2 The Probability Model

This chapter describes a well-accepted model for the analysis of random experiments which we refer to as the *Probability Model*. We also define a set algebra suitable for defining sets of events, and describe how measures of likelihood or *probabilities* are assigned to these events. Probabilities provide quantitative numerical values to the likelihood of occurrence of events.

Events do not always occur independently. In fact, it is the very *lack* of independence that allows us to infer one fact from another. Here we give a mathematical meaning to the concept of independence and further develop relations to deal with probabilities when events are or are not independent.

Several illustrations and examples are given throughout this chapter on basic probability. In addition, a number of applications of the theory to some basic electrical engineering problems are given to provide motivation for further study of this topic and those to come.

2.1 The Algebra of Events

We have seen in Chapter 1 that the collection of all possible outcomes of a random experiment comprise the *sample space*. Outcomes are members of the sample space and events of interest are represented as *sets* (see Fig. 2.1). In order to describe these events

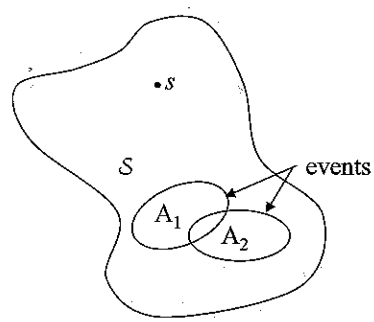


Figure 2.1 Abstract representation of the sample space $\mathcal{S}$ with element s and sets A_1 and A_2 representing events.

and compute their probabilities in a consistent manner it is necessary to have a formal representation for operations involving events. More will be said about representing the sample space in Section 2.1.2; for now, we shall focus on the methods for describing relations among events.

2.1.1 Basic operations

In analyzing the outcome of a random experiment, it is usually necessary to deal with events that are derived from other events. For example, if A is an event, then A^c, known as the *complement* of A, represents the event that "A did not occur." The complement of the sample space is known as the *null event*, $\emptyset = \mathcal{S}^c$. The operations of multiplication and addition will be used to represent certain combinations of events (known as intersections and unions in set theory). The statement "$A_1 \cdot A_2$," or simply "A_1A_2" represents the event that *both* event A_1 and event A_2 have occurred (intersection), while the statement "$A_1 + A_2$" represents the event that *either* A_1 or A_2 *or both* have occurred (union).[1]

Since complements and combinations of events are themselves events, a formal structure for representing events and derived events is needed. This formal structure is in fact a set of sets known in mathematics as an *algebra* or a *field* and referred to here as the *algebra of events*. Table 2.1 lists the two postulates that define an algebra $\mathcal{A}$.

1. *If* $A \in \mathcal{A}$ *then* $A^c \in \mathcal{A}$.
2. *If* $A_1 \in \mathcal{A}$ *and* $A_2 \in \mathcal{A}$ *then* $A_1 + A_2 \in \mathcal{A}$.

Table 2.1 Postulates for an algebra of events.

Table 2.2 lists seven axioms that define the properties of the operations. Together these

$A_1{A_1}^c = \emptyset$	Mutual exclusion
$A_1\mathcal{S} = A_1$	Inclusion
$({A_1}^c)^c = A_1$	Double complement
$A_1 + A_2 = A_2 + A_1$	Commutative law
$A_1 + (A_2 + A_3) = (A_1 + A_2) + A_3$	Associative law
$A_1(A_2 + A_3) = A_1A_2 + A_1A_3$	Distributive law
$(A_1A_2)^c = {A_1}^c + {A_2}^c$	DeMorgan's law

Table 2.2 Axioms of operations on events.

tables can be used to show all of the properties of the algebra of events. For example, the postulates state that the event $A_1 + A_2$ is included in the algebra. The postulates in conjunction with the last axiom (DeMorgan's law) show that the event "A_1A_2" is also included in the algebra. Table 2.3 lists some other handy identities that can be derived from the axioms and the postulates. You will find that you use many of the results in Tables 2.2 and 2.3 either implicitly or explicitly in solving problems involving events and their probability. Notice especially the two distributive laws; addition is distributive over multiplication (Table 2.3) as well as *vice versa* (Table 2.2).

Since the events "$A_1 + A_2$" and "A_1A_2" are included in the algebra, it is easy to

[1] The operations represented as multiplication and addition are commonly represented with the intersection $\cap$ and union $\cup$ symbols. Except for the case of multiple such operations, we will adhere to the former notation introduced above.

$\mathcal{S}^c = \emptyset$	
$A_1 + \emptyset = A_1$	Inclusion
$A_1A_2 = A_2A_1$	Commutative law
$A_1(A_2A_3) = (A_1A_2)A_3$	Associative law
$A_1 + (A_2A_3) = (A_1 + A_2)(A_1 + A_3)$	Distributive law
$(A_1 + A_2)^c = {A_1}^c{A_2}^c$	DeMorgan's law

Table 2.3 Additional identities in the algebra of events.

show by induction for any finite number of events A_i, $i = 1, 2, \ldots, N$, that the events

$$\bigcup_{i=1}^{N} A_i = A_1 + A_2 + \cdots + A_N$$

and

$$\bigcap_{i=1}^{N} A_i = A_1A_2 \cdots A_N$$

are also included in the algebra. In many cases it is important that the sum and product of a countably infinite number of events have a representation in the algebra. For example, suppose an experiment consists of measuring a random voltage, and the events A_i are defined as "$i - 1 \leq$ voltage $< i$; $i = 1, 2, \ldots$." Then the (infinite) sum of these events, which is the event "voltage ≥ 0," should be in the algebra. An algebra that includes the sum and product of an infinite number of events, that is,

$$\bigcup_{i=1}^{\infty} A_i = A_1 + A_2 + A_3 + \cdots$$

and

$$\bigcap_{i=1}^{\infty} A_i = A_1A_2A_3 \cdots$$

is called a sigma-algebra or a sigma-field. The algebra of events is defined to be such an algebra.

Since the algebra of events can be thought of as an algebra of sets, events are often represented as Venn diagrams. Figure 2.2 shows some typical Venn diagrams for a sample space and its events. The notation '$\subset$' is used to mean one event is "contained" in another and is defined by

$$A_1 \subset A_2 \Longleftrightarrow A_1{A_2}^c = \emptyset \tag{2.1}$$

2.1.2 Representation of the sample space

Students of probability may at first have difficulty in defining the sample space for an experiment. It is thus worthwhile to spend a little more time on this concept.

We begin with two more ideas from the algebra of events. Let $A_1, A_2, A_3, \ldots$ be a finite or countably infinite set of events with the following properties:

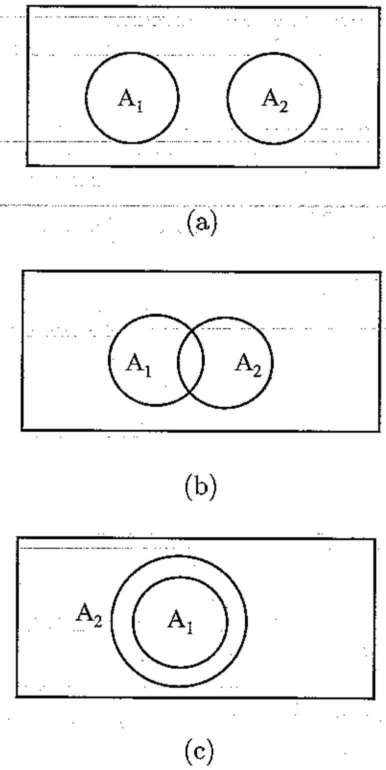


Figure 2.2 Venn diagram for events. (a) Events with no commonality ($A_1A_2 = \emptyset$). (b) Events with some commonality ($A_1A_2 \neq \emptyset$). (c) One event contained in another ($A_1 \subset A_2$).

