11.3 Randomized Complete Block Designs 528
- 11.3.1 Model and Distribution Assumptions 530
- 11.3.2 Treatment Contrasts 532
- 11.3.3 Simultaneous Estimation and Testing 537
- 11.3.4 Partitioning Sums of Squares 539
- 11.3.5 Implications of Random Blocking 541

Exercises 543

Miscellanea 551

12 Linear Regression 554

12.1 Introduction 554

12.2 Simple Linear Regression 557
- 12.2.1 Least Squares: A Mathematical Solution 557
- 12.2.2 Best Linear Unbiased Estimation: A Statistical Solution 560
- 12.2.3 Models and Distribution Assumptions 564
- 12.2.4 Estimation and Testing with Normal Errors 567
- 12.2.5 Estimation and Prediction at a Specified $x = x_0$ 574
- 12.2.6 Simultaneous Estimation and Confidence Bands 577

12.3 Regression with Errors in Variables 581
- 12.3.1 Functional and Structural Relationships 583
- 12.3.2 A Least Squares Solution 584
- 12.3.3 Maximum Likelihood Estimation 586
- 12.3.4 Confidence Sets 592

Exercises 595

Miscellanea 603

Tables 607
- 1 Normal 608
- 2 Student's t 609
- 3 Chi Squared 610
- 4 Snedecor's F 612

Distributions 624
- 1 Catalog 624
- 2 Relationships 630

References 631

Author Index 641

Subject Index 643

1 Probability Theory

"Take time to consider. The smallest point may be the most essential."

Sherlock Holmes
The Adventure of the Red Circle

The subject of probability theory is the foundation upon which all of statistics is built, providing a means for modeling populations, experiments, or almost anything else that could be considered a random phenomenon. Through these models, statisticians are able to draw inferences about populations, inferences based on examination of only a part of the whole.

The theory of probability has a long and rich history, dating back at least to the seventeenth century when, at the request of their friend, the Chevalier de Meré, Pascal and Fermat developed a mathematical formulation of gambling odds.

The aim of this chapter is not to give a thorough introduction to probability theory; such an attempt would be foolhardy in so short a space. Rather, we attempt to outline some of the basic ideas of probability theory that are fundamental to the study of statistics.

Just as statistics builds upon the foundation of probability theory, probability theory in turn builds upon set theory, which is where we begin.

1.1 Set Theory

One of the main objectives of a statistician is to draw conclusions about a population of objects by conducting an experiment. The first step in this endeavor is to identify the possible outcomes or, in statistical terminology, the sample space.

DEFINITION 1.1.1: The set, S, of all possible outcomes of a particular experiment is called the *sample space* for the experiment.

If the experiment consists of tossing a coin, the sample space contains two outcomes, heads and tails; thus,

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

If, on the other hand, the experiment consists of observing the reported SAT scores of randomly selected students at a certain university, the sample space would be the set of positive integers between 200 and 800 that are multiples of ten—that is, $S = \{200, 210, 220, \ldots, 780, 790, 800\}$. Finally, consider an experiment where the observation is reaction time to a certain stimulus. Here, the sample space would consist of all positive numbers, that is, $S = (0, \infty)$.

We can classify sample spaces into two types according to the number of elements they contain. Sample spaces can be either countable or uncountable; if the elements of a sample space can be put into 1–1 correspondence with a subset of the integers, the sample space is countable. Of course, if the sample space contains only a finite number of elements, it is countable. Thus, the coin-toss and SAT score sample spaces are both countable (in fact, finite), whereas the reaction time sample space is uncountable, since the positive real numbers cannot be put into 1–1 correspondence with the integers. If, however, we measured reaction time to the nearest second, then the sample space would be (in seconds) $S = \{0, 1, 2, 3, \ldots\}$, which is then countable.

This distinction between countable and uncountable sample spaces is important only in that it dictates the way in which probabilities can be assigned. For the most part, this causes no problems, although the mathematical treatment of the situations is different. On a philosophical level, it might be argued that there can only be countable sample spaces, since measurements cannot be made with infinite accuracy. (A sample space consisting of, say, all ten-digit numbers is a countable sample space.) While in practice this is true, probabilistic and statistical methods associated with uncountable sample spaces are, in general, less cumbersome than those for countable sample spaces, and provide a close approximation to the true (countable) situation.

Once the sample space has been defined, we are in a position to consider collections of possible outcomes of an experiment.

