CHAPTER ONE

events, sample space, and probability

Our study of probability theory begins with three closely related topics:

1 *The algebra of events* (*or sets or areas or spaces*) will provide us with a common language which is never ambiguous. This algebra will be defined by a set of seven axioms.

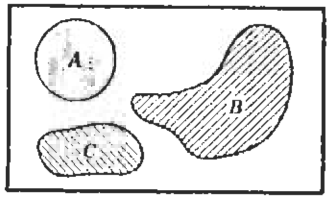
2 *Sample space* is the vital picture of a model of an "experiment" (any nondeterministic process) and all of its possible outcomes.

3 *Probability measure* is the relevant measure in sample space. This measure is established by three more axioms and it provides for the measurement of events in sample space.

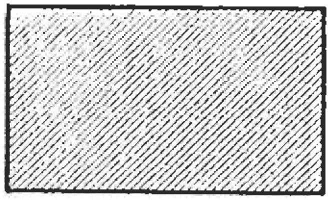
1-1 A Brief Introduction to the Algebra of Events

We first introduce an explicit language for our discussion of probability theory. Such a language is provided by the algebra of events. We wish to develop an understanding of those introductory aspects of this algebra which are essential to our study of probability theory and its applications.

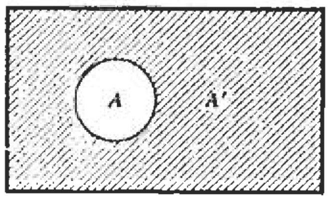
Let's begin with a brief tour of some of the definitions, concepts, and operations of the algebra of events.



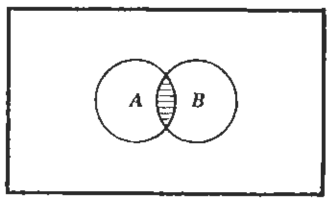
Events (or sets) are collections of points or areas in a space. The physical interpretation of this space will be provided in the following section.



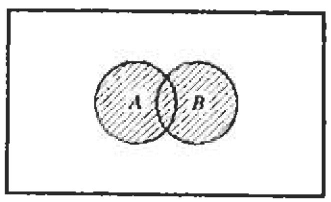
The collection of all points in the entire space is called U, the *universal set* or the *universal event*.



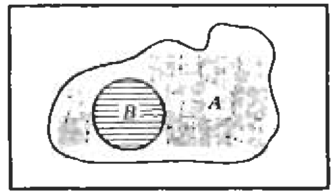
Event A', the *complement* of event A, is the collection of all points in the universal set which are not included in event A. The *null set* ϕ contains no points and is the complement of the universal set.



The *intersection* of two events A and B is the collection of all points which are contained both in A and in B. For the intersection of events A and B we shall use the simple notation AB.



The *union* of two events A and B is the collection of all points which are either in A or in B or in both. For the union of events A and B we shall use the notation $A + B$.



If all points of U which are in B are also in A, then event B is said to be *included* in event A.

Two events A and B are said to be *equal* if every point of U which is in A is also in B and every point of U which is in A' is also in B'. Another way to state the condition for the equality of two events would be to say that two events are equal if and only if each event is included in the other event.

We have sampled several of the notions of the algebra of events. Pictures of events in the universal set, such as those used above, are known as *Venn diagrams*. Formally, the following is a set of laws (*axioms*) which fully defines the algebra of events:

1 $A + B = B + A$	Commutative law	
2 $A + (B + C) = (A + B) + C$	Associative law	
3 $A(B + C) = AB + AC$	Distributive law	
4 $(A')' = A$		**6** $AA' = \phi$
5 $(AB)' = A' + B'$		**7** $AU = A$

The Seven Axioms of the Algebra of Events

Technically, these seven axioms define everything there is to know about the algebra of events. The reader may consider visualizing each of these axioms in a Venn diagram. Our selection of a list of axioms is not a unique one. Alternative sets of axioms could be stated which would lead to the same results.

Any relation which is valid in the algebra of events is subject to proof by use of the seven axioms and with no additional information. Some representative relations, each of which may be interpreted easily on a Venn diagram, are

$$A + A = A \qquad A + U = U$$
$$A + AB = A \qquad A + BC = (A + B)(A + C)$$
$$A + A'B = A + B \qquad A\phi = \phi$$
$$A + A' = U \qquad A(BC) = (AB)C$$

If it were our intention to prove (or test) relations in the algebra of events, we would find it surprisingly taxing to do so using only the seven axioms in their given form. One could not take an "obvious" step such as using $CD = DC$ without proving the validity of this relation from the axioms. For instance, to show that $CD = DC$, we may proceed

$C' + D' = D' + C'$	Axiom (1) with $A = C'$ and $B = D'$
$(C' + D')' = (D' + C')'$	Take complement of both sides
$\therefore CD = DC$	By use of axioms (5) and (4)

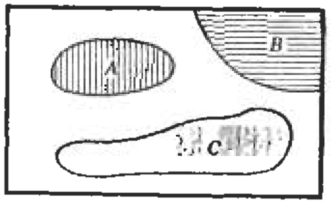
The task of proving or testing relations becomes easier if we have relations such as the above available for direct use. Once such relations are proved via the seven axioms, we call them *theorems* and use them with the same validity as the original axioms.

We are already prepared to note those definitions and properties of the algebra of events which will be of value to us in our study of probability.

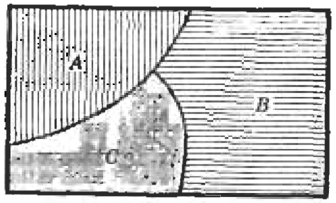
A list of events $A_1, A_2, \ldots, A_N$ is said to be composed of *mutually exclusive* events if and only if

$$A_iA_j = \begin{cases} A_i & \text{if } i = j \\ \phi & \text{if } i \neq j \end{cases} \qquad i = 1, 2, \ldots, N; \quad j = 1, 2, \ldots, N$$

A list of events is composed of mutually exclusive events if there is no point in the universal set which is included in more than one event in the list. A Venn diagram for three mutually exclusive events A, B, C could be



or

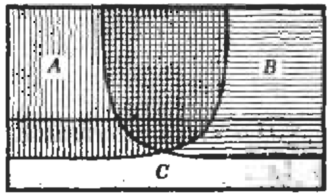


but in no case may there be any overlap of the events.

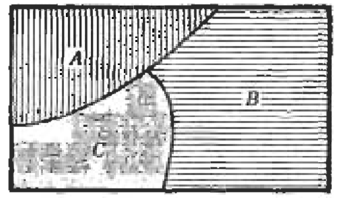
A list of events $A_1, A_2, \ldots, A_N$ is said to be *collectively exhaustive* if and only if

$$A_1 + A_2 + \cdots + A_N = U$$

A list of events is collectively exhaustive if each point in the universal set is included in at least one event in the list. A Venn diagram for three collectively exhaustive events A, B, C could be



or



but in no case may there be any point in the universal set which is included in none of the events. A list of events may be mutually exclusive, collectively exhaustive, both, or neither. After the discussion of sample space in the next section, we shall have many opportunities to consider lists of events and familiarize ourselves with the use of these definitions.

We note two additional matters with regard to the algebra of events. One is that we do not make errors in the algebra of events if we happen to include the same term several times in a union. If we are trying to collect all the points in a rather structured event, we need only be sure to include every appropriate point at least once—multiple inclusion will do no harm.

Another consideration is of particular importance in dealing with actual problems. The algebra of events offers a language advantage in describing complex events, *if* we are careful in defining all relevant simple events. Since we shall be making conscious and subconscious use of the algebra of events in all our work, one necessary warning should be sounded. "He who would live with the algebra of events had better know exactly with which events he is living." The original defined events should be simple and clear. If this is the case, it is then an easy matter to assemble expressions for complex events from the original definitions. For instance:

NEVER Event A: Neither Tom nor Mary goes without Harry unless they see Fred accompanied by Phil or it is raining and . . . .

BETTER Event A: Tom goes. Event B: Mary goes.
Event C: Harry goes. Etc.

1-2 Sample Spaces for Models of Experiments

In this book, we use the word "experiment" to refer to any process which is, to some particular observer, nondeterministic. It makes no difference whether the observer's uncertainty is due to the nature of the process, the state of knowledge of the observer, or both.

The first six chapters of this book are concerned with the analysis of abstractions, or *models* of actual physical experiments. Our last chapter is concerned with the relation of the model to the actual experiment.

It is probably safe to state that there are more variables associated with the outcome of any physical experiment than anybody could ever care about. For instance, to describe the outcome of an actual coin toss, we could be concerned with the height of the toss, the number of bounces, and the heating due to impact, as well as the more usual consideration of which face is up after the coin settles. For most purposes, however, a reasonable model for this experiment would involve a simple nondeterministic choice between a head and a tail.

Many of the "trick" probability problems which plague students are based on some ambiguity in the problem statement or on an inexact formulation of the model of a physical situation. The precise statement of an appropriate *sample space*, resulting from a detailed description of the model of an experiment, will do much to resolve common difficulties. In this text, we shall literally live in sample space.

Sample space: The finest-grain, mutually exclusive, collectively exhaustive listing of all possible outcomes of a model of an experiment

The "finest-grain" property requires that all possible distinguishable outcomes allowed by the model be listed separately. If our model for the flip of a coin is simply a nondeterministic selection between possible outcomes of a head and a tail, the sample space for this model of the experiment would include only two items, one corresponding to each possible outcome.

