Chapter 2
Probability and probability distributions

2.1 INTRODUCTION

The methods for reducing and analyzing data described in the previous chapter are not only useful in helping understand what happened in the past. Patterns, trends, and characteristics identified through studies of past data can be of assistance in establishing the likelihood of future events.

For example, life insurance companies invest heavily in studies of past patterns of mortality. The studies have a certain historical interest, but their main use is in allowing the companies to estimate the number of their customers that are expected to die next year, the year after, etc. These estimates, in turn, influence the premiums which the companies must charge if they are to meet the claims arising from these deaths in the future and still make a profit.

In this chapter, we turn away from studying the past for its own sake, and, among other issues, consider how it may be used to estimate the likelihood of future events.

2.2 PROBABILITY

It is impossible to think of any course of action, decision, or choice whose consequences do not depend on the outcomes of some "random process." By "random process"—and here we make no attempt to be precise—we understand any physical or social process the outcomes of which cannot be predicted with certainty. For example, the toss of a coin may be described as a random process having two possible outcomes—"heads" and "tails." The roll of a die, the spin of a roulette wheel, the draw of a bridge hand or of a lottery ticket, tomorrow's closing price of a certain stock at the stock exchange, the number of units of a particular product demanded by customers during a one-month period—all these can be viewed as processes with random outcomes.

In many cases, there is some indication of the likelihood or *probability* that a particular outcome will occur. The term "probability" is part of our everyday speech. We say, for example, that the probability of heads in the

© *Peter Tryfos, 2001.*

toss of an ordinary coin is 1/2; we hear a weather forecaster announcing that the probability of showers tomorrow is 30%; or we may read that the probability of a 40-year-old male surviving the next five years is 98%.

"Probability" describes a person's assessment of the likelihood of occurrence of a particular outcome. It has certain familiar properties: it is expressed as a number between 0 and 1; a 0 indicates an impossible outcome; a 1 indicates that an outcome is certain to occur; probabilities between 0 and 1 indicate various degrees of likelihood, ranging from "very unlikely" to "very likely."

In assessing the likelihood of the occurrence of an outcome, at least three approaches may be distinguished.

The equal likelihood approach. If there are n mutually exclusive and collectively exhaustive outcomes, and if it is reasonable to consider these outcomes equally likely, then it is also reasonable to set the probability of one of them occurring as $1/n$. For example, if the outcomes "heads" and "tails" of a toss of a coin are considered equally likely, then the probability that tails will occur should be 1/2; the probability of heads should also be 1/2. The probability that a six will occur when an ordinary die is rolled should be 1/6, if each of the six faces of the die is considered equally likely to show up. The probability that an ace of hearts will be drawn from an ordinary deck of cards is 1/52, since the ace of hearts is one of 52 cards, each of which—we may agree—has an equal likelihood of being drawn.

Note that the term "equally likely" is not further specified. It is regarded as an intuitive, "primitive," concept. When we say that the outcomes are equally likely, we express the belief that were we to observe the random process a large number of times, we should expect to find the outcomes occurring with about equal relative frequencies.

The relative frequency approach. The approach based on equal likelihood obviously cannot be applied when the outcomes of a random process are not regarded as equally likely. An insurance company, for example, would like to estimate the probability that a 40-year-old man applying for life insurance coverage will die during or survive the following five-year period. There are two outcomes, death and survival, but these are not equally likely, as we know from intuition and from numerous studies. Indeed, survival is much more likely than death. Recent studies indicate that about 1.75% of men who reached age 40 died before they turned 45, while the remaining 98.25% survived. If the insurance company has no reason to believe that the mortality rate in the near future will differ from that observed in the near past, it may assume—as indeed it does—that, in the future, 1.75% of 40-year-old men will die before they turn 45. It may treat the number 0.0175 as the probability that any one of these persons will die before turning 45.

Essentially, then, under the relative frequency approach, the probability

of an outcome is set equal to the relative frequency of its occurrence in a large number of past observations. Implied in this approach is the assumption that the process remains stable—an essential condition for using past experience as a guide to future action.

The judgmental approach. The two previous approaches cannot be called strictly objective, if by "objective" we mean that the judgment of the person assessing the probabilities does not enter. In the approach based on equal likelihood, a judgment must be made that the possible outcomes are equally likely; for how else can we claim that the coin or die is "ordinary," and hence that the two sides are equally likely? In the frequency approach, a judgment must be made that the process is stable, and hence that the past history of the process can be used in forecasting its future performance. Judgment is also necessary in the all-too-frequent cases wherein neither of the two approaches can be applied, but nevertheless reasonable persons feel that an assessment of probabilities is meaningful.

Consider, for example, a weather forecast of the form: "The probability of rain tomorrow is 30%." What precisely is meant by this forecast, and what is the meaning of the number 0.30? It cannot be claimed that the probability is based on past frequencies; the weather characteristics are so numerous that it is difficult to regard today's configuration as the last repetition of a large number of occurrences of the same random process. Nor is the concept of equal likelihood relevant to this situation. Perhaps the best way to interpret the forecast is to regard it as expressing the forecaster's belief that showers would occur in 30% of all days in the future in which today's weather configuration will be observed.

Subjective probabilities need not be arbitrary. In the previous example, the forecaster comes to an informed judgment based on his experience with numerous similar—though not identical—past weather observations. The quality control supervisor may use his experience with similar raw materials in order to assess the probabilities associated with the various quality grades of a material used for the first time. A market analyst may assess the probabilities associated with the levels of sales for a product about to be introduced on the basis of experience with similar products, his estimate of the market situation in general, a test-promotional campaign, and any other relevant information.

The three approaches to the assessment of probabilities are thus not mutually exclusive. Prior information, judgment, and reasonable assumptions may be given various weights as the situation and common sense require.

In the majority of cases, examples, and illustrations that follow we shall have in mind the frequency approach—that is, we shall assume that a reasonably large number of repetitions of a stable random process were recorded and that the observed relative frequencies can be used as estimates of the probabilities of the outcomes. In many important cases (especially in sam-

pling), we shall rely on the principle of equal likelihood for the assessment of probabilities. In all cases, however, we shall interpret the probabilities as the expected relative frequencies of the outcomes in a large number of future repetitions of the random process, if such repetitions are possible.

2.3 PROBABILITY DISTRIBUTIONS AND THEIR CHARACTERISTICS

A random process can often be described by one or more variables or attributes, and its outcomes by their numerical values or categories. Such variables or attributes will be referred to as *random variables* or *random attributes*.

A *probability distribution* is a list showing the possible values of a random variable (or the possible categories of a random attribute) and the associated probabilities.

Example 2.1 A machine produces items in batches of five. If an item meets the technical specifications, it is called Good; if it does not, Defective. The following probability distribution could be based on past inspection results:

Number of defective items	Probability
0	0.60
1	0.20
2	0.10
3	0.05
4 or more	0.05
Total	1.00

For example, the probability of finding 0 defectives in a batch is 60%; we expect to find no defectives in 60% of future batches.

We may note in passing that the probabilities need not be equal exactly to the observed relative frequencies. Suppose, for example, that a machine is used for the first time. The new machine is very similar to the old one but is expected to function a little better. In estimating the probabilities of 0, 1, 2, ... defectives we may wish to modify the observed relative frequencies to make them reflect the anticipated quality improvement—for instance, we may wish to set them equal to 0.70, 0.25, and 0.05 respectively.

Example 2.2 The following probability distribution of the punctuality of flight arrivals is based on historical records:

Flight arrival	Probability
On or ahead of time	0.95
Delayed	0.05
	1.00

For example, the probability of a delayed arrival is 5%; in our interpretation, 5% of future flight arrivals are expected to be delayed.

Example 2.3 The probability distribution of travel time for a bus on a certain route is:

Travel time (minutes)	Probability
Under 20	0.2
20 to 25	0.6
25 to 30	0.1
Over 30	0.1
	1.0

The probability that travel time will exceed 20 minutes is 0.8.

We shall always assume that the values, intervals, or categories listed are mutually exclusive and collectively exhaustive—when describing the outcomes of a random process, we shall take care that the list includes all possible outcomes and that the outcomes do not overlap. This is always possible, but may require some care occasionally. Provided that this requirement is met, the probability of this *or* that outcome is simply equal to the sum of their probabilities. In Example 2.1, for instance, the probability of 0 or 1 defective items is equal to 0.60 plus 0.20, or 0.80. This "addition rule" for probabilities is eminently reasonable if we think of probabilities as future relative frequencies, and can be extended in a straightforward way. In Example 2.1 again, the probability of 0 or 1 or 2 defectives—i.e., the probability of 2 or fewer defective items—is 0.60 + 0.20 + 0.10 or 0.90.

In view of the similarity between relative frequencies and probabilities, it is not surprising that nearly all the concepts and measures of relative frequency distributions carry over to probability distributions. In fact, such characteristics as the mean, variance, standard deviation, joint and conditional distribution, independence, correlation coefficient—all these are defined and calculated in exactly the same way as for relative frequencies.

For the sake of completeness we shall briefly restate these definitions and illustrate their interpretation with a few examples, but the reader will

observe that nearly *all the expressions of this chapter are identical to those in the previous one,* the only difference being that probabilities, represented by $p(x)$, $p(x,y)$, $p(x|y)$, ... , replace relative frequencies, represented by $r(x)$, $r(x,y)$, $r(x|y)$, ...

If the random variable X takes the values x_1, x_2, ... , x_m with probabilities $p(x_1)$, $p(x_2)$, ... , $p(x_m)$, the *mean* or *expected value* of X is

$$\mu = E(X) = x_1 p(x_1) + x_2 p(x_2) + \cdots + x_m p(x_m) = \sum xp(x). \tag{2.1}$$

The notation is not uniform in the literature—sometimes the Greek letter μ is used, at other times the symbol $E(X)$. $E(X)$ can be interpreted as the expected average value of X in a large number of future repetitions of the random process.

The *variance* of a probability distribution—sometimes denoted by the Greek symbol σ^2, sometimes by $Var(X)$—is defined exactly as in relative frequency distributions:

$$\sigma^2 = Var(X) = \sum (x-\mu)^2 p(x). \tag{2.2}$$

An alternative expression is often more convenient for calculations by hand:

$$Var(X) = \sum x^2 p(x) - [E(X)]^2 \tag{2.3}$$

This formula is derived from Equation (2.2) in exactly the same way as in the case of relative frequency distributions.

$Var(X)$ can be interpreted as a measure of the expected dispersion of the values of the random variable about the mean (μ) in a large number of future repetitions of the random process.

The *standard deviation* of X is simply the square root of the variance:

$$\sigma = Sd(X) = +\sqrt{Var(X)}. \tag{2.4}$$

Example 2.4 The probability distribution of the random variable X is given in columns (1) and (2) of the following table.

x (1)	$p(x)$ (2)	$xp(x)$ (3)	$(x-\mu)$ (4)	$(x-\mu)^2$ (5)	$(x-\mu)^2 p(x)$ (6)	x^2 (7)	$x^2 p(x)$ (8)
0	0.5	0.0	−0.9	0.81	0.405	0	0.0
1	0.2	0.2	0.1	0.01	0.002	1	0.2
2	0.2	0.4	1.1	1.21	0.242	4	0.8
3	0.1	0.3	2.1	4.41	0.441	9	0.9
	1.0	$\mu = 0.9$			$\sigma^2 = 1.090$		1.9

The expected value of X is 0.9, as calculated in column (3). The variance may be calculated either using Equation (2.2), as shown in column (6), or using Equation (2.3):

$$\sigma^2 = (1.9) - (0.9)^2 = 1.09.$$

It follows that the standard deviation of X is $\sqrt{1.09}$ or 1.04.

Apart from the mean, variance, and standard deviation, other measures are occasionally encountered (such as the *mode, median, quartiles,* etc.) describing certain characteristics of a probability distribution. They are defined very much as in relative frequency distributions.

Example 2.5 Figure 2.1 shows the probability distribution of the age at death for newborn males, as currently assessed by government actuaries.

We can see, for example, that the probability that a male born now will die between 55 and 60 years from now is about 0.06. Differently put, about 6% of a large number of males born now are expected to die in that time interval. The modal (most likely) five-year age interval at death is 75 to 80.

The probability of dying before a given age, and its complement, the probability of surviving a given age, are shown in Figure 2.2.

It can be seen that, for example, the probability of a male dying before his 50th birthday is about 12%, while that of surviving this birthday is $1 - 0.12$, or 88%.

The data used to plot Figures 2.1 and 2.2 allow us to calculate the expected age at death of newborn males; this figure (often called "the life expectancy at birth") is 68.8 years. The median and the other two quartiles of the distribution are also shown in Figure 2.2; they are approximately 72.5, 62, and 81.5 years respectively. In other words, 25%, 50%, and 75% of a large number of males born now are expected to die before 62, 72.5, and 81.5 years from now respectively.

Note that these distributions refer to newborn males—females tend to live longer. Also, the distributions for those who have already survived a given age are quite different. To mention just one difference, the probability that a 50-year-old man will die before turning 50 is obviously zero, and not 12%, the probability applicable to newborn males.

2.4 JOINT PROBABILITY DISTRIBUTIONS

It may be that a random process can be described according to more than one variable or attribute. The *joint probability distribution* of *two* variables or attributes can be specified in the form of a table, the rows of which show the possible values or categories of the first variable or attribute, the columns those of the second one, and the cells the probabilities of occurrence of the row and column entries.