1. The events are *mutually exclusive*. This means that only one event can occur at a time, i.e., the occurrence of one event precludes the occurrence of other events. Equivalently,

$$A_iA_j = \emptyset \quad \text{for } i \neq j$$

2. The events are *collectively exhaustive*. In other words, one of the events A_i must always occur. That is,

$$A_1 + A_2 + A_3 + \cdots = \mathcal{S}$$

A set of events that has *both* properties is referred to as a *partition*.

Now, for an experiment with discrete outcomes, the following provides a working definition of the sample space [1]:

> **The Sample Space is represented by the finest-grain, mutually exclusive, collectively exhaustive set of outcomes for an experiment.**

You can see that the elements of the sample space have the properties of a partition; however, the outcomes defining the sample space must also be "*finest-grain.*" This is important, since without this property it may not be possible to represent all of the events of interest in the experiment. A Venn diagram is generally not sufficient to represent the sample space in solving problems, because the representation usually needs to be more explicit.

A discrete sample space may be just a listing of the possible outcomes (see Example

2.1) or could take the form of some type of diagram. For example, consider the rolling of a pair of dice. The sample space might be drawn as shown in Fig. 2.3.

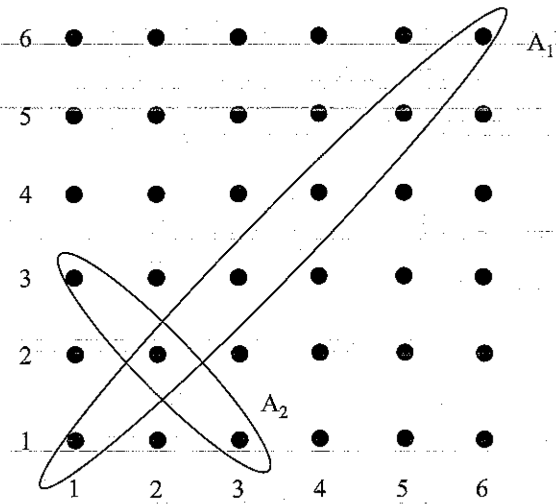


Figure 2.3 Sample space corresponding to roll of the dice. A_1 is the event "rolling doubles"; A_2 is the event "rolling a '4'."

The black dots represent the outcomes of the experiment, which are mutually exclusive and collectively exhaustive. Some more complex events, such as "rolling doubles" are also shown in the figure. It will be seen later that the probabilities of such more complicated events can be computed by simply adding the probabilities of the outcomes of which they are comprised.

For an experiment in which the outcome is a real number or a set of real numbers, the sample space is usually chosen as a subset of the real line or a subset of N-dimensional Euclidean space ($\mathcal{R}^N$), as appropriate. This is the case for some of the examples in Chapter 1. If the outcome of the experiment were complex numbers, then you would probably define the sample space as a subspace of the space of complex numbers ($\mathcal{C}^N$). These are examples of *continuous* sample spaces. We shall emphasize that in the solution of most problems involving probability, a first step is *to find an appropriate representation for the sample space.*

2.2 Probability of Events

2.2.1 Defining probability

We have seen that probability represents the likelihood of occurrence of events. The probability model, when properly formulated, can be used to show that the *relative frequency* for the occurrence of an event in a large number of repetitions of the experiment, defined as

$$\text{relative frequency} = \frac{\text{number of occurrences of the event}}{\text{number of repetitions of the experiment}}$$

converges to the probability of the event. Although probability could be *defined* in this way, it is more common to use the axiomatic development given below.

Probability is conveniently represented in a Venn diagram if you think of the area covered by events as measures of probability. For example, if the area of the sample space $\mathcal{S}$ is normalized to one, then the area of overlap of events A_1 and A_2 in Fig. 2.2(b) can be thought of as representing the probability of the event A_1A_2. If the probability of this joint event were larger, the events might be drawn to show greater overlap.

Probability can be defined formally by the following axioms:

(I) The probability of any event is nonnegative.

$$\Pr[\mathrm{A}] \geq 0 \tag{2.2}$$

(II) The probability of the universal event (i.e., the entire sample space) is 1.

$$\Pr[\mathcal{S}] = 1 \tag{2.3}$$

(III) If A_1 and A_2 are mutually exclusive, then

$$\Pr[A_1 + A_2] = \Pr[A_1] + \Pr[A_2] \quad (\text{if} \quad A_1A_2 = \emptyset) \tag{2.4}$$

(IV) If $\{A_i\}$ represent a countably infinite set of mutually exclusive events, then

$$\Pr\left[\bigcup_{i=1}^{\infty} A_i\right] = \sum_{i=1}^{\infty} \Pr[A_i] \quad (\text{if} \quad A_iA_j = \emptyset \quad i \neq j) \tag{2.5}$$

Although the additivity of probability for any finite set of disjoint events follows from (III), the property has to be stated explicitly for an infinite set in (IV). These axioms and the algebra of events can be used to show a number of other properties, some of which are discussed below.

From axioms (II) and (III), the probability of the complement of an event is

$$\Pr[A^c] = 1 - \Pr[A] \tag{2.6}$$

Since by (I) the probability of any event is greater than or equal to zero, it follows from (2.6) that $\Pr[A] \leq 1$; thus

$$0 \leq \Pr[A] \leq 1 \tag{2.7}$$

for any event A.

If $A_1 \subset A_2$ then A_2 can be written as $A_2 = A_1 + {A_1}^c A_2$ (see Fig. 2.2(c)). Since the events A_1 and ${A_1}^c A_2$ are mutually exclusive, it follows from (III) and (I) that

$$\Pr[A_2] \geq \Pr[A_1]$$

From (2.6) and axiom (II), it follows that the probability of the null event is zero:

$$\Pr[\emptyset] = 0 \tag{2.8}$$

Thus it also follows that if A_1 and A_2 are mutually exclusive, then $A_1A_2 = \emptyset$ and consequently

$$\Pr[A_1A_2] = 0$$

If events A_1 and A_2 are not mutually exclusive, i.e., they may occur together, then one has the general relation

$$\Pr[A_1 + A_2] = \Pr[A_1] + \Pr[A_2] - \Pr[A_1A_2] \tag{2.9}$$

This is not an addition property; rather it can be derived using axioms (I) through (IV) and the algebra of events. It can be intuitively justified on the grounds that in summing the probabilities of the event A_1 and the event A_2, one has counted the common event "A_1A_2" twice (see Fig. 2.2(b)). Thus the probability of the event "A_1A_2" must be subtracted to obtain the probability of the event "$A_1 + A_2$".

