2. Basic Concepts of Probability Theory

The following basic concepts will be presented.

- First, set theory is used to specify the sample space and the events of a random experiment.
- Second, the axioms of probability specify rules for computing the probabilities of events.
- Third, the notion of conditional probability allows us to determine how partial information about the outcome of an experiment affects the probabilities of events. Conditional probability also allows us to formulate the notion of “independence” of events and of experiments.
- Finally, we consider “sequential” random experiments that consist of performing a sequence of simple random subexperiments. We show how the probabilities of events in these experiments can be derived from the probabilities of the simpler subexperiments.

Throughout the book it is shown that complex random experiments can be analyzed by decomposing them into simple subexperiments.

2.1 Specifying Random Experiments

A random experiment is specified by stating an experimental procedure and a set of one or more measurements or observations.

- An unambiguous statement of exactly what is measured or observed.
- It may involve more than one measurement or observation.
- It may consist of a sequence of experiments and measurements.

Example 2.1

Experiment E1: Select a ball from an urn containing balls numbered 1 to 50. Note the number of the ball.

Experiment E2: Select a ball from an urn containing balls numbered 1 to 4. Suppose that balls 1 and 2 are black and that balls 3 and 4 are white. Note the number and color of the ball you select.

Experiment E3: Toss a coin three times and note the sequence of heads and tails.

Experiment E4: Toss a coin three times and note the number of heads.

Experiment E5: Count the number of voice packets containing only silence produced from a group of N speakers in a 10-ms period.
Experiment E6: A block of information is transmitted repeatedly over a noisy channel until an error-free block arrives at the receiver. Count the number of transmissions required.

Experiment E7: Pick a number at random between zero and one.

Experiment E8: Measure the time between page requests in a Web server.

Experiment E9: Measure the lifetime of a given computer memory chip in a specified environment.

Experiment E10: Determine the value of an audio signal at time

Experiment E11: Determine the values of an audio signal at times and

Experiment E12: Pick two numbers at random between zero and one.

Experiment E13: Pick a number X at random between zero and one, then pick a number Y at random between zero and X.

Experiment E14: A system component is installed at time t=0. For t≥0 let X(t)=1 as long as the component is functioning, and X(t)=0 let after the component fails.

2.1.1 The Sample Space

The sample space S of a random experiment is defined as the set of all possible outcomes.

- Sample spaces may be discrete.
- Sample spaces may be continuous.
- Sample spaces may be a multi-dimensional composed both discrete and continuous subspaces.

The sample space S can be specified compactly by using set notation. It can be visualized by drawing tables, diagrams, intervals of the real line, or regions of the plane.

Two basic ways to specify a set:

1. List all the elements, separated by commas, inside a pair of braces. For example:

   \[ A = \{1,2,3,\ldots,13\} \]

2. Give a property that specifies the elements of the set. For example:

   \[ A = \{x : x \text{ is an integer such that } 0 \leq x \leq 3\} \]

Sample space examples for the experiments of example 2.1 are shown in Example 2.2.
Some more interesting sample spaces are shown in the following figure.

Experiment E7: Pick a number at random between zero and one.

Experiment E9: Measure the lifetime of a given computer memory chip in a specified environment.

Experiment E12: Pick two numbers at random between zero and one.

Experiment E13: Pick a number X at random between zero and one, then pick a number Y at random between zero and X.

### 2.1.2 Events

We are usually not interested in the occurrence of specific outcomes, but rather in the occurrence of some event, typically a subset of the sample spaces.

We say that A is a subset of B if every element of A also belongs to B.

Two events of special interest are the certain event, S, which consists of all outcomes and hence always occurs, and the impossible or null event, Ø, which contains no outcomes and hence never occurs.
2.1.3 Review of Set Theory

A set is a collection of objects and will be denoted by capital letters: S, A, B, …

We define U as the universal set that consists of all possible objects of interest in a given setting or application.

In the context of random experiments we refer to the universal set as the sample space.

A set A is a collection of objects from U, and these objects are called the elements or points of the set A and will be denoted by lowercase letters. We use the notation:

\[ x \in A \quad \text{or} \quad y \notin A \]

read as “x is an element of A” and “y is not an element of A”.

We say A is a subset of B if every element of A also belongs to B, that is, if \( x \in A \) implies \( x \in B \). We say that “A is contained in B” and we write:

\[ A \subseteq B \]

Venn diagrams of set operations and relations.
Equality

We say sets $A$ and $B$ are equal if they contain the same elements. Since every element in $A$ is also in $B$, then $x \in A$ implies $x \in B$ so $A \subseteq B$. Similarly every element in $B$ is also in $A$, so $x \in B$ implies $x \in A$ and so $B \subseteq A$. Therefore:

$$A = B \text{ if and only if } A \subseteq B \text{ and } B \subseteq A$$

The standard method to show that two sets, $A$ and $B$, are equal is to show that $A \subseteq B$ and $B \subseteq A$. A second method is to list all the items in $A$ and all the items in $B$, and to show that the items are the same.

Sum or Union

The union of two sets $A$ and $B$ is denoted by $A \cup B$ and is defined as the set of outcomes that are either in $A$ or in $B$, or both:

$$A \cup B = \{x : x \in A \text{ or } x \in B\}$$

- The sum or union of sets results in a set that contains all of the elements that are elements of every set being summed.

$$S = A_1 \cup A_2 \cup A_3 \cdots \cup A_N$$

- Laws for Unions

$$A \cup B = B \cup A$$

$$A \cup A = A$$

$$A \cup \emptyset = A$$

$$A \cup S = S$$

$$A \cup B = A, \text{ if } B \subseteq A$$
Products or Intersection

The intersection of two sets $A$ and $B$ is denoted by $A \cap B$ and is defined as the set of outcomes that are in both $A$ and $B$:

$$A \cap B = \{x : x \in A \text{ and } x \in B\}$$

- The product or intersection of sets results in a set that contains all of the elements that are present in every one of the sets.

$$S \cap \emptyset = \emptyset$$

- Laws for Intersections

$$A \cap B = B \cap A$$

$$A \cap A = A$$

$$A \cap \emptyset = \emptyset$$

$$A \cap S = A$$

$$A \cap B = B, \text{ if } B \subset A$$

Mutually Exclusive or Disjoint Sets

Two sets are said to be disjoint or mutually exclusive if their intersection is the null set,

$$A \cap B = \emptyset$$

- NOTE: The intersection of two disjoint sets is a set … the null set.
Complement

The complement of a set $A$ is denoted by $A^C$ and is defined as the set of all elements not in $A$:

$$A^C = \{ x : x \notin A \}$$

- Laws for Complement

$\emptyset = S$

$S = \emptyset$

$\overline{A} = A$

$A \subseteq B$, if $B \subseteq A$

$A = B$, if $B = A$

Differences

The relative complement or difference of sets $A$ and $B$ is the set of elements in $A$ that are not in $B$:

$$A - B = \{ x : x \in A \text{ and } x \notin B \}$$

$$A - B = A \cap B^C$$

- Laws for Differences

$(A - B) \cup B \neq B$

$(A \cup A) - A = \emptyset$

$(A - A) \cup A = A$

$A - \emptyset = A$

$A - S = \emptyset$

$S - A = \overline{A}$

$S - A = A^C$
Properties of Sets

Commutative properties:

\[ A \cup B = B \cup A \quad \text{and} \quad A \cap B = B \cap A \]

Associative properties:

\[ A \cup (B \cup C) = (A \cup B) \cup C \quad \text{and} \quad A \cap (B \cap C) = (A \cap B) \cap C \]

Distributive properties:

\[ A \cup (B \cap C) = (A \cup B) \cap (A \cup C) \quad \text{and} \quad A \cap (B \cup C) = (A \cap B) \cup (A \cap C) \]

DeMorgan’s rules:

\[ (A \cup B)^C = A^C \cap B^C \quad \text{and} \quad (A \cap B)^C = A^C \cup B^C \]
### Set Algebra Table

<table>
<thead>
<tr>
<th>Equation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$F \cup G = G \cup F$</td>
<td>commutative law</td>
</tr>
<tr>
<td>$F \cup (G \cup H) = (F \cup G) \cup H$</td>
<td>associative law</td>
</tr>
<tr>
<td>$F \cap (G \cup H) = (F \cap G) \cup (F \cap H)$</td>
<td>distributive law</td>
</tr>
<tr>
<td>$(F^c)^c = F$</td>
<td></td>
</tr>
<tr>
<td>$F \cap F^c = \emptyset$</td>
<td></td>
</tr>
<tr>
<td>$(F \cap G)^c = F^c \cup G^c$</td>
<td>DeMorgan’s “law”</td>
</tr>
<tr>
<td>$F \cap \Omega = F$</td>
<td></td>
</tr>
<tr>
<td>$F \cap G = G \cap F$</td>
<td>commutative law</td>
</tr>
<tr>
<td>$F \cap (G \cap H) = (F \cap G) \cap H$</td>
<td>associative law</td>
</tr>
<tr>
<td>$(F \cup G)^c = F^c \cap G^c$</td>
<td>DeMorgan’s other “law”</td>
</tr>
<tr>
<td>$F \cup F^c = \Omega$</td>
<td></td>
</tr>
<tr>
<td>$F \cup \emptyset = F$</td>
<td></td>
</tr>
<tr>
<td>$F \cup (F \cap G) = F = F \cap (F \cup G)$</td>
<td></td>
</tr>
<tr>
<td>$F \cup \Omega = \Omega$</td>
<td></td>
</tr>
<tr>
<td>$F \cap \emptyset = \emptyset$</td>
<td></td>
</tr>
<tr>
<td>$F \cap G = F \cup (F^c \cap G) = F \cup (G - F)$</td>
<td></td>
</tr>
<tr>
<td>$F \cup (G \cap H) = (F \cup G) \cap (F \cup H)$</td>
<td>distributive law</td>
</tr>
<tr>
<td>$\Omega^c = \emptyset$</td>
<td></td>
</tr>
<tr>
<td>$F \cup F = F$</td>
<td></td>
</tr>
<tr>
<td>$F \cap F = F$</td>
<td></td>
</tr>
</tbody>
</table>

**Table A.1 Set algebra**

2.1.4 Event Classes (special – optional definition)

We have introduced the sample space $S$ as the set of all possible outcomes of the random experiment. We have also introduced events as subsets of $S$.