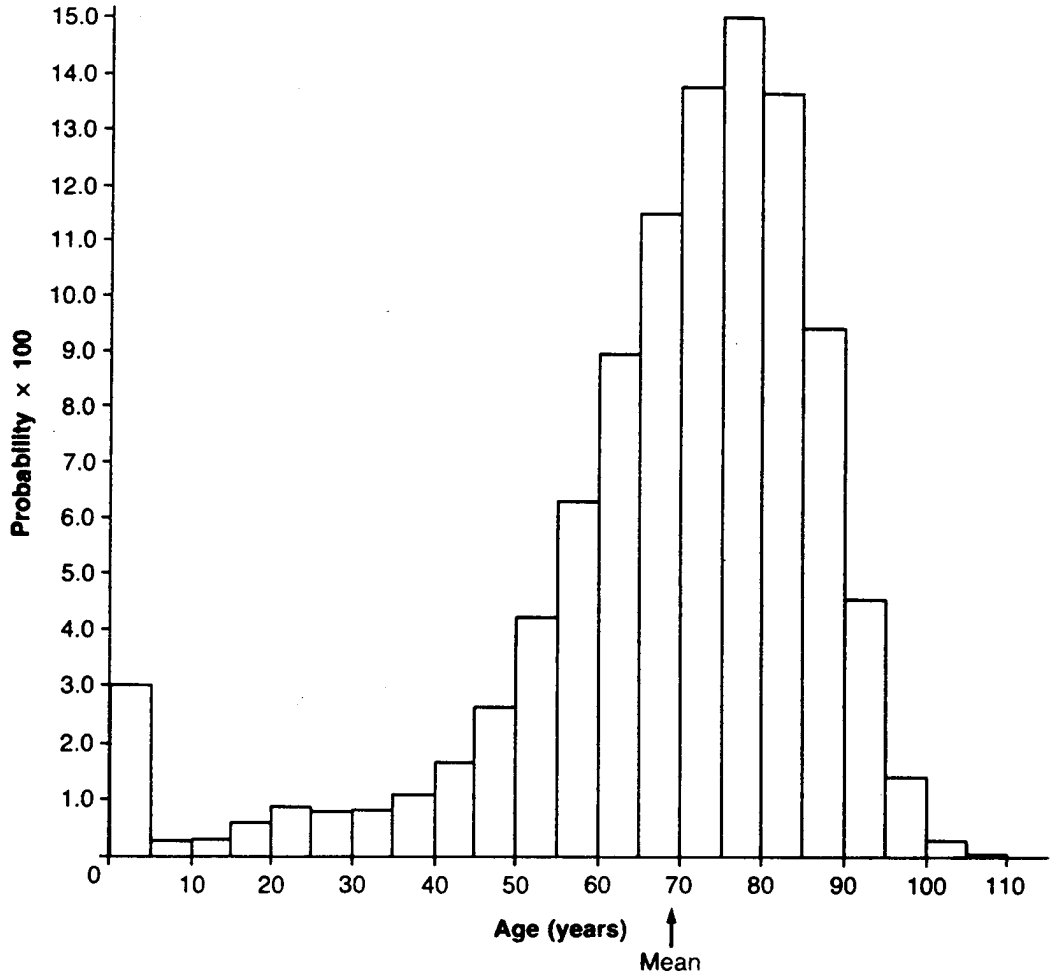


Figure 2.1
Probability distribution of age at death, newborn males

Example 2.6 Electric drills are inspected for defects in the motor (X) and finish (Y). The following joint probability distribution is based on past inspection results and is abridged for simplicity.

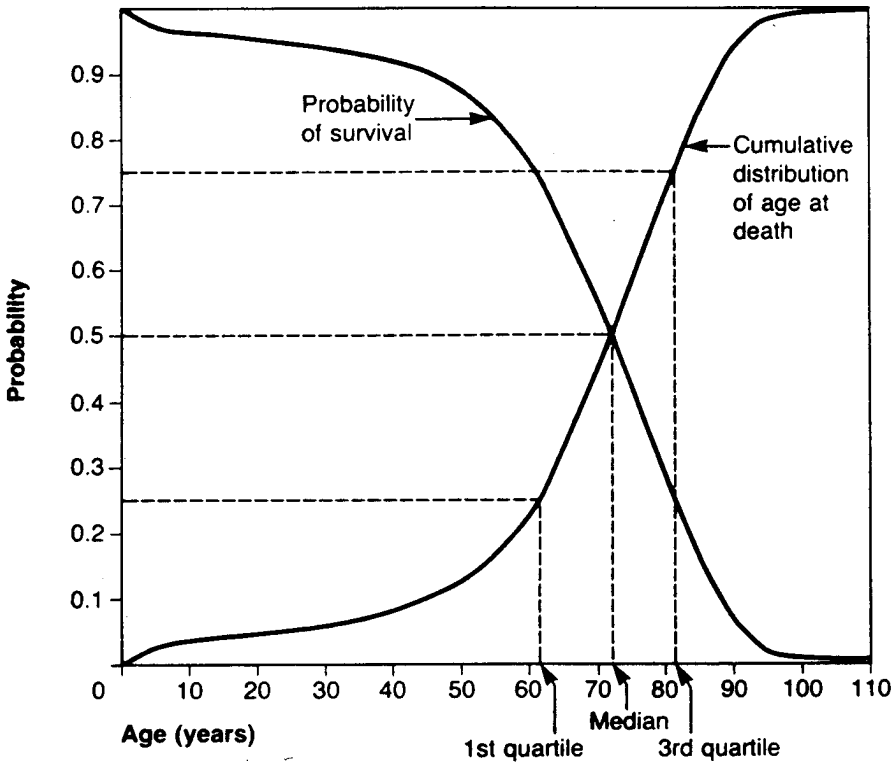


Figure 2.2
Cumulative distributions af age at death, newborn males

Number of defects in motor (X)	Number of defects in finish (Y) 2	3	Total
0	0.1	0.3	0.4
1	0.2	0.4	0.6
Total	0.3	0.7	1.0

Thus, the probability is 10% that a drill will have 0 motor defects *and* 2 defects in the finish, 30% that it will have 0 motor defects and 3 finish defects, and so on. The probability of no motor defects is 40%, as shown in the right margin of the table; the probability that a drill will have 2 finish defects is 30%, as shown in the bottom margin; and so on.

In general, our notation for a joint probability distribution of X and Y is as shown in the following table:

		Y		
X	$\cdots$	y	$\cdots$	Total
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
x	$\cdots$	$p(x,y)$	$\cdots$	$p(x)$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
Total	$\cdots$	$p(y)$	$\cdots$	1.0

The notation is similar to that used for relative frequencies. In the above table, x is a representative value (or category) of X, y one of Y, and $p(x,y)$ denotes the probability that $X = x$ and $Y = y$. The margins show the marginal (univariate) probability distributions of X and Y; for example, $p(x)$ is the probability that $X = x$ regardless of Y. (As in joint relative frequency distributions, we assume that the lists of possible values or categories of X and Y are mutually exclusive and collectively exhaustive.) The marginal probabilities are equal to the sum of the joint probabilities in the row or column:

$$p(x) = \sum_y p(x,y), \tag{2.5}$$

$$p(y) = \sum_x p(x,y). \tag{2.6}$$

Example 2.7 A card drawn at random from a full deck can be described according to its suit (X) and denomination (Y). Since there are 52 cards in the deck and each of these is equally likely to be drawn, the joint distribution of X and Y is as follows:

	Denomination, Y						
Suit, X	A	2	$\cdots$	J	Q	K	Total
Club	1/52	1/52	$\cdots$	1/52	1/52	1/52	13/52
Diamond	1/52	1/52	$\cdots$	1/52	1/52	1/52	13/52
Heart	1/52	1/52	$\cdots$	1/52	1/52	1/52	13/52
Spade	1/52	1/52	$\cdots$	1/52	1/52	1/52	13/52
Total	4/52	4/52	$\cdots$	4/52	4/52	4/52	52/52

For example, the probability of an ace of hearts is 1/52; the probability of a diamond is 13/52; the probability of a king is 4/52; etc.

The means, variances, and standard deviations of the random variables X and Y can be calculated in a straightforward way using the marginal probabilities.

Example 2.6 (Continued) The characteristics of the joint distribution of the number of motor defects (X) and finish defects (Y) are as follows.

$$\mu_x = E(X) = \sum xp(x) = (0)(0.4) + (1)(0.6) = 0.6$$
$$\mu_y = E(Y) = \sum yp(y) = (2)(0.3) + (3)(0.7) = 2.7$$
$$\sigma_x^2 = Var(X) = \sum x^2p(x) - [E(X)]^2 = (0)^2(0.4) + (1)^2(0.6) - (0.6)^2 = 0.24$$
$$\sigma_y^2 = Var(Y) = \sum y^2p(y) - [E(Y)]^2 = (2)^2(0.3) + (3)^2(0.7) - (2.7)^2 = 0.21$$
$$\sigma_x = Sd(X) = \sqrt{Var(X)} = \sqrt{0.24} = 0.49$$
$$\sigma_y = Sd(Y) = \sqrt{Var(Y)} = \sqrt{0.21} = 0.46$$

The calculations are as in relative frequency distributions.

The *correlation coefficient* of two random variables X and Y—denoted by ρ or $Cor(X, Y)$—is calculated in exactly the same way as for relative frequencies:

$$\rho = Cor(X, Y) = \frac{\sum\sum(x - \mu_x)(y - \mu_y)p(x, y)}{\sqrt{\sum(x - \mu_x)^2 p(x)}\sqrt{\sum(y - \mu_y)^2 p(y)}}. \tag{2.7}$$

The double summation symbol ($\sum\sum$) indicates that the numerator consists of the sum of terms $(x - \mu_x)(y - \mu_y)p(x, y)$ calculated for all *pairs* of values (x, y).

The correlation coefficient measures the degree to which two random variables are linearly related. It is always a number between -1 and $+1$. A positive value of ρ indicates positive correlation between X and Y (that is, a tendency for high values of X to be associated with high values of Y, and vice versa). A negative ρ indicates negative correlation (a tendency for high values of X to be associated with low values of Y, and vice versa).

The denominator of Equation (2.7) is the product of the standard deviations of X and Y, and is always positive. The sign of ρ, therefore, depends on the sign of the numerator in (2.7). This numerator is called the *covariance* of X and Y, and denoted by σ_{xy} or $Cov(X, Y)$:

$$\sigma_{xy} = Cov(X, Y) = \sum\sum(x - \mu_x)(y - \mu_y)p(x, y). \tag{2.8}$$

It can be calculated more easily from

$$\sigma_{xy} = Cov(X,Y) = \sum\sum xyp(x,y) - \mu_x\mu_y. \tag{2.9}$$

The correlation coefficient (2.7) can also be written as

$$\rho = Cor(X,Y) = \frac{Cov(X,Y)}{Sd(X)Sd(Y)} = \frac{\sigma_{xy}}{\sigma_x\sigma_y}. \tag{2.10}$$

The covariance finds its way into a number of useful expressions, and for this reason deserves some attention.

Example 2.6 (Continued) The covariance of the number of motor and finish defects can be calculated using Equation (2.9) (recall that $E(X) = 0.6$ and $E(Y) = 2.7$):

$$\begin{aligned} Cov(X,Y) &= \sum\sum xyp(x,y) - E(X)E(Y) \\ &= (0)(2)(0.1) + (0)(3)(0.3) + (1)(2)(0.2) + (1)(3)(0.4) - (0.6)(2.7) \\ &= -0.02. \end{aligned}$$

The correlation coefficient of X and Y is

$$Cor(X,Y) = \frac{-0.02}{(0.49)(0.46)} = -0.09.$$

The number of motor defects (X) and finish defects (Y) are negatively but very weakly correlated.

2.5 CONDITIONAL PROBABILITY DISTRIBUTIONS AND INDEPENDENCE

Very much as in the case of relative frequencies, a conditional probability distribution is a list showing the possible values of a variable (or categories of an attribute) and the probabilities of their occurrence *given* that the other variable or attribute takes a specified value or category.

Example 2.8 The accident records of a large number of insured drivers over two consecutive years were examined. Let X represent a driver's number of accidents in the first year (Year 1), and Y the number of accidents in the second year (Year 2). The joint relative frequency distribution of X and Y was as follows:

Number of accidents in Year 1, X	Number of accidents in Year 2, Y 0	1	Total
0	0.5	0.1	0.6
1	0.1	0.3	0.4
Total	0.6	0.4	1.0

Thus, 50% of the drivers had no accidents in Year 1 *and* none in Year 2, etc. If it is reasonable to assume that this pattern will hold in any future pair of years, the above table also provides the joint probability distribution of X and Y, where now Year 1 and Year 2 refer to *any* pair of consecutive future years.

To form the conditional probability distribution of the number of accidents "next" year given that the driver has no accidents "this" year, we could argue as follows: Out of every—say—100 drivers, 60 are expected to have no accidents this year; out of these 60, 50 are expected to have no accidents (and 10 to have one accident) next year. Therefore, the chances are 50/60 or 0.833 that a driver with no accidents in a given year will have no accidents in the next year; the chances are 10/60 or 0.167 that such a driver will have one accident next year. This conditional probability distribution is shown in columns (1) and (2) below.

Number of accidents next year, Y (1)	Conditional probabilities, $p(y\|x)$ $p(y\|X=0)$ (2)	$p(y\|X=1)$ (3)
0	$0.5/0.6 = 0.833$	$0.1/0.4 = 0.25$
1	$0.1/0.6 = 0.167$	$0.3/0.4 = 0.75$
Total	1.000	1.00

Columns (1) and (3) show the conditional probability distribution of Y for drivers with one accident this year. (The conditional distributions of X for each value of Y can also be calculated, but are obviously of little practical interest in this example.)

In general, given the joint probabilities, $p(x, y)$, the conditional probability that $X = x$ given that $Y = y$ is denoted by $p(x|y)$ and defined as

$$p(x|y) = \frac{p(x, y)}{p(y)}. \tag{2.11}$$

Similarly,

$$p(y|x) = \frac{p(x,y)}{p(x)}. \tag{2.12}$$

In the above expressions, $p(x)$ and $p(y)$ are the marginal probabilities corresponding to the "given" row or column of the table of joint probabilities. Equations (2.11) and (2.12) merely say that in order to calculate a conditional probability one divides the joint probability by the appropriate marginal one. These expressions can also be solved for $p(x,y)$ and written as

$$p(x,y) = p(x)p(y|x) = p(y)p(x|y). \tag{2.13}$$

In words, the probability that $X = x$ and $Y = y$ is found by multiplying the (marginal) probability that $X = x$ by the conditional probability that $Y = y$ given that $X = x$; or, alternatively, by multiplying the probability that $Y = y$ by the conditional probability that $X = x$ given that $Y = y$. This sounds complicated, but in fact it provides a simple method for the development of some useful results, as will soon be illustrated.

Two jointly distributed random variables or attributes, X and Y, are said to be *independent* of (or *unrelated* to) one another if all the conditional distributions of any variable or attribute are identical. This is the same definition of independence as that explained in the context of relative frequencies. Simply stated, it says that two random variables are unrelated to one another if the conditional probability that one variable will take any specified value does not depend on the value of the other variable. A similar interpretation applies to attributes or to a pair of variable and attribute.

Example 2.7 (Continued) We know that the probability of drawing an ace of spades (that is, an ace and a spade) is 1/52. We also know that the probability of a spade is 1/4. Now suppose that a card is drawn at random, and you are told it is a spade. What is the probability that it is an ace?

The conditional probability of an ace given that the card is a spade is

$$Pr(Ace|Spade) = \frac{Pr(Spade, Ace)}{Pr(Spade)} = \frac{1/52}{1/4} = \frac{1}{13}.$$

The conditional probability of a 2 given that the card is a spade is also 1/13, and so is the conditional probability of any other denomination. Thus, the conditional probability distribution of any card's denomination given that the card is a spade is

Denomination, y:	A	2	$\cdots$	J	Q	K
Cond. probability, $p(y\|Spade)$:	1/13	1/13	$\cdots$	1/13	1/13	1/13

In the same manner, it is easy to show that the conditional distribution of a card's denomination is the same for every other suit. We conclude, therefore, that suit and denomination are independent of one another: knowing the suit of a card makes no difference in assessing the probability of a denomination. We could even say that knowledge of the suit is not *informative* concerning the denomination of a card. Similarly, knowledge of the denomination provides no information concerning the suit.