To avoid unnecessarily cumbersome statements, we shall often use the word experiment in place of the phrase "model of an experiment." Until the final chapter, the reader is reminded that all our work is concerned with abstractions of actual physical situations. When we wish to refer to the real world, we shall speak of the "physical experiment."

Soon we shall consider several experiments and their sample spaces. But first we take note of two matters which account for our interest in these spaces. First, the universal set with which we deal in the study of probability theory will always be the sample space for an experiment. The second matter is that *any* event described in terms

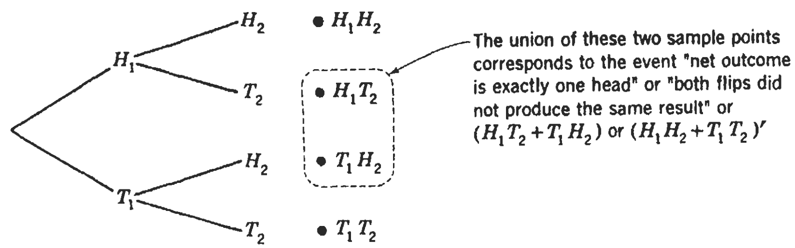
of the outcome of a performance of some experiment can be identified as a collection of events in sample space. In sample space one may collect the members of any event by taking a union of mutually exclusive points or areas from a collectively exhaustive space. The advantage of being able to collect any event as a union of mutually exclusive members will become clear as we learn the properties of probability measure in the next section.

A sample space may look like almost anything, from a simple listing to a multidimensional display of all possible distinguishable outcomes of an experiment. However, two types of sample spaces seem to be the most useful.

One common type of sample space is obtained from a sequential picture of an experiment in terms of its most convenient parameters. Normally, this type of sample space is not influenced by our particular interests in the experimental outcome. Consider the sequential sample space for a reasonable model of two flips of a coin. We use the notation

$$\text{Event} \begin{Bmatrix} H_n \\ T_n \end{Bmatrix} : \begin{Bmatrix} \text{Heads} \\ \text{Tails} \end{Bmatrix} \text{ on the } n\text{th toss of the coin}$$

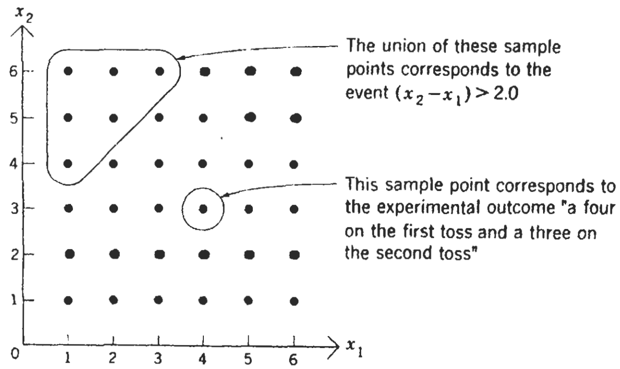
This leads to the sequential sample space



Above we picture the experiment proceeding rightward from the left origin. Each sample point, located at the end of a *terminal* tree branch, represents the event corresponding to the intersection of all events encountered in tracing a path from the left origin to that sample point. On the diagram, we have noted one example of how an event may be collected as a union of points in this sample space.

Formally, the four sample points and their labels constitute the sample space for the experiment. However, when one speaks of a *sequential* sample space, he normally pictures the entire generating tree as well as the resulting sample space.

For an experiment whose outcomes may be expressed numerically, another useful type of sample space is a coordinate system on which is displayed a finest-grain mutually exclusive collectively exhaustive set of points corresponding to every possible outcome. For instance, if we throw a six-sided die (with faces labeled 1, 2, 3, 4, 5, 6) twice and call the value of the up face on the nth toss x_n, we have a sample space with 36 points, one for each possible experimental outcome,



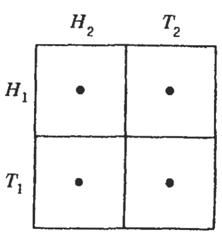
Often we shall have reason to abbreviate a sample space by working with a mutually exclusive collectively exhaustive listing of all possible outcomes which is not finest-grain. Such spaces, possessing all attributes of sample spaces other than that they may not list separately all possible distinguishable outcomes, are known as *event spaces*. If a list of events $A_1, A_2, \ldots, A_N$ forms an event space, each possible finest-grain experimental outcome is included in exactly one event in this list. However, more than one distinguishable outcome may be included in any event in the list.

We now present some sample and event spaces for several experiments.

1 Experiment: Flip a coin twice.

Notation: Let $\begin{Bmatrix} H_n \\ T_n \end{Bmatrix}$ be the event $\begin{Bmatrix} \text{heads} \\ \text{tails} \end{Bmatrix}$ on the nth toss.

We already have seen the sequential sample space for this experiment, but let's consider some event spaces and some other displays of the sample space. The sample space in a different form is



The sample space displayed as a simple listing is

$\bullet H_1H_2 \quad \bullet H_1T_2 \quad \bullet T_1H_2 \quad \bullet T_1T_2$

An example of an event space but not a sample space for this experiment is

$\bullet H_1 \quad \bullet T_1$

Another example of an event space but not a sample space is

$\bullet H_1T_2 \quad \bullet (H_1T_2)'$

The following is neither an event space nor a sample space:

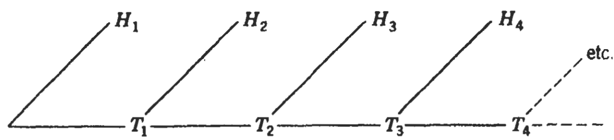
$\bullet H_1 \quad \bullet T_1 \quad \bullet H_2 \quad \bullet T_2$

2 Experiment: Flip a coin until we get our first head.

Notation: Let $\begin{Bmatrix} H_n \\ T_n \end{Bmatrix}$ be the event $\begin{Bmatrix} \text{heads} \\ \text{tails} \end{Bmatrix}$ on the nth toss.

A sample space generated by a sequential picture of the experiment is

$\bullet H_1 \quad \bullet T_1H_2 \quad \bullet T_1T_2H_3 \quad \bullet T_1T_2T_3H_4$



A sample space which is a simple listing for this experiment, with N representing the number of the toss on which the first head occurred, is

$\bullet N = 1 \quad \bullet N = 2 \quad \bullet N = 3 \quad \bullet N = 4 \quad \cdots$

Note that this is still a finest-grain description and that one can specify the exact sequence of flips which corresponds to each of these sample points.

An event space but not a sample space for this experiment is

$\bullet N > 4 \quad \bullet N \le 4$

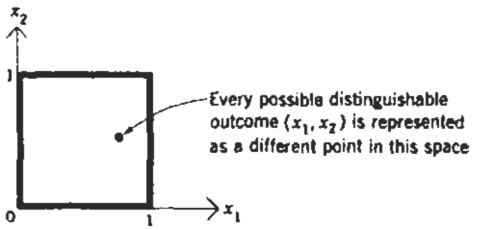
Neither a sample space nor an event space for this experiment is

$\bullet N \ge 1 \quad \bullet N \ge 2 \quad \bullet N \ge 3 \quad \bullet N \ge 4 \quad \cdots$

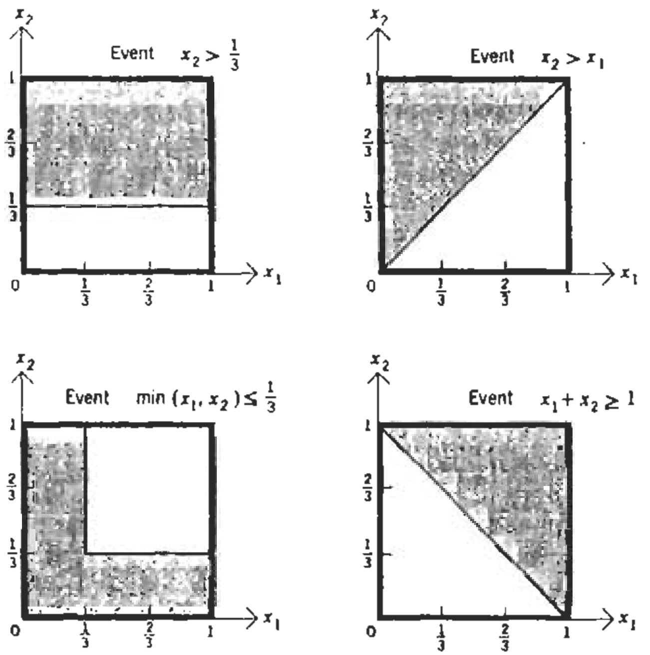
3 Experiment: Spin a wheel of fortune twice. The wheel is continuously calibrated, not necessarily uniformly, from zero to unity.

Notation: Let x_n be the exact reading of the wheel on the nth spin.

A sample space for this experiment is



Let's collect several typical events in this sample space.



Any point in a sample space corresponds to one possible outcome of a performance of the experiment. For brevity, we shall often refer to the sample point as though it were the event to which it corresponds.

1-3 Probability Measure and the Relative Likelihood of Events

To complete our specification of the model for a physical experiment, we wish to assign *probabilities* to the events in the sample space of the experiment. The probability of an event is to be a number, representing the chance, or "relative likelihood," that a performance of the experiment will result in the occurrence of that event.