Probability theory also requires that we state the class $F$ of events of interest. Only events in this class are assigned probabilities. We expect that any set operation on events in $F$ will produce a set that is also an event in $F$.

When the sample space $S$ is finite or countable, we simply let $F$ consist of all subsets of $S$ and we can proceed without further concerns about $F$.

However, when $S$ is the real line $\mathbb{R}$ (or an interval of the real line), we cannot let $F$ be all possible subsets of $\mathbb{R}$ and still satisfy the axioms of probability. Fortunately, we can obtain all the events of practical interest by letting $F$ be of the class of events obtained as complements and countable unions and intersections of intervals of the real line, e.g., $(a, b]$ or $(-\infty, b]$. We will refer to this class of events as the Borel field.

_In the remainder of the book, we will refer to the event class from time to time. For the introductory-level course in probability you will not need to know more than what is stated in this paragraph._

---

**Example 2.8**

Let $S = \{T, H\}$ be the outcome of a coin toss. Let every subset of $S$ be an event. Find all possible events of $S$.

An event is a subset of $S$, so we need to find all possible subsets of $S$. These are:

$$S = \{\emptyset, \{H\}, \{T\}, \{H, T\}\}.$$  

Note that $S$ includes both the empty set and $S$. Let $i_T$ and $i_H$ be binary numbers where $i = 1$ indicates that the corresponding element of $S$ is in a given subset. We generate all possible subsets by taking all possible values of the pair $i_T$ and $i_H$. Thus $i_T = 0, i_H = 1$ corresponds to the set $\{H\}$. Clearly there are $2^2$ possible subsets as listed above.

---

For a finite sample space, $S = \{x_1, x_2, \ldots, x_k\}$, we usually allow all subsets of $S$ to be events. This class of events is called the power set of $S$ and we will denote it by $\mathcal{S}$. We can index all possible subsets of $S$ with binary numbers $i_1, i_2, \ldots, i_k$, and we find that the power set of $S$ has $2^k$ members.
2.2 The Axioms of Probability

Probabilities are numbers assigned to events that indicate how “likely” it is that the events will occur when an experiment is performed. A probability law for a random experiment is a rule that assigns probabilities to the events of the experiment that belong to the event class $F$. Thus a probability law is a function that assigns a number to sets (events).

Let $E$ be a random experiment with sample space $S$ and event class $F$. A probability law for the experiment $E$ is a rule that assigns to each event $A \in F$ a number $P[A]$, called the probability of $A$, that satisfies the following axioms:

- **Axiom I**  
  \[ 0 \leq P[A] \]

- **Axiom II**  
  \[ P[S] = 1 \]

- **Axiom III**  
  If $A \cap B = \emptyset$, then  
  \[ P[A \cup B] = P[A] + P[B] \]

- **Axiom III’**  
  If $A_1, A_2, \ldots$ is a sequence of events such that  
  $A_i \cap A_j = \emptyset$ for all $i \neq j$, then  
  \[ P\left[ \bigcup_{k=1}^{\infty} A_k \right] = \sum_{k=1}^{\infty} P[A_k] \]

**Corollaries based on the axioms (can use Venn diagrams to show graphically)**

- **Corollary 1**  
  \[ P[A^c] = 1 - P[A] \]

- **Corollary 2**  
  \[ P[A] \leq 1 \]

- **Corollary 3**  
  \[ P[\emptyset] = 1 - P[S] = 0 \]

- **Corollary 4**  
  If $A_1, A_2, \ldots, A_n$ are pairwise mutually exclusive, then  
  \[ P\left[ \bigcup_{k=1}^{n} A_k \right] = \sum_{k=1}^{n} P[A_k] \]

- **Corollary 5**  
  \[ P[A \cup B] = P[A] + P[B] - P[A \cap B] \]

  \[ P\left[ \bigcup_{k=1}^{n} A_k \right] = \sum_{k=1}^{n} P[A_k] - \sum_{j<k} P[A_j \cap A_k] + \sum_{j<k<d} P[A_j \cap A_k \cap A_d] - \cdots \]

  \[ + (-1)^{n+1} \cdot P[A_1 \cap A_2 \cap \cdots \cap A_n] \]

- **Corollary 7**  
  If $A \subset B$, then $P[A] \leq P[B]$
2.2.1 Discrete Sample Spaces

In this section we show that the probability law for an experiment with a countable sample space can be specified by giving the probabilities of the elementary events.

Example 2.9

An urn contains 10 identical balls numbered 0, 1, 2, …, 9. A random experiment involves selecting a ball from the urn and noting the number of the ball. Find the probability of the following events:

A = “number of ball selected is odd,”
B = “number of ball selected is a multiple of 3,”
C = “number of ball selected is less than 5,”

in addition, find $A \cup B$ and $A \cup B \cup C$

$S = \{0,1,2,3,4,5,6,7,8,9\}$


$B = \{3,6,9\}$ \hspace{1cm} $P[B] = P[3] + P[6] + P[9] = \frac{3}{10}$

$C = \{0,1,2,3,4\}$ \hspace{1cm} $P[C] = P[0] + P[1] + P[2] + P[3] + P[4] = \frac{5}{10}$

$A \cup B = \{1,3,5,6,7,9\}$ \hspace{1cm} $P[A \cup B] = P[A] + P[B] - P[A \cap B]$

$A \cup B = \{1,3,5,7,9\}$ \hspace{1cm} $P[A \cup B] = P[A] + P[B] - P[3] - P[9]$

$P[A \cup B] = \frac{5}{10} + \frac{3}{10} - \frac{2}{10} = \frac{6}{10}$

$A \cup B \cup C = \{0,1,2,3,4,5,6,7,9\}$

$P[A \cup B \cup C] = P[A] + P[B] + P[C] - P[A \cap B] - P[A \cap C] - P[B \cap C] + P[A \cap B \cap C]$


$P[A \cup B \cup C] = \frac{5}{10} + \frac{3}{10} + \frac{5}{10} - \frac{2}{10} - \frac{2}{10} - \frac{1}{10} + \frac{1}{10} = \frac{9}{10}$

Example 2.10

Suppose that a coin is tossed three times. If we observe the sequence of heads and tails, then there are eight possible outcomes \( S_3 = \{ \text{HHH}, \text{HHT}, \text{HTH}, \text{THH}, \text{THT}, \text{TTH}, \text{HTT}, \text{TTT} \} \). If we assume that the outcomes of \( S_3 \) are equiprobable, then the probability of each of the eight elementary events is \( 1/8 \). This probability assignment implies that the probability of obtaining two heads in three tosses is, by Corollary 3,

\[
P[\text{“2 heads in 3 tosses”}] = P[\{ \text{HHT}, \text{HTH}, \text{THH} \}] = P[\{ \text{HHT} \}] + P[\{ \text{HTH} \}] + P[\{ \text{THH} \}] = \frac{3}{8}.
\]

Now suppose that we toss a coin three times but we count the number of heads in three tosses instead of observing the sequence of heads and tails. The sample space is now \( S_4 = \{0, 1, 2, 3\} \). If we assume the outcomes of \( S_4 \) to be equiprobable, then each of the elementary events of \( S_4 \) has probability \( 1/4 \). This second probability assignment predicts that the probability of obtaining two heads in three tosses is

\[
P[\text{“2 heads in 3 tosses”}] = P[\{2\}] = \frac{1}{4}.
\]

The first probability assignment implies that the probability of two heads in three tosses is \( 3/8 \), and the second probability assignment predicts that the probability is \( 1/4 \). Thus the two assignments are not consistent with each other.

As far as the theory is concerned, either one of the assignments is acceptable. It is up to us to decide which assignment is more appropriate. Later in the chapter we will see that only the first assignment is consistent with the assumption that the coin is fair and that the tosses are “independent.” This assignment correctly predicts the relative frequencies that would be observed in an actual coin tossing experiment.
Example 2.11

A fair coin is tossed repeatedly until the first heads shows up; the outcome of the experiment is the number of tosses required until the first heads occurs. Find a probability law for this experiment.

It is conceivable that an arbitrarily large number of tosses will be required until heads occurs, so the sample space is $S = \{1, 2, 3, \ldots\}$. Suppose the experiment is repeated $n$ times. Let $N_j$ be the number of trials in which the $j$th toss results in the first heads. If $n$ is very large, we expect $N_j$ to be approximately $n/2$ since the coin is fair. This implies that a second toss is necessary about $n - N_j \approx n/2$ times, and again we expect that about half of these—that is, $n/4$—will result in heads, and so on, as shown in Fig. 2.5. Thus for large $n$, the relative frequencies are

$$f_j \approx \frac{N_j}{n} = \left(\frac{1}{2}\right)^{j} \quad j = 1, 2, \ldots$$

We therefore conclude that a reasonable probability law for this experiment is

$$P[j \text{ tosses till first heads}] = \left(\frac{1}{2}\right)^{j} \quad j = 1, 2, \ldots$$  \hspace{1cm} (2.16)

We can verify that these probabilities add up to one by using the geometric series with $\alpha = 1/2$:

$$\sum_{j=1}^{\infty} \alpha^j = \frac{\alpha}{1 - \alpha} \bigg|_{\alpha=1/2} = 1.$$
2.2.2 Continuous Sample Spaces

Continuous sample spaces arise in experiments in which the outcomes are numbers that can assume a continuum of values, so we let the sample space $S$ be the entire real line $R$ (or some interval of the real line).

We could consider letting the event class consist of all subsets of $R$. But it turns out that this class is “too large” and it is impossible to assign probabilities to all the subsets of $R$. Fortunately, it is possible to assign probabilities to all events in a smaller class that includes all events of practical interest. This class denoted by $B$ is called the Borel field and it contains all open and closed intervals of the real line as well as all events that can be obtained as countable unions, intersections, and complements.