DEFINITION 1.1.2: An *event* is any collection of possible outcomes of an experiment, that is, any subset of S (including S itself).

Let A be an event, a subset of S. We say the event A occurs if the outcome of the experiment is in the set A. When speaking of probabilities, we generally speak of the probability of an event, rather than a set. But we may use the terms interchangeably.

We first need to define formally the following two relationships, which allow us to order and equate sets:

$$A \subset B \Leftrightarrow x \in A \Rightarrow x \in B; \qquad \text{(containment)}$$

$$A = B \Leftrightarrow A \subset B \text{ and } B \subset A. \qquad \text{(equality)}$$

Given any two events (or sets) A and B, we have the following elementary set operations:

Union: The union of A and B, written $A \cup B$, is the set of elements that belong to either A or B or both:

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

Intersection: The intersection of A and B, written $A \cap B$, is the set of elements that belong to both A and B:

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

Complementation: The complement of A, written A^c, is the set of all elements that are not in A:

$$A^c = \{x : x \notin A\}.$$

Example 1.1.1: Consider the experiment of selecting a card at random from a standard deck, and noting its suit: clubs (C), diamonds (D), hearts (H), or spades (S). The sample space is

$$S = \{\text{C}, \text{D}, \text{H}, \text{S}\},$$

and some possible events are

$$A = \{\text{C}, \text{D}\} \text{ and } B = \{\text{D}, \text{H}, \text{S}\}.$$

From these events we can form

$$A \cup B = \{\text{C}, \text{D}, \text{H}, \text{S}\}, A \cap B = \{\text{D}\}, \text{ and } A^c = \{\text{H}, \text{S}\}.$$

Furthermore, notice that $A \cup B = S$ (the event S), and $(A \cup B)^c = \emptyset$, where $\emptyset$ denotes the *empty set* (the set consisting of no elements). ||

The elementary set operations can be combined, somewhat akin to the way addition and multiplication can be combined. As long as we are careful, we can treat sets as if they were numbers. We can now state the following useful properties of set operations.

THEOREM 1.1.1: For any three events A, B, and C defined on a sample space S,

1. Commutativity $\quad A \cup B = B \cup A,$ $\quad A \cap B = B \cap A;$
2. Associativity $\quad A \cup (B \cup C) = (A \cup B) \cup C,$ $\quad A \cap (B \cap C) = (A \cap B) \cap C;$
3. Distributive Laws $\quad A \cap (B \cup C) = (A \cap B) \cup (A \cap C),$ $\quad A \cup (B \cap C) = (A \cup B) \cap (A \cup C);$
4. DeMorgan's Laws $\quad (A \cup B)^c = A^c \cap B^c,$ $\quad (A \cap B)^c = A^c \cup B^c.$

Proof: The proof of much of this theorem is left as Exercise 1.3. Also, Exercises 1.4 and 1.5 generalize the theorem. To illustrate the technique, however, we will prove the Distributive Law

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C).$$

(You might be familiar with the use of Venn diagrams to "prove" theorems in set theory. We caution that although Venn diagrams are sometimes helpful in visualizing a situation, they do not constitute a formal proof.) To prove that two sets are equal, it must be demonstrated that each set contains the other. Formally, then

$$A \cap (B \cup C) = \{x \in S : x \in A \text{ and } x \in (B \cup C)\};$$
$$(A \cap B) \cup (A \cap C) = \{x \in S : x \in (A \cap B) \text{ or } x \in (A \cap C)\}.$$

We first show that $A \cap (B \cup C) \subset (A \cap B) \cup (A \cap C)$. Let $x \in (A \cap (B \cup C))$. By the definition of intersection, it must be that $x \in (B \cup C)$, that is, either $x \in B$ or $x \in C$. Since x also must be in A, we have that either $x \in (A \cap B)$ or $x \in (A \cap C)$; therefore,

$$x \in ((A \cap B) \cup (A \cap C)),$$

and the containment is established.