This measure of events in sample space is known as *probability measure*. By combining three new axioms with the seven axioms of the algebra of events, one obtains a consistent system for associating real nonnegative numbers (probabilities) witb events in sample space and for computing the probabilities of more complex events. If we use the notation $P(A)$ for the probability measure associated with event A, then the three additional axioms required to establish probability measure are:

1 For any event A, $P(A) \geq 0$
2 $P(U) = 1$
3 If $AB = \phi$, then $P(A + B) = P(A) + P(B)$

The Three Axioms of Probability Measure

The first axiom states that all probabilities are to be nonnegative. By attributing a probability of unity to the universal event, the second axiom provides a normalization for probability measure. The third axiom states a property that most people would hold to be appropriate to a reasonable measure of the relative likelihood of events. If we accept these three axioms, all of conventional probability theory follows from them.

The original assignment of probability measure to a sample space for an experiment is a matter of how one chooses to model the physical experiment. Our axioms do not tell us how to compute the probability of any experimental outcome "from scratch." Before one can begin operating under the rules of probability theory, he must first provide an assignment of probability measure to the events in the sample space.

Given a physical experiment to be modeled and analyzed, there is no reason why we would expect two different people to agree as to what constitutes a reasonable probability assignment in the sample space. One would expect a person's modeling of a physical experiment to be influenced by his previous experiences in similar situations and by any other information which might be available to him. Probability theory will operate "correctly" on *any* assignment of probability to the sample space of an experiment which is consistent with the three axioms. Whether any such assignment and the results obtained from it are of any physical significance is another matter.

If we consider the definition of sample space and the third axiom of probability measure, we are led to the important conclusion,

Given an assignment of probability measure to the finest-grain events in a sample space, the probability of any event A may be computed by summing the probabilities of all finest-grain events included in event A.

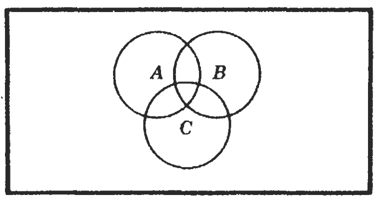
One virtue of working in sample space is that, for any point in the sample space there must be a "yes" or "no" answer as to whether the point is included in event A. Were there some event in our space for which this was not the case, then either the space would not be a sample space (because it did not list separately certain distinguishable outcomes) or our model is inadequate for resolution between events A and A'. We shall return to this matter when we discuss the sample-space interpretation of conditional probability.

One can use the seven axioms of the algebra of events and the three axioms of probability measure to prove various relations such as

$$P(A') = 1 - P(A) \qquad P(A + B) = P(A) + P(B) - P(AB)$$
$$P(\phi) = 0 \qquad P(A + B + C) = 1 - P(A'B'C')$$

Because we shall live in sample space, few, if any, such relations will be required formally for our work. We shall always be able to employ directly the third axiom of probability measure to compute the probability of complex events, since we shall be expressing such events as unions of mutually exclusive members in a sample space. In computing the probability of any event in sample space, that axiom states that we must include the probability of every sample point in that event exactly *once*. Multiple counting caused no error in taking a union in the algebra of events to describe another event, but one must carefully confine himself to the axioms of probability measure when determining the probability of a complex event.

There are many ways to write out a "formula" for the probability of an event which is a union such as $P(A + B + C)$. Using the third axiom of probability measure and looking at a Venn diagram,



we may, for instance, write any of the following:

$$P(A + B + C) = P(A) + P(A'B) + P(A'B'C)$$
$$P(A + B + C) = P(A) + P(B) + P(C) - P(AB) - P(AC) - P(BC) + P(ABC)$$
$$P(A + B + C) = P(AB'C') + P(A'BC') + P(A'B'C) + P(AB) + P(AB'C) + P(A'BC)$$
$$P(A + B + C) = 1 - P(A'B'C')$$

Obviously, one could continue listing such relations for a very long time.

Now that we have considered both the algebra of events and probability measure, it is important to recall that:

1 Events are combined and operated upon only in accordance with the seven axioms of the algebra of events.
2 The probabilities of events are numbers and can be computed only in accordance with the three axioms of probability measure.
3 Arithmetic is something else.

$$\underbrace{P[A + CD + B(A + C'D)]}$$

The algebra of events applies inside the brackets. It operates upon events and has nothing to do with their measure.

The axioms of probability theory are used in obtaining the numerical value of this quantity.

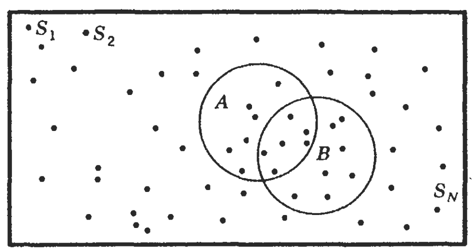
In practice, we shall usually evaluate such probabilities by collecting the event as a union of mutually exclusive points in sample space and summing the probabilities of all points included in the union.

1-4 Conditional Probability and Its Interpretation in Sample Space

Assume that we have a fully defined experiment, its sample space, and an initial assignment of probability to each finest-grain event in the sample space. Let two events A and B be defined on the sample space of this experiment, with $P(B) \neq 0$.

We wish to consider the situation which results if the experiment is performed once and we are told only that the experimental outcome has attribute B. Thus, if B contains more than one sample point, we are considering the effect of "partial information" about the experimental outcome.

Let's look at a sample-space picture of this situation. Consider a sample space made up of sample points $S_1, S_2, \ldots, S_N$.



Given that event B has occurred, we know that the sample point representing the experimental outcome must be in B and cannot be in B'. We have no information which would lead us to alter the *relative* probabilities of the sample points in B. Since we know one of the sample points in B must represent the experimental outcome, we scale up their original probabilities by a constant, $1/P(B)$, such that they now add to unity, to obtain *conditional probabilities* which reflect the influence of our partial information.

We formalize these ideas by defining $P(S_j \mid B)$, the "conditional probability of S_j given B," to be

$$P(S_j \mid B) = \begin{Bmatrix} \dfrac{P(S_j)}{P(B)} & \text{if } S_j \text{ in } B \\ 0 & \text{if } S_j \text{ in } B' \end{Bmatrix} = \frac{P(S_jB)}{P(B)}$$

The conditional probability of any other event, such as A, is to be the sum of the conditional probabilities of the sample points included in A, leading to the common definition of the conditional probability of event A given B.

$$P(A \mid B) = \frac{P(AB)}{P(B)} \qquad \text{defined only for } P(B) \neq 0$$

which may be obtained from our previous statements via

$$P(A \mid B) = \sum_{\text{all } j \text{ in } A} P(S_j \mid B) = \sum_{\text{all } j \text{ in } A} \frac{P(S_jB)}{P(B)} = \frac{P(AB)}{P(B)}$$

We may conclude that one way to interpret conditional probability is to realize that a conditioning event (some partial information about the experimental outcome) allows one to move his analysis from the original sample space into a new conditional space. Only those finest-grain events from the original sample space which are included in the conditioning event appear in the new conditional space with a nonzero assignment of conditional probability measure. The original ("a priori") probability assigned to each of these finest-grain events is multiplied by the same constant such that the sum of the conditional ("a posteriori") probabilities in the conditional space is unity. *In the resulting conditional sample space, one uses and interprets these a posteriori probabilities exactly the same way he uses the a priori probabilities in the original sample space.* The conditional probabilities obtained by the use of some partial information will, in fact, serve as initial probabilities in the new sample space for any further work.

We present one simple example. A fair coin is flipped twice, and Joe, who saw the experimental outcome, reports that "at least one toss

resulted in a head." Given this partial information, we wish to determine the conditional probability that both tosses resulted in heads. Using the notation $\begin{Bmatrix} H_n \\ T_n \end{Bmatrix}$ and the problem statement, we may draw a sample space for the experiment and indicate, in the $P(\bullet)$ column, our a priori probability for each sample point.

	$P(\bullet)$	A	B	AB
$\bullet H_1H_2$	0.25	✓	✓	✓
$\bullet H_1T_2$	0.25	✓		
$\bullet T_1H_2$	0.25	✓		
$\bullet T_1T_2$	0.25			
		$P(A) = 0.75$	$P(B) = 0.25$	$P(AB) = 0.25$

For the above display, we defined the events A and B to be "at least one head" and "two heads," respectively. In the A, B, and AB columns we check those sample points included in each of these events. The probability of each of these events is simply the sum of the probabilities of the sample points included in the event. (The only use of the B column in this example was to make it easier to identify the sample points associated with the complex event AB.)

The desired answer $P(B \mid A)$ may be found either directly from the definition of conditional probability

$$P(B \mid A) = \frac{P(AB)}{P(A)} = \frac{0.25}{0.75} = \frac{1}{3}$$

or by noting that, in the conditional sample space given our event A, there happen to be three equally likely sample points, only one of which has attribute B.

The solution to this problem could have been obtained in far less space. We have taken this opportunity to introduce and discuss a type of sample-space display which will be valuable in the investigation of more complex problems.

It is essential that the reader realize that certain operations (such as collecting events and conditioning the space) which are always simple in a sample space may not be directly applicable in an arbitrary event space. Because of the *finest-grain* property of a sample space, any sample point must be either wholly excluded from or wholly included in any arbitrary event defined within our model. However, in an event space, an event point A might be *partially* included in B, some other event of interest. Were this the case, the lack of detail in the event space would make it impossible to collect event B in this event space or to condition the event space by event B.