There is actually an easier way to *determine* if two random variables or attributes are independent. As explained in the context of relative frequencies, two random variables or attributes are independent if all joint probabilities are equal to the product of the corresponding marginals. That is, X and Y are independent if $p(x, y) = p(x)p(y)$ for all x and y.* If we look at the table showing the joint probability distribution of suit and denomination of a card in Example 2.7 we will observe that all the joint probabilities $(1/52)$ equal the product of the marginals for the corresponding row and column, $(4/52) \times (13/52)$. Therefore, we can say that suit and denomination are independent of one another without actually calculating any conditional probabilities. In Example 2.8, however, this equality does not hold; therefore, the number of accidents next year is related to the driver's number of accidents this year.

It can also be shown that if two random variables are independent, their covariance and correlation coefficient are zero. That is, if X and Y are independent, then

$$Cov(X, Y) = Cor(X, Y) = 0.$$

The reverse, however, is not always true: that is, the covariance or correlation coefficient of two variables may be zero without the variables' being independent.

2.6 SAMPLING WITH AND SAMPLING WITHOUT REPLACEMENT

A simple example will demonstrate the nature of the two sampling methods.

Example 2.9 Think of a box containing 10 manufactured items. The items are inspected to determine if they meet certain technical specifications; 4 of these items do meet the specifications (we will call them "good" items), while the remaining 6 do not (we will call these "defective" items). The box, therefore, contains 4 good and 6 defective items.

---

* To see this, observe that if $p(x, y) = p(x)p(y)$, $p(x|y) = p(x, y)/p(y) = p(x)p(y)/p(y) = p(x)$ for all x and y; that is, all the conditional disstributions of X are identical. Likewise, $p(y|x) = p(y)$ for all x and y.

We *plan* to select two items at random in one of two ways. (a) The items will first be thoroughly mixed. The first item will be selected, examined, and replaced in the box. Then the items will again be mixed before the second item is selected and examined. We call this sampling method *random sampling with replacement.*

(b) The procedure is the same as (a) except that the first item is not replaced in the box before the second item is selected. We shall call this method of sampling *random sampling without replacement.*

The question now is: If we were to select a random sample—with or without replacement—what are the possible sample outcomes and what are the probabilities of their occurrence?

Let X represent the quality of the first selected item, and Y that of the second; both X and Y are random attributes, which will be either good (G for short) or defective (D).

(a) Sampling with replacement. There are 10 items to begin with—4 good and 6 defective. The probability of a good item in the first draw is 4/10 (applying the principle of equal likelihood); that of a defective is 6/10. Since the item is replaced after it is examined, there will be again 10 items in the box prior to the second draw, of which 4 are good and 6 defective. Therefore, the probability is 4/10 that the second item will be good and 6/10 that it will be defective regardless of the outcome of the first draw. X and Y are independent.

The *probability tree* in Figure 2.3 shows all the possible sample outcomes and the calculation of their probabilities.

The branches of the tree show the outcomes of the first and second draw. There are altogether 4 sample outcomes: the first item is good and the second good (G,G), first good and second defective (G,D), etc. Along the branches we write the *conditional probabilities* of the outcomes. For example, the probability that the second item will be good given that the first is good is 4/10; the probability that the second item will be defective given that the first is good is 6/10. (Since there is no draw preceding the first, the probabilities of the first draw are not conditional.) The probabilities of the sample outcomes are calculated by multiplying the probabilities along the branches. For example, the probability that the first item is good and the second good is (4/10)(4/10) or 16/100, and the probability that the first is good and the second defective is (4/10)(6/10) or 24/100. The justification for this operation is Equation (2.13), which for the first case can be translated as

$$p(G, G) = p(G)p(G|G) = (4/10)(4/10) = 16/100,$$

and for the second case,

$$p(G, D) = p(G)p(D|G) = (4/10)(6/10) = 24/100.$$

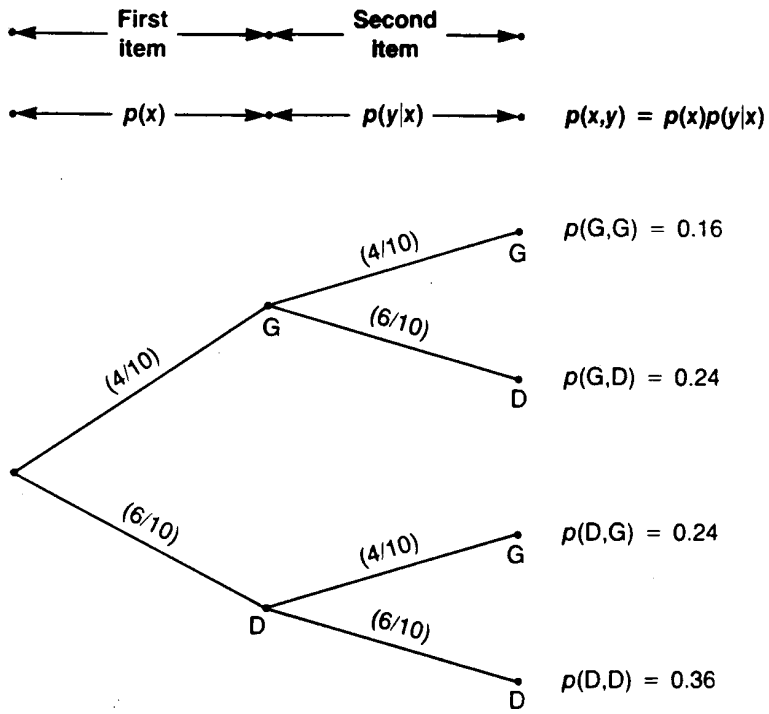


Figure 2.3
Probability tree, sampling with replacement

The joint distribution of X and Y can also be shown in the more familiar format of a table.

Sample of two items with replacement

First item,	Second item, Y		
X	G	D	Total
G	0.16	0.24	0.40
D	0.24	0.36	0.60
Total	0.40	0.60	1.00

(b) Sampling without replacement. The probability of selecting a good item in the first draw is 4/10, and that of a defective is 6/10—the same as in sampling with replacement. However, the probability of the outcome of the second draw clearly depends on the outcome of the first draw. If, for example, the first item is good, 3 of the remaining 9 items will be good and 6 defective; the probability that the second item will be good given that the first is good is therefore 3/9. By contrast, the probability that the

second item will be good given that the first is defective is 4/9, because 4 of the 9 items remaining after the first draw are good. Figure 2.4 shows the probability tree for a sample without replacement.

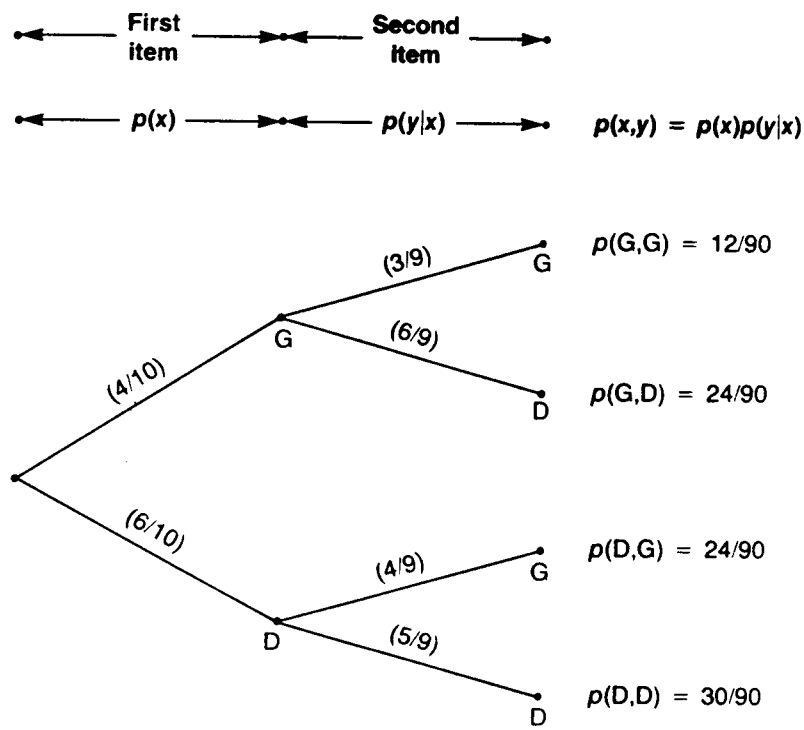


Figure 2.4
Probability tree, sampling without replacement

There are again 4 possible sample outcomes, and their probabilities are calculated by multiplying the probabilities along the corresponding branch of the tree. For example, the probability that the first item will be good and the second good is

$$p(G, G) = p(G)p(G|G) = (4/10)(3/9) = 12/90.$$

The joint distribution of X and Y is:

Sample of two items without replacement

First item, X	Second item, Y G	D	Total
G	12/90	24/90	36/90 = 0.4
D	24/90	30/90	54/90 = 0.6
Total	36/90 = 0.4	54/90 = 0.6	1.0

Compare this table with that for sampling with replacement.

As demonstrated in this simple illustration, sampling with and sampling without replacement differ in two respects: first, the draws are independent if sampling is with replacement but dependent if it is without replacement; second, the probabilities of the sample outcomes are different under the two methods.

In business practice, sampling is nearly always without replacement. The following chapters are devoted to the objectives, properties, and uses of sampling without replacement.

2.7 FUNCTIONS OF RANDOM VARIABLES

This topic could have been brought up earlier in the context of data analysis, but it is especially useful in that of probabilities. Very simply the problem is this. We know the probability distribution of a random variable or attribute X, and would like to determine the probability distribution of another random variable or attribute W which is a function of X (that is, for every value or category of X there corresponds one of W). Or, we know the *joint* distribution of X and Y and would like to determine the distribution of W, where W is a function of X and Y.

The solution to this problem is very simple: we merely list all possible values of X (or pairs of values of X and Y), and write down the corresponding probabilities and values of W; the probability distribution of W is then found by listing each possible value of W and adding up the probabilities of its occurrence. The following example illustrates the general approach.

Example 2.10 The weekly demand for a product (X) has the probability distribution shown in columns (1) and (2) below. It is the policy of the firm to start each week with an inventory of 2 units; no additional units can be ordered during the week. Weekly sales (W) is clearly a function of demand: if demand is 2 units or less, sales equal demand; if demand is greater than 2, sales equal 2, since it is not possible to sell more units than are available.

Demand, x (1)	Probability, $p(x)$ (2)	Sales, w (3)
0	0.4	0
1	0.3	1
2	0.2	2
3	0.1	2
	1.0	

Weekly sales could be 0, 1, or 2 units. Sales equal 0 when demand is 0, and the probability of this occurrence is 0.4; sales equal 1 when demand is 1, and this occurs with probability 0.3; finally, sales equal 2 when demand is 2 or 3, and the probability of this is 0.2 + 0.1 or 0.3. Therefore, the probability distribution of sales is:

Sales, w	Probability, $p(w)$
0	0.4
1	0.3
2	0.3
	1.0

Of course, having obtained the distribution of W, it is straightforward to calculate its mean and variance. In this example,

$$E(W) = (0)(0.4) + (1)(0.3) + (2)(0.3) = 0.9,$$
$$Var(W) = (0)^2(0.4) + (1)^2(0.3) + (2)^2(0.3) - (0.9)^2 = 0.69.$$

We shall make use of these last results shortly.

Sometimes, we are interested only in the mean or variance of the distribution of W, and not in the entire distribution. The question then arises as to whether it is possible to express the mean and variance of W as a simple function of the mean and variance of X. In general, the answer is negative, but in the special case where W is a *linear function* of X the solution is very simple. If $W = a + bX$, then

$$\begin{aligned} E(W) &= a + bE(X), \\ Var(W) &= b^2 Var(X), \end{aligned} \tag{2.14}$$

where a and b are given constants.

To see why, note that for every value of X there corresponds only one value of $W = a + bX$. The probability of this value of W is the probability

of the corresponding value of X, and $p(w) = p(x)$. The expected value of W—found, as usual, by multiplying the values of W by their probabilities and adding up the products—is

$$E(W) = \sum wp(w) = \sum (a+bx)p(x) = a\sum p(x)+b\sum xp(x) = a+bE(X).$$

Similarly, the variance of W is

$$\begin{aligned} Var(W) &= \sum [w - E(W)]^2 p(w) \\ &= \sum [a + bx - a - bE(X)]^2 p(x) \\ &= b^2 \sum [x - E(X)]^2 p(x) \\ &= b^2 Var(X). \end{aligned}$$

Example 2.11 The distance travelled daily (X, in miles) by trucks of a delivery company is a random variable with mean $E(X) = 300$ and variance $Var(X) = 250$. The total daily operating cost of a truck consists of a part that does not vary with distance traveled (insurance, depreciation, driver's salary, etc.) and of another that does (gas, oil, tires, maintenance, etc.). The fixed part is estimated at \$75 per day; the variable part is estimated to be about \$0.20 per mile. The total daily operating cost, W, is therefore a linear function of distance, $X : W = 75+0.20X$. It follows that the expected daily operating cost is $E(W) = 75+(0.20)E(X) = (75)+(0.2)(300)$ or \$135, while its variance is $Var(W) = (0.20)^2 Var(X) = (0.2)^2(250) = 10$.

To construct the probability distribution of a function W of *two* jointly distributed random variables X and Y, we list all possible pairs of values of X and Y, together with their probabilities and corresponding values of W. The distribution of W is then found by listing each possible value of W and adding up the probabilities of its occurrence.

Example 2.6 (Continued) Given the joint distribution of the number of motor (X) and finish (Y) defects, we wish to determine the distribution of the *total* number of defects, $W = X + Y$.

x	y	$p(x,y)$	$w = x + y$
0	2	0.1	2
0	3	0.3	3
1	2	0.2	3
1	3	0.4	4
		1.0	

If, for example, $X = 0$ and $Y = 2$, then $W = 2$; the probability of this value of W is that of the pair $(X = 0, Y = 2)$, which is 0.1. The distribution of W is obtained from the last two columns.

w	$p(w)$
2	0.1
3	0.5
4	0.4
	1.0

The possible values of W are 2, 3, and 4, and these will occur with probabilities 0.1, 0.5 (the sum of the two pairs of (x, y) values that yield $W = 3$), and 0.4 respectively.

A useful special case is that of a *linear function* of X and Y, $W = a + bX + cY$, where a, b, and c are some constants. Note that when $a = 0$ and $b = c = 1$, W is the *sum* of X and Y; when $a = 0$, $b = 1$, and $c = -1$, W is the *difference* of X and Y; and when $a = 0$, $b = c = 1/2$, W is the *average* of X and Y.

It can be shown that the mean and variance of W are simply related to the means, variances, and covariance of X and Y. If $W = a + bX + cY$, then

$$E(W) = a + bE(X) + cE(Y), \tag{2.15}$$

and

$$Var(W) = b^2Var(X) + c^2Var(Y) + 2bcCov(X, Y). \tag{2.16}$$

Example 2.6 (Continued) The following characteristics of the joint distribution of the number of motor defects (X) and finish defects (Y) were calculated earlier: $E(X) = 0.06$, $E(Y) = 2.7$, $Var(X) = 0.24$, $Var(Y) = 0.21$, and $Cov(X, Y) = -0.02$.