Now assume $x \in ((A \cap B) \cup (A \cap C))$. This implies that $x \in (A \cap B)$ or $x \in (A \cap C)$. If $x \in (A \cap B)$ then x is in both A and B. Since $x \in B, x \in (B \cup C)$ and thus $x \in (A \cap (B \cup C))$. If, on the other hand, $x \in (A \cap C)$, the argument is similar, and we again conclude that $x \in (A \cap (B \cup C))$. Thus, we have established $(A \cap B) \cup (A \cap C) \subset A \cap (B \cup C)$, showing containment in the other direction and, hence, proving the Distributive Law. □

The operations of union and intersection can be extended to infinite collections of sets as well. If $A_1, A_2, A_3, \ldots$ is a collection of sets, all defined on a sample space S, then

$$\bigcup_{i=1}^{\infty} A_i = \{x \in S : x \in A_i \text{ for some } i\},$$
$$\bigcap_{i=1}^{\infty} A_i = \{x \in S : x \in A_i \text{ for all } i\}.$$

For example, let $S = (0, 1]$ and define $A_i = [(1/i), 1]$. Then

$$\bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{\infty} [(1/i), 1] = \{x \in (0, 1] : x \in [(1/i), 1] \text{ for some } i\}$$
$$= \{x \in (0, 1]\} = (0, 1];$$

$$\bigcap_{i=1}^{\infty} A_i = \bigcap_{i=1}^{\infty} [(1/i), 1] = \{x \in (0, 1] : x \in [(1/i), 1] \text{ for all } i\}$$
$$= \{x \in (0, 1] : x \in [1, 1]\}$$
$$= \{1\} \qquad \text{(the point 1).}$$

It is also possible to define unions and intersections over uncountable collections of sets. If Γ is an index set (a set of elements to be used as indices) then

$$\bigcup_{a \in \Gamma} A_a = \{x \in S : x \in A_a \text{ for some } a\},$$
$$\bigcap_{a \in \Gamma} A_a = \{x \in S : x \in A_a \text{ for all } a\}.$$

If, for example, we take $\Gamma = \{\text{all positive real numbers}\}$ and $A_a = (0, a]$, then $\cup_{a \in \Gamma} A_a = (0, \infty)$ is an uncountable union. While uncountable unions and intersections do not play a major role in statistics, they sometimes provide a useful mechanism for obtaining an answer (see Section 8.2.4).

Finally, we discuss the idea of a partition of the sample space.

DEFINITION 1.1.3: Two events A and B are *disjoint* (or *mutually exclusive*) if $A \cap B = \emptyset$. The events $A_1, A_2, \ldots$ are *pairwise disjoint* (or *mutually exclusive*) if $A_i \cap A_j = \emptyset$ for all $i \neq j$.

Disjoint sets are sets with no points in common. If we draw a Venn diagram for two disjoint sets, the sets do not overlap. The collection

$$A_i = [i, i+1), \quad i = 0, 1, 2, \ldots$$

consists of pairwise disjoint sets. Note further that $\cup_{i=0}^{\infty} A_i = [0, \infty)$.

DEFINITION 1.1.4: If $A_1, A_2, \ldots$ are pairwise disjoint and $\cup_{i=1}^{\infty} A_i = S$, then the collection $A_1, A_2, \ldots$ forms a *partition* of S.

The sets $A_i = [i, i+1)$ form a partition of $[0, \infty)$. In general, partitions are very useful, allowing us to divide the sample space into small, nonoverlapping pieces.

1.2 Probability Theory

When an experiment is performed, the realization of the experiment is an outcome in the sample space. If the experiment is performed a number of times, different outcomes may occur each time or some outcomes may repeat. This "frequency of occurrence" of an outcome can be thought of as a probability. More probable outcomes occur more frequently. If the outcomes of an experiment can be described probabilistically, we are on our way to analyzing the experiment statistically.

In this section we describe some of the basics of probability theory. We do not define probabilities in terms of frequencies but instead take the mathematically simpler axiomatic approach. As will be seen, the axiomatic approach is not concerned with the interpretations of probabilities, but is concerned only that the probabilities are defined by a function satisfying the axioms. Interpretations of the probabilities are quite another matter. The "frequency of occurrence" of an event is one example of a particular interpretation of probability. Another possible interpretation is a subjective one, where rather than thinking of probability as frequency, we can think of it as a belief in the chance of an event occurring.

1.2.1 Axiomatic Foundations

For each event A in the sample space S we want to associate with A a number between zero and one which will be called the probability of A, denoted by $P(A)$. It would seem natural to define the domain of P (the set where the arguments of the function $P(\cdot)$ are defined) as all subsets of S; that is, for each $A \subset S$ we define $P(A)$ as the probability that A occurs. Unfortunately, matters are not that simple. There are some technical difficulties to overcome. We will not dwell on these technicalities; although they are of importance, they are usually of more interest to probabilists than to statisticians. However, a firm understanding of statistics requires at least a passing familiarity with the following.