Base on axiom III’

Axiom III’  If $A_1, A_2, \ldots$ is a sequence of events such that $A_i \cap A_j = \emptyset$ for all $i \neq j$, then

$$P\left[ \bigcup_{k=1}^{\infty} A_k \right] = \sum_{k=1}^{\infty} P[A_k]$$

Let $A_1, A_2, \ldots$ be a sequence of mutually exclusive events that are represented by intervals of the real line, then

$$P\left[ \bigcup_{k=1}^{\infty} A_k \right] = \sum_{k=1}^{\infty} P[A_k]$$

where each $P[A_k]$ is specified by the probability law. For this reason, probability laws in experiments with continuous sample spaces specify a rule for assigning numbers to intervals of the real line.
Example 2.12

Consider the random experiment “pick a number x at random between zero and one.” The sample space S for this experiment is the unit interval [0, 1], which is uncountably infinite. If we suppose that all the outcomes S are equally likely to be selected, then we would guess that the probability that the outcome is in the interval [0, 1/2] is the same as the probability that the outcome is in the interval [1/2, 1]. We would also guess that the probability of the outcome being exactly equal to 1/2 would be zero since there are an uncountably infinite number of equally likely outcomes.

Consider the following probability law: “The probability that the outcome falls in a subinterval of S is equal to the length of the subinterval,” that is,

\[ P([a, b]) = (b - a), \quad \text{for } 0 \leq a \leq b \leq 1 \]

where by \( P([a, b]) \) we mean the probability of the event corresponding to the interval \([a, b]\).

Therefore we have

\[
P([0, 0.5]) = (0.5 - 0) = 0.5
\]

\[
P([0.5, 1]) = (1 - 0.5) = 0.5
\]

\[
P([0.5, 0.5]) = (0.5 - 0.5) = 0
\]

Notice also that

\[
P([0.3, 0.5] \cup [0.4, 0.6]) = P([0.3, 0.5]) + P([0.4, 0.6]) - P([0.3, 0.5] \cap [0.4, 0.6])
\]

\[
P([0.3, 0.5] \cup [0.4, 0.6]) = (0.5 - 0.3) + (0.6 - 0.4) - (0.5 - 0.4) = 0.3
\]
Example 2.13

Suppose that the lifetime of a computer memory chip is measured, and we find that “the proportion of chips whose lifetime exceeds t decreases exponentially at a rate $\alpha$. Find an appropriate probability law.

Let the sample space in this experiment be $S = (0, \infty)$. If we interpret the above finding as “the probability that a chip’s lifetime exceeds t decreases exponentially at a rate $\alpha$” we then obtain the following assignment of probabilities to events of the form $(t, \infty)$:

$$P[(t, \infty)] = \exp(-\alpha \cdot t), \quad \text{for } t > 0$$

where $\alpha > 0$.

This formula satisfies axiom I as the value is between 0 and 1 or

$$0 \leq P[(t, \infty)] = \exp(-\alpha \cdot t), \quad \text{for } t > 0$$

It also satisfies axiom II as $P[(0, \infty)] = 1$

Then, the probability that the lifetime is in the interval $(r,s]$ can be found. First, note that

$$(r,s) \cup (s, \infty) = (r, \infty)$$

That is,

$$P[(r,s)] + P[(s, \infty)] = P[(r, \infty)]$$

or rearranging … we get the result

$$P[(r,s)] = P[(r, \infty)] - P[(s, \infty)] = \exp(-\alpha \cdot r) - \exp(-\alpha \cdot s)$$

Mathematically this is equivalent to a factored version …

$$P[(r,s)] = \exp(-\alpha \cdot r) \cdot [1 - \exp(-\alpha \cdot (s - r))]$$

This has an interesting interpretation that will be discussed later …
2.3  *Computing Probabilities Using Counting Methods

In many experiments with finite sample spaces, the outcomes can be assumed to be equiprobable. The probability of an event is then the ratio of the number of outcomes in the event of interest to the total number of outcomes in the sample space. The calculation of probabilities reduces to counting the number of outcomes in an event. In this section, we develop several useful counting (combinatorial) formulas.

The following sections establish the mathematics and representation used/required for discrete sample space probability.

2.3.1 Sampling with Replacement and with Ordering

Suppose we choose k objects from a set A that has n distinct objects, with replacement—that is, after selecting an object and noting its identity in an ordered list, the object is placed back in the set before the next choice is made. We will refer to the set A as the “population.” The experiment produces an ordered k-tuple

\[(x_1, x_2, \ldots, x_k)\]

where \(x_i \in A\) and \(i = 1, 2, \ldots, k\) with \(n_1, n_2, \ldots, n_k = n\) implies that the number of distinct ordered k-tuples is \(n^k\).

**Example 2.15**

An urn contains five balls numbered 1 to 5. Suppose we select two balls from the urn with replacement. How many distinct ordered pairs are possible? What is the probability that the two draws yield the same number?

The above equation states that the number of ordered pairs is \(5^2 = 25\). Table 2.1 shows the 25 possible pairs. Five of the 25 outcomes have the two draws yielding the same number; if we suppose that all pairs are equiprobable, then the probability that the two draws yield the same number is \(\frac{5}{25} \cdot \frac{1}{5} = \frac{5}{25} = 0.2\).

<table>
<thead>
<tr>
<th>TABLE 2.1 Enumeration of possible outcomes in various types of sampling of two balls from an urn containing five distinct balls.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) Ordered pairs for sampling with replacement.</td>
</tr>
<tr>
<td>(1,1)</td>
</tr>
<tr>
<td>(2,1)</td>
</tr>
<tr>
<td>(3,1)</td>
</tr>
<tr>
<td>(4,1)</td>
</tr>
<tr>
<td>(5,1)</td>
</tr>
</tbody>
</table>

### 2.3.2 Sampling without Replacement and with Ordering

Suppose we choose $k$ objects in succession *without replacement* from a population $A$ of $n$ distinct objects. Clearly, $k \leq n$. The number of possible outcomes in the first draw is $n_1 = n$; the number of possible outcomes in the second draw is $n_2 = n - 1$, namely all $n$ objects except the one selected in the first draw; and so on, up to $n_k = n - (k - 1)$ in the final draw. This then gives the number of distinct ordered $k$-tuples

\[
n_1 \cdot n_2 \ldots n_k = n \cdot (n - 1) \ldots (n - (k - 1)) = \frac{n!}{(n - k)!}
\]

**Example 2.16**

An urn contains five balls numbered 1 to 5. Suppose we select two balls in succession without replacement. How many distinct ordered pairs are possible? What is the probability that the first ball has a number larger than that of the second ball?

The above equation states that the number of ordered pairs is $n_1 \cdot n_2 = 5 \cdot 4 = 20$. The 20 possible ordered pairs are shown in Table 2.1(b). Ten ordered pairs in Tab. 2.1(b) have the first number larger than the second number; thus the probability of this event is $\frac{10}{20} = 0.5$. Alternately, enumerating each possible event

\[
\begin{align*}
\frac{1}{5} \cdot \frac{4}{4} + \frac{1}{5} \cdot \frac{3}{4} + \frac{1}{5} \cdot \frac{2}{4} + \frac{1}{5} \cdot \frac{1}{4} + \frac{1}{5} \cdot 0 &= \frac{1}{5} \left( \frac{4 + 3 + 2 + 1}{4} \right) = \frac{10}{20} = 0.5 \\
(b) \text{ Ordered pairs for sampling without replacement.} \\
(1,2) & (1,3) & (1,4) & (1,5) \\
(2,1) & (2,3) & (2,4) & (2,5) \\
(3,1) & (3,2) & (3,4) & (3,5) \\
(4,1) & (4,2) & (4,3) & (4,5) \\
(5,1) & (5,2) & (5,3) & (5,4)
\end{align*}
\]

The number of possible outcomes can also be described as $n^2 - n = n \cdot (n - 1)$
Example 2.17

An urn contains five balls numbered 1, 2, 3, 4, 5. Suppose we draw three balls with replacement. What is the probability that all three balls are different?

There are $5^3 = 125$ possible outcomes, which we will suppose are equiprobable. The number of these outcomes for which the three draws are different is given by:

$$5 \cdot 4 \cdot 3 = 60.$$  

Thus the probability that all three balls are different is $\frac{60}{125} = 0.48$.

Alternately,

$$\frac{5 \cdot 4 \cdot 3}{5 \cdot 5 \cdot 5} = \frac{60}{125} = 0.48$$

Always think … (acceptable outcomes)/(total possible outcomes)=probability.

They sometimes make up two separate problems when the circumstances are complicated.

2.3.3 Permutations of n Distinct Objects

Consider sampling without replacement with $k = n$. This is simply drawing objects from an urn containing $n$ distinct objects until the urn is empty. Thus, the number of possible orderings (arrangements, permutations) of $n$ distinct objects is equal to the number of ordered $n$-tuples in sampling without replacement with $k = n$.

$$n_1 \cdot n_2 \ldots \cdot n_n = n \cdot (n-1) \cdot \ldots \cdot (n-(n-1)) = n!$$

This is a factorial.

Example 2.18

Find the number of permutations of three distinct objects $\{1,2,3\}$. Equation (2.22) gives $3! = 3 \cdot 2 \cdot 1 = 6$. The six permutations are

$$123, 312, 231, 132, 213, 321$$

Step 1: Identify the number of outcomes,

Step 2: Determine what they all are. [If you don’t do step 1, you may miss some!]
Example 2.19

Suppose that 12 balls are placed at random into 12 cells, where more than 1 ball is allowed to occupy a cell. What is the probability that all cells are occupied?

The placement of each ball into a cell can be viewed as the selection of a cell number between 1 and 12. Equation (2.20) implies that there are $12^{12}$ possible placements of the 12 balls in the 12 cells.

In order for all cells to be occupied, the first ball selects from any of the 12 cells, the second ball from the remaining 11 cells, and so on. Thus the number of placements that occupy all cells is $12!$.

If we suppose that all possible placements are equiprobable, we find that the probability that all cells are occupied is

$$
\frac{12!}{12^{12}} \approx \frac{479001600}{8.9161e+12} \approx 5.3723e-05
$$

Reinterpreting, if there are 12 students in a classroom, what is the probability that, as a group, there is one and only one birthday every month of a year.