For instance, suppose that for the coin example above we were given only the event space

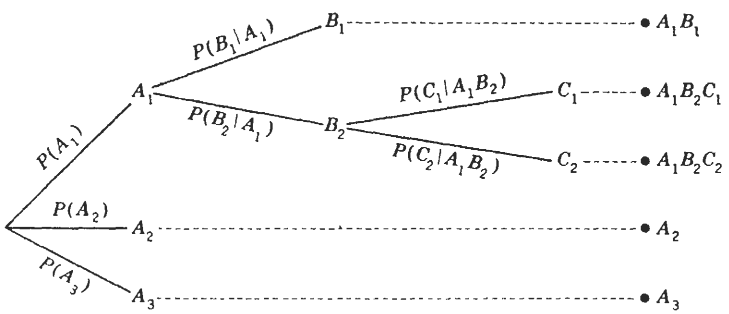
$\bullet H_1 \qquad \bullet T_1H_2 \qquad \bullet T_1T_2$

its probability assignment, and no other details of the experiment. Because this event space lacks adequate detail, it would be impossible to calculate the probability of event H_2 or to condition this space by event H_2.

When we are given a *sample* space for an experiment and the probability assigned to each sample point, we can answer all questions with regard to any event defined on the possible experimental outcomes.

1-5 Probability Trees for Sequential Experiments

Sequential sample and event spaces were introduced in Sec. 1-2. Such spaces, with all branches labeled to indicate the probability structure of an experiment, are often referred to as *probability trees*. Consider the following example:



This would be a sample space for an experiment in which, for instance, the "C_1 or C_2" trial occurs only if A_1B_2 has resulted from the earlier stages of the sequential experiment. As before, one sample point appears for each terminal branch of the tree, representing the intersection of all events encountered in tracing a path from the left origin to a terminal node.

Each branch is labeled such that the product of all branch probabilities from the left origin to any node equals the probability that the event represented by that node will be the outcome on a particular performance of the experiment. Only the first set of branches leaving the origin is labeled with a priori probabilities; all other branches must be labeled with the appropriate conditional probabilities. The sum of the probabilities on the branches leaving any nonterminal node must sum to unity; otherwise the terminal nodes could not represent a collectively exhaustive listing of all possible outcomes of the experiment.

Of course, any tree sample space which contains the complete set of finest-grain events for an experiment is a "correct" sample space. In any physical situation, however, the model of the experiment will usually specify the sequential order of the tree if we wish to label all branches with the appropriate conditional probabilities without any calculations.

Sometimes, a complete picture of the sample space would be too large to be useful. But it might still be of value to use an "outline" of the actual sample space. These outlines may be "trimmed" probability trees for which we terminate uninteresting branches as soon as possible. But once we have substituted such an event space for the sample space, we must again realize that we may be unable to perform certain calculations in this event space.

1-6 The Independence of Events

Thus far, our structure for probability theory includes seven axioms for the algebra of events, three more for probability measure, and the definition and physical interpretation of the concept of conditional probability. We shall now formalize an intuitive notion of the *independence of events*. This definition and its later extensions will be of considerable utility in our work.

In an intuitive sense, if events A and B are defined on the sample space of a particular experiment, we might think them to be "independent" if knowledge as to whether or not the experimental outcome had attribute B would not affect our measure of the likelihood that the experimental outcome also had attribute A. We take a formal statement of this intuitive concept to be our definition of the independence of two events.

Two events A and B are defined to be independent if and only if

$P(A \mid B) = P(A)$

From the definition of conditional probability, as long as $P(A) \neq 0$ and $P(B) \neq 0$, we may write

$$P(AB) = P(A)P(B \mid A) = P(B)P(A \mid B)$$

When we substitute the condition for the independence of A and B into this equation, we learn both that $P(A \mid B) = P(A)$ requires that

$P(B \mid A) = P(B)$ and that an alternative statement of the condition for the (mutual) independence of two events is $P(AB) = P(A)P(B)$.

If A and B are other than trivial events, it will rarely be obvious whether or not they are independent. To test two events for independence, we collect the appropriate probabilities from a sample space to see whether or not the definition of independence is satisfied. Clearly, the result depends on the original assignment of probability measure to the sample space in the modeling of the physical experiment.

We have *defined* conditional probability such that (as long as none of the conditioning events is of probability zero) the following relations always hold:

$$P(AB) = P(A \mid B)P(B) = P(B \mid A)P(A)$$

$$P(ABC) = P(A)P(BC \mid A) = P(B)P(C \mid B)P(A \mid BC) = P(AC)P(B \mid AC) = \cdots$$

but only when two events are independent may we write

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

We extend our notion of independence by defining the mutual independence of N events $A_1, A_2, \ldots, A_N$.

N events $A_1, A_2, \ldots, A_N$ are defined to be mutually independent if and only if

$$P(A_i \mid A_jA_k \cdots A_p) = P(A_i) \qquad \text{for all } i \neq j, k, \ldots p; \quad 1 \leq i, j, k, \ldots p \leq N$$

This is equivalent to requiring that the probabilities of *all* possible intersections of these (different) events taken any number at a time [such as $P(A_1A_3A_4A_9)$] be given by the products of the individual event probabilities [such as $P(A_1)P(A_3)P(A_4)P(A_9)$]. Pairwise independence of the events on a list, as defined at the beginning of this section, does not necessarily result in the mutual independence defined above.

One should note that there is no reason why the independence or dependence of events need be preserved in going from an a priori sample space to a conditional sample space. Similarly, events which are mutually independent in a particular conditional space may or may not be mutually independent in the original universal set or in another conditional sample space. Two events A and B are said to be *conditionally independent*, given C, if it is true that

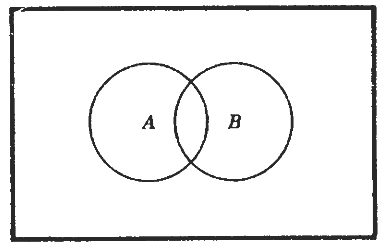
$$P(AB \mid C) = P(A \mid C)P(B \mid C)$$

This relation does not require, and is not required by, the separate condition for the unconditional independence of events A and B,

$$P(AB) = P(A)P(B)$$

We close this section with a consideration of the definition of the independence of two events from a sample-space point of view. One statement of the condition for the independence of events A and B, with $P(B) \neq 0$, is

$$P(A \mid B) = P(A) \qquad \text{or, equivalently,} \qquad P(A) = \frac{P(AB)}{P(B)}$$



Thus, the independence requirement is that event AB is assigned a fraction of the probability measure of event B which is numerically equal to $P(A)$. As one would expect, we see (from the above diagram and the last equation) that, as long as $P(A) \neq 0$ and $P(B) \neq 0$, events A and B cannot be independent if they are mutually exclusive.

1-7 Examples

example 1 Our first example is simply an exercise to review the properties of probability measure, conditional probability, and some definitions from the algebra of events. The reader is encouraged to work out these examples for himself before reading through our discussion of the solutions.

Suppose that we are given three lists of events, called lists 1, 2, and 3. All the events in the three lists are defined on the same experiment, and none of the events is of probability zero.

List 1 contains events $A_1, A_2, \ldots, A_k$ and is mutually exclusive and collectively exhaustive.

List 2 contains events $B_1, B_2, \ldots, B_l$ and is mutually exclusive and collectively exhaustive.

List 3 contains events $C_1, C_2, \ldots, C_m$ and is mutually exclusive but *not* collectively exhaustive.

Evaluate each of the following quantities numerically. If you cannot evaluate them numerically, specify the *tightest upper and lower* numerical bounds you can find for each quantity.

(a) $\sum_{i=1}^{m} P(C_i)$

(b) $\sum_{j=1}^{k} P(A_2 \mid A_j)$

(c) $\sum_{i=1}^{k} \sum_{j=1}^{k} P(A_iA_j)$

(d) $\sum_{j=1}^{k} P(C_2 \mid A_jC_3)$

(e) $\sum_{j=1}^{k} P(A_j')$

(f) $\sum_{i=1}^{k} \sum_{j=1}^{l} P(A_i)P(B_j \mid A_i)$

(g) $\sum_{j=1}^{k} P(B_1 \mid A_j)$

Part (a) requires that we sum the probability measure associated with each member of a mutually exclusive but not collectively exhaustive list of events defined on a particular experiment. Since the events are mutually exclusive, we are calculating the probability of their union. According to the conditions of the problem, this union represents an event of nonzero probability which is not the universal set; so we obtain

$$0.0 < \sum_{i=1}^{m} P(C_i) < 1.0$$

For part (b), the A_i's are mutually exclusive so we note that $P(A_2 \mid A_j)$ is zero unless $j = 2$. When $j = 2$, we have $P(A_2 \mid A_2)$, which is equal to unity. Therefore, we may conclude that

$$\sum_{j=1}^{k} P(A_2 \mid A_j) = 1.0$$

Again in part (c), the mutually exclusive property of the A_i's will require $P(A_iA_j) = 0$ unless $j = i$. If $j = i$, we have $P(A_jA_j)$, which is equal to $P(A_j)$; so upon recalling that the A_i list is also collectively exhaustive, there follows

$$\sum_{i=1}^{k} \sum_{j=1}^{k} P(A_iA_j) = \sum_{i=1}^{k} \sum_{\substack{j=1 \\ i \neq j}}^{k} P(A_iA_j) + \sum_{i=1}^{k} P(A_i) = 0 + 1 = 1.0$$

