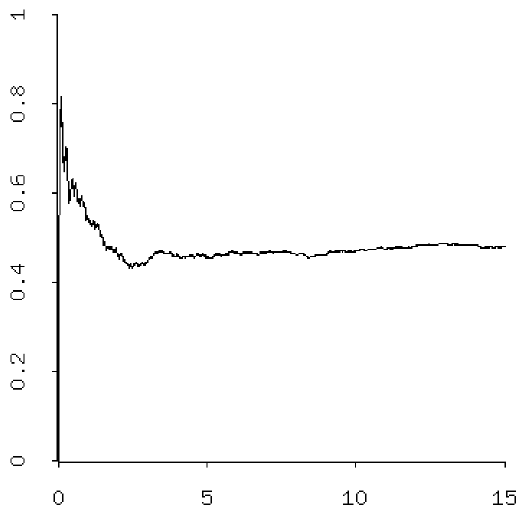


Figure 2.1: Illustration of the law of large numbers. Computer generated plot of the cumulative fraction of 1500 single coin-tosses turning up heads. The horizontal axis gives the number of tosses in hundreds and the vertical axis the fraction turning up heads.

Purely subjective probability measures tend to be useful in business situations where the person or organization that stands to lose from an incorrect assignment of probabilities is the one making the assignment. If I attach a probability of 0.1 that a recession will occur next year and govern my investment plans accordingly, I am the one who stands to gain or lose if the event 'recession' occurs and I will be the loser over the long run if my probability assignments tend to be out of line with the frequency with which the event occurs. Since I stand to gain or lose, my probability assessment is 'believable' to an outside observer—there is no strategic gain to me from 'rigging' it. On the other hand, if the issue in question is the amount my industry will lose from free trade, then a probability assignment I might make to the set of sample points comprising the whole range of losses that could be incurred should not be taken seriously by policy makers in deciding how much compensation, if any, to give to my industry. Moreover, outside observers' subjective probability assignments are also suspect because one does not know what their connection might happen to be to firms and industries affected by proposed policy actions.

2.6 Probability Assignment in Bivariate Sample Spaces

Probability assignment in bivariate sample spaces can be easily visualized using the following table, which further extends our previous example of world-wide parts delivery to oil drilling sites.

		Time of Delivery			
		S	N	M	Sum
Order	C	.600	.24	.120	.96
Status	I	.025	.01	.005	.04
	Sum	.625	.25	.125	1.00

Probabilities have been assigned to the six elementary events either purely subjectively or using frequency data. Those probabilities, represented by the numbers in the central enclosed rectangle must sum to unity because they cover the entire sample space—at least one of the sample points must occur. They are called *joint probabilities* because each is an intersection of two events—an 'order status' event (C or I) and a 'delivery time' event (S, N, or M). The probabilities in the right-most column and and along the bottom row are called *marginal probabilities.* Those in the right margin give the probabilities of the events 'correct' and 'incorrect'. They are the unions of the joint probabilities along the respective rows and they must sum to unity because the order delivered must be either correct or incorrect. The marginal probabilities along the bottom row are the probabilities of the events 'same day delivery' (S), 'next day delivery' (N) and 'more than one day to deliver' (M). They are the intersections of the joint probabilities in the respective columns and must also sum to unity because all orders are delivered eventually. You can read from the table that the probability of the correct order being delivered in less than two days is $.60 + .24 = .84$ and the probability of unsatisfactory performance (either incorrect order or two or more days to deliver) is $(.12 + .025 + .01 + .005) = .16 = (1 - .84)$.

2.7 Conditional Probability

One might ask what the probability is of sending the correct order when the delivery is made on the same day. Note that this is different than the probability of both sending the correct order *and* delivering on the same day. It is the probability of getting the order correct *conditional upon* delivering on the same day and is thus called a *conditional probability.* There are two things that can happen when delivery is on the same day—the order sent can be correct, or the incorrect order can be sent. As you can see from the table a probability weight of .600 + .025 = .625 is assigned to same-day delivery. Of this probability weight, the fraction .600/.625 = .96 is assigned to the event 'correct order' and the fraction .25/.625 = .04 is assigned to the event 'incorrect order'. The probability of getting the order correct conditional upon same day delivery is thus .96 and we define the conditional probability as