As birthdays would be thought to occur randomly in a year, they do not tend to occur uniformly spaced in time for the group. Next, what is the probability that two or more students would be born in the same month …

Again, make it into two problems … what is the numerator, what is the denominator.
2.3.4 Sampling without Replacement and without Ordering

Suppose we pick \( k \) objects from a set of \( n \) distinct objects without replacement and that we record the result without regard to order. (You can imagine putting each selected object into another jar, so that when the \( k \) selections are completed we have no record of the order in which the selection was done.) We call the resulting subset of \( k \) selected objects a “combination of size \( k \).”

From Eq. (2.22), there are \( k! \) possible orders in which the \( k \) objects in the second jar could have been selected.

Thus if \( \binom{n}{k} \) denotes the number of combinations of size \( k \) from a set of size \( n \), then the total number of distinct ordered samples of \( k \) objects, which is given by Eq. (2.21)

\[
\binom{n}{k} \cdot k! = n \cdot (n-1) \cdot \ldots \cdot (n-(k-1)) = \frac{n!}{(n-k)!}
\]

Then by definition

\[
\binom{n}{k} = \frac{n!}{k!(n-k)!} \equiv \binom{n}{k}
\]

This expression is referred to as a binomial coefficient and is read “\( n \) choose \( k \).”

Note that choosing \( k \) objects out of a set of \( n \) is equivalent to choosing the \( n-k \) objects that are to be left out of selection. That is

\[
\binom{n}{k} = \binom{n}{n-k} = \frac{n!}{k!(n-k)!} = \binom{n}{n-k} = \binom{n}{n-k}
\]

**Example 2.20**

Find the number of ways of selecting two objects from \( A = \{1,2,3,4,5\} \) without regard to order.

Equation (2.25) gives

\[
\binom{5}{2} = \frac{5!}{2!(3)!} = \frac{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{(2 \cdot 1) \cdot (3 \cdot 2 \cdot 1)} = \frac{5 \cdot 4}{2} = 10
\]
Example 2.21

Find the number of distinct permutations of \( k \) white balls and \( n-k \) black balls. This problem is equivalent to the following sampling problem: Put \( n \) tokens (numbered 1 to \( n \)) in an urn, where each token represents a position in the arrangement of balls; pick a combination of \( k \) tokens and put the \( k \) white balls in the corresponding positions. Each combination of size \( k \) leads to a distinct arrangement (permutation) of \( k \) white balls and \( n-k \) black balls. Thus the number of distinct permutations of \( k \) white balls and \( n-k \) black balls is \( \binom{n}{k} \).

As a specific example let \( n=4 \) and \( k=2 \). The number of combinations of size 2 from a set of four distinct objects is

\[
\binom{4}{2} = \frac{4!}{2!(4-2)!} = \frac{4 \cdot 3 \cdot 2 \cdot 1}{2 \cdot 1} = \frac{4 \cdot 3}{2} = 6
\]

The 6 distinct permutations with 2 whites (zeros) and 2 blacks (ones) are

\[1100, 1010, 1001, 0110, 0101, 0011\]

Example 2.22 Quality Control

A batch of 50 items contains 10 defective items. Suppose 10 items are selected at random and tested. What is the probability that exactly 5 of the items tested are defective?

The number of ways of selecting 10 items out of a batch of 50 is the number of combinations of size 10 from a set of 50 objects:

\[
\binom{50}{10} = \frac{50!}{10!(50-10)!}
\]

The number of ways of selecting 5 defective and 5 nondefective items from the batch of 50 is the product \( N_1 \times N_2 \) where \( N_1 \) is the number of ways of selecting the 5 items from the set of 10 defective items, and \( N_2 \) is the number of ways of selecting 5 items from the 40 nondefective items. Thus the probability that exactly 5 tested items are defective is

\[
\frac{C_{10}^{10} \cdot C_{40}^{5}}{C_{50}^{10}} = \frac{\frac{10!}{5!5!} \cdot \frac{40!}{5!35!}}{\frac{50!}{10!40!}} = \frac{252 \cdot 658008}{10272278170} = 0.01614
\]
2.3.5 Sampling with Replacement and without Ordering

Suppose we pick \( k \) objects from a set of \( n \) distinct objects with replacement and we record the result without regard to order. This can be done by filling out a form which has \( n \) columns, one for each distinct object. Each time an object is selected, an “x” is placed in the corresponding column.

The number of different ways of picking \( k \) objects from a set of \( n \) distinct objects with replacement and without ordering is given by

\[
\binom{n-1+k}{k} = \binom{n-1+k}{n-1} = \frac{(n-1+k)!}{k!(n-1)!}
\]

Recap of equations

- Sampling with replacement with ordering
- Sampling without replacement with ordering
- Sampling without replacement without ordering
  - typical card games – cards are drawn one at a time
- Sampling with replacement without ordering
2.4 Conditional Probability

Quite often we are interested in determining whether two events, \( A \) and \( B \), are related in the sense that knowledge about the occurrence of one, say \( B \), alters the likelihood of occurrence of the other, \( A \). This requires that we find the conditional probability, \( P[A | B] \), of event \( A \) given that event \( B \) has occurred. The conditional probability is defined by

\[
P[A | B] = \frac{P[A \cap B]}{P[B]}, \text{ for } P(B) > 0
\]

Note also that

\[
Pr(A \cap B) = Pr(A | B) \cdot Pr(B), \text{ for } Pr(B) > 0
\]

Note also that

\[
Pr(A \cap B) = Pr(B | A) \cdot Pr(A), \text{ for } Pr(A) > 0
\]

Using a Venn diagram this is relatively easy to observe.
Extensions based on subsets:

If $A$ is a subset of $B$, then the conditional probability must be

$$
Pr(A \mid B) = \frac{Pr(A \cap B)}{Pr(B)} = \frac{Pr(A)}{Pr(B)}, \text{ for } A \subseteq B
$$

Therefore, it can be said that

$$
Pr(A \mid B) = \frac{Pr(A \cap B)}{Pr(B)} = \frac{Pr(A)}{Pr(B)} \geq Pr(A), \text{ for } A \subseteq B
$$

If $B$ is a subset of $A$, then the conditional probability becomes

$$
Pr(A \mid B) = \frac{Pr(A \cap B)}{Pr(B)} = \frac{Pr(B)}{Pr(B)} = 1, \text{ for } B \subseteq A
$$

If $A$ and $B$ are mutually exclusive,

$$
Pr(A \mid B) = \frac{Pr(A \cap B)}{Pr(B)} = \frac{0}{Pr(B)} = 0, \text{ for } B \cap A = \emptyset
$$

These concepts are useful when performing sequential experiments and defining structures based on tree diagrams.
**Example: Marbles**

Bag of marbles: 3-blue, 2-red, one-yellow

- **Objects:** Marbles
- **Attributes:** Color (Blue, Red, Yellow)
- **Experiment:** Draw two marble, with replacement
- **Sample Space:** {BB, BR, RB, RR, RY, YR, YY}
- **Probability (relative frequency method)**

<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>R</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>( \frac{3}{6} \cdot \frac{3}{6} = \frac{9}{36} )</td>
<td>( \frac{3}{6} \cdot \frac{2}{6} = \frac{6}{36} )</td>
<td>( \frac{3}{6} \cdot \frac{1}{6} = \frac{3}{36} )</td>
</tr>
<tr>
<td>R</td>
<td>( \frac{2}{6} \cdot \frac{3}{6} = \frac{6}{36} )</td>
<td>( \frac{2}{6} \cdot \frac{2}{6} = \frac{4}{36} )</td>
<td>( \frac{2}{6} \cdot \frac{1}{6} = \frac{2}{36} )</td>
</tr>
<tr>
<td>Y</td>
<td>( \frac{1}{6} \cdot \frac{3}{6} = \frac{3}{36} )</td>
<td>( \frac{1}{6} \cdot \frac{2}{6} = \frac{2}{36} )</td>
<td>( \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36} )</td>
</tr>
</tbody>
</table>

What if we want RB and know that we drew R as the first marble?

\[
\Pr(R, B) \text{ given that R occurred } = \frac{\Pr(RB)}{\Pr(R)} = \frac{\frac{6}{36}}{\frac{2}{6}} = \frac{6}{2} \cdot \frac{3}{6} = \frac{3}{6}
\]
Example 2.25

An urn contains two black balls and three white balls. Two balls are selected at random from the urn without replacement and the sequence of colors is noted. Find the probability that both balls are black.

This experiment consists of a sequence of two subexperiments. We can imagine working our way down the tree shown in Fig. 2.10 from the topmost node to one of the bottom nodes:

- We reach node 1 in the tree if the outcome of the first draw is a black ball; then the next subexperiment consists of selecting a ball from an urn containing one black ball and three white balls.
- On the other hand, if the outcome of the first draw is white, then we reach node 2 in the tree and the second subexperiment consists of selecting a ball from an urn that contains two black balls and two white balls.

Thus if we know which node is reached after the first draw, then we can state the probabilities of the outcome in the next subexperiment.

In general, the probability of any sequence of colors is obtained by multiplying the probabilities corresponding to the node transitions in the tree in Fig. 2.10.

What is the probability of selecting two black balls given that the first ball selected is black?

\[
P[A \mid B] = \frac{P[A \cap B]}{P[B]} = \frac{\frac{2}{5} \cdot \frac{1}{4}}{\frac{2}{5}} = \frac{\frac{1}{10}}{\frac{2}{5}} = \frac{1}{4}
\]
Example 2.26 Binary Communication System

Many communication systems can be modeled in the following way. First, the user inputs a 0 or a 1 into the system, and a corresponding signal is transmitted. Second, the receiver makes a decision about what was the input to the system, based on the signal it received.

Suppose that the user sends 0s with probability 1-p and 1s with probability p, and suppose that the receiver makes random decision errors with probability $\varepsilon$. For $i=0,1$ let $A_i$ be the event “input was $i$,” and let $B_i$ be the event “receiver decision was $i$.“ Find the probabilities $P[A_i \cap B_j]$ for $i=0,1$ and $j=0,1$.