In part (d), we know that C_2 and C_3 are mutually exclusive. Therefore, C_2 and C_3 can never describe the outcome of the same performance of the experiment on which they are defined. So, with no attention to the properties of the A_j's (assuming that we can neglect any pathological cases where the conditioning event A_jC_3 would be of probability zero and the conditional probability would be undefined), we have

$$\sum_{j=1}^{k} P(C_2 \mid A_jC_3) = 0.0$$

Part (e) is most easily done by direct substitution,

$$\sum_{j=1}^{k} P(A_j') = \sum_{j=1}^{k} [1 - P(A_j)] = k - \sum_{j=1}^{k} P(A_j)$$

and since we are told that the A_j's are mutually exclusive and collectively exhaustive, we have

$$\sum_{j=1}^{k} P(A_j') = k - 1.0$$

For part (f), we can use the definition of conditional probability and the given properties of lists 1 and 2 to write

$$\sum_{i=1}^{k} \sum_{j=1}^{l} P(A_i)P(B_j \mid A_i) = \sum_{i=1}^{k} \sum_{j=1}^{l} P(A_iB_j) = \sum_{i=1}^{k} P(A_i) = 1.0$$

The quantity to be evaluated in part (*g*) involves the summation of conditional probabilities each of which applies to a different conditioning event. The value of this sum need not be a probability. By decomposing the universal set into $A_1, A_2, \ldots, A_k$ and considering a few special cases (such as $B_1 = U$) consistent with the problem statement, we see

$$0 < \sum_{j=1}^{k} P(B_1 \mid A_j) \leq k$$

example 2 To the best of our knowledge, with probability 0.8 Al is guilty of the crime for which he is about to be tried. Bo and Ci, each of whom knows whether or not Al is guilty, have been called to testify.

Bo is a friend of Al's and will tell the truth if Al is innocent but will lie with probability 0.2 if Al is guilty. Ci hates everybody but the judge and will tell the truth if Al is guilty but will lie with probability 0.3 if Al is innocent.

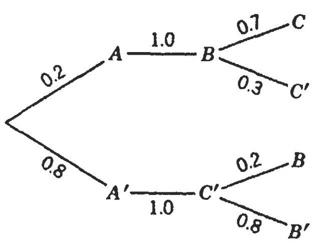
Given this model of the physical situation:

(a) Determine the probability that the witnesses give conflicting testimony.
(b) Which witness is more likely to commit perjury?
(c) What is the conditional probability that Al is innocent, given that Bo and Ci gave conflicting testimony?
(d) Are the events "Bo tells a lie" and "Ci tells a lie" independent? Are these events conditionally independent to an observer who knows whether or not Al is guilty?

We begin by establishing our notation for the events of interest:

Event A: Al is innocent.
Event B: Bo testifies that Al is innocent.
Event C: Ci testifies that Al is innocent.
Event X: The witnesses give conflicting testimony.
Event Y: Bo commits perjury.
Event Z: Ci commits perjury.

Now we'll draw a sample space in the form of a probability tree and collect the sample points corresponding to the events of interest.



Sample points	$P(\bullet)$	X	Y	Z	AX
• ABC	0.14				
• ABC'	0.06	✓		✓	✓
• $A'BC'$	0.16	✓	✓		
• $A'B'C'$	0.64				
		$P(X) = 0.22$	$P(Y) = 0.16$	$P(Z) = 0.06$	$P(AX) = 0.06$

To find the probability of any event in sample space, we simply sum the probabilities of all sample points included in that event. (It is because of the *mutually exclusive* property of the sample space that we may follow this procedure.) Now, to answer our questions,

a $P(X) = P(BC' + B'C) = P(BC') + P(B'C) = 0.22$

b $P(Y) = P(AB' + A'B) = P(AB') + P(A'B) = 0.00 + 0.16 = 0.16$

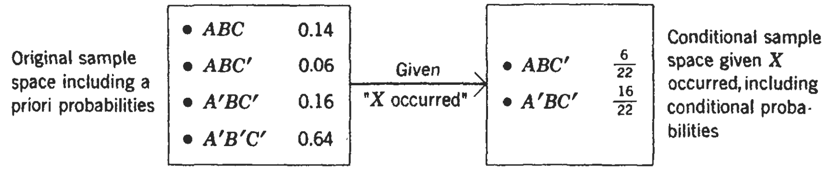
$P(Z) = P(AC' + A'C) = P(AC') + P(A'C) = 0.06 + 0.00 = 0.06$

Therefore Bo is the more likely of the witnesses to commit perjury.

c $P(A \mid X) = \dfrac{P(AX)}{P(X)} = \dfrac{0.06}{0.06 + 0.16} = \dfrac{3}{11}$

Conflicting testimony is more likely to occur if Al is innocent than if he is guilty; so given X occurred, it should increase our probability that Al is innocent. This is the case, since $3/11 > 1/5$.

c (One other method of solution.) Given X occurred, we go to a conditional space containing only those sample points with attribute X. The conditional probabilities for these points are found by scaling up the original a priori probabilities by the same constant $[1/P(X)]$ so that they add to unity.



Now, in the conditional space we simply sum the conditional probabilities of all sample points included in any event to determine the conditional probability of that event.

$P(A \mid X) = P(ABC' \mid X) = 3/11$

d To determine whether "Bo tells a lie" and "Ci tells a lie" are independent in the original sample space, we need only test $P(YZ) \stackrel{?}{=} P(Y)P(Z)$. Since $P(YZ) = 0$ but $P(Y) > 0$ and $P(Z) > 0$, we see that events Y and Z are not independent, in fact they are mutually exclusive.

To determine whether Y and Z are conditionally independent given A or A', we must test $P(YZ \mid A) \stackrel{?}{=} P(Y \mid A)P(Z \mid A)$ and $P(YZ \mid A') \stackrel{?}{=} P(Y \mid A')P(Z \mid A')$. From the sample space, we find that the left-hand side and one term on the right-hand side of each of these tests is equal to zero, so events Y and Z are conditionally independent to one who knows whether or not Al is innocent.

Since the testimony of the two witnesses depends only on whether or not Al is innocent, this is a reasonable result. If we don't know whether or not Al is innocent, then Y and Z are dependent because the occurrence of one of these events would give us partial information about Al's innocence or guilt, which would, in turn, change our probability of the occurrence of the other event.

example 3 When we find it necessary to resort to games of chance for simple examples, we shall often use four-sided (tetrahedral) dice to keep our problems short. With a tetrahedral die, one reads the "down" face either by noticing which face *isn't* up or by looking up at the bottom of the die through a glass cocktail table. Suppose that a fair four-sided die (with faces labeled 1, 2, 3, 4) is tossed twice and we are told only that the product of the resulting two down-face values is less than 7.

(a) What is the probability that at least one of the face values is a two?

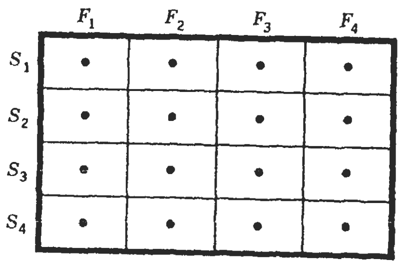
(b) What is the probability that the sum of the face values is less than 7?

(c) If we are told that both the product and the sum of the face values is less than 7 and that at least one of the face values is a two, determine the probability of each possible value of the other face.

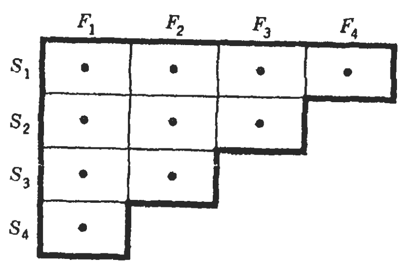
We use the notation

Event F_i: Value of down face on first throw is equal to i
Event S_j: Value of down face on second throw is equal to j

to construct the sample space



in which, from the statement of the problem ("fair die"), all 16 sample points are equally likely. Given that the product of the down faces is less than 7, we produce the appropriate conditional space by eliminating those sample points not included in the conditioning event and scaling up the probabilities of the remaining sample points. We obtain

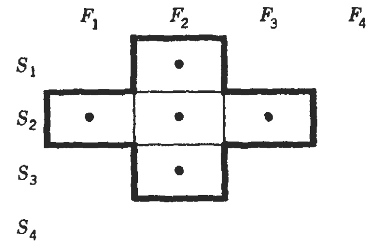


in which each of the remaining 10 sample points is still an equally likely outcome in the conditional space.

a There are five sample points in the conditional space which are included in the event "At least one down face is a two"; so the conditional probability of this event is $5 \times 0.1 = 0.5$.

b All the outcomes in the conditional space represent experimental outcomes for which the sum of the down-face values is less than 7; so the conditional probability of this event is unity.

c Given all the conditioning events in the problem statement, the resulting conditional sample space is simply



and the conditional probability that the other face value is unity is 2/5; the same applies for the event that the other face value is 3, and there is a conditional probability of 1/5 that the other face value is also a two.

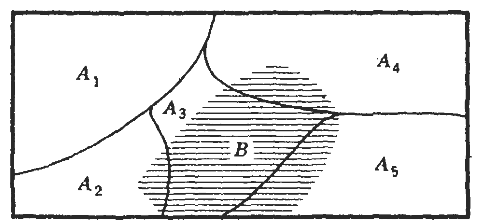
For Example 3 we have considered one approach to a simple problem which may be solved by several equivalent methods. As we become familiar with random variables in the next chapter, we shall employ extensions of the above methods to develop effective techniques for the analysis of experiments whose possible outcomes may be expressed numerically.