$$P(C|S) \quad = \quad \frac{P(C \cap S)}{P(S)}. \tag{2.12}$$

where $P(C|S)$ is the probability of C occurring conditional upon the occurrence of S, $P(C \cap S)$ is the joint probability of C and S (the probability that both C and S will occur), and $P(S)$ is the marginal or unconditional probability of S (the probability that S will occur whether or not C occurs). The definition of conditional probability also implies, from manipulation of (2.12), that

$$P(C \cap S) \quad = \quad P(C|S)P(S). \tag{2.13}$$

Thus, if we know that the conditional probability of C given S is equal to .96 and that the marginal probability of C is .625 but are not given the joint probability of C and S, we can calculate that joint probability as the product of .625 and .96 —namely .600.

2.8 Statistical Independence

From application of (2.12) to the left-most column in the main body of the table we see that the conditional probability distribution of the event 'order status' given the event 'same day delivery' is

$$\begin{array}{rl} P(C|S) & .96 \\ P(I|S) & .04 \end{array}$$

which is the same as the marginal probability distribution of the event 'order status'. Further calculations using (2.12) reveal that the probability distributions of 'order status' conditional upon the events 'next day delivery' and 'more than one day delivery' are

$$\begin{array}{rl} P(C|N) & .24/.25 = .96 \\ P(I|N) & .01/.25 = .04 \end{array}$$

and

$$\begin{array}{rl} P(C|M) & .120/.125 = .96 \\ P(I|M) & .005/.125 = .04 \end{array}$$

which are the same as the marginal or unconditional probability distribution of 'order status'. Moreover, the probability distributions of 'time of delivery' conditional upon the events 'correct order' and 'incorrect order' are, respectively

$$\begin{array}{rl} P(S|C) & .60/.96 = .625 \\ P(N|C) & .24/.96 = .25 \\ P(M|C) & .12/.96 = .125 \end{array}$$

and

$$\begin{array}{rl} P(S|I) & .025/.04 = .625 \\ P(N|I) & .010/.04 = .25 \\ P(M|I) & .005/.04 = .125 \end{array}$$

which are the same as the marginal or unconditional probability distribution of 'time of delivery'. Since the conditional probability distributions are the same as the corresponding marginal probability distributions, the probability of getting the correct order is the same whether delivery is on the same day or on a subsequent day—that is, independent of the day of delivery. And the probability of delivery on a particular day is independent of whether or not the order is correctly filled. Under these conditions the two events 'order status' and 'time of delivery' are said to be *statistically independent.* Statistical independence means that the marginal and conditional probabilities are the same, so that

$$P(C|S) \quad = \quad P(C). \tag{2.14}$$

The case where two events are not statistically independent can be illustrated using another example. Suppose that we are looking at the behaviour of two stocks listed on the New York Stock Exchange—Stock A and Stock B—to observe whether over a given interval the prices of the stocks increased, decreased or stayed the same. The sample space, together with the probabilities assigned to the sample points based on several years of data on the price movements of the two stocks can be presented in tabular form as follows:

Stock B		Stock A Increase A_1	No Change A_2	Decrease A_3	Sum
Increase	B_1	.20	.05	.05	.30
No Change	B_2	.15	.10	.15	.40
Decrease	B_3	.05	.05	.20	.30
	Sum	.40	.20	.40	1.00

The conditional probability that the price of stock A will increase, given that the price of stock B increases is

$$\begin{aligned} P(A_1|B_1) &= \frac{P(A_1 \cap B_1)}{P(B_1)} \\ &= \frac{.20}{.30} = .666 \end{aligned}$$

which is greater than the unconditional probability of an increase in the price of stock A, the total of the A_1 column, equal to .4. This says that the probability that the price of stock A will increase is greater if the price of stock B also increases. Now consider the probability that the price of stock A will fall, conditional on a fall in the price of stock B. This equals

$$\begin{aligned} P(A_3|B_3) &= \frac{P(A_3 \cap B_3)}{P(B_3)} \\ &= \frac{.20}{.30} = .666 \end{aligned}$$