The tree diagram for this experiment is shown in Fig. 2.11. We then readily obtain the desired probabilities:

\[
P[A_0 \cap B_0] = (1-p) \cdot (1-\varepsilon) \\
P[A_0 \cap B_1] = (1-p) \cdot \varepsilon \\
P[A_1 \cap B_0] = p \cdot \varepsilon \\
P[A_1 \cap B_1] = p \cdot (1-\varepsilon)
\]

What is the probability of successful bit transmission and reception? No bit errors …

\[
P[A_0 \cap B_0] + P[A_1 \cap B_1] = (1-p) \cdot (1-\varepsilon) + p \cdot (1-\varepsilon) = 1-\varepsilon
\]
**Total Probability**

For a space, $S$, that consists of multiple mutually exclusive events, the probability of a random event, $B$, occurring in space $S$, can be described based on the conditional probabilities associated with each of the possible events. (see Venn diagram Fig. 1-7)

Proof:

$$S = A_1 \cup A_2 \cup A_3 \cdots \cup A_n$$

and

$$A_i \cap A_j = \emptyset, \text{ for } i \neq j$$

$$B = B \cap S = B \cap (A_1 \cup A_2 \cup A_3 \cdots \cup A_n) = (B \cap A_1) \cup (B \cap A_2) \cup (B \cap A_3) \cdots \cup (B \cap A_n)$$

$$\Pr(B) = \Pr(B \cap A_1) + \Pr(B \cap A_2) + \Pr(B \cap A_3) \cdots + \Pr(B \cap A_n)$$

But

$$\Pr(B \cap A_i) = \Pr(B \mid A_i) \cdot \Pr(A_i), \text{ for } \Pr(A_i) > 0$$

Therefore

$$\Pr(B) = \Pr(B \mid A_1) \cdot \Pr(A_1) + \Pr(B \mid A_2) \cdot \Pr(A_2) + \cdots + \Pr(B \mid A_n) \cdot \Pr(A_n)$$

![Venn Diagram](image)

**Figure 2.12**
A partition of $S$ into $n$ disjoint sets.

Note: text reverses the A and B representation above.
Example 2.28

A manufacturing process produces a mix of “good” memory chips and “bad” memory chips. The lifetime of good chips follows the exponential law introduced in Example 2.13, with a rate of failure $\alpha$. The lifetime of bad chips also follows the exponential law, but the rate of failure is 1000 x $\alpha$. Suppose that the fraction of good chips is 1-p and of bad chips, p. Find the probability that a randomly selected chip is still functioning after t seconds.

Let C be the event “chip still functioning after t seconds,” and let G be the event “chip is good,” and B the event “chip is bad.” By the theorem on total probability we have

$$P[C] = P[C \mid G] \cdot P[G] + P[C \mid B] \cdot P[B]$$

$$P[C \mid G] = \exp(-\alpha \cdot t) \quad \text{and} \quad P[C \mid B] = \exp(-1000 \cdot \alpha \cdot t)$$

$$P[G] = 1 - p \quad \text{and} \quad P[B] = p$$

$$P[C] = \exp(-\alpha \cdot t) \cdot (1 - p) + \exp(-1000 \cdot \alpha \cdot t) \cdot p$$
2.4.1 **A Priori and A Posteriori Probability**

The probabilities defined for the expected outcomes, \( \Pr(A_i) \), are referred to as *a priori* probabilities (before the event). They describe the probability before the actual experiment or experimental results are known.

After an event has occurred, the outcome B is known. Then, the probability of the event belonging to one of the expected outcomes can be defined as

\[
\Pr(A_i | B)
\]

Using mathematics from before

\[
\Pr(A_i \cap B) = \Pr(B | A_i) \cdot \Pr(A_i) = \Pr(A_i | B) \cdot \Pr(B)
\]

\[
\Pr(A_i | B) = \frac{\Pr(B | A_i) \cdot \Pr(A_i)}{\Pr(B)}, \text{ for } \Pr(B) > 0
\]

and the final form using total probability

\[
\Pr(A_i | B) = \frac{\Pr(B | A_i) \cdot \Pr(A_i)}{\Pr(B | A_1) \cdot \Pr(A_1) + \Pr(B | A_2) \cdot \Pr(A_2) + \cdots + \Pr(B | A_n) \cdot \Pr(A_n)}
\]

This probability is referred to as the *a posteriori* probability (after the event).

The probability that event \( A_i \) occurred given that outcome B was observed.

It is also referred to as **Bayes Theorem or Bayes’ Rule**.
Example 2.29 Binary Communication System

In the binary communication system in Example 2.26, find which input is more probable given that the receiver has output a 1. Assume that, a priori, the input is equally likely to be 0 or 1.

Let $A_k$ be the event that the input was $k = 0, 1$, then $A_0$ and $A_0$ are a partition of the sample space of input-output pairs. Let $B_1$ be the event “receiver output was a 1.” The probability is

$$\Pr(B_1) = \Pr(B_1 | A_0) \cdot \Pr(A_0) + \Pr(B_1 | A_1) \cdot \Pr(A_1)$$

$$\Pr(B_1) = \varepsilon \cdot (1 - p) + (1 - \varepsilon) \cdot p$$

Applying Bayes’ rule, we obtain the a posteriori probabilities

$$\Pr(A_0 | B_1) = \frac{\Pr(B_1 | A_0) \cdot \Pr(A_0)}{\Pr(B_1)} = \frac{\varepsilon \cdot (1 - p)}{\varepsilon \cdot (1 - p) + (1 - \varepsilon) \cdot p}$$

$$\Pr(A_1 | B_1) = \frac{\Pr(B_1 | A_1) \cdot \Pr(A_1)}{\Pr(B_1)} = \frac{(1 - \varepsilon) \cdot p}{\varepsilon \cdot (1 - p) + (1 - \varepsilon) \cdot p}$$

For $p = 0.5$

$$\Pr(A_0 | B_1) = \frac{\Pr(B_1 | A_0) \cdot \Pr(A_0)}{\Pr(B_1)} = \frac{\varepsilon \cdot 0.5}{0.5} = \varepsilon$$

$$\Pr(A_1 | B_1) = \frac{\Pr(B_1 | A_1) \cdot \Pr(A_1)}{\Pr(B_1)} = \frac{(1 - \varepsilon) \cdot 0.5}{0.5} = 1 - \varepsilon$$


33 of 55
Example

More Resistors

<table>
<thead>
<tr>
<th></th>
<th>Bin 1</th>
<th>Bin 2</th>
<th>Bin 3</th>
<th>Bin 4</th>
<th>Bin 5</th>
<th>Bin 6</th>
<th>Subtotal</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 ohm</td>
<td>500</td>
<td>0</td>
<td>200</td>
<td>800</td>
<td>1200</td>
<td>1000</td>
<td>3700</td>
</tr>
<tr>
<td>100 ohm</td>
<td>300</td>
<td>400</td>
<td>600</td>
<td>200</td>
<td>800</td>
<td>0</td>
<td>2300</td>
</tr>
<tr>
<td>1000 ohm</td>
<td>200</td>
<td>600</td>
<td>200</td>
<td>600</td>
<td>0</td>
<td>1000</td>
<td>2600</td>
</tr>
<tr>
<td>Subtotal</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
<td>1600</td>
<td>2000</td>
<td>2000</td>
<td>8600</td>
</tr>
</tbody>
</table>

What is the probability of selecting a 10 ohm resistor from a random bin? (a priori)

A random bin … \( \Pr(Bin#) = \frac{1}{6} \)

\[
\begin{align*}
\Pr(10\Omega \mid Bin1) &= \frac{500}{1000} \\
\Pr(10\Omega \mid Bin2) &= \frac{0}{1000} \\
\Pr(10\Omega \mid Bin3) &= \frac{200}{1000} \\
\Pr(10\Omega \mid Bin4) &= \frac{800}{1600} \\
\Pr(10\Omega \mid Bin5) &= \frac{1200}{2000} \\
\Pr(10\Omega \mid Bin6) &= \frac{1000}{2000} \\
\end{align*}
\]

\[
\Pr(B) = \Pr(B \mid A_1) \cdot \Pr(A_1) + \Pr(B \mid A_2) \cdot \Pr(A_2) + \cdots + \Pr(B \mid A_n) \cdot \Pr(A_n)
\]

\[
\Pr(B) = \frac{500}{1000} \cdot \frac{1}{6} + \frac{0}{1000} \cdot \frac{1}{6} + \frac{200}{1000} \cdot \frac{1}{6} + \frac{800}{1600} \cdot \frac{1}{6} + \frac{1200}{2000} \cdot \frac{1}{6} + \frac{1000}{2000} \cdot \frac{1}{6}
\]

\[
\Pr(B) = 0.0833 + 0.0000 + 0.1250 + 0.4917 + 0.6000 + 0.5000 = 2.3333 = 0.3833
\]
Assuming a 10 ohm resistor is selected, what is the probability it came from bin 3? (a posteriori)

\[
\Pr(A_i \mid B) = \frac{\Pr(B \mid A_i) \cdot \Pr(A_i)}{\Pr(B \mid A_i) \cdot \Pr(A_i) + \Pr(B \mid A_2) \cdot \Pr(A_2) + \cdots + \Pr(B \mid A_n) \cdot \Pr(A_n)}
\]

\[
\Pr(Bin3 \mid 10\Omega) = \frac{\Pr(10\Omega \mid Bin3) \cdot \Pr(Bin3)}{\Pr(10\Omega \mid Bin1) \cdot \Pr(Bin1) + \cdots + \Pr(10\Omega \mid Bin6) \cdot \Pr(Bin6)}
\]

\[
\Pr(Bin3 \mid 10\Omega) = \frac{\frac{2}{10} \cdot \frac{1}{6}}{0.3833} = 0.0870
\]

Note: The marginal probabilities of getting a 10 ohm resistor.

\[
\Pr(10\Omega \mid Bin1) = \frac{500}{1000} = 0.50 \quad \Pr(10\Omega \mid Bin2) = \frac{0}{1000} = 0.00 \quad \Pr(10\Omega \mid Bin3) = \frac{200}{1000} = 0.20
\]

\[
\Pr(10\Omega \mid Bin4) = \frac{800}{1600} = 0.50 \quad \Pr(10\Omega \mid Bin5) = \frac{1200}{2000} = 0.60 \quad \Pr(10\Omega \mid Bin6) = \frac{1000}{2000} = 0.50
\]

\[
\Pr(Bin1 \mid 10\Omega) = 0.2174 \quad \Pr(Bin2 \mid 10\Omega) = 0.0000 \quad \Pr(Bin3 \mid 10\Omega) = 0.0870
\]

\[
\Pr(Bin4 \mid 10\Omega) = 0.2174 \quad \Pr(Bin5 \mid 10\Omega) = 0.2609 \quad \Pr(Bin6 \mid 10\Omega) = 0.2174
\]
### 2.5 Independence of Events

Two events, A and B, are independent if and only if

$$\Pr(A \cap B) = \Pr(A) \cdot \Pr(B)$$

Independence is typically assumed when there is no apparent physical mechanism by which the two events could depend on each other. For events derived from independent elemental events, their independence may not be obvious but may be able to be derived.