1-8 Bayes' Theorem

The relation known as *Bayes' theorem* results from a particular application of the definition of conditional probability. As long as the conditioning events are not of probability zero, we have

$$P(AB) = P(A)P(B \mid A) = P(B)P(A \mid B)$$

We wish to apply this relation to the case where the events $A_1, A_2, \ldots, A_N$ form a mutually exclusive and collectively exhaustive list of events. For this case where the events $A_1, A_2, \ldots, A_N$ form an event space, let's consider the universal set and include some other event B.



For the case of interest, assume that we know $P(A_i)$ and $P(B \mid A_i)$ for all $1 \leq i \leq N$ and we wish to determine the $P(A_i \mid B)$'s.

An example of this situation follows: Let A_i represent the event that a particular apple was grown on farm i. Let B be the event that an apple turns blue during the first month of storage. Then the quantity $P(B \mid A_i)$ is the probability that an apple will turn blue during the first month of storage given that it was grown on farm i. The question of interest: Given an apple which did turn blue during its first month of storage, what is the probability it came from farm i?

We return to the problem of determining the $P(A_i \mid B)$'s. As long as $P(B) \neq 0$ and $P(A_i) \neq 0$ for $i = 1, 2, \ldots, N$, we may substitute A_i for A in the above definition of conditional probability to write

$$P(A_i \mid B) = \frac{P(A_i)P(B \mid A_i)}{P(B)}$$

$$P(B) = P(UB) = P[(A_1 + A_2 + \cdots + A_N)B]$$

$$= \sum_{i=1}^{N} P(A_iB) = \sum_{i=1}^{N} P(A_i)P(B \mid A_i)$$

and we substitute the above expression for $P(B)$ into the equation for $P(A_i \mid B)$ to get

Bayes' theorem:

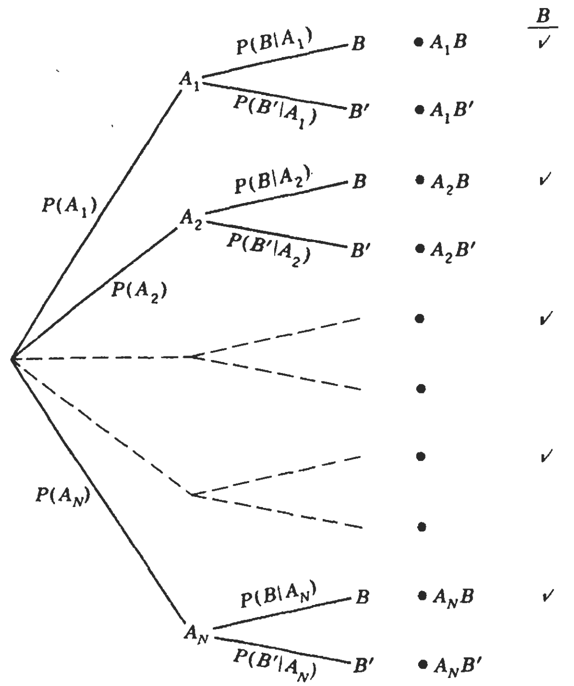
$$P(A_i \mid B) = \frac{P(A_i)P(B \mid A_i)}{\sum_{i=1}^{N} P(A_i)P(B \mid A_i)}$$

with $P(B) \neq 0$ and $A_1, A_2, \ldots, A_N$ an event space

Bayes' theorem could also have been obtained directly from our sample space interpretation of conditional probability. Note that, using a sequential sample space (shown at the top of the following page) this interpretation would allow us to derive Bayes' theorem [the expression for $P(A_i \mid B)$] by inspection.

A knowledge of Bayes' theorem by name will not affect our approach to any problems. However, the theorem is the basis of much of the study of statistical inference, and we shall discuss its application for that purpose in the last chapter of this book.

In the literature, one finds appreciable controversy about the "validity" of certain uses of Bayes' theorem. We have learned that the theorem is a simple consequence of the definition of conditional probability. Its application, of course, must always be made with conscious knowledge of just what model of reality the calculations represent. The trouble is generated, not by Bayes' theorem, but by



the attempt to assign a priori probabilities in the sample space for a model when one has very little information to assist in this assignment. We'll return to this matter in Chap. 7.

One last note regarding conditional probability is in order. No matter how close to unity the quantity $P(A \mid B)$ may be, it would be foolish to conclude from this evidence alone that event B is a "cause" of event A. *Association and physical causality may be different phenomena.* If the probability that any man who consults a surgeon about lung cancer will pass away soon is nearly unity, few of us would conclude "Aha! Lung surgeons are the cause of these deaths!" Sometimes, equally foolish conclusions are made in the name of "statistical reasoning." We wish to keep an open mind and develop a critical attitude toward physical conclusions based on the mathematical analyses of probabilistic models of reality.

1-9 Enumeration in Event Space: Permutations and Combinations

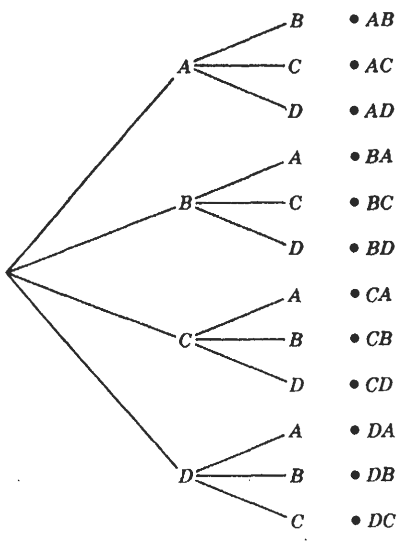
Sample spaces and event spaces play a key role in our presentation of introductory probability theory. The value of such spaces is that they enable one to display events in a mutually exclusive, collectively exhaustive form. Most problems of a combinatorial nature, whether probabilistic or not, also require a careful listing of events. It is the purpose of this section to demonstrate how combinatorial problems may be formulated effectively in event spaces.

Given a set of N distinguishable items, one might wonder how many distinguishable orderings (or arrangements) may be obtained by using these N items K at a time. For instance, if the items are the four events A, B, C, D, the different arrangements possible, subject to the requirement that each arrangement must include exactly two of the items, are

AB	BA	CA	DA
AC	BC	CB	DB
AD	BD	CD	DC

Such arrangements of N distinguishable items, taken K at a time, are known as *K-permutations* of the items. In the above example, we have found by enumeration that there are exactly 12 different 2-permutations of 4 distinguishable items.

To determine the number of K-permutations which may be formed from N distinguishable items, one may consider forming the permutations sequentially. One begins by choosing a first item, then a second item, etc. For instance, for the above example, this process may be shown on a sequential sample space,



There were N choices for the first member, any such choice results in $N - 1$ possible choices for the second member, and this continues until there are $N - (K + 1)$ possible choices for the Kth (and last) member of the permutation. Thus we have

Number of K-permutations of N distinguishable items

$$= N(N-1)(N-2)\cdots(N-K+1) = \frac{N!}{(N-K)!} \qquad N \geq K$$

A selection of K out of N distinguishable items, without regard to ordering, is known as a *K-combination.* Two K-combinations are identical if they both select exactly the same set of K items out of the original list and no attention is paid to the order of the items. For the example considered earlier, AB and BA are different permutations, but both are included in the same combination.

By setting $N = K$ in the above formula, we note that any combination containing K distinguishable items includes $K!$ permutations of the members of the combination. To determine the number of K-combinations which may be formed from N distinguishable items, we need only divide the number of K-permutations by $K!$ to obtain

Number of K-combinations of N distinguishable items

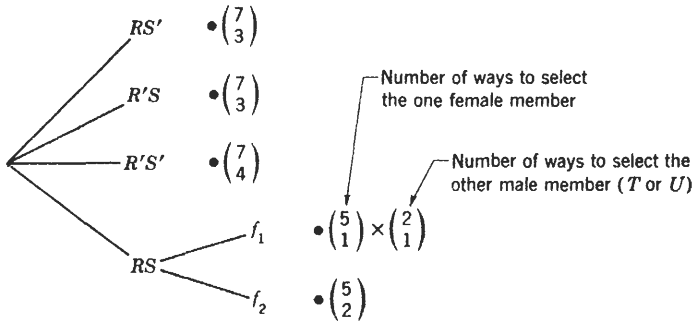
$$= \frac{N!}{K!(N-K)!} \equiv \binom{N}{K} \qquad N \geq K$$

Many enumeration and probability problems require the orderly collection of complex combinatorial events. When one attacks such problems in an appropriate sequential event space, the problem is reduced to several simple counting problems, each of which requires nothing more than minor (but careful) bookkeeping. The event-space approach provides the opportunity to deal only with the collection of mutually exclusive events in a collectively exhaustive space. This is a powerful technique and, to conclude this section, we note one simple example of its application.

Suppose that an unranked committee of four members is to be formed from a group of four males R, S, T, U and five females V, W, X, Y, Z. It is also specified that R and S cannot be on the same committee unless the committee contains at least one female. We first wish to determine the number of different such committees which may be formed.

We draw an event space for the enumeration of these committees. Our only methodology is to decompose the counting problem into smaller pieces, always branching out into a *mutually exclusive* decomposition which includes all members of interest. (We may omit items which are not relevant to our interests.)

NOTATION Event X: X is on the committee.
Event f_n: Exactly n females are on the committee.