which is greater than the 0.4 unconditional probability of a decline in the price of stock A given by the total at the bottom of the A_3 column. The probability that the price of stock A will decline conditional upon the price of stock B not declining is

$$\begin{aligned}\frac{P(A_3 \cap B_1) + P(A_3 \cap B_2)}{P(B_1) + P(B_2)} &= \frac{.05 + .15}{.30 + .40} \\ &= \frac{20}{70} = .286\end{aligned}$$

which is smaller than the 0.4 unconditional probability of the price of stock A declining regardless of what happens to the price of stock B. The price of stock A is more likely to decline if the price of stock B declines and less likely to decline if the price of stock B does not decline. A comparison of these conditional probabilities with the relevant unconditional ones make it clear that the prices of stock A and stock B move together. They are not statistically independent.

There is an easy way to determine if the two variables in a bivariate sample space are statistically independent. From the definition of statistical independence (2.14) and the definition of conditional probability as portrayed in equation (2.13) we have

$$P(C \cap S) = P(C|S)P(S) = P(C)P(S). \tag{2.15}$$

This means that when there is statistical independence the joint probabilities in the tables above can be obtained by multiplying together the two relevant marginal probabilities. In the delivery case, for example, the joint probability of 'correct order' and 'next day' is equal to the product of the two marginal probabilities .96 and .25, which yields the entry .24. The variables 'order status' and 'time of delivery' are statistically independent. On the other hand, if we multiply the marginal probability of A_1 and the marginal probability of B_1 in the stock price change example we obtain $.30 \times .40 = .12$ which is less than .20, the actual entry in the joint probability distribution table. This indicates that the price changes of the two stocks are not statistically independent.

2.9 Bayes Theorem

Many times when we face a problem of statistical inference about a population from a sample, we already have some information prior to looking at the sample. Suppose, for example, that we already know that the probabilities that an offshore tract of a particular geological type contains no gas (A_1), a minor gas deposit (A_2) or a major gas deposit (A_3) are .7, .25 and .05 respectively. Suppose further that we know that a test well drilled in a tract like the one in question will yield no gas (B_1) if none is present and will yield gas (B_2) with probability .3 if a minor deposit is present and with probability .9 if a major deposit is present. A sensible way to proceed is to begin with the information contained in the probability distribution of gas being present in the tract and then upgrade that probability distribution on the basis of the results obtained from drilling a test well. Our procedure can be organized as follows:

	Prior Probability $P(A_i)$	$P(B_2\|A_i)$	Joint Probability $P(A_i \cap B_2)$	Posterior Probability $P(A_i\|B_2)$
No Gas (A_1)	0.70	0.00	0.000	0.000
Minor Deposit (A_2)	0.25	0.30	0.075	0.625
Major Deposit (A_3)	0.05	0.90	0.045	0.375
Total	1.00		0.120	1.000

Suppose that our test well yields gas (otherwise it's game over!). We begin with our prior probabilities $P(A_i)$ and then use the fact that the joint probability distribution $P(B_2 \cap A_i)$ equals the prior probabilities multiplied by the conditional probabilities $P(B_2|A_i)$ that gas will be obtained, given the respective A_i,

$$P(B_2 \cap A_i) = P(B_2|A_i)P(A_i).$$

These probabilities are entered in the second column from the right. Their sum gives the probability of finding gas, which equals .12 (the probability of finding gas and there being no gas (0.000) plus the probability of finding gas and there being a minor deposit (0.075) plus the probability of finding gas and there being a major deposit (0.045)). It then follows that the probability of there being no gas conditional upon gas being found in the test well is $0.000/.12 = 0.000$, the probability of there being a minor deposit conditional upon the test well yielding gas is $.075/.12 = .625$ and the probability of there being a major deposit conditional upon gas being found in the test well is

.045/.12 = .375. Since the test well yielded gas, these latter probabilities are the posterior (post-test or post-sample) probabilities of there being no gas, a minor deposit and a major deposit. They are entered in the column on the extreme right. When we are finished we can say that there is a .625 probability that the tract contains a minor gas deposit and a .375 probability that it contains a major deposit.