These various derived properties are summarized in Table 2.4 below. It is a useful excercise to depict these properties (and the axioms as well) as Venn diagrams.

$\Pr[A^c] = 1 - \Pr[A]$
$0 \leq \Pr[A] \leq 1$
If $A_1 \subseteq A_2$ then $\Pr[A_1] \leq \Pr[A_2]$
$\Pr[\emptyset] = 0$
If $A_1A_2 = \emptyset$ then $\Pr[A_1A_2] = 0$
$\Pr[A_1 + A_2] = \Pr[A_1] + \Pr[A_2] - \Pr[A_1A_2]$

Table 2.4 Some corollaries derived from the axioms of probability.

As a final consideration, let $A_1, A_2, A_3, \ldots$ be a finite or countably infinite set of mutually exclusive and collectively exhaustive events (see Section 2.1.2). Recall that such a set of events is referred to as a *partition*. The probabilities of the events in a partition satisfy the relation

$$\sum_i \Pr[A_i] = 1 \tag{2.10}$$

and if B is any other event, then

$$\boxed{\sum_i \Pr[A_iB] = \Pr[B]} \tag{2.11}$$

The latter result is referred to as the *principle of total probability* and is frequently used in solving problems. The relation (2.11) is illustrated by a Venn diagram in Fig. 2.4. The event B is comprised of all of the pieces that represent intersections or overlap of event B with the events A_i.

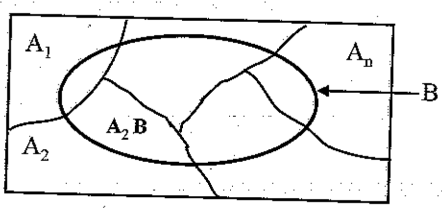


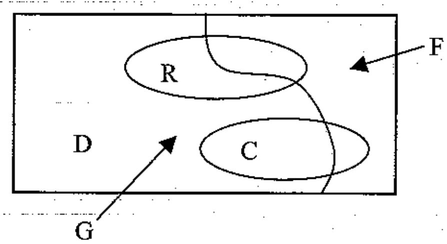
Figure 2.4 Venn diagram illustrating the principle of total probability.

Let us consider the following example to illustrate the formulas in this section.

Example 2.1: Simon's Surplus Warehouse has large barrels of mixed electronic components (parts) that you can buy by the handful or by the pound. You are not allowed to select parts individually. Based on your previous experience, you have determined that in one barrel, 29% of the parts are bad (faulted), 3% are bad resistors, 12% are good resistors, 5% are bad capacitors, and 32% are diodes. You decide to assign probabilities based on these percentages. Let us define the following events:

Event	Symbol
Bad (faulted) component	F
Good component	G
Resistor	R
Capacitor	C
Diode	D

A Venn diagram representing this situation is shown below along with probabilities of various events as given:



$$\Pr[F] = 0.29$$
$$\Pr[FR] = 0.03$$
$$\Pr[GR] = 0.12$$
$$\Pr[FC] = 0.05$$
$$\Pr[D] = 0.32$$

We can answer a number of questions.

1. What is the probability that a component is a resistor (either good *or* bad)?
   Since the events F and G form a partition of the sample space, we can use the principle of total probability (2.11) to write
   $$\Pr[R] = \Pr[GR] + \Pr[FR] = 0.12 + 0.03 = 0.15$$
2. You have no use for either defective parts or resistors. What is the probability that a part is either defective and/or a resistor?
   Using (2.9) and the previous result we can write
   $$\Pr[F + R] = \Pr[F] + \Pr[R] - \Pr[FR] = 0.29 + 0.15 - 0.03 = 0.41$$
3. What is the probability that a part is useful to you?
   Let U represent the event that the part is useful. Then (see (2.6))
   $$\Pr[U] = 1 - \Pr[U^c] = 1 - 0.41 = 0.59$$
4. What is the probability of a bad diode?
   Observe that the events R, C, and G form a partition, since a component has to be one and only one type of part. Then using (2.11) we write
   $$\Pr[F] = \Pr[FR] + \Pr[FC] + \Pr[FD]$$
   Substituting the known numerical values and solving yields
   $$0.29 = 0.03 + 0.05 + \Pr[FD] \quad \text{or} \quad \Pr[FD] = 0.21$$

□

It is worthwhile to consider what an appropriate representation of the sample space would be for this example. While the Venn diagram shown above represents the sample space in an abstract way, a more explicit representation is most useful. In this case, since a part can be a bad resistor, good resistor, bad capacitor, and so on, a suitable representation is the list of outcomes:

sample space: { FR GR FC GC FD GD }

(You may want to check that this satisfies the requirements discussed in Section 2.1.2.) In later sections of this chapter, you will see that the answers to the four questions posed in this example can be easily obtained if the probabilities of these six outcomes were specified or could be computed. In this example however, the probabilities of all of these experimental outcomes are not known, i.e., only partial information is given.

2.2.2 Statistical independence

There is one more concept that is frequently used when solving basic problems in probability, but cannot be derived from either the algebra of events or any of the axioms. Because of its practical importance in solving problems, we introduce this concept early in our discussion of probability:

> Two events A_1 and A_2 are said to be *statistically independent* if and only if
> $$\Pr[A_1A_2] = \Pr[A_1] \cdot \Pr[A_2] \tag{2.12}$$

That is, for two independent events, the probability of both occuring is the product of the probabilities of the individual events. Independence of events is not generally something that you are asked to *prove* (although it may be). More frequently it is an assumption made when the conditions of the problem warrant it. The idea can be extended to multiple events. For example, if A_1, A_2 and A_3 are said to be *mutually* independent if and only if

$$\Pr[A_1A_2A_3] = \Pr[A_1]\Pr[A_2]\Pr[A_3]$$

Note also that for independent events, (2.9) becomes

$$\Pr[A_1 + A_2] = \Pr[A_1] + \Pr[A_2] - \Pr[A_1]\Pr[A_2]$$

so the computation of probability for the union of two events is also simplified.

The concept of statistical independence, as we have already said, cannot be derived from anything else presented so far, and does not have a convenient interpretation in terms of a Venn diagram. However it can be argued in terms of the relative frequency interpretation of probability. Suppose two events are "independent" in that they arise from two different experiments that have nothing to do with each other. Let it be known that in N_1 repetitions of the first experiment there are k_1 occurrences of the event A_1 and in N_2 repetitions of the second experiment there are k_2 occurrences of the event A_2. If N_1 and N_2 are sufficiently large, then the relative frequencies k_i/N_i remain approximately constant as N_i is increased. Let us now perform both experiments together a total of N_1N_2 times. Consider the event A_1. Since it occurs k_1 times in N_1 repetitions of the experiment it will occur N_2k_1 times in N_1N_2 repetitions of the experiment. Now consider those N_2k_1 cases where A_1 occured. Since event A_2 occurs k_2 times in N_2 repetitions, it will occur k_1k_2 times in these N_2k_1 cases where A_1 has occured. In other words the two events occur together k_1k_2 times in all of these N_1N_2 repetitions of the experiments. The relative frequency for the occurrence of the two events together is therefore

$$\frac{k_1k_2}{N_1N_2} = \frac{k_1}{N_1} \cdot \frac{k_2}{N_2}$$

which is the product of the relative frequencies of the individual events. So given the relative frequency interpretation of probability, the definition (2.12) makes good sense.