The mean and variance of the distribution of the total number of defects, $W = X + Y$, can be determined directly from Equations (2.15) and (2.16), without first calculating the distribution of W:

$$E(W) = E(X) + E(Y) = 0.6 + 2.7 = 3.3,$$

$$\begin{aligned} Var(W) &= Var(X) + Var(Y) + 2Cov(X, Y) \\ &= (0.24) + (0.21) + (2)(-0.02) \\ &= 0.41. \end{aligned}$$

These results may be confirmed from the distribution of W:

w	$p(w)$	$wp(w)$	$w^2p(w)$
2	0.1	0.2	0.4
3	0.5	1.5	4.5
4	0.4	1.6	6.4
Totals	1.0	3.3	11.3

which gives

$$E(W) = \sum wp(w) = 3.3,$$

$$Var(W) = \sum w^2p(w) - [E(W)]^2 = (11.3) - (3.3)^2 = 0.41,$$

as claimed.

Example 2.12 An automobile insurance policy is actually three policies combined in one, providing coverage for (1) legal liability for bodily injury to or death of any person, or damage to property ("third-party liability"); (2) bodily injury or death of the insured ("accident benefits"); and (3) loss of or damage to the insured car ("comprehensive," "all perils," etc.). The total premium is the sum of the premiums of the three separate coverages. We shall examine here how an insurance company determines the annual premium of the first of these coverages, but the procedure is very similar for those of the other two.

The experience of a large insurance company with third-party liability during the most recent calendar year is summarized in Table 2.1. The company had 286,309 policies in force during the year. All policies were identical with respect to minimum and maximum limits: there was no deductible, and the maximum allowable claim was \$1 million.

Of the 286,309 policies in force, 265,188 or 92.623% made no claim (alternatively, claimed \$0); 19,592 or 6.843% made a claim of between \$0 and \$1,000 during the year; 948 or 0.3311% claimed between \$1,000 and \$5,000; and so on. These are *total* claims during the year: it is possible (though rather rare) for more than one claim to be made against a policy in one year.

Note how skew is the distribution of claim size—the first three intervals account for nearly all claims. However, it would be a mistake to overlook large claims since they have a significant effect on the calculation of the premium.

In order to determine the probability distribution of claim size next year, the company must consider whether or not its most recent experience, summarized in Table 2.1, should be modified. If the general level of car repair costs, court settlements, or inflation is expected to change, then some

Table 2.1
Probability distribution of claim size

Size of claim ($000) (1)	Number of policies (2)	Average claim ($000) (3)	Rel. frequ., and probability (4)
0	265,188	0.000	0.926230
> 0 to 1	19,592	0.744	0.068430
1 to 5	948	2.779	0.003311
5 to 10	300	7.740	0.001048
10 to 25	181	16.872	0.000632
25 to 50	61	36.847	0.000213
50 to 100	27	72.269	0.000094
100 to 250	8	129.306	0.000029
250 to 500	3	304.269	0.000010
500 to 1,000	1	563.174	0.000003
Total	286,309		1.000000

adjustments are necessary. If not, the most recent experience may be taken as indicative of the likely experience next year, in which case the probability distribution of the size of the claim is given by the relative frequency distribution of claim size last year. This latter case is assumed here, and the probability distribution is shown in columns (1) and (4) of Table 2.1.

The expected claim size can be approximated by determining the midpoint of each interval, multiplying it by the corresponding probability, and adding the products. Remember that, in the case of relative frequency distributions, this type of calculation produces the exact mean if the midpoint of each interval equals the average value of the observations in the interval. This applies to probability distributions as well. Column (3) of Table 2.1 shows the average claim for each interval. The exact expected claim size per policy, $E(X)$, is therefore equal to

$$E(X) = (0)(0.92623)+(0.744)(0.06843)+\cdots+(563.174)(0.000003) = 0.1023,$$

or $102.30.

In the language of the industry, the expected claim per policy is the "pure premium." If each insured were to pay this amount for third-party liability coverage, the company's revenue from a policy would equal the expected claim of that policy. Insurance companies, however, must cover their other expenses, pay dividends, and maintain reserves for contingencies. In practice, therefore, the pure premium is adjusted by a "mark-up factor." If the company's mark-up factor is, say, 40%, the annual premium for third-party liability will be $102.30 $\times$1.40, or $143.22.

Let us now consider the premium calculation for the same type of coverage but *with a $10,000 limit.* This means that if the total annual claim exceeds this limit, the company will pay $10,000; the insured is responsible for the difference.

The company's payment (W) is a function of the claim size (X). Refer to Table 2.1. If the claim is any amount less than $10,000, the payment equals the claim; if the claim exceeds $10,000, the payment equals $10,000. The probability that the payment will be $10,000 is the probability that the claim will exceed $10,000, or $(0.000632 + \cdots + 0.000003)$, or 0.000981. The probability distribution of the payment under this policy is shown in Table 2.2.

Table 2.2
Probability distribution of payment,
limit of $10,000

Payment ($000)	Probability
0	0.926230
> 0 to 1	0.068430
1 to 5	0.003311
5 to 10	0.001048
10	0.000981
	1.000000

The expected payment, $E(W)$, is calculated using the data of Table 2.1, as follows:

$$E(W) = (0)(0.926230) + (0.744)(0.068430) + \cdots + (10)(0.000981) = 0.07801,$$

or $78.01.

The annual pure premium of third-party liability coverage with a limit of $10,000, therefore, is $24.29 less than that with a limit of $1 million.

2.8 SPECIAL DISTRIBUTIONS: BINOMIAL, HYPERGEOMETRIC, POISSON

So far, a probability distribution was defined in the form of a list showing the possible values and probabilities of the variable. In some cases, however, the distribution can be described by a mathematical formula, from which the tabular representation can be obtained. Numerous such *special distributions* can be found in the statistical literature. In this section and the next we examine briefly some of these special distributions.

We shall say that a random variable X has a *binomial distribution* with parameters n and p if

$$p(x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \quad (x = 0, 1, \ldots, n) \tag{2.17}$$

Translated, this definition simply says that the possible values of X are 0, 1, 2, ... , up to and including n, and that the probabilities of these values are given by Equation (2.17). The notation $a!$ ("a factorial") is shorthand for the product $(a)(a-1)(a-2)\ldots(2)(1)$. By definition, $0! = 1! = 1$.

Equation (2.17) actually defines not one but a family of distributions—one for each set of values of the *parameters* n and p.

For example, the statement "The distribution of X is binomial with parameters $n = 2$ and $p = 0.4$" means that X can take the values 0, 1, or 2, with probabilities given by

$$p(x) = \frac{2!}{x!(2-x)!} (0.4)^x (0.6)^{2-x}.$$

The probability that $X = 0$, for instance, can be calculated by replacing x above by 0:

$$p(0) = \frac{2!}{0!2!} (0.4)^0 (0.6)^2 = 0.36.$$

The reader can verify easily that $p(1) = 0.48$, and $p(2) = 0.16$, so that this distribution can be written in the more familiar tabular format as follows:

x	$p(x)$
0	0.36
1	0.48
2	0.16
	1.00

Computer programs for the calculation of binomial probabilities are widely available. For selected values of the parameters, the binomial distribution is tabulated in Appendixes 4A and 4B. *The introductory notes in Appendix 4 explain the use of all the tables in that Appendix.*

Before explaining the reason for our interest in this special distribution, let us examine another, slightly more complicated probability distribution.

We say that a random variable X has a *hypergeometric distribution* with parameters N, n, and k if

$$p(x) = \frac{\binom{N-k}{n-x}\binom{k}{x}}{\binom{N}{n}}, \tag{2.18}$$

where $x = a, a+1, a+2, \ldots, b$; $a = \max[0, n-(N-k)]$; and $b = \min[n, k]$.

Despite the formidable name and notation, this definition can be easily translated and applied. It says that X takes integer values in the range a to b, where a is the larger of the numbers 0 and $n-(N-k)$, and b equals the smaller of the numbers n and k (more about these parameters in a moment). The probabilities of these values are given by Equation (2.18), where it is understood that for any $c > d$

$$\binom{c}{d} = \frac{c!}{d!(c-d)!}. \tag{2.19}$$

As in the case of the binomial distribution, Equation (2.18) defines a family of probability distributions—one for each set of values of the parameters N, n, and k.

As an example, the statement "The distribution of X is hypergeometric with parameters $N = 10$, $n = 2$, and $k = 4$" means that X can take the values 0, 1, or 2, with probabilities given by

$$p(x) = \frac{\binom{6}{2-x}\binom{4}{x}}{\binom{10}{2}}.$$

The probability that $X = 0$ is found by replacing x above by 0:

$$p(0) = \frac{\binom{6}{2}\binom{4}{0}}{\binom{10}{2}} = \frac{6!}{2!4!} \times \frac{4!}{0!4!} \div \frac{10!}{2!8!} = \frac{6!8!}{4!10!} = \frac{30}{90} = 0.333.$$

The reader can verify that $p(1) = 0.533$ and $p(2) = 0.133$, so that the distribution of X is:

x	$p(x)$
0	0.333
1	0.533
2	0.133
	1.000

The calculation of these probabilities by hand can be very tedious and time-consuming. It can be done easily with the help of special tables (see Appendix 4J) or, better still, special computer programs.

The practical usefulness of these special distributions lies in the following properties of random samples—properties with important applications.

> Consider any population of N elements, and suppose that k of these elements belong to a certain category (class or interval) C with respect to a given attribute or variable. Suppose that a random sample of n elements will be selected from this population, and let W represent the number of elements in the sample that belong to category C. It can be shown that: (a) if the sample is with replacement, the probability distribution of W is binomial with parameters n (the sample size) and $p = k/N$ (the fraction of population elements that belong to C); (b) if the sample is without replacement, the probability distribution of W is hypergeometric with parameters N (the population size), n (the sample size), and k (the number of elements in the population that belong to C).

By "population" we understand any collection of elements from which a number are selected at random. We shall confirm these properties below with the help of a simple example.

Example 2.9 (Continued) The population of this example is the lot of $N = 10$ items described earlier. The items are classified into two categories, Good and Defective. We know that 4 of these items are good, and 6 are defective. A random sample of $n = 2$ items will be taken. We are interested in determining the probability distribution of the number of good items in the sample, W.

If the sample is *without* replacement, then, according to the property just stated, the probability distribution of W is hypergeometric with parameters $N = 10$, $n = 2$, and $k = 4$. Earlier in this section we showed that this distribution is:

w	$p(w)$
0	0.333
1	0.533
2	0.133
	1.000

We can easily confirm this result. The joint probability distribution of the quality of the first (X) and second (Y) item selected was derived earlier and is reproduced in a slightly different form in columns (1) to (3) below.

x (1)	y (2)	$p(x,y)$ (3)	w (4)	w (5)	$p(w)$ (6)
G	G	$12/90$	2	0	$30/90 = 0.333$
G	D	$24/90$	1	1	$48/90 = 0.533$
D	G	$24/90$	1	2	$12/90 = 0.133$
D	D	$30/90$	0		1.000

Let W represent the number of good items in the sample. W is a function of X and Y. For every pair of values of X and Y, there corresponds a value of W, as shown in column (4). For example, if the first item is good and the second good, the number of good items in the sample is 2, and the probability of this is 12/90. The probability distribution can be easily obtained from columns (3) and (4), and is listed in columns (5) and (6). It is indeed the hypergeometric distribution with parameters $N = 10$, $n = 2$, and $k = 4$.

If the sample is *with* replacement, the reader can easily verify in a similar manner that the distribution of W is binomial with parameters $n = 2$ and $p = 4/10 = 0.4$:

w	$p(w)$
0	0.36
1	0.48
2	0.16
	1.00

The Poisson distribution is the last special distribution to be examined in this section. We say that a random variable X has a *Poisson distribution* with parameter $m > 0$ if

$$p(x) = \frac{m^x e^{-m}}{x!} \qquad x = 0, 1, 2, 3, \ldots \tag{2.20}$$

Poisson probabilities are tabulated in Appendixes 4C and 4D. For example, the Poisson distribution with parameter $m = 0.20$ is, as the reader can easily verify,

x	$p(x)$
0	0.819
1	0.164
2	0.016
3 or greater	0.001
	1.000

In practice, the Poisson distribution is used as an approximation to the binomial or hypergeometric distributions for certain ranges of parameter values, although computer programs have reduced the need for such approximations. Certain conditions about the nature of the random process can be shown to imply a Poisson distribution. In *queueing theory*, for example, many elegant and useful results are based on such Poisson random processes. The interested reader will find a more complete description of these applications in operations research textbooks.

2.9 CONTINUOUS PROBABILITY DISTRIBUTIONS: EXPONENTIAL, NORMAL

Until now, we dealt with variables which take a finite number of values.*

Continuous variables, on the other hand, are those which—in principle at least—can take any value within a specified interval, no matter how small that interval may be. Variables representing time, temperature, length, weight, volume, etc. belong to this category.

The distribution of a continuous variable can be specified by a list showing the probabilities that the variable will fall into each of a set of *intervals* (see Example 2.3). A histogram of the distribution can then be constructed in the usual way. In such a histogram, the *area* of the bar is equal to the probability that the value of the variable will be in the corresponding interval.

If the intervals are narrow enough, it is possible to approximate the histogram by a smooth curve, as illustrated in Figure 2.5. In some cases, this smooth curve can be described by a mathematical formula, giving the *height* of the curve $p(x)$ at each point x. The probability that the random variable X will be in the interval from a to b is then equal to the *area* under $p(x)$ between a and b. Figure 2.6 illustrates this.

From the geometry of Figure 2.6, it is clear that

$$Pr(a \leq X \leq b) = Pr(X \geq a) - Pr(X \geq b) = Pr(X \leq b) - Pr(X \leq a).$$

* In the case of the Poisson distribution, the number of values is not finite, but the values themselves are integer.