Notice what we have done here. We have taken advantage of the fact that the joint probability distribution $P(A_i \cap B_j)$ can be obtained in two ways:

$$P(A_i \cap B_j) = P(A_i|B_j)\, P(B_j)$$

and

$$P(A_i \cap B_j) = P(B_j|A_i)\, P(A_i).$$

Subtracting the second of these from the first, we obtain

$$P(A_i|B_j)\, P(B_j) \quad = \quad P(B_j|A_i)\, P(A_i)$$

which implies

$$P(A_i|B_j) \quad = \quad P(B_j|A_i)\, \frac{P(A_i)}{P(B_j)} \tag{2.16}$$

We can then use the fact that

$$P(B_j) \quad = \quad \sum_i P(B_j \cap A_i) \quad = \quad \sum_i [P(B_j|A_i)\, P(A_i)] \tag{2.17}$$

to express (2.16) as

$$P(A_i|B_j) \quad = \quad \frac{P(B_j|A_i)\, P(A_i)}{\sum_i [P(B_j|A_i)\, P(A_i)]} \tag{2.18}$$

This latter equation is called *Bayes Theorem*. Given the prior probability distribution $P(A_i)$ (the marginal or unconditional probabilities of gas being present) plus the conditional probability distribution $P(B_j|A_i)$ (the probabilities of finding gas conditional upon it being not present, present in a minor deposit or present in a major deposit), we can calculate the posterior probability distribution (probabilities of no deposit or a minor or major deposit being present conditional upon the information obtained from drilling a test hole).

The operation of Bayes Theorem can perhaps best be understood with reference to a tabular delineation of the sample space of the sort used in the parts delivery case.

Type of Deposit	Test Drill No Gas (B_1)	Gas (B_2)	Prior Probability Distribution
(A_1)		0.000	0.70
(A_2)		0.075	0.25
(A_3)		0.045	0.05
Total		0.120	1.00

On the basis of our previous calculations we are able to fill in the right-most two columns. The column on the extreme right gives the prior probabilities and the second column from the right gives the joint probabilities obtained by multiplying together the prior probabilities and the probabilities of finding gas in a test well conditional upon its absence or minor or major presence in the tract. We can fill in the missing column by subtracting the second column from the right from the right-most column. This yields

Type of Deposit	Test Well No Gas (B_1)	Gas (B_2)	Prior Probability Distribution
(A_1)	0.700	0.000	0.70
(A_2)	0.175	0.075	0.25
(A_3)	0.005	0.045	0.05
Total	0.880	0.120	1.00

We can now see from the bottom row that the probability of not finding gas in a test well drilled in this type of tract is .88. The posterior probabilities conditional upon finding no gas or gas, respectively, in the test well can be calculated directly from the table by taking the ratios of the numbers in the two columns to the unconditional probabilities at the bottoms of those columns. The posterior probabilities are therefore

Type of Deposit	Posterior Probabilities No Gas $(A_i \mid B_1)$	Gas $(A_i \mid B_2)$	Prior Probability Distribution
(A_1)	0.795	0.000	0.70
(A_2)	0.199	0.625	0.25
(A_3)	0.006	0.375	0.05
Total	1.000	1.000	1.00

Notice how the prior probabilities are revised as a consequence of the test results. The prior probability of no gas being present is .70. If the test well yields no gas, that probability is adjusted upward to .795 and if the test well yields gas it is adjusted downward to zero. The prior probability that there is a minor deposit in the tract is .25. If the test well yields no gas this is adjusted downward to less than .2 while if gas is found in the test well this probability is adjusted upward to .625. Note that it is possible for gas to be present even if the test well yields no gas (gas could be present in another part of the tract) while if there is no gas present the test well will not find any. Finally, the prior probability of there being a major deposit present is adjusted upward from .05 to .375 if the test well yields gas and downward to .006 if the test well finds no gas.

2.10 The AIDS Test

Now consider another application of Bayes Theorem. You go to your doctor for a routine checkup and he tells you that you have just tested positive for HIV. He informs you that the test you have been given will correctly identify an AIDS carrier 90 percent of the time and will give a positive reading for a non-carrier of the virus only 1 percent of the time. He books you for a second more time consuming and costly but absolutely definitive test for Wednesday of next week.