Note that for independent events

$$\Pr[A \mid B] = \Pr[A]$$
$$\Pr[B \mid A] = \Pr[B]$$

Independence can be extended to more than two events, for example three, A, B, and C. The conditions for independence of three events is

$$\Pr(A \cap B) = \Pr(A) \cdot \Pr(B) \quad \Pr(B \cap C) = \Pr(B) \cdot \Pr(C) \quad \Pr(A \cap C) = \Pr(A) \cdot \Pr(C)$$

$$\Pr(A \cap B \cap C) = \Pr(A) \cdot \Pr(B) \cdot \Pr(C)$$

Note that it is not sufficient to establish pair-wise independence; the entire set of equations is required.

For multiple events, every set of events from n down must be verified. This implies that $2^n - (n + 1)$ equations must be verified for n independent events. (Remember number of sets is $2^n$ including the null set. Then remove the null set and the n individual elements from the total for $2^n - (n + 1)$).

### Important Properties of Independence

**Unions**

$$\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$$

$$\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A) \cdot \Pr(B)$$

**Intersection with a Union**

$$\Pr(A \cap (B \cup C)) = \Pr(A) \cdot \Pr(B \cup C)$$
Also note:

If two events have nonzero probability and are mutually exclusive, then they cannot be independent. For suppose they were independent and mutually exclusive; then

\[ \Pr(A \cap B) = \Pr(A) \cdot \Pr(B) = 0 \]

which implies that at least one of the events must have zero probability.
Example 2.32

Two numbers $x$ and $y$ are selected at random between zero and one. Let the events $A$, $B$, and $C$ be defined as follows:

$$A = \{x > 0.5\}, \quad B = \{y > 0.5\}, \quad \text{and} \quad C = \{x > y\}$$

Are the events $A$ and $B$ independent? Are $A$ and $C$ independent?

Figure 2.13 shows the regions of the unit square that correspond to the above events.

Using Eq. (2.32a), we have

$$P(A \mid B) = \frac{\Pr(A \cap B)}{\Pr(B)} \quad \text{for} \quad \Pr(B) > 0$$

$$P(A \mid B) = \frac{\frac{1}{4}}{\frac{1}{2}} = \frac{1}{2} = P(A)$$

so events $A$ and $B$ are independent. Again we have that the “proportion” of outcomes in $S$ leading to $A$ is equal to the “proportion” in $B$ that lead to $A$.

Using Eq. (2.32b), we have

$$P(A \mid C) = \frac{\Pr(A \cap C)}{\Pr(C)} \quad \text{for} \quad \Pr(C) > 0$$

$$P(A \mid C) = \frac{\frac{3}{8}}{\frac{1}{2}} = \frac{3}{4} \neq P(A)$$

so events $A$ and $C$ are not independent. Indeed from Fig. 2.13(b) we can see that knowledge of the fact that $x$ is greater than $y$ increases the probability that $x$ is greater than 0.5.
Example 2.34

Suppose a fair coin is tossed three times and we observe the resulting sequence of heads and tails. Find the probability of the elementary events.

The sample space of this experiment is
\[ B = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}. \]

The assumption that the coin is fair means that the outcomes of a single toss are equiprobable, that is, \( P[H] = P[T] = 0.5 \). If we assume that the outcomes of the coin tosses are independent, then
\[
P[HHH] = P[H] \cdot P[H] \cdot P[H] = 0.5^3 = \frac{1}{8}
\]
and identical for all other elements of the sample space.

Example 2.35 System Reliability

A system consists of a controller and three peripheral units. The system is said to be “up” if the controller and at least two of the peripherals are functioning. Find the probability that the system is up, assuming that all components fail independently.

Define the following events: \( A \) is “controller is functioning” and \( B_i \) is “peripheral \( i \) is functioning” where \( i = 1, 2, 3 \). The event \( F \), “two or more peripheral units are functioning,” occurs if all three units are functioning or if exactly two units are functioning. Thus
\[
F = (B_1 \cap B_2 \cap B_3) \cup (B_1^C \cap B_2 \cap B_3) \cup (B_1 \cap B_2^C \cap B_3) \cup (B_1 \cap B_2 \cap B_3^C)
\]

Note that the events in the above union are mutually exclusive. Thus
\[
P[F] = P[B_1] \cdot P[B_2] \cdot P[B_3] + P[B_1^C] \cdot P[B_2] \cdot P[B_3]
+ P[B_1] \cdot P[B_2^C] \cdot P[B_3] + P[B_1] \cdot P[B_2] \cdot P[B_3^C]
\]

We assume that each peripheral fails with probability \( a \), so that \( P[B_i] = 1 - a \) and \( P[B_i^C] = a \). Then,
\[
P[F] = (1 - a)^3 + 3 \cdot (1 - a)^2 \cdot a
\]
\[
P[F] = (1 - 3 \cdot a + 3 \cdot a^2 - a^3) + (3 \cdot a - 6 \cdot a^2 + 3 \cdot a^3)
\]
\[
P[F] = (1 - 3 \cdot a^2 + 2 \cdot a^3)
\]
Let \( a = 10\% \) then all three peripherals are functioning \( (1 - a)^3 = 72.9\% \) of the time and two are functioning and one is “down” \( 3 \cdot (1 - a)^2 \cdot a = 24.3\% \) of the time. Thus two or more peripherals are functioning 97.2\% of the time.

\[
P[F] = 0.729 + 0.243 = 0.972
\]

The event “system is up” is then \( A \cap F \). If we assume that the controller fails with probability \( p \), then

\[
P[A \cap F] = P[A] \cdot P[F] = (1 - p) \cdot (1 - 3 \cdot a^2 + 2 \cdot a^3)
\]

Suppose that the controller is not very reliable, say \( p = 20\% \) then the system is up only 77.8\% of the time, mostly because of controller failures.

\[
P[A \cap F] = P[A] \cdot P[F] = 0.80 \cdot 0.972 = 0.7776
\]

If a redundant controller is added so that the system is up when at least one of the controllers is functional, the up time is significantly increased.

\[
R = (A_1 \cap A_2) \cup (A_1^c \cap A_2) \cup (A_1 \cap A_2^c)
\]

\[
P[R] = P[A_1] \cdot P[A_2] + P[A_1^c] \cdot P[A_2] + P[A_1] \cdot P[A_2^c]
\]

\[
P[R] = (1 - p)^2 + 2 \cdot (1 - p) \cdot p = (1 - 2 \cdot p + p^2) + 2 \cdot p - 2 \cdot p^2
\]

\[
P[R] = 1 - p^2 = 0.96
\]

Therefore

\[
P[R \cap F] = P[R] \cdot P[F] = (1 - p^2) \cdot (1 - 3 \cdot a^2 + 2 \cdot a^3)
\]

\[
P[R \cap F] = P[R] \cdot P[F] = 0.96 \cdot 0.972 = 0.933
\]

Adding redundancy in a system significantly improves the up time and overall reliability. This is an important application of probability.

Note:

- 99\% reliable computers are down 3.65 days per year
- 99.9\% reliable computers are down 8.76 hours per year
- 99.99\% reliable computers are down approximately 53 minutes per year
- which system would you want your bank to have?
2.6 Sequential Experiments

Many random experiments can be viewed as sequential experiments that consist of a sequence of simpler subexperiments. These subexperiments may or may not be independent. In this section we discuss methods for obtaining the probabilities of events in sequential experiments.

2.6.1 Sequences of Independent Experiments

We can usually determine, because of physical considerations, when the subexperiments are independent, in the sense that the outcome of any given subexperiment cannot affect the outcomes of the other subexperiments. Let $A_1, A_2, \ldots, A_n$ be events such that $A_k$ concerns only the outcome of the $k$th subexperiment. If the subexperiments are independent, then it is reasonable to assume that the above events are independent. Thus,

$$P[A_1 \cap A_2 \cap \cdots \cap A_n] = P[A_1] \cdot P[A_2] \cdots P[A_n]$$

Example 2.36

Suppose that 10 numbers are selected at random from the interval $[0, 1]$. Find the probability that the first 5 numbers are less than $1/4$ and the last 5 numbers are greater than $1/2$. Then the events of interest are

$$A_k = \{x_k : x_k < \frac{1}{4}\}, \text{ for } k = 1,2,3,4,5$$

$$A_k = \{x_k : x_k > \frac{1}{2}\}, \text{ for } k = 6,7,8,9,10$$

For a sequence of independent trials

$$P[A_1 \cap A_2 \cap \cdots \cap A_n] = P[A_1] \cdot P[A_2] \cdots P[A_n]$$

$$P[A_1 \cap A_2 \cap \cdots \cap A_n] = \left(\frac{1}{4}\right)^5 \cdot \left(\frac{1}{2}\right)^5 = \frac{1}{2^{15}} = \frac{1}{32768}$$
2.6.2 The Binomial Probability Law

A Bernoulli trial involves performing an experiment once and noting whether a particular event A occurs. The outcome of the Bernoulli trial is said to be a “success” if A occurs and a “failure” otherwise. In this section we are interested in finding the probability of k successes in n independent repetitions of a Bernoulli trial.

We can view the outcome of a single Bernoulli trial as the outcome of a toss of a coin for which the probability of heads (success) is $p = P[A]$. The probability of k successes in n Bernoulli trials is then equal to the probability of k heads in n tosses of the coin.

$$P[A_1 \cap A_2 \cap \cdots \cap A_n] = (p)^k \cdot (1 - p)^{n-k}$$

Example 2.37

Suppose that a coin is tossed three times. The sample space of this experiment is $B = \{HHH, HHT, HTH, THH, HTT, TTH, THT, TTT\}$.