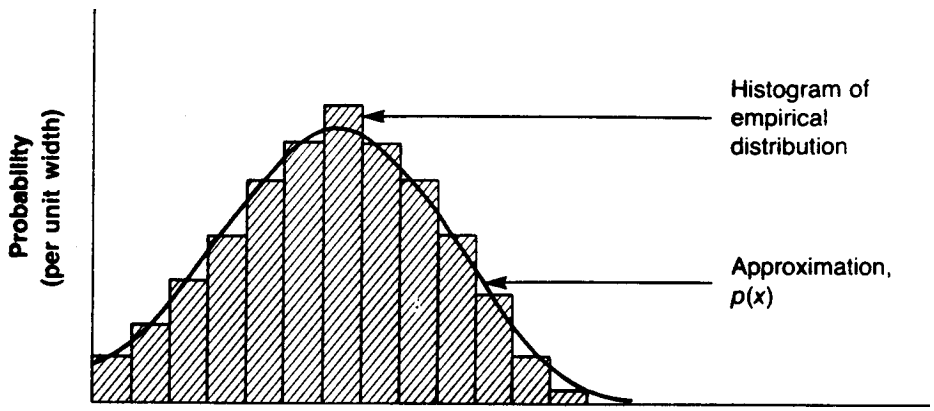


Figure 2.5
Approximation of empirical distribution

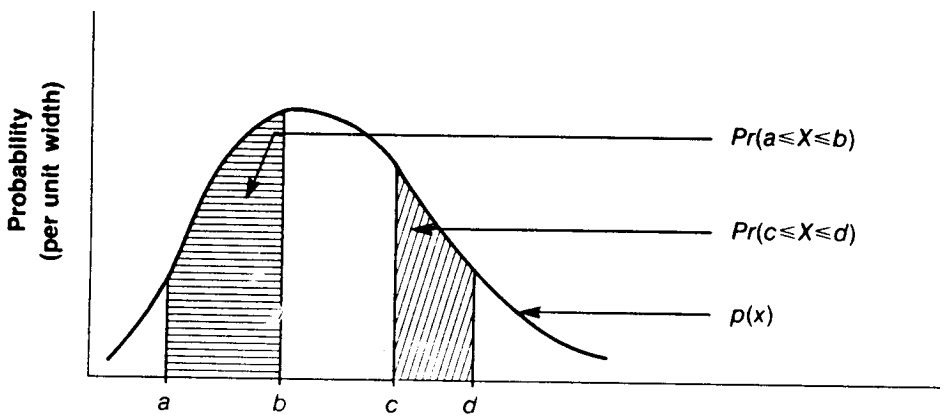


Figure 2.6
Probability equals area under $p(x)$

In words, the area under $p(x)$ between a and b is equal to the area to the right of a minus the area to the right of b, or, alternatively, the area to the left of b minus the area to the left of a.† The total area under $p(x)$ is, of course, equal to 1.

† Since continuous random variables can take an infinite number of values, the probability that $X = x$ *exactly* (for example, the probability that X equals 0 precisely and not 0.0000000000000000...01 or some such number) is practically zero. Therefore, for all practical purposes, $Pr(X < a) = Pr(X \leq a)$, and so on for similar expressions.

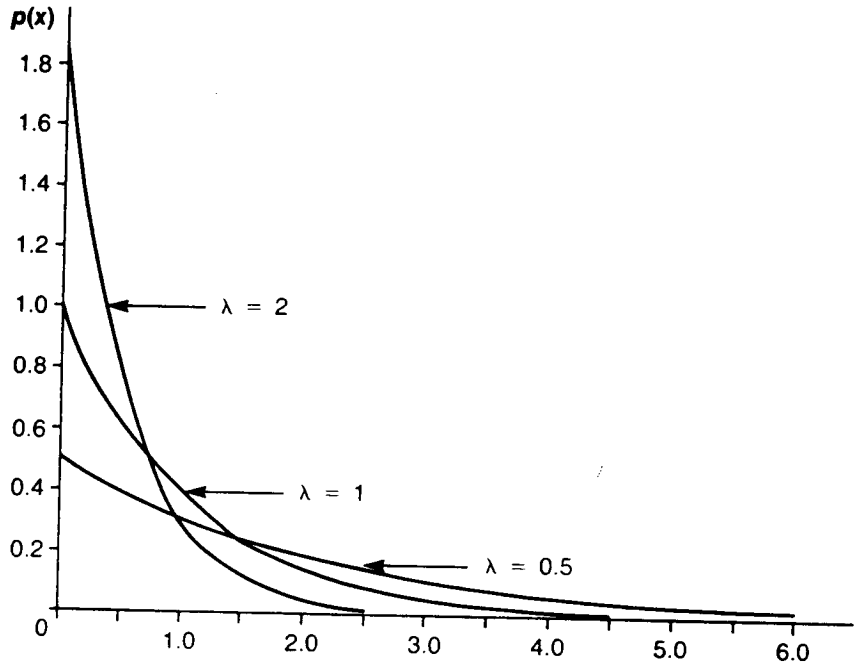


Figure 2.7
Exponential distributions

Also following from the geometry of Figure 2.6,

$$Pr(a \leq X \leq b \text{ or } c \leq X \leq d) = Pr(a \leq X \leq b) + Pr(c \leq X \leq d),$$

for non-overlapping intervals $[a, b]$ and $[c, d]$.

Many special continuous distributions can be found in the literature. We examine briefly two of these.

We say that the distribution of a continuous random variable X is *exponential* with parameter $\lambda > 0$ if

$$p(x) = \lambda e^{-\lambda x} \qquad (x > 0) \tag{2.21}$$

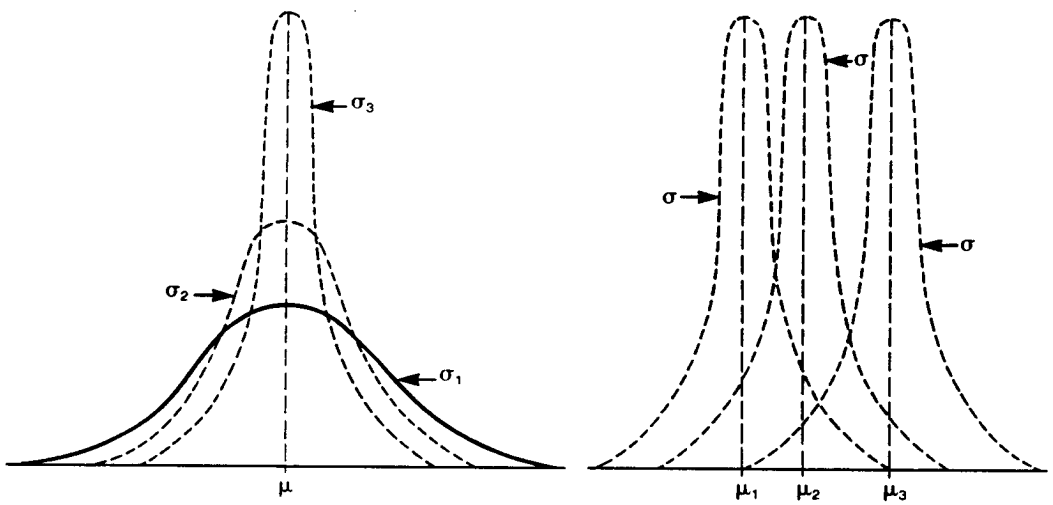
This expression defines a family of distributions, one for each value of the parameter λ. Some exponential distributions are plotted in Figure 2.7 for selected values of λ.

The exponential distribution is sometimes used as an approximation to empirical distributions having this characteristic "inverse-J" shape. Like the Poisson distribution, to which it is related, the exponential finds applications in queueing theory and other topics of operations research.

The second special continuous distribution to be examined here is the normal. We say that the distribution of X is *normal* with parameters μ and $\sigma > 0$ if

$$p(x) = \frac{1}{\sqrt{2\pi}} exp\{-\frac{(x-\mu)^2}{2\sigma^2}\} \qquad (-\infty < x < +\infty) \tag{2.22}$$

The notation $exp\{a\}$ is another way of writing e^a. Figure 2.8 shows how the form of the distribution depends on the parameters μ and σ.



(a) Normal Distributions Having the Same μ but Different σ ($\sigma_1 > \sigma_2 > \sigma_3$)

(b) Normal Distributions Having the Same σ but Different μ ($\mu_1 < \mu_2 < \mu_3$)

Figure 2.8
Normal distributions

The normal distribution is bell-shaped and symmetric about μ. Note that a change in μ—with σ held constant—shifts the distribution to the right or left without affecting its shape. A change in σ—with μ held constant—changes the "spread" of the distribution without affecting its location: the larger the value of σ, the "flatter" the distribution.

The normal distribution is used as an approximation to empirical distributions and to some special distributions (e.g., binomial, hypergeometric, Poisson). The normal distribution, despite its complicated appearance, is very amenable to mathematical analysis. In sampling and regression (two major areas of statistical analysis treated later in this text) there exist powerful results for cases where the population distribution of the variable of

interest is or can be assumed to be normal. Even more importantly, however, the normal distribution is the approximate distribution of many estimators when the sample size is large. We cannot be more specific at this stage; the reader will hear more about the uses of the normal distribution later in this text.

Equation (2.22) defines a family of distributions, one for each pair of values of the parameters μ and σ. The normal distribution with $\mu = 0$ and $\sigma = 1$ is called the *unit* or *standard normal distribution* and is used to evaluate probabilities for all normally distributed random variables.

To see how, suppose that X is normal with parameters μ and σ. Applying two basic rules of inequalities* we find

$$\begin{aligned} Pr(a \leq X \leq b) &= Pr(a - \mu \leq X - \mu \leq b - \mu) \\ &= Pr(\frac{a-\mu}{\sigma} \leq \frac{X-\mu}{\sigma} \leq \frac{b-\mu}{\sigma}) \\ &= Pr(\frac{a-\mu}{\sigma} \leq U \leq \frac{b-\mu}{\sigma}), \end{aligned}$$

where $U = (X - \mu)/\sigma$. In words, the probability that X will take a value between a and b is equal to the probability that U (a function of X) will take a value between $(a - \mu)/\sigma$ and $(b - \mu)/\sigma$. It can be shown that:

> If the distribution of X is normal with parameters μ and σ, the distribution of
> $$U = \frac{X - \mu}{\sigma}$$
> is the standard normal.

This is a very convenient result because areas under the standard normal distribution are tabulated: Appendix 4F shows the probability that U will take a value greater than u, for the values of u shown in the margins of the table.

For example, suppose that X is normal with parameters $\mu = 10$ and $\sigma = 4$. The probability that X will take a value between 11 and 12 is equal

---

* If $a \leq X \leq b$ then $ca \leq cX \leq cb$ for $c > 0$, or $ca \geq cX \geq cb$ for $c < 0$. In words, multiplying all terms of an inequality by a positive number maintains the signs of the inequality, but multiplication by a negative number reverses the signs. Also, if $a \leq X \leq b$ then $a + c \leq X + c \leq b + c$ for any c. In words, adding a number (positive or negative) to all the terms preserves the inequality.

to the probability that U will take a value between $(11-10)/4 = 0.25$ and $(12-10)/4 = 0.50$. From Appendix 4F,

$$\begin{aligned} Pr(0.25 \leq U \leq 0.50) &= Pr(U \geq 0.25) - Pr(U \geq 0.50) \\ &= 0.4013 - 0.3085 \\ &= 0.0928, \end{aligned}$$

that is, $Pr(11 \leq X \leq 12) = 0.0928$. For the same parameter values, the probability that X will be greater than 8 equals the probability that U will be greater than $(8-10)/4 = -0.5$. Bearing in mind the symmetry of the normal distribution,

$$Pr(-0.5 \leq U) = 1 - Pr(U \geq 0.5) = 1 - 0.3085 = 0.6915,$$

that is, $Pr(X \geq 8) = 0.6915$.

2.10 THE MEAN AND VARIANCE OF SPECIAL DISTRIBUTIONS

The mean and variance of the special distributions depend on their parameters in a rather simple manner, as indicated in Table 2.3. The proof of these results is beyond the level of this text but may be found in any introductory mathematical statistics text.

Table 2.3
Means and variances of special distributions

Distribution	Parameters	Mean, $E(X)$	Variance, $Var(X)$
Binomial	n, p	np	$np(1-p)$
Hypergeometric	N, n, k $(p = k/N)$	np	$np(1-p)\frac{N-n}{N-1}$
Poisson	m	m	m
Exponential	λ	$1/\lambda$	$1/\lambda^2$
Normal	μ, σ	μ	σ^2

2.11 MULTIVARIATE PROBABILITY DISTRIBUTIONS

All the definitions, concepts and results for joint distributions of two variables can be extended to joint distributions of an arbitrary number of random variables or attributes. The extensions are straightforward, although the notation becomes a little more complicated. We briefly consider the most important of these extensions.

The joint distribution of n random variables or attributes X_1, X_2, ... , X_n will be denoted by $p(x_1, x_2, \ldots, x_n)$. It can be thought of as a list showing the possible *sets* of values of the variables (or categories of the attributes) and their probabilities:

x_1	x_2	...	x_n	$p(x_1, x_2, \ldots, x_n)$
...	...	...	...	...

The extension of Equation (2.13) reads as follows:

$$p(x_1, x_2, \ldots, x_n) = p(x_1)p(x_2|x_1)p(x_3|x_1, x_2) \cdots p(x_n|x_1, x_2 \ldots, x_{n-1}). \tag{2.23}$$

In words, the probability that $X_1 = x_1$ and $X_2 = x_2$ and $X_3 = x_3$ ... is equal to the probability that $X_1 = x_1$, times the probability that $X_2 = x_2$ given $X_1 = x_1$, times the probability that $X_3 = x_3$ given $X_1 = x_1$ and $X_2 = x_2$, ...

Example 2.9 (Continued) We would like to determine the probability distribution of the number of good items in a random sample of size $n = 3$ without replacement. The probability tree in Figure 2.9 shows the possible sample outcomes and their probabilities.

For example, the probability that all three items will be good equals the probability that the first item will be good (4/10), times the probability that the second item will be good given that the first is good (3/9), times the probability that the third item will be good given that the first and second items are good (2/8), or 24/720.

The probability distribution of the number of good items in the sample (W) can be easily obtained from the tree:

w	$p(w)$
0	$120/720 = 0.1667$
1	$360/720 = 0.5000$
2	$216/720 = 0.3000$
3	$24/720 = 0.0333$
	1.000

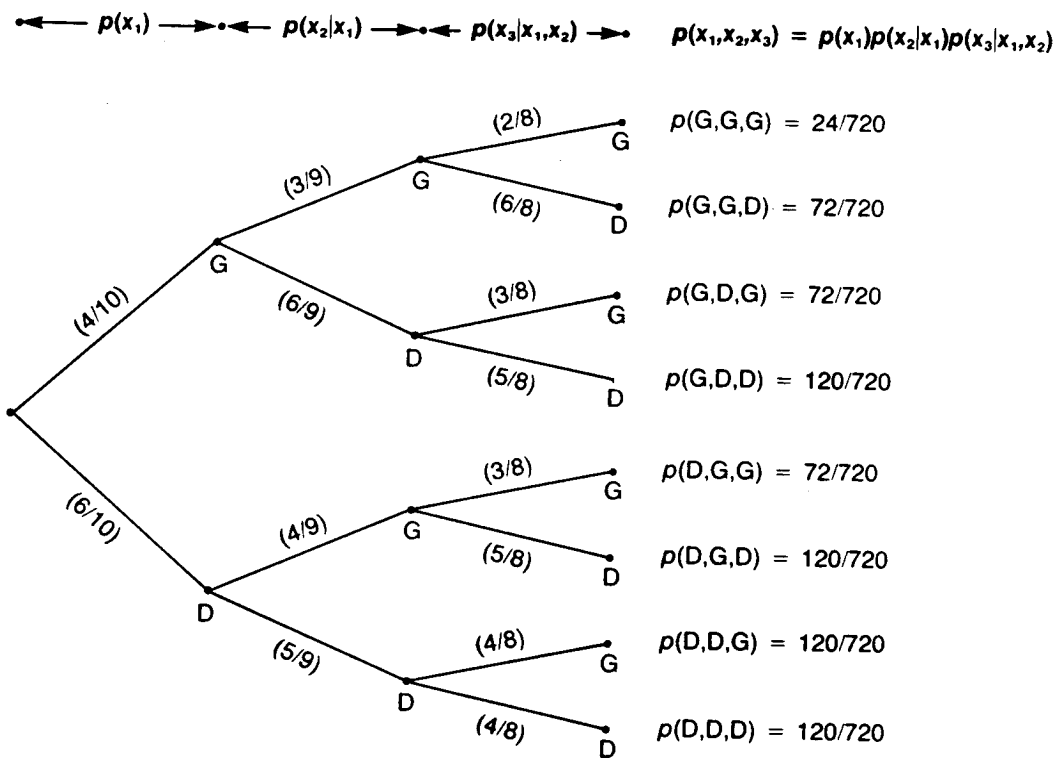


Figure 2.9
Probability tree, Example 2.9

This is, of course, a hypergeometric distribution with parameters $N = 10$, $n = 3$, and $k = 4$, and could have been more easily obtained directly from the tabulated probabilities in Appendix 4J.