The first question anyone would ask under these circumstances is "Does this mean that I have a 90 percent chance of being a carrier of HIV." On the way home from the doctor's office you stop at the library and rummage through some medical books. In one of them you find that only one person per thousand of the population in your age group is a carrier of aids.[1] You think "Am I so unfortunate to be one of these?" Then you remember about Bayes Theorem from your statistics class and decide to do a thorough analysis. You arrange the sample space as follows

An HIV Carrier?	Test Result Positive (T_1)	Negative (T_0)	Prior Probability Distribution
No (A_0)	0.0099		0.999
Yes (A_1)	0.0009		0.001
Total	0.0108		1.000

and make special note that the test results give you some conditional probabilities. In particular, the probability of a positive result conditional upon

[1] These numbers, indeed the entire scenario, should not be taken seriously—I am making everything up as I go along!

you being a carrier is $P(T_1|A_1) = .90$ and the probability of a positive result conditional upon you not being a carrier is $P(T_1|A_0) = .01$. You obtain the joint probability of being a carrier and testing positive by multiplying $P(T_1|A_1)$ by $P(A_1)$ to obtain $.90 \times .001 = .0009$ and enter it into the appropriate cell of the above table. You then obtain the joint probability of testing positive and not being a carrier by multiplying $P(T_1|A_0)$ by $P(A_0)$. This yields $.01 \times .999 = .0099$ which you enter appropriately in the above table. You then sum the numbers in that column to obtain the unconditional probability of testing positive, which turns out to be .0108. You can now calculate the posterior probability—that is, the probability of being a carrier conditional on testing positive. This equals $.0009/.0108 = .08333$. The information from the test the doctor gave you has caused you to revise your prior probability of .001 upward to .0833. You can now fill in the rest of the table by subtracting the joint probabilities already there from the prior probabilities in the right margin.

An HIV Carrier?	Test Result Positive (T_1)	Negative (T_0)	Prior Probability Distribution
No (A_0)	0.0099	.9891	0.999
Yes (A_1)	0.0009	.0001	0.001
Total	0.0108	.9892	1.000

Notice the importance to this problem of the 1% conditional probability of testing positive when you don't carry HIV. If that conditional probability were zero then the fact that the test will come up positive for a carrier 90% of the time is irrelevant. The joint probability of testing positive and not being a carrier is zero. A carrier of HIV will sometimes test negative but a non-carrier will never test positive. The above tabular representation of the bivariate sample space then becomes

An HIV Carrier?	Test Result Positive (T_1)	Negative (T_0)	Prior Probability Distribution
No (A_0)	0.0000	.999	0.999
Yes (A_1)	0.0009	.0001	0.001
Total	0.0009	.9991	1.000

The probability that you carry HIV conditional upon testing positive is now $.0009/.0009 = 1.000$. You are a carrier.

2.11 Basic Probability Theorems

This chapter concludes with a statement of some basic probability theorems, most of which have already been motivated and developed and all of which will be used extensively in the chapters that follow. These theorems are best understood with reference to the Venn diagram presented in Figure 2.2. The area inside the square denotes the sample space with each point representing a sample point. The circular areas E_1, E_2 and E_3 represent events—the points inside these areas are those points belonging to the sample space contained in the respective events. The letters A, B, C and D denote collections of sample points inside the respective events. For example the event E_1 consists of $A + B$, event E_2 consists of $B + C$. And the area D represents event E_3. The probability theorems below apply to any two events of a sample space.

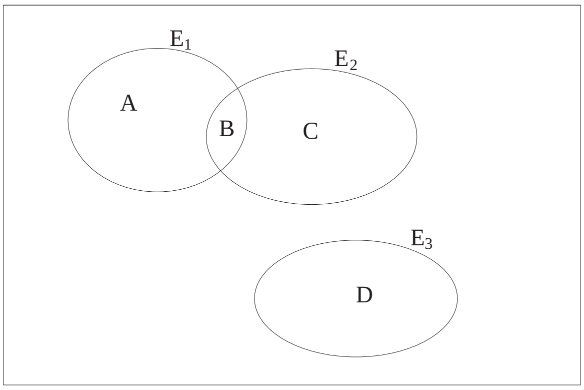


Figure 2.2: Venn diagram to illustrate basic probability theorems. The rectangle contains the sample space and the circular areas denote events E_1, E_2 and E_3.