2.3 Some Applications

Let us continue with some examples in which many of the ideas discussed so far in this chapter are illustrated.

2.3.1 Repeated independent trials

Many problems involve a repetition of independent events. As a typical example of this, consider the experiment of tossing a coin three times in succession. The result of the second toss is independent of the result of the first toss; likewise the result of the third toss is independent of the result of the first two tosses. Let us denote the probability of a "head" (H) on any toss by p and the probability of a "tail" (T) by $q = 1 - p$. (For a fair coin, $p = q = 1/2$, but let us be more general.) Since the results of the tosses are independent, the probability of any experimental outcome such as HHT is simply the product of the probabilities: $p \cdot p \cdot q = p^2q$. The sequence HTH has the same probability: $p \cdot q \cdot p = p^2q$. Experiments of this type are said to involve *repeated independent trials.*

An application which is familiar to electrical and computer engineers is the transmission of a binary sequence over a communication channel. In many practical cases the bits (1 or 0) can be modeled as independent events. Thus the probability of a bit sequence of any length 1 0 1 1 0 1 ... is simply equal to the product of probabilities: $p \cdot q \cdot p \cdot p \cdot q \cdot p \cdots$. This considerably simplifies the analysis of such systems.

An example is given below where the outcomes of the experiment are based on repeated independent trials. Once the sample space and probabilities of the outcomes have been specified, a number of other probabilistic questions can be answered.

Example 2.2: Diskettes selected from the bins at Simon's Surplus are as likely to be good as to be bad. If three diskettes are selected independently and at random, what is the probability of getting exactly *three* good diskettes? Exactly *two* good diskettes? How about *one* good diskette?

Evidently buying a diskette at Simon's is like tossing a coin. The sample space is represented by the listing of outcomes shown below: where G represents a good diskette

BBB	BBG	BGB	BGG	GBB	GBG	GGB	GGG
A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8

and B represents a bad one. Each outcome is labeled as an event A_i; note that the events A_i are mutually exclusive and collectively exhaustive.

Three good diskettes is represented by only the last event (A_8) in the sample space. Since the probability of selecting a good diskette and the probability of selecting a bad diskette are both equal to $\frac{1}{2}$, and the selections are independent, we can write

$$\Pr[3 \text{ good diskettes}] = \Pr[A_8] = \Pr[G]\Pr[G]\Pr[G] = \left(\tfrac{1}{2}\right)^3 = \tfrac{1}{8}$$

(see Section 2.2.2).

The result of two good diskettes is represented by the events A_4, A_6, and A_7. By a procedure similar to the above, each of these events has probability $\frac{1}{8}$. Since these three events are mutually exclusive, their probabilities add (see Section 2.2.1). That is,

$$\Pr[2 \text{ good diskettes}] = \Pr[A_4 + A_6 + A_7] = \Pr[A_4] + \Pr[A_6] + \Pr[A_7] = \tfrac{1}{8} + \tfrac{1}{8} + \tfrac{1}{8} = \tfrac{3}{8}$$

Finally, a single good diskette is represented by the events A_2, A_3, and A_5. By an identical procedure it is found that this result also occurs with probability $\frac{3}{8}$.

□

To be sure you understand the steps in this example, you should repeat this example for the case where the probability of selecting a good diskette is increased to $\frac{5}{8}$. In this case the probability of all of the events A_i are not equal. For example, the probability of the event A_6 is given by $\Pr[G]\Pr[B]\Pr[G] = \frac{5}{8} \cdot \frac{3}{8} \cdot \frac{5}{8} = \frac{75}{512}$. When you work through the example you will find that the probabilities of three and two good diskettes is increased to $\frac{125}{512}$ and $\frac{225}{512}$ respectively while the probability of just one good diskette is decreased to $\frac{135}{512}$.

2.3.2 Problems involving counting

Many important problems involve adding up the probabilities of a number of equally-likely events. These problems involve some basic combinatorial analysis, i.e., counting the number of possible events, configurations, or outcomes in an experiment.

Some discussion of combinatorial methods is provided in Appendix A. All of these deal with the problem of counting the number of pairs, triplets, or k-tuples of elements that can be formed under various conditions. Let us review the main results here.

Rule of product. In the formation of k-tuples consisting of k elements where there are N_i choices for the i^{th} element, the number of possible k-tuples is $\prod_{i=1}^{k} N_i$. An important special case is when there are the *same* number of choices N for each element. The number of k-tuples is then simply N^k.

Permutations. A *permutation* is a k-tuple formed by selecting from a set of N *distinct* elements, where each element can only be selected once. (Think of forming words from a finite alphabet where each letter can be used only once.) There are N choices for the first element, $N-1$ choices for the second element, $\ldots$, and $N-k+1$ choices for the k^{th} element. The number of such permutations is given by

$$N \cdot (N-1) \cdots (N-k+1) = \frac{N!}{(N-k)!}$$

For $k = N$ the result is simply $N!$.

Combinations. A *combination* is a k-tuple formed by selecting from a set of N distinct elements where the *order* of selecting the elements makes no difference. For example, the sequences ACBED and ABDEC would represent two different permutations, but only a single *combination* of the letters A through E. The number of combinations k from a possible set of N is given by the binomial coeffient

$$\binom{N}{k} = \frac{N!}{k!(N-k)!}$$

This is frequently read as "N *choose* k," which provides a convenient mnemonic for its interpretation.

Counting principles provide a way to assign or compute probability in many cases. This is illustrated in a number of examples below.

The following example illustrates use of some basic counting ideas.

Example 2.3: In sequences of k binary digits, 1's and 0's are equally likely. What is the probability of encountering a sequence with a single '1' (in any position) and all other digits zero?

Imagine drawing the sample space for this experiment consisting of all possible sequences. Using the rule of product we see that there are 2^k events in the sample space and they are all equally likely. Thus we assign probability $1/2^k$ to each outcome in the sample space.

Now, there are just k of these sequences that have exactly one '1'. Thus the probability is $k/2^k$.

□

The next example illustrates the use of permutation.

Example 2.4: IT technician Chip Gizmo has a cable with four twisted pairs running from each of four offices to the service closet; but he has forgotten which pair goes to which office. If he connects one pair to each of four telephone lines arbitrarily, what is the probability that he will get it right on the first try?

The number of ways that four twisted pairs could be assigned to four telephone lines is $4! = 24$. Assuming that each arrangement is equally likely, the probability of getting it right on the first try is $1/24 = 0.0417$.

□

The following example illustrates the use of permutations versus combinations.

Example 2.5: Five surplus computers are available for adoption. One is an IBM, another is an HP, and the rest are nondescript. You can request two of the suplus computers but cannot specify which ones. What is the probability that you get the IBM and the HP?