---

n random variables or attributes are *independent* if all pairs of variables or attributes are independent. In other words, a group of variables are independent if the probability that any given variable will take any specified value does not depend on the values of the other variables. A similar interpretation applies to groups of attributes or variables and attributes. In such cases,

$$p(x_1, x_2, \ldots, x_n) = p(x_1)p(x_2)\cdots p(x_n).$$

If a group of variables are independent, their covariances and correlation coefficients are equal to zero.

A *linear function* of the random variables X_1, X_2, ... , X_n is a function

of the form

$$W = k_0 + k_1X_1 + k_2X_2 + \cdots + k_nX_n, \tag{2.24}$$

where the k's are given constants. Note once again that special cases of Equation (2.24) include the *sum* of the X's ($k_0 = 0$, all other $k_i = 1$), and the *average* of the X's ($k_0 = 0$, all other $k_i = 1/n$). It can be shown that the expected value of W is the same linear function of the expected values of the X's:

$$E(W) = k_0 + k_1E(X_1) + k_2E(X_2) + \cdots + k_nE(X_n). \tag{2.25}$$

The variance of W is a function of the variances and covariances of the X's:

$$\begin{aligned} Var(W) = {} & k_1^2Var(X_1) + k_2^2Var(X_2) + \cdots + k_n^2Var(X_n) + \\ & + 2k_1k_2Cov(X_1, X_2) + 2k_1k_3Cov(X_1, X_3) + \cdots \\ & \cdots + 2k_{n-1}k_nCov(X_{n-1}, X_n). \end{aligned} \tag{2.26}$$

This formula looks complicated, but its meaning is easy to understand. Imagine arranging the k's and the variances and covariances of the variables as in the following table:

	k_1	k_2	$\cdots$	k_{n-1}	k_n
k_1	$Var(X_1)$	$Cov(X_1, X_2)$	$\cdots$	$Cov(X_1, X_{n-1})$	$Cov(X_1, X_n)$
k_2	$Cov(X_2, X_1)$	$Var(X_2)$	$\cdots$	$Cov(X_2, X_{n-1})$	$Cov(X_2, X_n)$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
k_{n-1}	$Cov(X_{n-1}, X_1)$	$Cov(X_{n-1}, X_2)$	$\cdots$	$Var(X_{n-1})$	$Cov(X_{n-1}, X_n)$
k_n	$Cov(X_n, X_1)$	$Cov(X_n, X_2)$	$\cdots$	$Cov(X_n, X_{n-1})$	$Var(X_n)$

This table is sometimes called the *variance-covariance matrix* of the variables X_1, X_2, ... , X_n. Note that the variances are placed on the diagonal of the table, while the covariances are symmetrically arranged off the diagonal.

To calculate $Var(W)$ according to Equation (2.26), we multiply each cell entry by the corresponding row and column k's and add the products. Remember that $Cov(X_i, X_j) = Cov(X_j, X_i)$.

Equations (2.25) and (2.26) reduce to (2.14) in the case where $n = 1$, and to (2.15) and (2.16) when $n = 2$.

In particular, the mean and variance of the *sum of n independent random variables*, $W = X_1 + X_2 + \cdots + X_n$, are obtained from Equations (2.25) and (2.26) by setting $k_0 = 0$, all other $k_i = 1$, and all covariances to zero:

$$E(W) = E(X_1) + E(X_2) + \cdots + E(X_n), \tag{2.27}$$

and

$$Var(W) = Var(X_1) + Var(X_2) + \cdots + Var(X_n). \tag{2.28}$$

Similarly, the mean and variance of the *average of n independent random variables*, $W = (X_1 + X_2 + \cdots + X_n)/n$, are obtained by setting in Equations (2.25) and (2.26) $k_0 = 0$, all other $k_i = 1/n$, and all covariances to zero:

$$E(W) = \frac{1}{n}\big[E(X_1) + E(X_2) + \cdots + E(X_n)\big], \tag{2.29}$$

and

$$Var(W) = \frac{1}{n^2}\big[Var(X_1) + Var(X_2) + \cdots + Var(X_n)\big]. \tag{2.30}$$

These expressions are useful in sampling theory, as will soon be demonstrated.

Example 2.13 A sum of $\$C$ is available for investment in a number of n different securities (stocks, bonds, etc.). Such an investment is called a *portfolio* of securities. The securities will be purchased at the current price, held over a period of time (for example, one year), and then sold at the then prevailing price. The return from each \$1 invested in a given security is the rate of return of that security and is defined as:

$$\text{Rate of return} = \frac{(\text{Final price} - \text{Current price}) + \text{Dividends}}{\text{Current price}}.$$

The exact return of the portfolio can be determined with certainty only at the time of its liquidation. At the time the investment is made, a decision as to how much to invest in each security must be made on the basis of anticipated return and anticipated risk. Let X_i be the anticipated rate of return of security i. Suppose $\$C_i$ is initially invested in security i. The return from an investment of $\$C_i$ in security i is $\$C_iX_i$, and the total return of the portfolio is $C_1X_1 + C_2X_2 + \cdots + C_nX_n$. The portfolio rate of return is

$$\begin{aligned} W &= \frac{1}{C}\big(C_1X_1 + C_2X_2 + \cdots + C_nX_n\big) \\ &= \frac{C_1}{C}X_1 + \frac{C_2}{C}X_2 + \cdots + \frac{C_n}{C}X_n. \end{aligned}$$

The portfolio rate of return, therefore, is a linear function of the rates of return of the individual securities:

$$W = k_1X_1 + k_2X_2 + \cdots + k_nX_n,$$

where $k_i = C_i/C$. The expected portfolio rate of return, $E(W)$, and variance, $Var(W)$, are given by Equations (2.25) and (2.26). In portfolio analysis, $Var(W)$ is referred to as the *risk* of the portfolio.

In order to determine the expected rate of return and the risk of a given portfolio (that is, one in which the k_i are given), we need to estimate the means $E(X_i)$, variances $Var(X_i)$, and covariances $Cov(X_i, X_j)$ of the rates of return of the individual securities. An initial estimate can be obtained from the joint relative frequency distribution of security returns in the past, which can then be modified to reflect any available relevant information affecting the future performance of the securities.

The *portfolio problem* is to *find* the values of the k_i which minimize the risk of the portfolio, subject to the condition that the expected portfolio rate of return not be short of a certain "desired" rate. More formally, the problem is to find non-negative k_1, k_2, ... , k_n which minimize $Var(W)$, subject to $E(W) \geq r$, and $k_1 + k_2 + \cdots + k_n = 1$, where r is the desired rate of return.

As a numerical illustration, let us suppose that an amount of $C = \$1,000$ is available to invest in $n = 3$ securities. It is desired to form a portfolio having at least a 10% expected rate of return. (In what follows, it will be convenient to express the rates of return as percentages; for example, as 10 instead of 0.10.)

The expected rate of return of security 1 is estimated as $E(X_1) = 17$ (percent), while those of securities 2 and 3 are $E(X_2) = 21$ and $E(X_3) = 3$. The estimated variances and covariances of the rates of return are shown in the following table:

	Security		
Security	1	2	3
1	44	34	38
2	34	97	62
3	38	62	137

For example, the estimated variance of the rate of return (always expressed as a percentage) of security 1 is 44, the covariance of securities 1 and 2 is 34, and so on. If k_1, k_2, and k_3 denote the proportions of the budget invested in securities 1, 2, and 3 respectively, the variance ("risk") of the portfolio rate of return is

$$\begin{aligned} Var(W) &= 44k_1^2 + 97k_2^2 + 137k_3^2 + 2(34)k_1k_2 + 2(38)k_1k_3 + 2(62)k_2k_3 \\ &= 44k_1^2 + 97k_2^2 + 137k_3^2 + 68k_1k_2 + 76k_1k_2 + 124k_2k_3. \end{aligned}$$

The expected portfolio rate of return is

$$E(W) = 17k_1 + 21k_2 + 3k_3.$$

Thus, the portfolio problem is to find the values of k_1, k_2, and k_3 which minimize $Var(W)$ subject to the constraints $E(W) \geq 10$, $k_1 + k_2 + k_3 = 1$, and all $k_i \geq 0$.

This problem is now discussed in most textbooks of financial analysis. It is a special case of a quadratic programming problem and can be solved with the help of widely available computer programs.

PROBLEMS

2.1 What are the possible outcomes of a roll of a die? If the die is "fair," what are the associated probabilities? What is the probability of rolling a number greater than 4? What is the probability of rolling a number less than or equal to 3?

2.2 Consider the following probability distribution:

x	$p(x)$
1	0.3
2	0.4
3	0.3
	1.0

(a) Calculate and interpret the mean (expected value) of X.
(b) Calculate and interpret the variance of X.
(c) Calculate and interpret the standard deviation of X.

2.3 The manager of a supermarket believes that the following data (derived from records of the shop) reflect accurately the probability distribution of the weekly demand for eggs:

Weekly demand (hundreds of dozen eggs)	Probability
1	0.20
2	0.40
3	0.30
4	0.10
	1.00

Assume that any eggs not sold at the end of the week must be thrown away. Eggs are bought at \$40 and sold at \$60 per hundred dozen. The supermarket has a standing order with a local supplier to have 2 hundred dozen eggs delivered at the beginning of every week.

(a) What is the probability distribution of weekly sales?

(b) What is the probability distribution of weekly lost sales? (Lost sales is the number of eggs short of demand.)

(c) What is the probability distribution of weekly revenue? What is the expected weekly revenue?

(d) The manager estimates that lost sales of one hundred dozen eggs is equivalent to an outright loss of \$20 because annoyed customers may stop shopping at

the supermarket. What is the probability distribution of weekly profit? What is the expected weekly profit?

(e) How would you determine—in principle—the optimal ordering policy?

2.4 The performance of light bulbs (and of other products subject to failure) in tests is often described in the form of a table, which shows the number of light bulbs still functioning after a number of periods of continuous use. A simplified table is shown below.

Number of periods	Number functioning
0	1,000
1	800
2	400
3	100
4	0

Thus, out of the 1,000 light bulbs tested, 800 survived after 1 period of use, 400 after 2 periods, 100 after 3 periods, and none after 4 periods of use.

(a) Determine the probability distribution of the life duration of new light bulbs. Assume that bulbs fail at the end of a period. Calculate the expected life duration of new light bulbs. Briefly interpret these results.

(b) Calculate the probability distribution of life duration of light bulbs which survive 2 periods of use.

(c) An office building has 1,000 light bulbs of the type described above. All 1,000 bulbs were installed at the same time. At the end of each period, the maintenance staff replaces the bulbs that failed with new ones. These replacement bulbs have the same characteristics as the original ones, that is, 80% survive one period, 40% two periods, and 10% three periods of use. When they fail, they too are replaced by new bulbs with the same characteristics. Show that the expected number of failures in period 1 is 200; in period 2, 440; in period 3, 468; and so on. Carry on with these calculations for a number of periods to show that the total number of failures in a period converges to a constant. In general, this constant is equal to $N/E(Y)$, where N is the number of light bulbs installed originally, and $E(Y)$ is the expected life duration of the light bulbs as calculated in part (a).

2.5 A finance company specializes in small, short-term comsumer loans, which are intended to assist the purchase of a car, appliance, or vacation, to overcome temporary financial difficulties, etc. A credit officer reviews the application, interviews the prospective client, and classifies the application as a Good, Fair, or Poor credit risk. Normally, an application is handled by one credit officer. However, in order to test whether two credit officers, A and B, apply consistent standards in evaluating applications, the loan manager selected a number of applications at random and had them evaluated independently by A and B. The following table shows the joint relative frequency distribution of the two evaluations.

A's	B's evaluation			
evaluation	Good	Fair	Poor	Total
Good	0.15	0.05	0.05	0.25
Fair	0.05	0.35	0.10	0.50
Poor	___	0.05	0.20	0.25
Total	0.20	0.45	0.35	1.00

Thus, for example, 15% of the applications were judged Good by both officers, 5% were judged Good by A and Fair by B, and so on. Use these joint relative frequencies as joint probabilities.

(a) What is the probability that an application will be judged Fair by A? That it will be judged Good by B?

(b) What is the probability that an application will be judged Fair by A and either Good or Fair by B? What is the probability that an application will be judged Fair by B and either Poor or Good by A?

(c) Construct and interpret the conditional probability distribution of A's evaluation given that B's evaluation is Fair. Construct and interpret the conditional probability distribution of B's evaluation given that A's evaluation is Poor.

(d) Are A and B consistent evaluators? Discuss.

2.6 Motor vehicle accidents are classified into three mutually exclusive and collectively exhaustive categories in increasing order of seriousness:

1. *Property damage only*: accidents resulting in property damage but not in injuries or deaths;
2. *Non-fatal injury only*: accidents resulting in the injury of one or more persons and perhaps in property damage, but not in deaths;
3. *Fatal*: accidents resulting in the death of one or more persons and perhaps in injuries and/or property damage.

Last year, 204,271 accidents were reported. The following is the joint frequency distribution of the seriousness of the accidents and the day of the week in which they occurred:

	Seriousness of accident			
Day of occurrence	Fatal	Non-fatal injury	Property damage	Total
Sunday	229	8,945	15,944	25,118
Monday	162	8,007	17,355	25,524
Tuesday	168	8,589	18,605	27,362
Wednesday	146	8,031	16,960	25,137
Thursday	176	9,080	19,730	28,986
Friday	255	11,422	24,232	35,909
Saturday	334	12,278	23,623	36,235
Total	1,470	60,352	136,449	204,271

Assuming that last year's conditions will also hold in the following years:

(a) What is the probability that an accident will be fatal and will occur on a Wednesday?

(b) What is the probability that a fatal accident will occur on the weekend (Saturday or Sunday)?

(c) What is the probability of a fatal accident? What is the probability that an accident—if one occurs—will occur on a Monday? Saturday?

(d) What is the conditional probability distribution of the seriousness of accidents for each day of the week? Interpret your results.

(e) Is the seriousness of accidents related to the day of the week? If not, why? If yes, in what manner?