1. **Addition**

$$P(E_1 \cup E_2) \quad = \quad P(E_1) + P(E_2) - P(E_1 \cap E_2) \tag{2.19}$$

The probabilities of the two events are added together and then the joint probability of the two events, given by the probability mass associated with the area B in Figure 2.2 is subtracted out to avoid double counting. If the events are mutually exclusive, as in the case of E_1 and E_3 the joint probability term will be zero,

$$P(E_1 \cup E_3) \quad = \quad P(E_1) + P(E_3). \tag{2.20}$$

2. **Complementation**

$$P(E_1) \quad = \quad 1 - P(E_1^c) \tag{2.21}$$

where E_1^c is the complementary event to E_1.

3. **Multiplication**

$$P(E_1 \cap E_2) \quad = \quad P(E_1)\, P(E_2|E_1). \tag{2.22}$$

This follows from the definition of conditional probability. In Figure 2.2, $P(E_2|E_1) = P(B)/(P(A) + P(B))$ (the proportion of the total weight in E_1 that also falls in E_2). And $P(E_1) = P(A)+P(B)$. So $P(E_1 \cap E_2) = [P(A)+P(B)][P(B)/(P(A) + P(B))] = P(B)$. If we know the joint probability and the marginal probability we can find the conditional probability. Similarly, if we know the conditional probability and the marginal probability we can find the joint probability.

2.12 Exercises

1. Suppose a random trial has three basic outcomes: o_1, o_2 and o_3. The probabilities of o_2 and o_3 are .5 and .4 respectively. Let E be the event consisting of basic outcomes o_2 and o_3. The probability of the complementary event to E is

a) .1

b) .9

c) .8

d) .2

e) none of the above.

2. Two marbles are drawn at random and without replacement from a box containing two blue marbles and three red marbles. Determine the probability of observing the following events.

a) Two blue marbles are drawn.

b) A red and a blue marble are drawn.

c) Two red marbles are drawn.

Hint: Organize the sample space according to a tree-diagram and then attach probabilities to the respective draws. Alternatively, you can organize the sample space in rectangular fashion with one draw represented as rows and the other as columns.

3. Three events, A, B, and C are defined over some sample space S. Events A and B are independent. Events A and C are mutually exclusive. Some relevant probabilities are $P(A) = .04$, $P(B) = .25$, $P(C) = .2$ and $P(B|C) = .15$. Compute the values of $P(A \cup B)$, $P(A \cup C)$, $P(A \cup B \cup C)$ and $P(C|B)$.

4. An experiment results in a sample space S containing five sample points and their associated probabilities of occurrence:

s_1	s_2	s_3	s_4	s_5
.22	.31	.15	.22	.10

The following events have been defined

- $E_1 = \{s_1, s_3\}$.
- $E_2 = \{s_2, s_3, s_4\}$.
- $E_3 = \{s_1, s_5\}$.

Find each of the following probabilities:

a) $P(E_1)$.

b) $P(E_2)$.

c) $P(E_1 \cap E_2)$.

d) $P(E_1|E_2)$.

e) $P(E_2 \cap E_3)$.

f) $P(E_3|E_2)$.

Consider each pair of events E_1 and E_2, E_1 and E_3 and E_2 and E_3. Are any of these events statistically independent? Why or why not? Hint: Are the joint probabilities equal to the products of the unconditional probabilities?

5. Roulette is a very popular game in Las Vegas. A ball spins on a circular wheel that is divided into 38 arcs of equal length, bearing the numbers 00, 0, 1, 2, . . . , 35, 36. The number of the arc on which the ball stops after each spin of the wheel is the outcome of one play of the game. The numbers are also coloured as follows:

Red: 1,3,5,7,9,12,14,16,18,19,21,23,25,27,30,32,34,36
Black: 2,4,6,8,10,11,13,15,17,20,22,24,26,28,29,31,33,35
Green: 00,0

Players may place bets on the table in a variety of ways including bets on odd, even, red, black, high, low, etc. Define the following events:

- A: Outcome is an odd number (00 and 0 are considered neither even nor odd).
- B: Outcome is a black number.
- C: Outcome is a low number, defined as one of numbers 1–18 inclusive.

a) What is the sample space here?

b) Define the event $A \cap B$ as a specific set of sample points.

c) Define the event $A \cup B$ as a specific set of sample points.

d) Find $P(A)$, $P(B)$, $P(A \cup B)$, $P(A \cap B)$ and $P(C)$.

e) Define the event $A \cap B \cap C$ as a specific set of sample points.

f) Find $P(A \cup B)$.

g) Find $P(A \cap B \cap C)$.

h) Define the event $A \cup B \cup C$ as a specific set of sample points.