If we assume that the tosses are independent and the probability of heads is $p$, then the probability for the sequences of heads and tails is

$$P[HHH] = P[H] \cdot P[H] \cdot P[H] = p^3$$
$$P[HHT] = P[H] \cdot P[H] \cdot P[T] = p^2 \cdot (1 - p)$$
$$P[HTH] = P[H] \cdot P[T] \cdot P[H] = p^2 \cdot (1 - p)$$
$$P[THH] = P[T] \cdot P[H] \cdot P[H] = p^2 \cdot (1 - p)$$
$$P[HTT] = P[H] \cdot P[T] \cdot P[T] = p \cdot (1 - p)^2$$
$$P[THT] = P[T] \cdot P[H] \cdot P[T] = p \cdot (1 - p)^2$$
$$P[TTH] = P[T] \cdot P[T] \cdot P[H] = p \cdot (1 - p)^2$$
$$P[TTT] = P[T] \cdot P[T] \cdot P[T] = (1 - p)^3$$

where we used the fact that the tosses are independent. Let $k$ be the number of heads in three trials, then

$$P[k = 0] = P[TTT] = (1 - p)^3$$
$$P[k = 1] = P[HHT, THT, TTH] = p \cdot (1 - p)^2$$
$$P[k = 2] = P[THHT, HTH, THH] = p^2 \cdot (1 - p)$$
$$P[k = 3] = P[HHH] = p^3$$

The number of possible permutations in Bernoulli trial can be computed as follows.
Binomial Probability Law

Let $k$ be the number of successes in $n$ independent Bernoulli trials, then the probabilities of $k$ are given by

$$p_n(k) = \binom{n}{k} \cdot (p)^k \cdot (1-p)^{n-k}, \text{ for } k = 0, 1, \ldots, n$$

where $p_n(k)$ is the probability of $k$ successes in $n$ trials, and

$$C_k^n = \binom{n}{k} = \frac{n!}{k!(n-k)!}$$

is the binomial coefficient. Note that by definition $0! = 1$.

As a note and why it is referred to as the binomial theorem

$$(a + b)^n = \sum_{k=0}^{n} \binom{n}{k} \cdot a^k \cdot b^{n-k}$$

Note that if we let $a = p$ and $b = 1-p$ in, we then obtain

$$(1)^n = 1 = \sum_{k=0}^{n} \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}$$

which proves that the sum of binomial probabilities is 1.

Example 2.39

Let $k$ be the number of active (nonsilent) speakers in a group of eight noninteracting (i.e., independent) speakers. Suppose that a speaker is active with probability $1/3$. Find the probability that the number of active speakers is greater than six.

For $i=1, 2, 3, 4, 5, 6, 7, 8$ let $A_i$ denote the event “ith speaker is active.” The number of active speakers is then the number of successes in eight Bernoulli trials with $p = 1/3$. Thus the probability that more than six speakers are active is

$$P[k = 7 \cup k = 8] = p_7(7) + p_8(7)$$

$$P[k = 7 \cup k = 8] = \binom{8}{7} \cdot (p)^7 \cdot (1-p)^1 + \binom{8}{8} \cdot (p)^8 \cdot (1-p)^0$$

$$P[k = 7 \cup k = 8] = 8 \cdot \left(\frac{1}{3}\right)^7 \cdot \frac{2}{3} + 1 \cdot \left(\frac{1}{3}\right)^8 = \frac{17}{6561} = 0.002591$$
Example 2.40 Error Correction Coding

A communication system transmits binary information over a channel that introduces random bit errors with probability $\varepsilon = 10^{-3}$. The transmitter transmits each information bit three times (k=3, 1 code), and a decoder takes a majority vote of the received bits to decide on what the transmitted bit was. Find the probability that the receiver will make an incorrect decision.

The receiver can correct a single error, but it will make the wrong decision if the channel introduces two or more errors. If we view each transmission as a Bernoulli trial in which a “success” corresponds to the introduction of an error, then the probability of two or more errors in three Bernoulli trials is

$$P[k \geq 2] = p_3(2) + p_3(3)$$

$$P[k \geq 2] = \binom{3}{2} \cdot (p)^2 \cdot (1-p)^1 + \binom{3}{3} \cdot (p)^3 \cdot (1-p)^0$$

$$P[k \geq 2] = 3 \cdot (0.001)^2 \cdot (0.999) + 1 \cdot (0.001)^3 = 2.998 \cdot 10^{-4}$$

This is a simple block code substitution that has been widely used in communications.
2.6.3 The Multinomial Probability Law

The binomial probability law can be generalized to the case where we note the occurrence of more than one event. Let $B_1, B_2, \ldots, B_M$ be a partition of the sample space $S$ of some random experiment and let $P[B_j]=p_j$. The events are mutually exclusive, so

$$p_1 + p_2 + \cdots + p_M = 1$$

Suppose that $n$ independent repetitions of the experiment are performed. Let $k_j$ be the number of times event $B_j$ occurs, then the vector $(k_1, k_2, \ldots, k_M)$ specifies the number of times each of the events $B_j$ occurs. The probability of the vector $(k_1, k_2, \ldots, k_M)$ satisfies the multinomial probability law:

$$P[(k_1, k_2, \ldots, k_M)] = \frac{n!}{k_1! k_2! \cdots k_M!} p_1^{k_1} p_2^{k_2} \cdots p_M^{k_M}$$

where $k_1 + k_2 + \cdots + k_M = n$.

Note that the binomial law is the $M=2$ case of the multinomial law.

Example 2.41

A dart is thrown nine times at a target consisting of three areas. Each throw has a probability of 0.2, 0.3, and 0.5 of landing in areas 1, 2, and 3, respectively. Find the probability that the dart lands exactly three times in each of the areas.

This experiment consists of nine independent repetitions of a subexperiment that has three possible outcomes. The probability for the number of occurrences of each outcome is given by the multinomial probabilities with parameters $n=9$, $p_1=0.2$, $p_2=0.3$, and $p_3=0.5$. Therefore:

$$P[(3,3,3)] = \frac{9!}{3!3!3!} \cdot (0.2)^3 \cdot (0.3)^3 \cdot (0.5)^3$$

$$P[(3,3,3)] = 1680 \cdot (0.008) \cdot (0.027) \cdot (0.125) = 0.4536$$
2.6.4 The Geometric Probability Law

Consider a sequential experiment in which we repeat independent Bernoulli trials until the occurrence of the first success. Let the outcome of this experiment be $m$, the number of trials carried out until the occurrence of the first success. The sample space for this experiment is the set of positive integers. The probability, $p(m)$, that $m$ trials are required is found by noting that this can only happen if the first $m-1$ trials result in failures and the $m$th trial in success.

$$p(m) = P[A_1^c \cap A_2^c \cap \cdots \cap A_{m-1}^c \cap A_m] = (1 - p)^{m-1} \cdot p$$

where $A_i$ is the event “success in $i$th trial.” The probability assignment specified is called the geometric probability law.

The probabilities sum to 1 if all trials $m=1$ to infinity are performed. That is

$$\sum_{m=1}^{\infty} p(m) = p \cdot \sum_{m=1}^{\infty} (1 - p)^{m-1} = p \cdot \frac{1}{1-(1-p)} = \frac{p}{p} = 1$$

where we have used the formula for the summation of a geometric series.

The probability that more than $K$ trials are required before a success occurs has a simple form:

$$P\{[p(m): m > K]\} = \sum_{m=K+1}^{\infty} p(m) = p \cdot \sum_{m=K+1}^{\infty} (1 - p)^{m-1} = p \cdot \frac{(1-p)^{K}}{1-(1-p)} = (1 - p)^{K}$$

Example 2.43 Error Control by Retransmission

Computer A sends a message to computer B over an unreliable radio link. The message is encoded so that B can detect when errors have been introduced into the message during transmission. If B detects an error, it requests A to retransmit it. If the probability of a message transmission error is $q = 0.1$, what is the probability that a message needs to be transmitted more than two times?

Each transmission of a message is a Bernoulli trial with probability of success $p=1-q$. The Bernoulli trials are repeated until the first success (error-free transmission). The probability that more than two transmissions are required is

$$P\{[p(m): m > 2]\} = \sum_{m=3}^{\infty} p(m) = q^2 = (1 - p)^2$$

$$P\{[p(m): m > 2]\} = (0.1)^2 = 10^{-2}$$
2.6.5 Sequences of Dependent Experiments

In this section we consider a sequence or “chain” of subexperiments in which the outcome of a given subexperiment determines which subexperiment is performed next. We first give a simple example of such an experiment and show how diagrams can be used to specify the sample space.

Example 2.44

A sequential experiment involves repeatedly drawing a ball from one of two urns, noting the number on the ball, and replacing the ball in its urn. Urn 0 contains a ball with the number 1 and two balls with the number 0, and urn 1 contains five balls with the number 1 and one ball with the number 0.

The urn from which the first draw is made is selected at random by flipping a fair coin. Urn 0 is used if the outcome is heads and urn 1 if the outcome is tails. Thereafter the urn used in a subexperiment corresponds to the number on the ball selected in the previous subexperiment.

The sample space of this experiment consists of sequences of 0s and 1s. Each possible sequence corresponds to a path through the “trellis” diagram shown in Fig. 2.15(a).

The nodes in the diagram denote the urn used in the nth subexperiment, and the labels in the branches denote the outcome of a subexperiment. Thus the path 0011 corresponds to the sequence: The coin toss was heads so the first draw was from urn 0; the outcome of the first draw was 0, so the second draw was from urn 0; the outcome of the second draw was 1, so the third draw was from urn 1; and the outcome from the third draw was 1, so the fourth draw is from urn 1.