2.7 (a) The random variable X has a Poisson probability distribution with parameter $m = 0.10$. Verify that the mean and the variance of the distribution both are equal to m.

(b) The random variable X has a binomial probability distribution with parameters $n = 5$ and $p = 0.1$. Verify that the mean of the distribution equals np and the variance $np(1-p)$.

(c) The random variable X has a hypergeometric probability distribution with parameters $N = 10$, $n = 1$, and $k = 3$. Verify that the mean of this distribution equals np, where $p = k/N$, and that the variance equals

$$np(1-p)\frac{N-n}{N-1}.$$

2.8 A lot of 10 manufactured items contains 3 defective and 7 good items. Calculate the joint probability distribution of the outcomes of the first and second draw for a random sample of two items drawn from the lot (a) with replacement, and (b) without replacement. Show that the marginal distributions are the same in both cases. Briefly interpret your results.

2.9 A lot contains 4 items, of which 1 is good and 3 defective. You plan to select from this lot a random sample of two items without replacement.

(a) Determine the probability distribution of the number of defective items in the sample.

(b) Show that this distribution is indeed hypergeometric with appropriate parameter values.

(c) Determine the probability distribution of the number of defective items in a sample of two items with replacement. Show that this distribution is binomial with appropriate parameter values.

2.10 Construct your own simple example to verify that if the values of X and Y are linearly related, $y = a + bx$, the correlation coefficient is $+1$ if $b > 0$, or -1 if $b < 0$.

2.11 For a certain project to be completed, two tasks, A and B, must be performed in sequence. The joint distribution of the times required to perform these tasks is given below:

Time for task B (days)	Time for task A (days) 1	2	Total
2	0.40	0.20	0.60
3	0.10	0.30	0.40
Total	0.50	0.50	1.00

Find the mean and the variance of the distribution of the time required to complete the *project*.

2.12 The joint probability distribution of the random variables X and Y is as follows:

	Y			
X	0	1	2	Total
0	0.4	0.1	0.1	0.6
1	0.3	0.1	0.0	0.4
Total	0.7	0.2	0.1	1.0

(a) Are X and Y independent? Why?

(b) Calculate the mean and variance of Y.

(c) Calculate the covariance and the correlation coefficient of X and Y. Interpret briefly.

(d) Determine the probability distribution of $W = XY$. Calculate the mean and variance of W.

(e) Determine the joint probability distribution of $W = XY$ and $V = X + Y$. Briefly interpret this distribution.

2.13 The joint probability distribution of the random variables X and Y is as follows:

	Y			
X	-1	0	$+1$	Total
-1	0.2	0.2	0.0	0.4
$+1$	0.4	0.1	0.1	0.6
Total	0.6	0.3	0.1	1.0

(a) Are X and Y independent? Why?

(b) Calculate the mean and variance of Y. Interpret these numbers.

(c) Calculate the covariance and the correlation coefficient of X and Y. Interpret briefly.

(d) Determine the probability distribution of $W = Y/X$. Calculate the mean and variance of W.

(e) Determine the joint distribution of $W = Y/X$ and $V = XY$. Briefly interpret this distribution.

2.14 A company operates two warehouses, one in city A and the other in city B. Both warehouses stock a single product. On the basis of past experience, the joint probability distribution of weekly demand in the two cities is estimated as follows:

Demand at A (Number of units)	Demand at B (Number of units) 0	1	2	Total
0	0.15	0.04	0.01	0.20
1	0.07	0.33	0.05	0.45
2	0.03	0.13	0.19	0.35
Total	0.25	0.50	0.25	1.00

In both warehouses, the policy is to begin every week with a stock of 1 unit. If demand during the week is greater than 1, the unsatisfied demand is lost. Clearly, in our simplified case, the total lost demand can be 0, 1, or 2 units.

(a) Construct the probability distribution of weekly lost demand. Explain your calculations.

(b) A proposal is made to close the warehouses at A and B, and to operate a single warehouse in a central location. The central warehouse will carry a stock of 2 units. The total demand will not be affected—clients at A and B will simply address themselves to the new location. Construct the probability distribution of lost demand for the central warehouse.

(c) Should the warehouses be centralized?

2.15 (a) The expected value of the product, W, of two random variables, X and Y, $W = XY$, is not in general equal to the product of their expected values. Construct a simple example to show that

$$E(W) = E(XY) = E(X)E(Y) + Cov(X, Y).$$

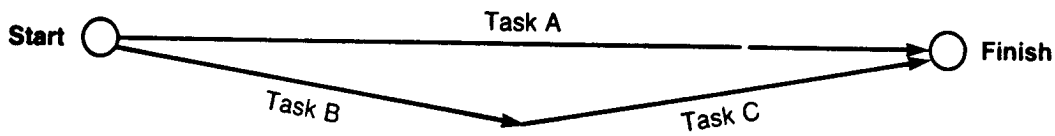
(b) The random variables X and Y are independent. Show that

$$E(XY) = E(X)E(Y).$$

Note: This property can be generalized. If X_1, X_2, ... , X_n are n independent random variables, the expected value of their product is equal to the product of their expected values:

$$E(X_1 X_2 \cdots X_n) = E(X_1)E(X_2) \cdots E(X_n).$$

2.16 Construction projects require the scheduling and coordination of a large number of tasks. Consider the following simplified example:



A project is made up of three tasks, designated as A, B, and C. Task B must be completed before C can start. Task A can be done in parallel, but both A and C must be completed before the project is considered finished. The time required to complete each task is uncertain, owing to weather conditions and other unpredictable factors. The probabilities assigned to task completion times are shown below:

Task	Completion time (weeks)	Probability
A	4	0.50
	6	0.50
		1.00
B	1	0.25
	3	0.75
		1.00
C	2	0.80
	4	0.20
		1.00

Assuming the task completion times are independent (for example, the time taken to complete B does not influence the time for C), and that each task will begin as early as possible, what is the probability that the time required to finish the project will be 6 weeks or more?

2.17 X_1 and X_2 are *independent* random variables with the following probability distributions:

x_1	$p(x_1)$	x_2	$p(x_2)$
0	0.4	2	0.3
1	0.6	3	0.7
	1.0		1.0

(a) Determine the joint probability distribution of $Y_1 = X_1 + X_2$ and $Y_2 = X_1 - X_2$.

(b) Calculate the mean and variance of Y_1 and of Y_2. Verify that $E(Y_1) = E(X_1) + E(X_2)$, $E(Y_2) = E(X_1) - E(X_2)$, and $Var(Y_1) = Var(Y_2) = Var(X_1) + Var(X_2)$.

(c) Calculate the covariance of Y_1 and Y_2. Show that

$$Cov(Y_1, Y_2) = Var(X_1) - Var(X_2).$$

(d) Briefly show why (c) is a special case of the following result (useful in time series and factor analysis). If $X_1, X_2, \ldots, X_m$ are independent random variables, and

$$Y_1 = a_1X_1 + a_2X_2 + \cdots + a_mX_m,$$

$$Y_2 = b_1X_1 + b_2X_2 + \cdots + b_mX_m,$$

where the a's and b's are given constants, then

$$Cov(Y_1, Y_2) = a_1b_1Var(X_1) + a_2b_2Var(X_2) + \cdots + a_mb_mVar(X_m).$$

2.18 An investor estimates the expected rate of return of security 1 to be 0.15 (i.e., 15%), and that for security 2 to be 0.10. The variances and the covariance of the rates of return are also estimated as follows:

$$Var(X_1) = 0.08$$
$$Var(X_2) = 0.02$$
$$Cov(X_1, X_2) = 0.05$$

The investor plans to allocate a certain fraction (k) of the budget to security 1, and the remainder $(1 - k)$ to security 2.

(a) Express the expected rate of return and the variance of the portfolio as a function of k.

(b) Formulate the "portfolio problem" for this special case.

2.19 An investor plans to invest 40% of the available capital in security A and 60% in security B. The expected rates of return of these securities are 15% (security A) and 22% (security B).

The variance of the rate of return of security A is 31, that of security B 52, and the covariance of the rates of return of securities A and B is -13.

Calculate the expected value and variance of the portfolio rate of return.

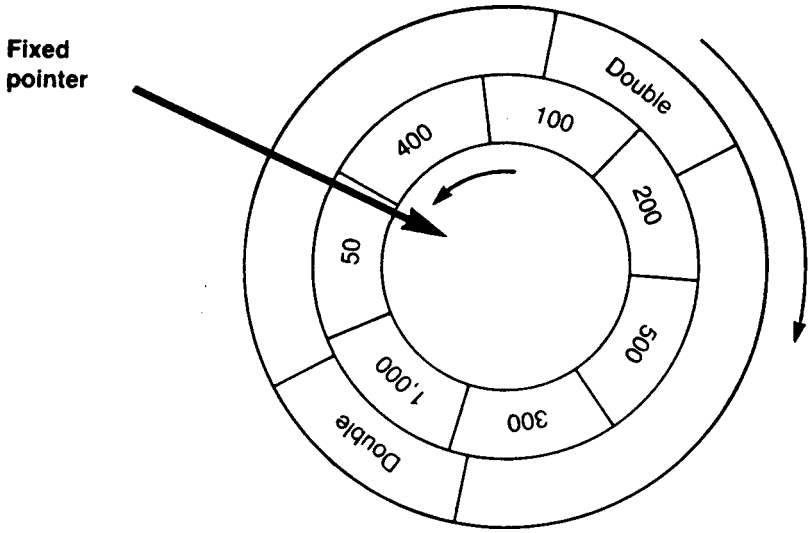


Figure 2.10
The "Wheels of Fortune," Problem 2.20

2.20 One feature of the popular television game show "Name That Tune" (NBC) is the following. The Master of Ceremonies spins two "Wheels of Fortune" as shown in Figure 2.10.

The two wheels are concentric and are spun independently. The outer wheel is divided into four sections; two of these are blank and the other two are labelled "double." The "double" sections each take one-seventh of the circumference of the wheel. The inner wheel is divided into seven equal sections marked \$400, \$100, \$200, \$500, \$300, \$1,000, and \$50. The MC first spins the outer wheel clockwise and then the inner wheel counterclockwise. Depending on which sections happen to come to rest against the fixed pointer, the payoff is determined. For example, if the pointer is in the \$100 section of the inner wheel and in one of the blank sections of the outer wheel, the payoff is \$100; if the pointer is in one of the sections marked "double" of the outer wheel, the payoff is \$200.

After the payoff is determined ("Contestants! You are now playing for \$ ... !"), the MC asks the two contestants a question of a musical nature. If a contestant believes he knows the answer, he presses a buzzer which blocks out the other contestant and gives him the right to answer the question. If the contestant answers the question correctly, he receives the payoff determined by the spinning of the two wheels; if the question is answered incorrectly, the payoff goes to the other contestant.

What is the probability distribution of the amount which the promoter of the show will have to pay? What is the expected value of this distribution? Interpret your answers.

2.21 One feature of the television program "The Price Is Right" is a wheel divided into 20 equal sections which are marked with the numbers 5, 10, 15, 20,

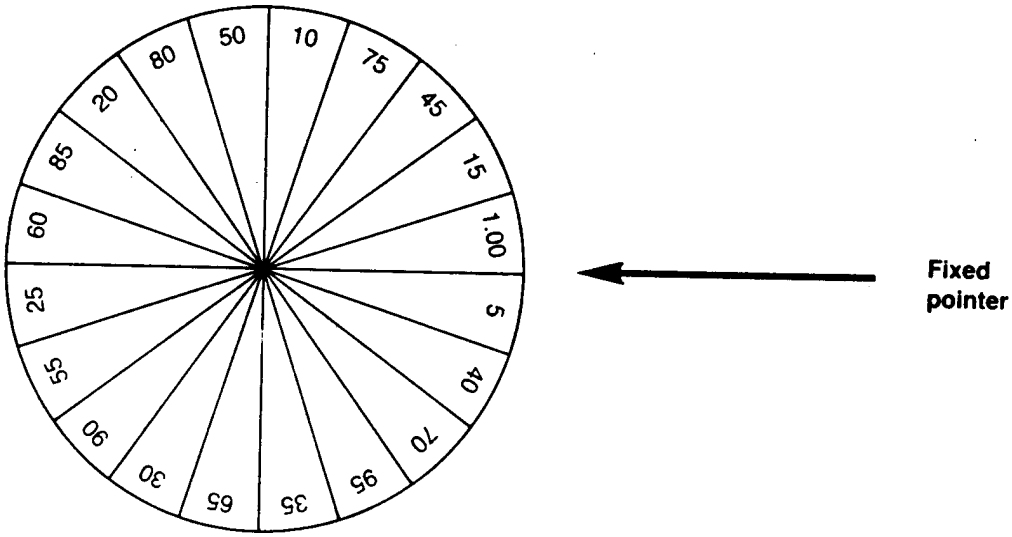


Figure 2.11
"The Price Is Right" wheel, Problem 2.21

. . . , 95, and 1.00 approximately as shown in Figure 2.11.

A contestant is allowed to spin the wheel up to two times. If the number "1.00" comes to rest against the pointer on the first spin, the contestant wins \$1,000, spins the wheel one more time, and may win more money; if not, the contestant wins nothing and leaves the game. If, on the second spin, the number "5" or "10" comes to rest against the pointer, the contestant wins \$5,000; if the number "1.00" shows up, the contestant wins \$10,000; in all other cases, the contestant wins nothing (these prizes, of course, are in addition to the \$1,000 won on the first spin).

What is the probability distribution of the payoff to this game? What is the mean and variance of this distribution?

2.22 (a) In the manner of Example 2.12, calculate the pure premium for third-party liability coverage with a \$5,000 limit.

(b) Using the data of Example 2.12, calculate the pure premium for third-party liability coverage up to \$1 million, but with a \$1,000 deductible. This means that if the claim is less than \$1,000, the company pays nothing; if the claim exceeds \$1,000, the company pays the difference between the claim and \$1,000.

2.23 The purpose of this exercise is to *verify* numerically some results applying to a model that has been found useful in automobile and other forms of insurance.

Consider an insured driver, and let X be the random variable representing the driver's total claim in dollars in a period of time—say, one year. X depends on the *number of claims* (Z) and the *size of the claims* made ($Y_1, Y_2, \ldots$) during the year as follows:

$$X = \begin{cases} 0, & \text{for } Z = 0, \\ Y_1 + Y_2 + \cdots + Y_Z & \text{for } Z = 1, 2, 3, \ldots . \end{cases}$$

For example, if a driver has two accidents in one year ($Z = 2$), the total claim (X) equals the sum ($Y_1 + Y_2$) of the claims for the two accidents.

Suppose that Z and Y_1, Y_2, Y_3, ... are independent and suppose that the Y's all have the same probability distribution, $p(y)$. It can be shown that:

$$E(X) = E(Z)E(Y), \tag{1}$$

$$Var(X) = E(Z)Var(Y) + Var(Z)[E(Y)]^2, \tag{2}$$

where $E(Y)$ and $Var(Y)$ denote the common mean and variance of the probability distributions of the Y's.