6. A bright young economics student at Moscow University in 1950 criticized the economic policies of the great leader Joseph Stalin. He was arrested and sentenced to banishment for life to a work camp in the east. In those days 70 percent of those banished were sent to Siberia and 30 percent were sent to Mongolia. It was widely known that a major difference between Siberia and Mongolia was that fifty percent of the men in Siberia wore fur hats, while only 10 percent of the people in Mongolia wore fur hats. The student was loaded on a railroad box car without windows and shipped east. After

many days the train stopped and he was let out at an unknown location. As the train pulled away he found himself alone on the prairie with a single man who would guide him to the work camp where he would spend the rest of his life. The man was wearing a fur hat. What is the probability he was in Siberia? In presenting your answer, calculate all joint and marginal probabilities. Hint: Portray the sample space in rectangular fashion with location represented along one dimension and whether or not a fur hat is worn along the other.

7. On the basis of a physical examination and symptoms, a physician assesses the probabilities that the patient has no tumour, a benign tumour, or a malignant tumour as 0.70, 0.20, and 0.10, respectively. A thermographic test is subsequently given to the patient. This test gives a negative result with probability 0.90 if there is no tumour, with probability 0.80 if there is a benign tumour, and with probability 0.20 if there is a malignant tumour.

a) What is the probability that a thermographic test will give a negative result for this patient?

b) Obtain the posterior probability distribution for the patient when the test result is negative?

c) Obtain the posterior probability distribution for the patient when the test result is positive?

d) How does the information provided by the test in the two cases change the physician's view as to whether the patient has a malignant tumour?

8. A small college has a five member economics department. There are two microeconomists, two macroeconomists and one econometrician. The World Economics Association is holding two conferences this year, one in Istanbul and one in Paris. The college will pay the expenses of one person from the department for each conference. The five faculty members have agreed to draw two names out of a hat containing all five names to determine who gets to go to the conferences. It is agreed that the person winning the trip to the first conference will not be eligible for the draw for the second one.

a) What is the probability that the econometrician will get to go to a conference?

b) What is the probability that macroeconomists will be the attendees at both conferences?

c) What is the probability that the attendees of the two conferences will be from different fields of economics?

d) The econometrician argued that a rule should be imposed specifying that both attendees could not be from the same field. She was outvoted. Would the provision have increased the probability that the econometrician would get to attend a conference?

Hint: Use a rectangular portrayal of the sample space with persons who can be chosen in the first draw along one axis and persons who can be chosen in the second draw along the other. Then blot out the diagonal on grounds that the same person cannot be chosen twice.

9. There is a 0.8 probability that the temperature will be below freezing on any winter's day in Toronto. Given that the temperature is below freezing my car fails to start 15 percent of the time. Given that the temperature is above freezing my car fails to start 5 percent of the time. Given that my car starts, what is the probability that the temperature is below freezing?

10. If a baseball player is hitting .250 (i.e., if averages one hit per four times at bat), how many times will he have to come up to bat to have a 90% chance of getting a hit? Hint: Ask yourself what the probability is of not getting a hit in n times at bat. Then take advantage of the fact that the event 'getting at least one hit in n times at bat' is the complementary event to the event of 'not getting a hit in n times at bat'.

11. A particular automatic sprinkler system for high-rise apartment buildings, office buildings, and hotels has two different types of activation devices on each sprinkler head. One type has a reliability of .91 (i.e., the probability that it will activate the sprinkler when it should is .91). The other type, which operates independently of the first type, has a reliability of .87. Suppose a serious fire starts near a particular sprinkler head.

a) What is the probability that the sprinkler head will be activated?

b) What is the probability that the sprinkler head will not be activated?

c) What is the probability that both activation devices will work properly?

d) What is the probability that only the device with reliability .91 will work properly?