DEFINITION 1.2.1: A collection of subsets of S is called a *Borel field* (or *sigma algebra*), denoted by $\mathcal{B}$, if it satisfies the following three properties:

1. $\emptyset \in \mathcal{B}$ (the empty set is contained in $\mathcal{B}$).
2. If $A \in \mathcal{B}$ then $A^c \in \mathcal{B}$ ($\mathcal{B}$ is closed under complementation).
3. If $A_1, A_2, \ldots \in \mathcal{B}$ then $\cup_{i=1}^{\infty} A_i \in \mathcal{B}$ ($\mathcal{B}$ is closed under countable unions).

The empty set $\emptyset$ is a subset of any set. Thus, $\emptyset \subset S$. Property (1) states that this subset is always in a Borel field. Since $S = \emptyset^c$, properties (1) and (2) imply that S is always in $\mathcal{B}$ also. In addition, from DeMorgan's Laws it follows that $\mathcal{B}$ is closed under countable intersections. If $A_1, A_2, \ldots \in \mathcal{B}$, then $A_1^c, A_2^c, \ldots \in \mathcal{B}$ by property (2), and therefore $\cup_{i=1}^{\infty} A_i^c \in \mathcal{B}$. However, using DeMorgan's Law (as in Exercise 1.4), we have

$$\left(\bigcup_{i=1}^{\infty} A_i^c\right)^c = \bigcap_{i=1}^{\infty} A_i. \tag{1.2.1}$$

Thus, again by property (2), $\cap_{i=1}^{\infty} A_i \in \mathcal{B}$.

Associated with sample space S we can have many different Borel fields. For example, the collection of the two sets $\{\emptyset, S\}$ is a Borel field, usually called the trivial Borel field. The only Borel field we will be concerned with is the smallest one that contains all of the open sets in a given sample space S.

Example 1.2.1: If S is finite or countable, then these technicalities really do not arise, for we define for a given sample space S

$$\mathcal{B} = \{\text{all subsets of } S, \text{ including } S \text{ itself}\}.$$

If S has n elements, there are 2^n sets in $\mathcal{B}$ (see Exercise 1.13). For example, if $S = \{1, 2, 3\}$, then $\mathcal{B}$ is the following collection of $2^3 = 8$ sets:

$$\begin{array}{lll} \{1\} & \{1,2\} & \{1,2,3\} \\ \{2\} & \{1,3\} & \emptyset \\ \{3\} & \{2,3\} & \end{array}$$

||

In general, if S is uncountable, it is not an easy task to describe $\mathcal{B}$. However, $\mathcal{B}$ is chosen to contain any set of interest.

Example 1.2.2: Let $S = (-\infty, \infty)$, the real line. Then $\mathcal{B}$ is chosen to contain all sets of the form

$$[a, b], \quad (a, b], \quad (a, b), \quad \text{and} \quad [a, b)$$

for all real numbers a and b. Also, from the properties of $\mathcal{B}$, it follows that $\mathcal{B}$ contains all sets that can be formed by taking (possibly countably infinite) unions and intersections of sets of the above varieties. ||

We are now in a position to define a probability function.

DEFINITION 1.2.2: Given a sample space S and an associated Borel field $\mathcal{B}$, a *probability function* is a function P with domain $\mathcal{B}$ that satisfies

1. $P(A) \geq 0$ for all $A \in \mathcal{B}$.
2. $P(S) = 1$.
3. If $A_1, A_2, \ldots \in \mathcal{B}$ are pairwise disjoint, then $P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$.

The three properties given in Definition 1.2.2 are usually referred to as the Axioms of Probability (or the Kolmogorov Axioms, after A. Kolmogorov, one of the fathers of probability theory). Any function P that satisfies the Axioms of Probability is called a probability function. The axiomatic definition makes no attempt to tell what particular function P to choose; it merely requires P to satisfy the axioms. For any sample space many different probability functions can be defined. Which one(s) reflect what is likely to be observed in a particular experiment is still to be discussed.