The quantity following each terminal node is the number of acceptable committees associated with the indicated event. For instance, for the RS' committees, note that there are as many acceptable committees containing R but not S as there are ways of selecting three additional members from T, U, V, W, X, Y, Z without regard to the order of selection. We have obtained, for the number of acceptable committees,

$$2\binom{7}{3} + \binom{7}{4} + \binom{5}{1}\binom{2}{1} + \binom{5}{2} = 125 \text{ possible acceptable committees}$$

Finally we determine, were the committees selected by lot, the probability that the first four names drawn would form an acceptable committee. Since any committee is as likely to be drawn as any other, we note that there are 125 acceptable committees out of $\binom{9}{4} = 126$ possible results of the draw. Thus there is a probability of 125/126 that a randomly drawn committee will meet the given constraint. Had we decided to solve by counting unacceptable committees and subtracting from 126, we would have noticed immediately that only one draw, $RSTU$, would be unacceptable.

The extension of our event-space approach to problems involving permutations, such as ranked committees, merely requires that we work in a more fine-grained event space.

PROBLEMS

1.01 Use the axioms of the algebra of events to prove the relations
a $U' = \phi$ **b** $A + B = A + A'B + ABC$ **c** $A + U = U$

1.02 Use Venn diagrams, the axioms of the algebra of events, or anything else to determine which of the following are valid relations in the algebra of events for arbitrary events A, B, and C:
a $(A + B + C)' = A' + B' + C'$
b $A + B + C = A + A'B + (A + A'B)'C$
c $(A + B)(A' + B') = AB' + A'B + A'BC'$
d $AB + AB' + A'B = (A'B')'$

1.03 Fully explain your answers to the following questions:
a If events A and B are mutually exclusive and collectively exhaustive, are A' and B' mutually exclusive?
b If events A and B are mutually exclusive but not collectively exhaustive, are A' and B' collectively exhaustive?
c If events A and B are collectively exhaustive but not mutually exclusive, are A' and B' collectively exhaustive?

1.04 We desire to investigate the validity of each of four proposed relations (called relations 1, 2, 3, and 4) in the algebra of events. Discuss what evidence about the validity of the appropriate relation may be obtained from each of the following observations. We observe that we can:
a Obtain a valid relation by taking the intersection of each side of relation 1 with some event E_1
b Obtain an invalid relation by taking the intersection of each side of relation 2 with some event E_2
c Obtain a valid relation by taking the union of each side of relation 3 with some event E_3
d Obtain an invalid relation by taking the complement of each side of relation 4

1.05 In a group of exactly 2,000 people, it has been reported that there are exactly:
612 people who smoke
670 people over 25 years of age
960 people who imbibe
86 smokers who imbibe
290 imbibers who are over 25 years of age

158 smokers who are over 25 years of age
44 people over 25 years of age each of whom both smokes and imbibes
250 people under 25 years of age who neither smoke nor imbibe

Determine whether or not this report is consistent.

1.06 Consider the experiment in which a four-sided die (with faces labeled 1, 2, 3, 4) is thrown twice. We use the notation

$$\text{Event } \begin{Bmatrix} F_n \\ S_n \end{Bmatrix}\text{: Down face value on } \begin{Bmatrix} \text{first} \\ \text{second} \end{Bmatrix} \text{ throw equals } n.$$

For each of the following lists of events, determine whether or not the list is (i) mutually exclusive, (ii) collectively exhaustive, (iii) a sample space, (iv) an event space:

a $F_1, F_2, (F_1 + F_2)'$
b $F_1S_1, F_1S_2, F_2, F_3, F_4$
c $F_1S_1, F_1S_2, (F_1)', (S_1 + S_2)'$
d $(F_1 + S_1), (F_2 + S_2), (F_3 + S_3), (F_4 + S_4)$
e $(F_1 + F_2)(S_1 + S_2 + S_3), (F_1 + F_2 + F_3)S_4,$
$(F_3 + F_4)(S_1 + S_2 + S_3), S_4F_4$

1.07 For three tosses of a fair coin, determine the probability of:

a The sequence HHH
b The sequence HTH
c A total result of two heads and one tail
d The outcome "More heads than tails"

Determine also the conditional probabilities for:

e "More heads than tails" given "At least one tail"
f "More heads than tails" given "Less than two tails"

1.08 Joe is a fool with probability of 0.6, a thief with probability 0.7, and neither with probability 0.25.

a Determine the probability that he is a fool or a thief but not both.
b Determine the conditional probability that he is a thief, given that he is not a fool.

1.09 Given $P(A) \neq 0, P(B) \neq 0, P(A + B) = P(A) + P(B) - 0.1$, and $P(A \mid B) = 0.2$. Either determine the exact values of each of the following quantities (if possible), or determine the tightest numerical bounds on the value of each:

a $P(ABC) + P(ABC')$ **b** $P(A' \mid B)$ **c** $P(B)$ **d** $P(A')$
e $P(A + B + A'B')$

1.10 If $P(A) = 0.4$, $P(B') = 0.7$, and $P(A + B) = 0.7$, determine:

a $P(B)$ **b** $P(AB)$ **c** $P(A' \mid B')$

1.11 A game begins by choosing between dice A and B in some manner such that the probability that A is selected is p. The die thus selected is then tossed until a white face appears, at which time the game is concluded.

Die A has 4 red and 2 white faces. Die B has 2 red and 4 white faces. After playing this game a great many times, it is observed that the probability that a game is concluded in exactly 3 tosses of the selected die is $\frac{7}{81}$. Determine the value of p.

1.12 Oscar has lost his dog in either forest A (with a priori probability 0.4) or in forest B (with a priori probability 0.6). If the dog is alive and not found by the Nth day of the search, it will die that evening with probability $N/(N + 2)$.

If the dog is in A (either dead or alive) and Oscar spends a day searching for it in A, the conditional probability that he will find the dog that day is 0.25. Similarly, if the dog is in B and Oscar spends a day looking for it there, he will find the dog that day with probability 0.15.

The dog cannot go from one forest to the other. Oscar can search only in the daytime, and he can travel from one forest to the other only at night.

All parts of this problem are to be worked separately.

a In which forest should Oscar look to maximize the probability he finds his dog on the first day of the search?
b Given that Oscar looked in A on the first day but didn't find his dog, what is the probability that the dog is in A?
c If Oscar flips a fair coin to determine where to look on the first day and finds the dog on the first day, what is the probability that he looked in A?
d Oscar has decided to look in A for the first two days. What is the a priori probability that he will find a live dog for the first time on the second day?
e Oscar has decided to look in A for the first two days. Given the fact that he was unsuccessful on the first day, determine the probability that he does not find a dead dog on the second day.
f Oscar finally found his dog on the fourth day of the search. He looked in A for the first 3 days and in B on the fourth day. What is the probability he found his dog alive?
g Oscar finally found his dog late on the fourth day of the search. The only other thing we know is that he looked in A for 2 days and and in B for 2 days. What is the probability he found his dog alive?

1.13 Considering the statement of Prob. 1.12, suppose that Oscar has decided to search each day wherever he is most likely to find the dog on that day. He quits as soon as he finds the dog.

a What is the probability that he will find the dog in forest A?
b What is the probability that he never gets to look in forest B?
c Given that he does get to look in forest B, what is the probability that he finds the dog during his first day of searching there?
d Given that he gets to look in forest B at least once, what is the probability that he eventually finds the dog in forest A?

1.14 Joe is an astronaut for project Pluto. Mission success or failure depends only on the behavior of three major systems. Joe decides the following assumptions are valid and apply to the performance of an entire mission: (1) The mission is a failure only if two or more of the major systems fail. (2) System I, the Gronk system, will fail with probability 0.1. (3) If at least one other system fails, no matter how this comes about, System II, the Frab system, will fail with conditional probability 0.5. If no other system fails, the Frab system will fail with probability 0.1. (4) System III, the Beer Cooler, fails with probability 0.5 if the Gronk system fails. Otherwise, the Beer Cooler cannot fail.
a What is the probability that the mission succeeds but that the Beer Cooler fails?
b What is the probability that all three systems fail?
c Given that more than one system failed, determine the conditional probabilities that:
i The Gronk did not fail.
ii The Beer Cooler failed.
iii Both the Gronk and the Frab failed.
d About the time when Joe was due back on Earth, you overhear a radio broadcast about Joe in a very noisy room. You are not positive what the announcer did say, but, based on all available information, you decide that it is twice as likely that he reported "Mission a success" as that he reported "Mission a failure." What now is the conditional probability (to you) that the Gronk failed?

1.15 A box contains two fair coins and one biased coin. For the biased coin, the probability that any flip will result in a head is $\frac{1}{3}$. Al draws two coins from the box, flips each of them once, observes an outcome of one head and one tail, and returns the coins to the box. Bo then draws one coin from the box and flips it. The result is a tail. Determine the probability that neither Al nor Bo removed the biased coin from the box.

1.16 Joe, the bookie, is attempting to establish the odds on an exhibition baseball game. From many years of experience, he has learned that his prime consideration should be the selection of the starting pitcher.

The Cardinals have only two pitchers, one of whom must start: C_1, their best pitcher; C_2, a worse pitcher. The other team, the Yankees, has only three pitchers, one of whom must start: Y_1, their best pitcher; Y_2, a worse one; Y_3, an even worse one.