Consider first the experiment of randomly selecting two computers. Let's call the computers A, B, C, D, and E. The sample space is represented by a listing of pairs

$$\mathrm{A,B} \quad \mathrm{B,A} \quad \mathrm{A,C} \quad \mathrm{C,A} \quad \cdots$$

representing the computers chosen. Each pair is a *permutation*, and there are $5!/(5-2)! = 5 \cdot 4 = 20$ such permutations that represent the outcomes in the sample space. Thus each outcome has a probability of $1/20$. We are interested in two of these outcomes, namely IBM,HP or HP,IBM. The probability is thus $2/20$ or $1/10$.

Another simpler approach is possible. Since we do not need to distinguish between ordering in the elements of a pair, we could choose our sample space to be

$$\mathrm{A,B} \quad \mathrm{A,C} \quad \mathrm{A,D} \quad \mathrm{A,E} \quad \cdots$$

where events such as B,A and C,A are not listed since they are equivalent to A,B and A,C. The number of pairs in this new sample space is the number of *combinations* of 5 objects taken 2 at a time:

$$\binom{5}{2} = \frac{5!}{2!(5-2)!} = 10$$

Thus each outcome in this sample space has probability $1/10$. We are interested only in the single outcome IBM,HP. Therefore this probability is again $1/10$.

□

The final example for this section illustrates a more advanced use of the combinatoric ideas.

Example 2.6: DEAL Computers Incorporated manufactures some of their computers in the US and others in Lower Slobbovia. The local DEAL factory store has a stock of 10 computers that are US made and 15 that are foreign made. You order five computers from the DEAL store which are randomly selected from this stock. What is the probability that two or more of them are US-made?

The number of ways to choose 5 computers from a stock of 25 is

$$\binom{25}{5} = \frac{25!}{5!(25-5)!} = 53130$$

This is the total number of possible outcomes in the sample space.

Now consider the number of outcomes where there are *exactly* 2 US-made computers in a selection of 5. Two US computers can be chosen from a stock of 10 in $\binom{10}{2}$ possible ways. For each such choice, three non-US computers can be chosen in $\binom{15}{3}$ possible ways. Thus the number of outcomes where there are exactly 2 US-made computers is given by

$$\binom{10}{2} \cdot \binom{15}{3}$$

Since the problem asks for "two or more" we can continue to count the number of ways there could be exactly 3, exactly 4, and exactly 5 out of a selection of five computers. Therefore the number of ways to choose 2 *or more* US-made computers is

$$\binom{10}{2}\binom{15}{3} + \binom{10}{3}\binom{15}{2} + \binom{10}{4}\binom{15}{1} + \binom{10}{5} = 36477$$

The probability of two or more US-made computers is thus the ratio 36477/53130 = 0.687.

□

2.3.3 Network reliability

Consider the set of communication links shown in Fig. 2.5. In both cases it is desired

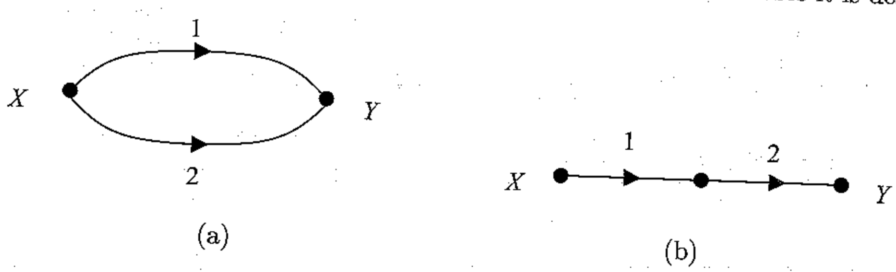


Figure 2.5 Connection of communication links. (a) Parallel. (b) Series.

to communicate between points X and Y. Let A_i represent the event that link i fails and F be the event that there is failure to communicate between X and Y. Further, assume that the link failures are *independent* events. Then for the parallel connection (Fig. 2.5(a))

$$\Pr[F] = \Pr[A_1A_2] = \Pr[A_1]\Pr[A_2]$$

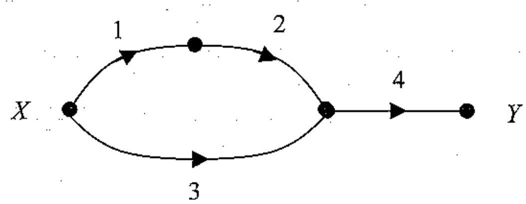
where the last equality follows from the fact that events A_1 and A_2 are independent. For the series connection (Fig. 2.5(b))

$$\Pr[F] = \Pr[A_1 + A_2] = \Pr[A_1] + \Pr[A_2] - \Pr[A_1A_2] = \Pr[A_1] + \Pr[A_2] - \Pr[A_1]\Pr[A_2]$$

where we have applied (2.9) and again used the fact that the events are independent.

The algebra of events and the rules for probability can be used to solve some additional simple problems such as in the following example.

Example 2.7: In the simple communication network shown below, link failures occur in-



dependently with probability p. What is the largest value of p that can be tolerated if the overall probability of failure of communication between X and Y is to be kept less than 10^{-3}?

Let F represent the failure of communication; this event can be expressed as $\mathrm{F} = (\mathrm{A}_1 + \mathrm{A}_2)\mathrm{A}_3 + \mathrm{A}_4$. The probability of this event can then be computed as follows:

$$\begin{aligned}
\Pr[\mathrm{F}] &= \Pr[(\mathrm{A}_1 + \mathrm{A}_2)\mathrm{A}_3 + \mathrm{A}_4] \\
&= \Pr[\mathrm{A}_1\mathrm{A}_3 + \mathrm{A}_2\mathrm{A}_3] + \Pr[\mathrm{A}_4] - \Pr[\mathrm{A}_1\mathrm{A}_3\mathrm{A}_4 + \mathrm{A}_2\mathrm{A}_3\mathrm{A}_4] \\
&= \Pr[\mathrm{A}_4] + \Pr[\mathrm{A}_1\mathrm{A}_3] + \Pr[\mathrm{A}_2\mathrm{A}_3] - \Pr[\mathrm{A}_1\mathrm{A}_2\mathrm{A}_3] \\
&\qquad - \Pr[\mathrm{A}_1\mathrm{A}_3\mathrm{A}_4] - \Pr[\mathrm{A}_2\mathrm{A}_3\mathrm{A}_4] + \Pr[\mathrm{A}_1\mathrm{A}_2\mathrm{A}_3\mathrm{A}_4] \\
&= p + 2p^2 - 3p^3 + p^4
\end{aligned}$$

To find the desired value of p, we set this expression equal to 0.001; thus we need to find the roots of the polynomial $p^4 - 3p^3 + 2p^2 + p - 0.001$. Using MATLAB, this polynomial is found to have two complex conjugate roots, one real negative root, and one real positive root $p = 0.001$, which is the desired value.