Example 1.2.3: Consider the simple experiment of tossing a fair coin, so $S = \{\text{H}, \text{T}\}$. By a "fair" coin we mean a balanced coin that is equally as likely to land heads up as tails up, and hence the reasonable probability function is the one that assigns equal probabilities to heads and tails, that is,

$$P(\{\text{H}\}) = P(\{\text{T}\}). \tag{1.2.2}$$

Note that (1.2.2) does not follow from the axioms of probability but rather is outside of the axioms. We have used a symmetry interpretation of probability (or just intuition) to impose the requirement that heads and tails be equally probable. Since $S = \{H\} \cup \{T\}$, we have, from axiom (2), $P(\{H\} \cup \{T\}) = 1$. Also, $\{H\}$ and $\{T\}$ are disjoint, so $P(\{H\} \cup \{T\}) = P(\{H\}) + P(\{T\})$ and

$$P(\{H\}) + P(\{T\}) = 1. \tag{1.2.3}$$

Simultaneously solving (1.2.2) and (1.2.3) shows that $P(\{H\}) = P(\{T\}) = \frac{1}{2}$.

Since (1.2.2) is based on our knowledge of the particular experiment, not the axioms, any nonnegative values for $P(\{H\})$ and $P(\{T\})$ that satisfy (1.2.3) define a legitimate probability function. For example, we might choose $P(\{H\}) = \frac{1}{9}$ and $P(\{T\}) = \frac{8}{9}$. ||

The physical reality of the experiment might dictate the probability assignment, as the next example illustrates. Of course, the assignment must satisfy the Kolmogorov Axioms.

Example 1.2.4: The game of darts is played by throwing a dart at a board, and receiving a score corresponding to the number assigned to the region in which the dart lands. For a novice player, it seems reasonable to assume that the probability of the dart hitting a particular region is proportional to the area of the region. Thus, a bigger region has a higher probability of being hit.

Referring to Figure 1.2.1, the dart board has radius r, and the distance between rings is $r/5$. If we make the assumption that the board is always hit (see Exercise 1.42 for a variation on this), then we have

$$P(\text{scoring } i \text{ points}) = \frac{\text{Area of region } i}{\text{Area of dart board}}.$$

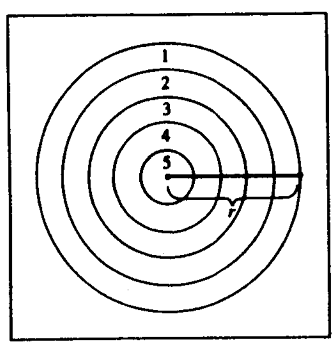


FIGURE 1.2.1 Dart board for Example 1.2.4

For example

$$P(\text{scoring 1 point}) = \frac{\pi r^2 - \pi(4r/5)^2}{\pi r^2} = 1 - \left(\frac{4}{5}\right)^2.$$

It is easy to derive the general formula, and we find that

$$P(\text{scoring } i \text{ points}) = \frac{(6-i)^2 - (5-i)^2}{5^2}, \quad i = 1, \ldots, 5$$

independent of π and r. It is straightforward to verify that this is a probability function. (See Exercise 1.11 for further details.) ||

Before we leave the axiomatic development of probability, there is one further point to consider. Axiom (3) of Definition 1.2.2, which is commonly known as the Axiom of Countable Additivity, is not universally accepted among statisticians. Indeed, it can be argued that axioms should be simple, self-evident statements. Comparing axiom (3) to the other axioms, which are simple and self-evident, may lead us to doubt whether it is reasonable to assume the truth of axiom (3).

The Axiom of Countable Additivity is rejected by a school of statisticians led by De Finetti (1972), who chooses to replace this axiom with the Axiom of Finite Additivity.

Axiom of Finite Additivity: If $A \in \mathcal{B}$ and $B \in \mathcal{B}$ are disjoint, then

$$P(A \cup B) = P(A) + P(B).$$

While this axiom may not be entirely self-evident, it is certainly simpler than the Axiom of Countable Additivity.

Assuming only finite additivity, while perhaps more plausible, can lead to unexpected complications in statistical theory—complications that, at this level, do not necessarily enhance understanding of the subject. We therefore proceed under the assumption that the Axiom of Countable Additivity holds.

1.2.2 The Calculus of Probabilities

From the Axioms of Probability we can build up many properties of the probability function, properties that are quite helpful in the calculation of more complicated probabilities. Some of these manipulations will be discussed in detail in this section; others will be left as exercises.

We start with some (fairly self-evident) properties of the probability function when applied to a single event.