By carefully weighting information from various sources, Joe has decided to make the following assumptions:
The Yankees will not start Y_1 if C_1 does not start. If C_1 does start, the Yankees will start Y_1 with probability $\frac{1}{3}$.
The Cardinals are equally likely to start C_1 or C_2, no matter what the Yankees do.
Y_2 will pitch for the Yankees with probability $\frac{3}{4}$ if Y_1 does not pitch, no matter what else occurs.
The probability the Cardinals will win given C_1 pitches is $m/(m+1)$, where m is the number (subscript) of the Yankee pitcher.
The probability the Yankees will win given C_2 pitches is $1/m$, where m is the number of the Yankee pitcher.
a What is the probability that C_2 starts?
b What is the probability that the Cardinals will win?
c Given that Y_2 does not start, what is the probability that the Cardinals will win?
d If C_2 and Y_2 start, what is the probability that the Yankees win?

1.17 Die A has five olive faces and one lavender face; die B has three faces of each of these colors. A fair coin is flipped once. If it falls heads, the game continues by throwing die A alone; if it falls tails, die B alone is used to continue the game. However awful their face colors may be, it is known that both dice are fair.
a Determine the probability that the nth throw of the die results in olive.
b Determine the probability that both the nth and $(n + 1)$st throw of the die results in olive.
c If olive readings result from all the first n throws, determine the conditional probability of an olive outcome on the $(n + 1)$st toss. Interpret your result for large values of n.

1.18 Consider events A, B, and C with $P(A) > P(B) > P(C) > 0$. Events A and B are mutually exclusive and collectively exhaustive. Events A and C are independent. Can C and B be mutually exclusive?

1.19 We are given that the following six relations hold:

$$P(A) \neq 0 \qquad P(B) \neq 0 \qquad P(C) \neq 0$$
$$P(AB) = P(A)P(B) \qquad P(AC) = P(A)P(C) \qquad P(BC) = P(B)P(C)$$

Subject to these six conditions, determine, for each of the following entries, whether it must be true, it might be true, or it cannot be true:

a $P(ABC) = P(A)P(B)P(C)$ **b** $P(B \mid A) = P(B \mid C)$
c $P(AB \mid C) = P(A \mid C)P(B \mid C)$
d $P(A + B + C) < P(A) + P(B) + P(C)$
e If $W = AB$ and $V = AC$, $WV = \phi$

1.20 Events E, F, and G form a list of mutually exclusive and collectively exhaustive events with $P(E) \neq 0$, $P(F) \neq 0$, and $P(G) \neq 0$. Determine, for each of the following statements, whether it must be true, it might be true, or it cannot be true:

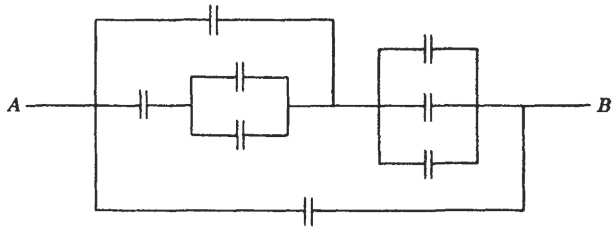
a E', F', and G' are mutually exclusive.
b E', F', and G' are collectively exhaustive.
c E' and F' are independent.
d $P(E') + P(F') > 1.0$.
e $P(E' + EF' + EFG') = 1.0$.

1.21 Bo and Ci are the only two people who will enter the Rover Dog Food jingle contest. Only one entry is allowed per contestant, and the judge (Rover) will declare the one winner as soon as he receives a suitably inane entry, which may be never.

Bo writes inane jingles rapidly but poorly. He has probability 0.7 of submitting his entry first. If Ci has not already won the contest, Bo's entry will be declared the winner with probability 0.3. Ci writes slowly, but he has a gift for this sort of thing. If Bo has not already won the contest by the time of Ci's entry, Ci will be declared the winner with probability 0.6.

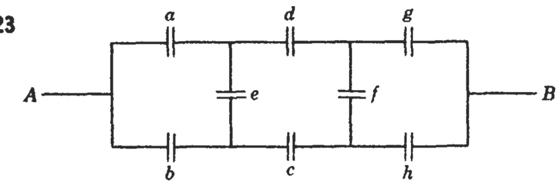
a What is the probability that the prize never will be awarded?
b What is the probability that Bo will win?
c Given that Bo wins, what is the probability that Ci's entry arrived first?
d What is the probability that the first entry wins the contest?
e Suppose that the probability that Bo's entry arrived first were P instead of 0.7. Can you find a value of P for which "First entry wins" and "Second entry does not win" are independent events?

1.22



Each ⊣⊢ represents one communication link. Link failures are independent, and each link has a probability of 0.5 of being out of service. Towns A and B can communicate as long as they are connected in the communication network by at least one path which contains only in-service links. Determine, in an efficient manner, the probability that A and B can communicate.

1.23



In the above communication network, link failures are independent, and each link has a probability of failure of p. Consider the physical situation before you write anything. A can communicate with B as long as they are connected by at least one path which contains only in-service links.

a Given that exactly five links have failed, determine the probability that A can still communicate with B.
b Given that exactly five links have failed, determine the probability that either g or h (*but not both*) is still operating properly.
c Given that a, d, and h have failed (but no information about the condition of the other links), determine the probability that A can communicate with B.

1.24 The events X, Y, and Z are defined on the same experiment. For each of the conditions below, state whether or not each statement following it must necessarily be true if the condition is true, and give a reason for your answer. (Remember that a simple way to demonstrate that a statement is not necessarily true may be to find a counterexample.)

a If $P(X + Y) = P(X) + P(Y)$, then:
 i $P(XY) = P(X)P(Y)$ **iii** $P(Y) = 0$
 ii $X + Y = X + X'Y$ **iv** $P(XY) = 0$
b If $P(XY) = P(X)P(Y)$ and $P(X) \neq 0$, $P(Y) \neq 0$, then:
 i $P(X \mid Y) = P(Y \mid X)$ **iii** $P(XY) \neq 0$
 ii $P(X + Y) = P(X) + P(Y) - P(XY)$ **iv** $X \neq Y$ unless $X = U$
c If $P(XY \mid Z) = 1$, then:
 i $P(XY) = 1$ **iii** $P(XYZ) = P(Z)$
 ii $Z = U$ **iv** $P(X' \mid Z) = 0$
d If $P[(X + Y) \mid Z] = 1$ and $X \neq \phi$, $Y \neq \phi$, $Z \neq \phi$, then:
 i $P(XY \mid Z) = 0$ **iii** $P(X'Y' \mid Z) = 0$
 ii $P(X + Y) < 1$ **iv** $P(X \mid Z) < 1$
e Define the events $R = X$, $S = X'Z$, $T = X'Y'Z'$. If the events R,

S, and T are collectively exhaustive, then:

i R, S, and T are mutually exclusive

ii $P(RST) = P(S)P(RT \mid S)$

iii $P(X'YZ') = 0$

iv $P[(R + S + T) \mid XYZ] = 1$

1.25 Only at midnight may a mouse go from either of two rooms to the other. Each midnight he stays in the same room with probability 0.9, or he goes to the other room with probability 0.1. We can't observe the mouse directly, so we use the reports of an unreliable observer, who each day reports correctly with probability 0.7 and incorrectly with probability 0.3. On day 0, we installed the mouse in room 1. We use the notation:

Event $R_k(n)$: Observer reports mouse in room k on day n.

Event $S_k(n)$: Mouse is in room k on day n.

a What is the a priori probability that on day 2 we shall receive the report $R_2(2)$?

b If we receive the reports $R_1(1)$ and $R_2(2)$, what is the conditional probability that the observer has told the truth on both days?

c If we receive the reports $R_1(1)$ and $R_2(2)$, what is the conditional probability of the event $S_2(2)$? Would we then believe the observer on the second day? Explain.

d Determine the conditional probability $P[R_2(2) \mid R_1(1)]$, and compare it with the conditional probability $P[S_2(2) \mid S_1(1)]$.

1.26 **a** The Pogo Thorhead rocket will function properly only if all five of its major systems operate simultaneously. The systems are labeled A, B, C, D, and E. System failures are independent, and each system has a probability of failure of 1/3. Given that the Thorhead fails, determine the probability that system A is *solely* at fault.

b The Pogo Thorhead II is an improved configuration using the same five systems, each still with a probability of failure of 1/3. The Thorhead II will fail if system A fails or if (at least) any two systems fail. Alas, the Thorhead II also fails. Determine the probability that system A is *solely* at fault.

1.27 The individual shoes from eight pairs, each pair a different style, have been put into a barrel. For $1, a customer may randomly draw and keep two shoes from the barrel. Two successive customers do this. What is the probability that at least one customer obtains two shoes from the same pair? How much would an individual customer improve his chances of getting at least one complete pair by investing $2 to be allowed to draw and keep four shoes? (Some combinatorial problems are quite trying. Consider, if you dare, the solution to this problem for eight customers instead of two.)

1.28 Companies A, B, C, D, and E each send three delegates to a conference. A committee of four delegates, selected by lot, is formed. Determine the probability that:

a Company A is not represented on the committee.

b Company A has exactly one representative on the committee.

c Neither company A nor company E is represented on the committee.

1.29 A box contains N items, K of which are defective. A sample of M items is drawn from the box at random. What is the probability that the sample includes at least one defective item if the sample is taken:

a With replacement

b Without replacement

1.30 The Jones family household includes Mr. and Mrs. Jones, four children, two cats, and three dogs. Every six hours there is a Jones family stroll. The rules for a Jones family stroll are:

Exactly five things (people + dogs + cats) go on each stroll.

Each stroll must include at least one parent and at least one pet.

There can never be a dog and a cat on the same stroll unless both parents go.

All acceptable stroll groupings are equally likely.

Given that exactly one parent went on the 6 P.M. stroll, what is the probability that Rover, the oldest dog, also went?