□

An alternative method can be used to compute the probability of failure in this example. You list the possible outcomes (the sample space) and their probabilities as shown in the table below and put a check (√) next to each outcome that results in failure to communicate.

outcome	probability	F
$\mathrm{A}_1\mathrm{A}_2\mathrm{A}_3\mathrm{A}_4$	p^4	√
$\mathrm{A}_1\mathrm{A}_2\mathrm{A}_3\mathrm{A}_4{}^c$	$p^3(1-p)$	√
$\mathrm{A}_1\mathrm{A}_2\mathrm{A}_3{}^c\mathrm{A}_4$	$p^3(1-p)$	√
$\mathrm{A}_1\mathrm{A}_2\mathrm{A}_3{}^c\mathrm{A}_4{}^c$	$p^2(1-p)^2$	
⋮	⋮	
$\mathrm{A}_1{}^c\mathrm{A}_2{}^c\mathrm{A}_3{}^c\mathrm{A}_4{}^c$	$(1-p)^4$	

Then you simply add the probabilities of the outcomes that comprise the event F. The procedure is straightforward but slightly tedious because of the algebraic simplification required to get to the answer (see Problem 2.19).

2.4 Conditional Probability and Bayes' Rule

2.4.1 Conditional probability

If A_1 and A_2 are two events, then the probability of the event A_1 when it is known that the event A_2 has occured is defined by the relation

$$\Pr[A_1|A_2] = \frac{\Pr[A_1A_2]}{\Pr[A_2]} \tag{2.13}$$

$\Pr[A_1|A_2]$ is called the probability of "A_1 conditioned on A_2" or simply the probability of "A_1 *given* A_2." Note that in the special case that A_1 and A_2 are statistically independent, it follows from (2.13) and (2.12) that $\Pr[A_1|A_2] = \Pr[A_1]$. Thus when two events are independent, conditioning one upon the other has no effect.

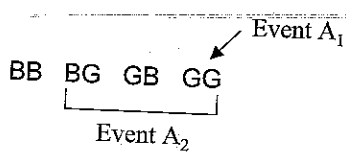
The use of conditional probability is illustrated in the following simple example.

Example 2.8: Remember Simon's Surplus and the diskettes? A diskette bought at Simon's is equally likely to be good or bad. Simon decides to sell them in packages of two and guarrantees that in each package, at least one will be good. What is the probability that when you buy a single package, you get two good diskettes?

Define the following events:

A_1 : Both diskettes are good.
A_2 : At least one diskette is good.

The sample space and these events are illustrated below:



The probability we are looking for is

$$\Pr[A_1|A_2] = \frac{\Pr[A_1A_2]}{\Pr[A_2]}$$

Recall that since all events in the sample space are equally likely, the probability of A_2 is 3/4. Also, since A_1 is included in A_2, it folows that $\Pr[A_1A_2] = \Pr[A_1]$, which is equal to 1/4. Therefore

$$\Pr[A_1|A_2] = \frac{1/4}{3/4} = \frac{1}{3}$$

□

It is meaningful to interpret conditional probability as the Venn diagram of Fig. 2.6. *Given* the event A_2, the only portion of A_1 that is of concern is the intersection that

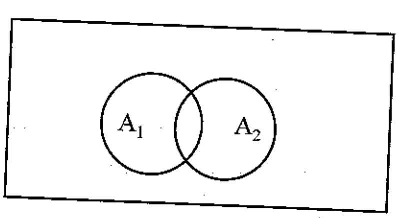


Figure 2.6 Venn diagram illustrating conditional probability.

A_1 has with A_2. It is as if A_2 becomes the new sample space. In defining conditional probability $\Pr[A_1|A_2]$, the probability of event A_1A_2 is therefore "renormalized" by dividing by the probability of A_2.

Equation 2.13 can be rewritten as

$$\Pr[A_1A_2] = \Pr[A_1|A_2]\Pr[A_2] = \Pr[A_2|A_1]\Pr[A_1] \tag{2.14}$$

where the second equality follows from the fact that $\Pr[A_1A_2] = \Pr[A_2A_1]$. Thus *the joint probability of two events can always be written as the product of a conditional probability and an unconditional probability.* Now let $\{A_i\}$ be a (finite or countably infinite) set of mutually exclusive collectively exhaustive events, and B be some other event of interest. Recall the *principle of total probability* introduced in Section 2.2 and expressed by (2.11). This equation can be rewritten using (2.14) as

$$\boxed{\Pr[B] = \sum_i \Pr[B|A_i]\Pr[A_i]} \tag{2.15}$$

Although both forms are equivalent, (2.15) is likely the more useful one to remember. This is because the information given in problems is more frequently in terms of conditional probabilities rather than joint probabilities. (And you must be able to recognize the difference!)

Let us consider one final fact about conditional probability before moving on. Again let $\{A_i\}$ be a (finite or countably infinite) set of mutually exclusive collectively exhaustive events. Then the probabilities of the A_i conditioned on *any* event B sum to one. That is,

$$\sum_i \Pr[A_i|B] = 1 \tag{2.16}$$

The proof of this fact is straightforward. Using the definition (2.13), we have

$$\sum_i \Pr[A_i|B] = \sum_i \frac{\Pr[A_iB]}{\Pr[B]} = \frac{\sum_i \Pr[A_iB]}{\Pr[B]} = \frac{\Pr[B]}{\Pr[B]} = 1$$

where in the next to last step we used the principle of total probability in the form (2.11).

As an illustration of this result, consider the following brief example.

Example 2.9: Consider the situation in Example 2.8. What is the probability that when you buy a package of two diskettes, only one is good?

Since Simon guarantees that there will be *at least* one good diskette in each package, we have the event A_2 defined in Example 2.8. Let A_3 represent the set of outcomes {BG GB} (only one good diskette). This event has probability 1/2. The probability of only one good diskette given the event A_2 is thus

$$\Pr[A_3|A_2] = \frac{\Pr[A_3A_2]}{\Pr[A_2]} = \frac{\Pr[A_3]}{\Pr[A_2]} = \frac{1/2}{3/4} = \frac{2}{3}$$

The events A_3 and A_1 are mutually exclusive and collectively exhaustive given the event A_2. Hence their probabilites, 2/3 and 1/3, sum to one.

□

2.4.2 Event trees

As stated above, often the information for problems in probability is stated in terms of *conditional* probabilities. An important technique for solving some of these problems is to draw the sample space by constructing a tree of dependent events and to use the information in the problem to determine the probabilities of compound events.

The idea is illustrated in Fig. 2.7. In this figure, A is assumed to be an event whose

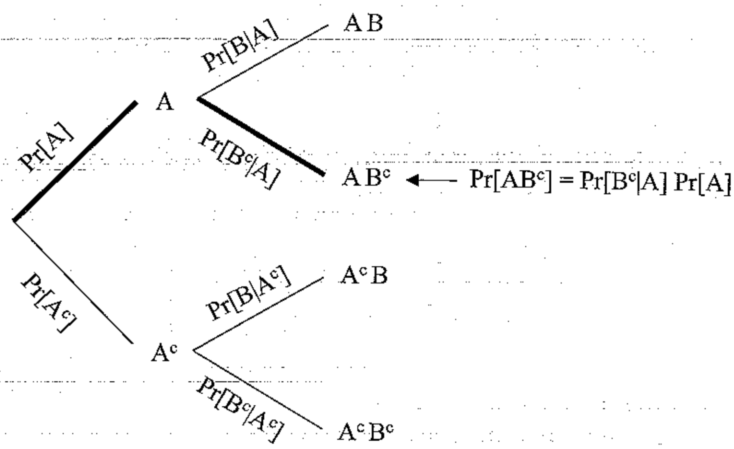


Figure 2.7 Sample space constructed using an event tree.

probability is known and does not depend on any other event. The probability of event B however depends on whether or not A occurred. These conditional probabilities are written on the branches of the tree. The endpoints of the tree comprise the sample space for the problem; they are a set of mutually exclusive collectively exhaustive events which represent the outcomes of the experiment. The probabilities of these events are computed using (2.14), which is equivalent to multiplying probabilities along the branches of the tree that form a path to the event (see figure). Once the probabilities of these elementary events are determined, you can find the probabilities for other compound events by adding the appropriate probabilities. The technique is best illustrated by an example.