THEOREM 1.2.1: If P is a probability function and A is any set in $\mathcal{B}$, then

a. $P(\emptyset) = 0$, where $\emptyset$ is the empty set;
b. $P(A) \leq 1$;
c. $P(A^c) = 1 - P(A)$.

Proof: It is easiest to prove (c) first. The sets A and A^c form a partition of the sample space, that is, $S = A \cup A^c$. Therefore,

$$P(A \cup A^c) = P(S) = 1, \tag{1.2.4}$$

by the second axiom. Also, A and A^c are disjoint, so by the third axiom,

$$P(A \cup A^c) = P(A) + P(A^c). \tag{1.2.5}$$

Combining (1.2.4) and (1.2.5) gives (c).

Since $P(A^c) \geq 0$, (b) is immediately implied by (c). To prove (a), we use a similar argument on $S = S \cup \emptyset$. (Recall that both S and $\emptyset$ are always in $\mathcal{B}$.) Since S and $\emptyset$ are disjoint, we have

$$1 = P(S) = P(S \cup \emptyset) = P(S) + P(\emptyset),$$

and thus $P(\emptyset) = 0$. □

Theorem 1.2.1 contains properties that are so basic that they also have the flavor of axioms, although we have formally proved them using only the original three Kolmogorov Axioms. The next theorem, which is similar in spirit to Theorem 1.2.1, contains statements that are not so self-evident.

THEOREM 1.2.2: If P is a probability function and A and B are any sets in $\mathcal{B}$, then

a. $P(B \cap A^c) = P(B) - P(A \cap B)$;
b. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$;
c. If $A \subset B$ then $P(A) \leq P(B)$.

Proof: To establish (a) note that for any sets A and B we have

$$B = \{B \cap A\} \cup \{B \cap A^c\},$$

and therefore

$$P(B) = P(\{B \cap A\} \cup \{B \cap A^c\}) = P(B \cap A) + P(B \cap A^c), \tag{1.2.6}$$

where the last equality in (1.2.6) follows from the fact that $B \cap A$ and $B \cap A^c$ are disjoint. Rearranging (1.2.6) gives (a).

To establish (b), we use the identity

$$A \cup B = A \cup \{B \cap A^c\}. \tag{1.2.7}$$

A Venn diagram will show why (1.2.7) holds, although a formal proof is not difficult (see Exercise 1.2). Using (1.2.7) and the fact that A and $B \cap A^c$ are disjoint (since A and A^c are), we have

$$P(A \cup B) = P(A) + P(B \cap A^c) = P(A) + P(B) - P(A \cap B), \tag{1.2.8}$$

from (a).

If $A \subset B$ then $A \cap B = A$. Therefore, using (a) we have

$$0 \leq P(B \cap A^c) = P(B) - P(A),$$

establishing (c). □

Formula (b) of Theorem 1.2.2 gives a useful inequality for the probability of an intersection. Since $P(A \cup B) \leq 1$ we have from (1.2.8), after some rearranging,

$$P(A \cap B) \geq P(A) + P(B) - 1. \tag{1.2.9}$$

This inequality is a special case of what is known as *Bonferroni's Inequality* [Miller (1981) is a good reference]. Bonferroni's Inequality allows us to bound the probability of a simultaneous event (the intersection) in terms of the probabilities of the individual events.

Example 1.2.5: Bonferroni's Inequality is particularly useful when it is difficult (or even impossible) to calculate the intersection probability, but some idea of the size of this probability is desired. Suppose A and B are two events and each has probability .95. Then the probability that both will occur is bounded below by

$$P(A \cap B) \geq P(A) + P(B) - 1 = .95 + .95 - 1 = .90.$$

Note that unless the probabilities of the individual events are sufficiently large, the Bonferroni bound is a useless (but correct!) negative number. ||

We close this section with a theorem that gives some useful results for dealing with a collection of sets.