As a simple numerical illustration, suppose that the probability distribution of the *number* of accidents in one year is:

z	:	0	1	2
$p(z)$	:	0.7	0.2	0.1

and that the probability distribution of the size of the claim (in \$) is:

y	:	100	200
$p(y)$	:	0.6	0.4

(a) Show that $E(Z) = 0.4$, $Var(Z) = 0.44$, $E(Y) = 140$, and $Var(Y) = 2{,}400$.

(b) Calculate the expected value and variance of the total claim, X, according to formulas (1) and (2) above.

(c) Complete the missing entries in the probability tree shown in Figure 2.12. The numbers in parentheses are probabilities. For example, the tree shows that if a driver has two accidents (and the probability of two accidents is 0.1), the total claim could be \$200, \$300 (with probability 0.48), or \$400; the probability that the driver will have two accidents *and* that the total claim will be \$300 is 0.048.

(d) Using your answer in (c), determine the probability distribution of X, the total dollar claim in one year. Calculate the mean, $E(X)$, and variance, $Var(X)$, of this probability distribution. Compare your results with those of (b).

2.24 A die has six faces marked with 1 to 6 dots. Dice used in casinos are made to exacting engineering specifications to ensure that the six faces will show up with equal relative frequencies in the long run.

Suppose *two* fair dice are rolled. Find the probability distribution of the sum (the total number of dots) that will show up. (For example, if one die shows up a 3 and the other a 2, the sum is 5.)

2.25 *Craps* is the name of a gambling game played with two dice. If the player gets a *total* of 7 or 11 dots on the first roll of the two dice, he wins his bet; if the total is 2, 3, or 12, he loses his bet; if the total is any other number (4, 5, 6, 8, 9, or 10), that number becomes the player's "point," the bet stands, and the player throws the dice a second time. If the total on the second roll equals the player's point, the player wins his bet; if it is a 7, he loses; if it is any other number, the bet stands and the player throws the dice once more. The player continues to throw until he makes his point (in which case he wins), or rolls a 7 (in which case he loses his bet). In the version of the game played in Nevada casinos, the house acts as a bank, accepting all bets placed by players.

(a) Using your answers to Problem 2.24, complete the probability tree in Figure 2.13. Figure 2.13 shows the possible outcomes of the game *for the first*

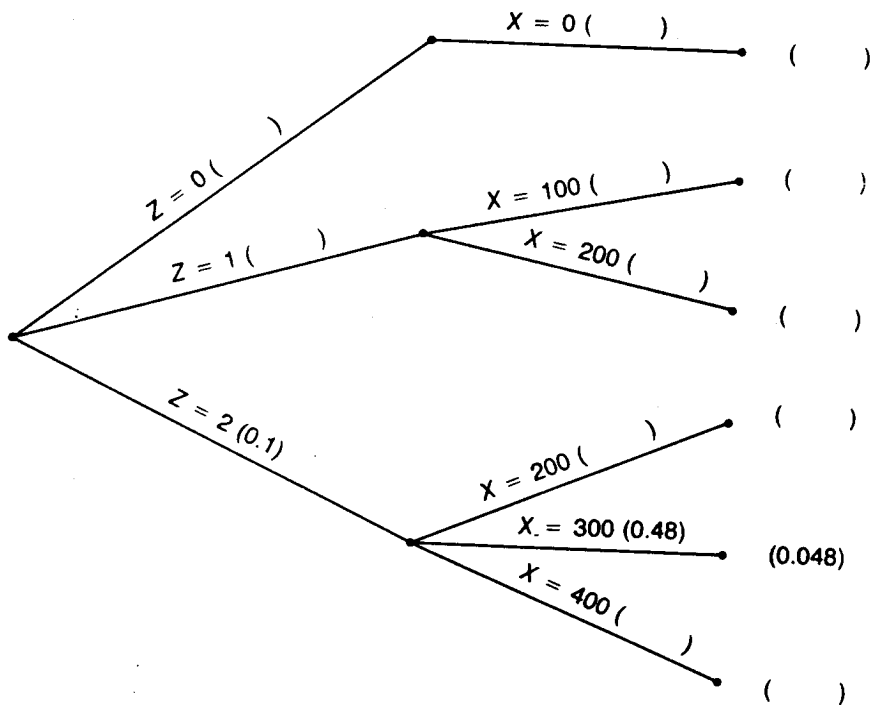


Figure 2.12
Worksheet, Problem 2.23

two rolls only. Show all conditional probabilities within the parentheses on the branches of the tree.

(b) Calculate the probability of winning on the *second* roll. Calculate the probability of losing on the second roll.

(c) What is the probability of winning on the *first or second* roll? What is the probability of losing on the first or second roll?

(d) What is the probability of having to throw the dice three or more times before the bet is resolved?

(e) Suppose that the player's point is 4. Show that the probability of winning on the second roll is 3/36; on the third roll, $(27/36) \times (3/36)$; on the fourth roll, $(27/36)^2(3/36)$; and that, in general, the probability of winning in exactly k rolls after the first is $(27/36)^{k-1}(3/36)$. Show also that the probability of winning eventually is

$$\sum_{k=1}^{\infty}(\frac{27}{36})^{k-1}(\frac{3}{36}),$$

or 12/36. *Hint:* $1 + a + a^2 + a^3 + \cdots = 1/(1-a)$ if $0 < a < 1$.

(f) Again assuming that the player's point is 4, show that the probability of losing the bet eventually is 24/36.

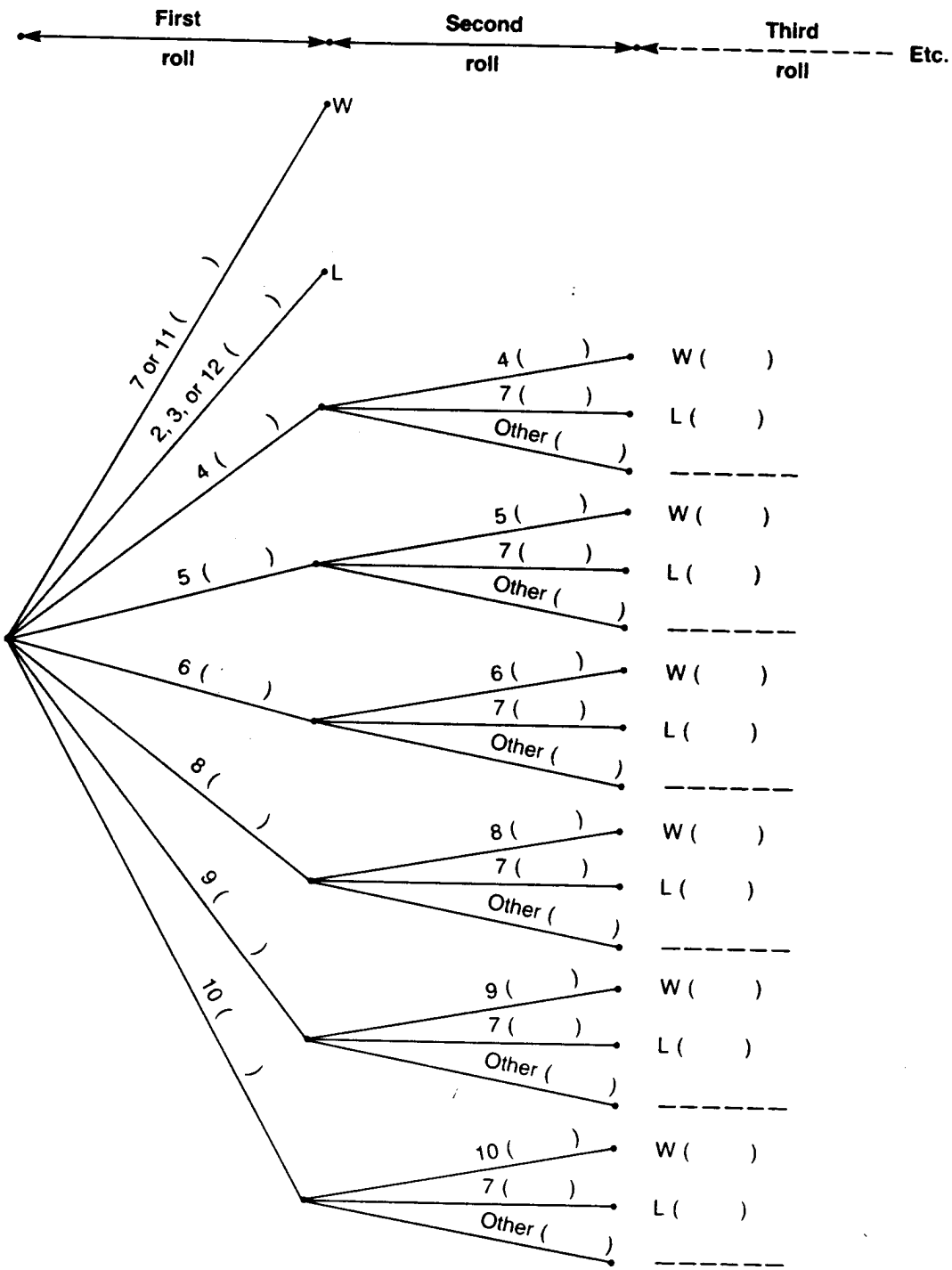


Figure 2.13
Worksheet, Problem 2.25

(g) Show that the probabilities of eventually winning and losing the bet, given that the player's point is any other number (5, 6, 8, 9, or 10) are $p_W/(1-p_C)$ and $p_L/(1-p_C)$ respectively, where p_W is the probability of getting the point, p_L the probability of getting a 7, and p_C the probability of getting a number other than the point or 7 in any one roll of two dice.

(h) Calculate the overall probability of winning and losing a bet. For every dollar bet, how much can a player expect to win or lose in the long run? (Alternatively, what is the profit/loss margin of the casino in this game?)

2.26 Following a winter of little snowfall, disastrous sales, and large inventories, the manufacturers of Bull—a popular make of snowthrowers—launched their *There's No Risk* fall advertising campaign. A portion of an advertisement which appeared in newspapers and magazines is shown in Table 2.4.

Table 2.4
Advertisement for Bull snowthrowers

IF IT DOESN'T SNOW WE'LL RETURN YOUR DOUGH!
AND YOU KEEP THE SNOWTHROWER!

If it snows less than:	You keep the snowthrower and you receive:
20% of AVERAGE SNOWFALL	100% REFUND of suggested retail price
30% of AVERAGE SNOWFALL	70% REFUND of suggested retail price
40% of AVERAGE SNOWFALL	60% REFUND of suggested retail price
50% of AVERAGE SNOWFALL	50% REFUND of suggested retail price

The campaign became widely known and was imitated by competitors and manufacturers of other winter equipment.

"Just buy a new Bull snowthrower," said the ads, "before December 25 and forget about how much snow we're going to have. ... If the snowfall in your area is less than 20% of average, you will be refunded 100% of Bull's suggested retail price for the unit. Even if it snows less than 50% of average, you get a 50% refund ... *And you keep the unit!*"

The offer applied to a number of models ranging in price from about \$500 to \$2,200. "Average" snowfall was defined as the official 30-year average. A year's snowfall is the total snowfall from July 1 to next June 30. The buyers' eligibility for refund, therefore, would not be determined until June 30 of next year.

Table 2.5 shows the official monthly and annual snowfall in the most recent 30-year period. (No snowfall was ever recorded in months not listed in the table.)

Table 2.5
Monthly and annual snowfall,
most recent 30-year period (in centimeters)

Year	Oct.	Nov.	Dec.	Jan.	Feb.	Mar.	Apr.	May	Total
1	0.0	0.5	11.9	42.2	31.0	11.9	19.8	0.0	117.3
2	0.0	27.4	26.2	32.8	19.6	10.2	4.6	0.0	120.8
3	0.0	15.5	21.3	30.5	43.7	39.9	0.0	0.5	151.4
4	0.0	1.8	13.5	63.5	63.0	42.4	5.3	0.0	189.5
5	0.0	3.8	15.2	27.7	30.2	41.4	15.7	0.0	134.0
6	2.0	1.8	24.9	26.7	77.5	2.5	5.3	0.0	140.7
7	0.0	1.3	49.5	19.6	16.3	20.1	13.2	1.8	121.8
8	0.0	3.3	23.9	15.7	33.8	32.3	10.2	0.0	119.2
9	1.3	3.6	15.0	55.4	41.9	47.0	10.4	0.0	174.6
10	0.0	9.4	20.3	70.9	13.5	13.5	24.6	1.3	153.5
11	0.0	2.3	16.0	49.0	48.3	28.7	16.3	0.5	161.1
12	0.0	7.6	50.5	51.1	7.9	39.9	0.0	0.0	157.0
13	12.2	3.8	35.3	15.0	14.0	9.9	0.0	0.0	90.2
14	0.0	4.6	67.8	37.3	14.5	16.3	2.8	0.0	143.3
15	0.0	14.7	33.8	33.5	42.2	38.4	1.3	0.0	163.9
16	0.0	14.2	49.5	31.0	47.5	40.9	11.9	0.0	195.0
17	0.0	3.1	58.0	10.4	24.9	14.2	1.3	0.0	111.9
18	0.0	1.5	27.4	30.5	25.7	16.5	0.8	0.0	102.4
19	0.0	0.8	64.5	13.2	29.0	26.4	24.4	0.0	158.3
20	0.0	3.0	41.4	53.0	4.9	38.3	10.2	0.0	150.8
21	0.0	5.9	67.6	70.5	6.2	22.6	1.9	0.0	174.7
22	0.0	13.4	14.2	63.4	27.3	11.4	0.8	0.0	130.5
23	0.0	0.8	56.3	55.4	26.2	1.2	37.6	0.0	177.5
24	0.0	10.9	28.3	10.8	14.9	38.1	0.2	0.0	103.2
25	0.0	0.0	24.4	27.0	24.0	4.1	0.0	0.0	79.5
26	0.0	2.6	29.8	69.3	28.8	29.0	7.0	0.0	166.5
27	0.1	3.1	30.8	35.2	27.6	20.5	9.2	0.0	126.5
28	1.2	9.7	43.8	45.4	30.2	20.7	10.9	0.4	162.3
29	0.0	0.5	35.6	40.8	30.4	28.9	7.3	0.0	143.5
30	0.0	2.2	17.0	19.1	15.6	12.2	4.0	0.0	70.1
Average	0.6	5.8	33.8	38.2	28.7	24.0	8.6	0.1	139.7

(a) Estimate the probability that this winter's snowfall will be less than 20% of the average. Estimate the probabilities that it will be less than 30%, less than 40%, or less than 50% of the average. What are the chances of getting any refund? Comment.

(b) Consider the variation of the *There's No Risk* warranty shown in Table 2.6.

Calculate the expected cost of this warranty per unit sold of a snowthrower model retailing at $1,000.

(c) What would be the effect of this warranty if its terms were to apply to *a given month's* (say February's) snowfall?

Table 2.6
Bull warranty, new version

If winter's snowfall is less than:	Refund is:
20%*	100%†
30%	70%
40%	60%
50%	50%
60%	40%
70%	30%
80%	20%
*Of official average	†Of suggested retail price