Example 2.10: You listen to the morning weather report. If the weather person forecasts rain, then the probability of rain is 0.75. If the weather person forecasts "no rain," then the probability of rain is 0.15. You have listened to this report for well over a year now and have determined that the weather person forecasts rain 1 out of every 5 days regardless of the season. What is the probability that the weather report is wrong? Suppose you take an umbrella if and only if the weather report forecasts rain. What is the probability that it rains and you are caught without an umbrella?

To solve this problem, define the following events:

$$\mathrm{F} = \text{"rain is forecast"} \qquad \mathrm{R} = \text{"it actually rains"}$$

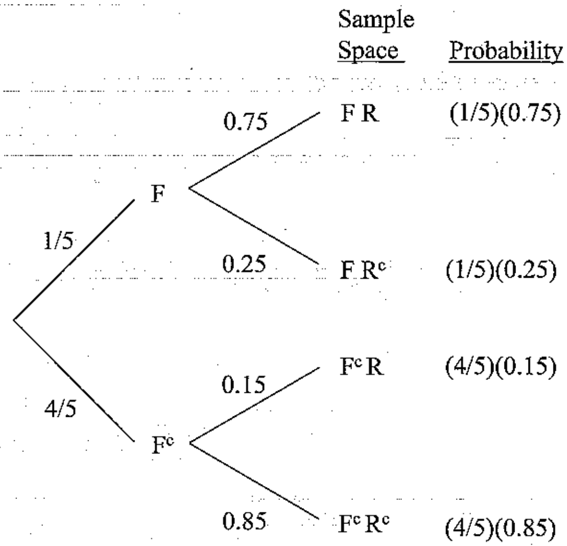
From the problem statement, we have the following information:

$$\Pr[\mathrm{R}|\mathrm{F}] = 0.75 \quad \Pr[\mathrm{R}|\mathrm{F}^c] = 0.15$$

$$\Pr[\mathrm{F}] = 1/5 \qquad \Pr[\mathrm{F}^c] = 4/5$$

The events and conditional probabilities are depicted in the event tree shown below. The event that the weather report is wrong is represented by the two elementary events FR^c and $\mathrm{F}^c\mathrm{R}$. Since these events are mutually exclusive, their probabilities can be added to find

$$\Pr[\text{wrong report}] = \tfrac{1}{5}(0.25) + \tfrac{4}{5}(0.15) = 0.17$$



The probability that it rains and you are caught without an umbrella is the event F^cR. The probability of this event is $\frac{4}{5}(0.15) = 0.12$.

□

2.4.3 Bayes' rule

One of the most important uses of conditional probability was developed by Bayes in the late 1800's. It follows if (2.14) is rewritten as

$$\Pr[A_1|A_2] = \frac{\Pr[A_2|A_1] \cdot \Pr[A_1]}{\Pr[A_2]} \tag{2.17}$$

This allows one conditional probability to be computed from the other. A particularly important case arises when $\{A_j\}$ is a (finite or countably infinite) set of mutually exclusive collectively exhaustive events and B is some other event of interest. From (2.17) the probability of one of these events conditioned on B is given by

$$\Pr[A_i|B] = \frac{\Pr[B|A_i] \cdot \Pr[A_i]}{\Pr[B]} \tag{2.18}$$

But since the $\{A_j\}$ are a set of mutually exclusive collectively exhaustive events, the principle of total probability can be used to express the probability of the event B. Thus, substituting (2.15) into the last equation yields

$$\boxed{\Pr[A_i|B] = \frac{\Pr[B|A_i] \cdot \Pr[A_i]}{\sum_j \Pr[B|A_j]\Pr[A_j]}} \tag{2.19}$$

This result is known as *Bayes' theorem* or *Bayes' rule*. It is used in a number of problems that commonly arise in communications or other areas where decisions or inferences are to be made from some observed signal or data. Because (2.19) is a more complicated formula, it is sufficient in problems of this type to remember the simpler result (2.18) and to know how to compute Pr[B] using the principle of total probability.

As an illustration of Bayes' rule, consider the following example.

Example 2.11: The US Navy is involved in a service-wide program to update memory in computers on-board ships. The Navy will buy memory modules only from American manufacturers known as A_1, A_2, and A_3. The probabilities of buying from A_1, A_2, and A_3 (based on availability and cost to the government) are given by 1/6, 1/3, and 1/2 repectively. The Navy doesn't realize, however, that the probability of failure for the modules from A_1, A_2, and A_3 is 0.006, 0.015, and 0.02 (respectively).

Back in the fleet, an enlisted technician upgrades the memory in a particular computer and finds that it fails. What is the probability that the failed module came from A_1? What is the probability that it came from A_3?

Let F represent the event that a memory module fails. Using (2.18) and (2.19) we can write

$$\Pr[A_1|F] = \frac{\Pr[F|A_1]\cdot\Pr[A_1]}{\Pr[F]} = \frac{\Pr[F|A_1]\cdot\Pr[A_1]}{\sum_{j=1}^{3}\Pr[F|A_j]\Pr[A_j]}$$

Then substituting the known probabilities yields

$$\Pr[A_1|F] = \frac{(0.006)1/6}{(0.006)1/6 + (0.015)1/3 + (0.02)1/2} = \frac{0.001}{0.016} = 0.0625$$

and likewise

$$\Pr[A_3|F] = \frac{(0.02)1/2}{(0.006)1/6 + (0.015)1/3 + (0.02)1/2} = \frac{0.01}{0.016} = 0.625$$

Thus in almost two-thirds of the cases the bad module comes from A_3.