Hint: Again use a rectangular portrayal of the sample space with the events 'type 1 activation (yes, no)' on one axis and 'type 2 activation (yes, no)' on the other.

12. At every one of the Toronto BlueJay's home games, little Johnny is there with his baseball mit. He wants to catch a ball hit into the stands. Years of study have suggested that the probability is .0001 that a person sitting in the type of seats Johnny and his dad sit in will have the opportunity to catch a ball during any game. Johnny is just turned six years old before the season started. If he goes to every one of the 81 home games from the start of the current season until he is 15 years old, what is the probability that he will have the opportunity to catch a ball.

13. A club has 100 members, 30 of whom are lawyers. Within the club, 25 members are liars and 55 members are neither lawyers nor liars. What proportion of the lawyers are liars?

14. The following is the probability distribution for an exam where students have to choose one of two questions. The pass mark is 3 points or more.

	5	4	3	2	1
Q1	.1	.1	.1	.2	0.0
Q2	0.0	.2	.1	.1	.1

a) Derive the marginal marks probability distribution.

b) What is the probability that a randomly selected student will pass? (.6)

c) Given that a randomly selected student got 4 marks, what is the probability that she did question 2?

15. Suppose you are on a game show and you are given the opportunity to open one of three doors and receive what ever is behind it. You are told that behind one of the doors is a brand new Rolls Royce automobile and behind the other two doors are goats. You pick a particular door—say door number 1—and before the host of the show, who knows what is behind each door, opens that door he opens one of the other doors—say door number 3—behind which is a goat. He then gives you the opportunity to stay with door number 1, which you originally chose, or switch your choice to door 2. Should you switch?

Answer:

This is a classic puzzle in statistics having a level of difficulty much greater than questions usually asked at the beginning level. Accordingly an effort is made here to present a detailed answer. One approach to answering this question is to examine the expected returns to "holding" (staying with the door originally picked) and "switching" to the other unopened door. Let us call the door you initially pick, which ever one it is, door A. Two mutually exclusive events are possible:

1) The car is behind door A —call this event AY.
2) The car is not behind door A —call this event AN.

If your initial guess is right (which it will be 1/3 of the time) you win the car by holding and lose it by switching. If your initial guess is wrong (which it will be 2/3 of the time) the host, by opening the door with the goat behind, reveals to you the door the car will be behind. You win by switching and lose by holding. If contestants in this game always switch they will win the car 2/3 of the time because their initial pick will be wrong 2/3 of the time. The expected payoff can be shown in tabular form. Let winning the car have a payoff of 1 and not winning it have a payoff of zero.

	Hold	Switch	Probability
AY	1	0	1/3
AN	0	1	2/3
Expected Payoff	$1/3 \times 1$ $+1/3 \times 0 = 1/3$	$1/3 \times 0$ $+2/3 \times 1 = 2/3$	

An alternative way to view the question is as a problem in Bayesian updating. Call the door you initially pick door A, the door the host opens door B, and the door you could switch to door C. On each play of the game the particular doors assigned the names A, B, and C will change as the doors picked by the contestant and opened by the host are revealed. The probabilities below are the probabilities that the car is behind the door in question.

Door	AY	AN	
	A	B	C
Prior Probability	1/3	1/3	1/3
Information From Host		$P(B \vert AN) = 0$	$P(C \vert AN) = 1$
Joint Probability		$P(B \cap AN) =$ $P(B \vert AN)(P(AN)) = 0$	$P(C \cap AN) =$ $P(C \vert AN)(P(AN)) = 2/3$
Posterior Probability	1/3	0	2/3

Keep in mind in looking at the above table that $P(AN) = P(B) + P(C) = 2/3$. The posterior probability of the car being behind the door the host leaves closed (i.e. the probability that it is behind door C conditional upon it not being behind door B) is 2/3. The posterior probability of the car being behind door A (i.e., the probability of it being behind door A conditional upon it not being behind door B) is 1/3, the same as the prior probability that it was behind door A. You should always switch!