Now suppose that we want to compute the probability of a particular sequence of outcomes, say s0,s1,s2. Denote this probability by \( P\{s_0\} \cap \{s_1\} \cap \{s_2\} \). Let \( A = \{s_2\} \) and \( B = \{s_0\} \cap \{s_1\} \), then since \( P[A \cap B] = P[A | B] \cdot P[B] \) we have

\[
P\{s_0\} \cap \{s_1\} \cap \{s_2\} = P\{s_2\} \cdot P\{s_0\} \cap \{s_1\} \cdot P[B]
\]

but
\[ P[\{s_0\} \cap \{s_1\}] = P[\{s_1\} | \{s_0\}] \cdot P[\{s_0\}] \]

therefore
\[ P[\{s_0\} \cap \{s_1\} \cap \{s_2\}] = P[\{s_2\} | \{s_0\} \cap \{s_1\}] \cdot P[\{s_1\} | \{s_0\}] \cdot P[\{s_0\}] \]

Now note that in the above urn example the probability \( P[\{s_n\} | \{s_0\} \cap \cdots \cap \{s_{n-1}\}] \) depends only on \( \{s_{n-1}\} \) since the most recent outcome determines which subexperiment is performed:
\[ P[\{s_n\} | \{s_0\} \cap \cdots \cap \{s_{n-1}\}] = P[\{s_n\} | \{s_{n-1}\}] \]

Therefore for the sequence of interest we have that
\[ P[\{s_0\} \cap \{s_1\} \cap \{s_2\}] = P[\{s_2\} | \{s_1\}] \cdot P[\{s_1\} | \{s_0\}] \cdot P[\{s_0\}] \]

Sequential experiments that satisfy this equation are called Markov chains. For these experiments, the probability of a sequence is given by
\[ P[s_0, s_1, \ldots, s_n] = P[s_n | s_{n-1}] \cdot P[s_{n-1} | s_{n-2}] \cdots P[s_1 | s_0] \cdot P[s_0] \]

where we have simplified notation by omitting braces. Thus the probability of the sequence \( s_0, s_1, \ldots, s_n \) is given by the product of the probability of the first outcome \( s_0 \) and the probabilities of all subsequent transitions, \( s_0 \) to \( s_1 \), \( s_1 \) to \( s_2 \) and so on.

Chapter 11 deals with Markov chains.

Aside: when decoding a sequence of symbols, a trellis is formed of the possible path of the previous symbols leading to the current symbol. As each symbol is received, the path with the smallest “error” is selected as the demodulated symbol sequence. (Note: a grossly simplified explanation.)
Example 2.45

Find the probability of the sequence 0011 for the urn experiment introduced in Example 2.44.

Recall that urn 0 contains two balls with label 0 and one ball with label 1, and that urn 1 contains five balls with label 1 and one ball with label 0. We can readily compute the probabilities of sequences of outcomes by labeling the branches in the trellis diagram with the probability of the corresponding transition as shown in Fig. 2.15(b).

Thus the probability of the sequence 0011 is given by

\[ P[0011] = P[1|1] \cdot P[1|0] \cdot P[0|0] \cdot P[0] \]

where the initial and transitional probabilities are given by

\[ P[0] = \frac{1}{2} \quad \text{and} \quad P[1] = \frac{1}{2} \]

\[ P[1|0] = \frac{1}{3} \quad \text{and} \quad P[0|0] = \frac{2}{3} \]

\[ P[1|1] = \frac{5}{6} \quad \text{and} \quad P[0|1] = \frac{1}{6} \]

Therefore

\[ P[0011] = \frac{5}{6} \cdot \frac{1}{3} \cdot \frac{2}{3} \cdot \frac{1}{2} = \frac{10}{108} \]

Applications

The two-urn model has been used to model the “bursty” behavior of the voice packets generated by a single speaker where bursts of active packets are separated by relatively long periods of silence. The model has also been used for the sequence of black and white dots that result from scanning a black and white image line by line.
2.7 *Synthesizing Randomness: Random Number Generators*

This section introduces the basic method for generating sequences of “random” numbers using a computer. Any computer simulation of a system that involves randomness must include a method for generating sequences of random numbers. These random numbers must satisfy long-term average properties of the processes they are simulating.

2.7.1 Pseudo-Random Number Generation

The preferred approach for the computer generation of random numbers involves the use of recursive formulas that can be implemented easily and quickly. These pseudorandom number generators produce a sequence of numbers that appear to be random but that in fact repeat after a very long period. The currently preferred pseudo-random number generator is the so-called Mersenne Twister, which is based on a matrix linear recurrence over a binary field.

From MATLAB

```matlab
>> help rand
rand Uniformly distributed pseudorandom numbers.
R = rand(N) returns an N-by-N matrix containing pseudorandom values drawn from the standard uniform distribution on the open interval (0,1). rand(M,N)
or rand([M,N]) returns an M-by-N matrix. rand(M,N,P,...) or rand([M,N,P,...]) returns an M-by-N-by-P-by-... array. rand returns a scalar. rand(SIZE(A)) returns an array the same size as A.
```

```matlab
>> help randi
randi Pseudorandom integers from a uniform discrete distribution.
R = randi(IMAX,N) returns an N-by-N matrix containing pseudorandom integer values drawn from the discrete uniform distribution on 1:IMAX. randi(IMAX,M,N) or randi(IMAX,[M,N]) returns an M-by-N matrix. randi(IMAX,M,N,P,...) or randi(IMAX,[M,N,P,...]) returns an M-by-N-by-P-by-... array.
randi(IMAX) returns a scalar. randi(IMAX,SIZE(A)) returns an array the same size as A.
```

```matlab
>> help randn
randn Normally distributed pseudorandom numbers.
R = randn(N) returns an N-by-N matrix containing pseudorandom values drawn from the standard normal distribution. randn(M,N) or randn([M,N]) returns an M-by-N matrix. randn(M,N,P,...) or randn([M,N,P,...]) returns an M-by-N-by-P-by-... array. randn returns a scalar. randn(SIZE(A)) returns an array the same size as A.
```
Example 2.46

```matlab
% Example 2.46
clear
close all

X = rand(10000,1);
K = 0.005:0.01:0.995;
figure
hist(X,K)
title('rand histogram')

figure
plot(K, empirical_cdf(K,X))
title('rand CDF')
```

Example 2.47 Bernoulli Trials and Binomial Probabilities

Generate the outcomes of 100 replications of a random experiment that counts the number of successes in 16 Bernoulli trials with probability of success $\frac{1}{2}$. Plot the histogram of the outcomes in the 100 experiments and compare to the binomial probabilities with $n=16$ and $p=1/2$.

```matlab
% Example 2.47
% clear;
close all;
num_trials = 16;
test_size = 10000;
p = 0.5;

X = rand(test_size,num_trials)<p; % Generate 100 rows of 16 Bernoulli trials with
```

Y=sum(X,Z); % Add the results of each row to obtain the
number of % successes in each experiment. Y contains the
100 % outcomes.
K=(0:num_trials);
Z=hist(Y,K)/test_size; % Find the relative frequencies of the
outcomes in Y.

figure
bar(K,Z) % Produce a bar graph of the relative
frequencies.
hold on % Retains the graph for next command.

ii=0;
pdf=zeros(1,num_trials);
pdftemp = binomial_coef(num_trials,K).*p^K.*(1-p)^(num_trials-K);

for kk = K
    ii=ii+1;
pdf(ii)=nchoosek(num_trials,kk)*p^kk*(1-p)^(num_trials-kk);
end
stem(K,pdf,'r') % Plot the binomial probabilities along
% with the corresponding relative frequencies.
axis_temp = axis;
axis([0 num_trials 0 1.1*max(pdf)])

Notes on Octave: see Wikipedia under GNU Octave

2.8 *Fine Points: Event Classes

2.9 *Fine Points: Probabilities of Sequences of Events

2.9.1 The Borel Field of Events

2.9.2 Continuity of Probability
SUMMARY

- A probability model is specified by identifying the sample space S, the event class of interest, and an initial probability assignment, a “probability law,” from which the probability of all events can be computed.

- The sample space S specifies the set of all possible outcomes. If it has a finite or countable number of elements, S is discrete; S is continuous otherwise.

- Events are subsets of S that result from specifying conditions that are of interest in the particular experiment. When S is discrete, events consist of the union of elementary events. When S is continuous, events consist of the union or intersection of intervals in the real line.

- The axioms of probability specify a set of properties that must be satisfied by the probabilities of events. The corollaries that follow from the axioms provide rules for computing the probabilities of events in terms of the probabilities of other related events.

- An initial probability assignment that specifies the probability of certain events must be determined as part of the modeling. If S is discrete, it suffices to specify the probabilities of the elementary events. If S is continuous, it suffices to specify the probabilities of intervals or of semi-infinite intervals.

- Combinatorial formulas are used to evaluate probabilities in experiments that have an equiprobable, finite number of outcomes.

- A conditional probability quantifies the effect of partial knowledge about the outcome of an experiment on the probabilities of events. It is particularly useful in sequential experiments where the outcomes of subexperiments constitute the “partial knowledge.”

- Bayes’ rule gives the a-posteriori probability of an event given that another event has been observed. It can be used to synthesize decision rules that attempt to determine the most probable “cause” in light of an observation.

- Two events are independent if knowledge of the occurrence of one does not alter the probability of the other. Two experiments are independent if all of their respective events are independent. The notion of independence is useful for computing probabilities in experiments that involve noninteracting subexperiments

- Many experiments can be viewed as consisting of a sequence of independent subexperiments. In this chapter we presented the binomial, the multinomial, and the geometric probability laws as models that arise in this context.

- A Markov chain consists of a sequence of subexperiments in which the outcome of a subexperiment determines which subexperiment is performed next. The probability of a sequence of outcomes in a Markov chain is given by the product of the probability of the first outcome and the probabilities of all subsequent transitions.

- Computer simulation models use recursive equations to generate sequences of pseudo-random numbers.
CHECKLIST OF IMPORTANT TERMS

- Axioms of Probability
- Bayes’ rule
- Bernoulli trial
- Binomial coefficient
- Binomial theorem
- Certain event
- Conditional probability
- Continuous sample space
- Discrete sample space
- Elementary event
- Event
- Event class
- Independent events
- Independent experiments
- Initial probability assignment
- Markov chain
- Mutually exclusive events
- Null event
- Outcome
- Partition
- Probability law
- Sample space
- Set operations
- Theorem on total probability
- Tree diagram