□

2.5 More Applications

This section illustrates the use of probability as it occurs in two problems involving digital communication. A basic digital communication system is shown in Fig. 2.8. The system has three basic parts: a transmitter which codes the message into some

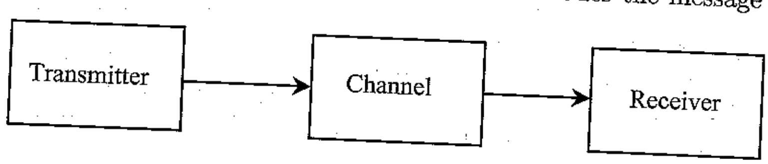


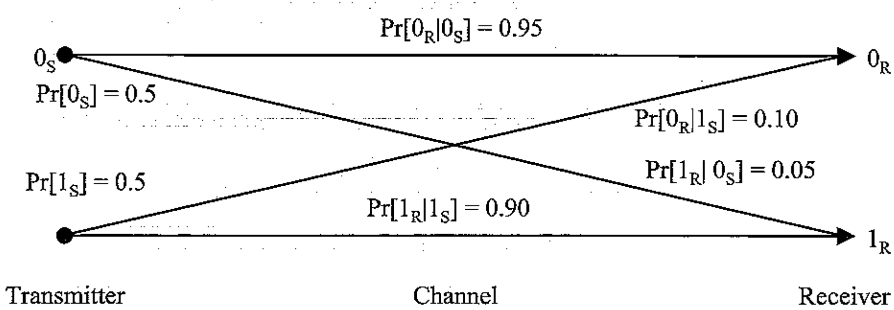
Figure 2.8 Digital communication system.

representation of binary data; a channel over which the binary data is transmitted; and a receiver, which decides whether a 0 or a 1 was sent and decodes the data. The simple binary communication channel discussed in Section 2.5.1 below is a probabilistic model for the binary communication system, which models the effect of sending one bit at a time. Section 2.5.2 discusses the digital communication system using the formal concept of "information," which is also a probabilistic idea.

2.5.1 The binary communication channel

A number of problems naturally involve the use of conditional probability and/or Bayes' rule. The binary communication channel, which is an abstraction for a communication system involving binary data, uses these concepts extensively. The idea is illustrated in the following example.

Example 2.12: The transmission of bits over a binary communication channel is represented in the drawing below, where we use notation like 0_S, 0_R ... to denote events "0



sent," "0 received," etc. When a 0 is transmitted, it is correctly received with probability 0.95 or incorrectly received with probability 0.05. That is $\Pr[0_R|0_S] = 0.95$ and $\Pr[1_R|0_S] = 0.05$. When a 1 is transmitted, it it is correctly received with probability 0.90 and incorrectly received with probability 0.10. The probabilities of sending a 0 or a 1 are denoted by $\Pr[0_S]$ and $\Pr[1_S]$ and are known as the *prior* probabilities. It is desired to compute the *probability of error* for the system.

This is an application of the principle of total probability. If two events A_1 and A_2 form a partition, then (2.15) can be written as

$$\Pr[B] = \Pr[B|A_1]\Pr[A_1] + \Pr[B|A_2]\Pr[A_2]$$

Since the two events 0_S and 1_S are mutually exclusive and collectively exhaustive, we can identify them with the events A_1 and A_2 and take the event B to be the event that an error occurs. It then follows that

$$\begin{aligned}\Pr[\text{error}] &= \Pr[\text{error}|0_S]\Pr[0_S] + \Pr[\text{error}|1_S]\Pr[1_S] \\ &= \Pr[1_R|0_S]\Pr[0_S] + \Pr[0_R|1_S]\Pr[1_S] \\ &= (0.05)(0.5) + (0.10)(0.5) = 0.075\end{aligned}$$

□

The probability of error is an overall measure of performance that is frequently used for a communication system. Notice that it involves not just the conditional error probabilities $\Pr[1_R|0_S]$ and $\Pr[0_R|1_S]$ but also the prior probabilities $\Pr[0_S]$ and $\Pr[1_S]$ for transmission of a 0 or a 1. One criterion for optimizing a communication system is to minimize the probability of error.

If a communication system is correctly designed, then the probability that a 1 was sent given that a 1 is received should be greater than the probability that a 0 was sent given that a 1 is received. In fact, as will be shown later in the text, this condition applied to both 0's and 1's leads to minimizing the probability of error. The computation of these "inverse" probabilities is illustrated in the next example.

Example 2.13: Assume for the communication system illustrated in the previous example that a 1 has been received. What is the probability that a 1 was sent? What is the probability that a 0 was sent?

This is an application of conditional probability and Bayes rule. For a 1, we have

$$\Pr[1_S|1_R] = \frac{\Pr[1_R|1_S]\Pr[1_S]}{\Pr[1_R]} = \frac{\Pr[1_R|1_S]\Pr[1_S]}{\Pr[1_R|1_S]\Pr[1_S] + \Pr[1_R|0_S]\Pr[0_S]}$$

Substituting the numerical values from the figure in Example 2.12 then yields

$$\Pr[1_S|1_R] = \frac{(0.9)(0.5)}{(0.9)(0.5) + (0.05)(0.5)} = 0.9474$$

For a 0, we have a similar analysis:

$$\begin{aligned}\Pr[0_S|1_R] &= \frac{\Pr[1_R|0_S]\Pr[0_S]}{\Pr[1_R|1_S]\Pr[1_S] + \Pr[1_R|0_S]\Pr[0_S]} \\ &= \frac{(0.05)(0.5)}{(0.9)(0.5) + (0.05)(0.5)} = 0.0526\end{aligned}$$

Note that $\Pr[1_S|1_R] > \Pr[0_S|1_R]$ as would be expected, and also that $\Pr[1_S|1_R] + \Pr[0_S|1_R] = 1$.

□

2.5.2 Measuring information and coding

The study of *Information Theory* is fundamental to understanding the trade-offs in design of efficient communication systems. The basic theory was developed by Shannon [2, 3, 4] and others in the late 1940's and '50's and provides fundamental results about what a given communication system can or cannot do. This section provides just a taste of the results which are based on a knowledge of basic probability.

Consider the digital communication system depicted in Fig. 2.8 and let the events A_1 and A_2 represent the transmission of two codes representing the symbols 0 and 1. To be specific, assume that the transmitter, or source, outputs the symbol to the communication channel with the following probabilities: $\Pr[A_1] = \frac{1}{8}$ and $\Pr[A_2] = \frac{7}{8}$. The *information* associated with the event A_i is defined as

$$I(A_i) = -\log \Pr[A_i]$$

The logarithm here is taken with respect to the base 2 and the resulting information is expressed in *bits*.[2] The information for each of the two symbols is thus

$$\begin{aligned}I(A_1) &= -\log \Pr[\tfrac{1}{8}] = 3 \text{ (bits)} \\ I(A_2) &= -\log \Pr[\tfrac{7}{8}] = 0.193 \text{ (bits)}\end{aligned}$$

Observe that events with *lower* probability have *higher* information. This corresponds to intuition. Someone telling you about an event that almost always happens provides little information. On the other hand, someone telling you about the occurrence of a very rare event provides you with much more information. The news media works on this principle in deciding what news to report and thus tries to maximize information.

The *average* information[3] H is given by the weighted sum

$$H = \sum_{i=1}^{2} \Pr[A_i] I(A_i) = \tfrac{1}{8} \cdot 3 + \tfrac{7}{8} \cdot 0.193 = 0.544$$

Notice that this average information is less than one bit, although it is not possible to transmit two symbols with less than one binary digit (bit).

Now consider the following scheme. Starting anywhere in the sequence, group together two consecutive bits and assign this pair to one of four possible codewords. Let

[2] Other less common choices for the base of the logarithm are 10, in which case the units of information are Hartleys, and e in which case the units are called nats.

[3] Average information is also known as the source *entropy* and is discussed further in Chapter 4.