THEOREM 1.2.3: If P is a probability function, then

a. $P(A) = \sum_{i=1}^{\infty} P(A \cap C_i)$ for any partition $C_1, C_2, \ldots$;
b. $P(\cup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} P(A_i)$ for any sets $A_1, A_2, \ldots$. (Boole's Inequality)

Proof: Since $C_1, C_2, \ldots$ forms a partition we have that $C_i \cap C_j = \emptyset$ for all $i \neq j$, and $S = \cup_{i=1}^{\infty} C_i$. Hence,

$$A = A \cap S = A \cap \left(\bigcup_{i=1}^{\infty} C_i \right) = \bigcup_{i=1}^{\infty} (A \cap C_i),$$

where the last equality follows from the Distributive Law (Theorem 1.1.1). We therefore have

$$P(A) = P\left(\bigcup_{i=1}^{\infty}(A \cap C_i)\right).$$

Now, since the C_i are disjoint, the sets $A \cap C_i$ are also disjoint, and from the properties of a probability function we have

$$P\left(\bigcup_{i=1}^{\infty}(A \cap C_i)\right) = \sum_{i=1}^{\infty} P(A \cap C_i),$$

establishing (a).

To establish (b) we first construct a disjoint collection $A_1^*, A_2^*, \ldots$, with the property that $\cup_{i=1}^{\infty} A_i^* = \cup_{i=1}^{\infty} A_i$. Define A_i^* by

$$A_1^* = A_1, \quad A_i^* = A_i \backslash \left(\bigcup_{j=1}^{i-1} A_j\right), \quad i = 2, 3, \ldots,$$

where the notation $A \backslash B$ denotes the part of A that does not intersect with B. In more familiar symbols, $A \backslash B = A \cap B^c$. It should be easy to see that $\cup_{i=1}^{\infty} A_i^* = \cup_{i=1}^{\infty} A_i$, and we therefore have

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = P\left(\bigcup_{i=1}^{\infty} A_i^*\right) = \sum_{i=1}^{\infty} P(A_i^*),$$

where the last equality follows since the A_i^* are disjoint. To see this, write

$$\begin{aligned}
A_i^* \cap A_k^* &= \left\{A_i \backslash \left(\bigcup_{j=1}^{i-1} A_j\right)\right\} \cap \left\{A_k \backslash \left(\bigcup_{j=1}^{k-1} A_j\right)\right\} && \text{(definition of } A_i^*) \\
&= \left\{A_i \cap \left(\bigcup_{j=1}^{i-1} A_j\right)^c\right\} \cap \left\{A_k \cap \left(\bigcup_{j=1}^{k-1} A_j\right)^c\right\} && \text{(definition of "}\backslash\text{")} \\
&= \left\{A_i \cap \bigcap_{j=1}^{i-1} A_j^c\right\} \cap \left\{A_k \cap \bigcap_{j=1}^{k-1} A_j^c\right\} && \text{(DeMorgan's Laws)}
\end{aligned}$$

Now if $i > k$ the first intersection above will contain the set A_k^c, which will have an empty intersection with A_k. If $k > i$ the argument is similar. Further, by construction $A_i^* \subset A_i$, so $P(A_i^*) \le P(A_i)$ and we have

$$\sum_{i=1}^{\infty} P(A_i^*) \le \sum_{i=1}^{\infty} P(A_i),$$

establishing (b). □

There is a similarity between Boole's Inequality and Bonferroni's Inequality. In fact, they are essentially the same thing. We could have used Boole's Inequality to derive (1.2.9). If we apply Boole's Inequality to A^c, we have

$$P\left(\bigcup_{i=1}^{n} A_i^c\right) \le \sum_{i=1}^{n} P(A_i^c),$$

and using the facts that $\cup A_i^c = (\cap A_i)^c$ and $P(A_i^c) = 1 - P(A_i)$, we obtain

$$1 - P\left(\bigcap_{i=1}^{n} A_i\right) \le n - \sum_{i=1}^{n} P(A_i).$$

This becomes, on rearranging terms,

$$P\left(\bigcap_{i=1}^{n} A_i\right) \ge \sum_{i=1}^{n} P(A_i) - (n-1), \tag{1.2.10}$$

which is a more general version of the Bonferroni Inequality of (1.2.9).

1.2.3 Counting

The elementary process of counting can become quite sophisticated when placed in the hands of a statistician. Most often, methods of counting are used in order to construct probability assignments on finite sample spaces, although they can be used to answer other questions also.

Example 1.2.6: For a number of years the New York state lottery operated according to the following scheme. From the numbers $1, 2, \ldots, 44$, a person may pick any six for her ticket. The winning number is then decided by randomly selecting six numbers from the forty-four. To be able to calculate the probability of winning we first must count how many different groups of six numbers can be chosen from the forty-four. ||

Example 1.2.7: In a single-elimination tournament, such as the U.S. Open tennis tournament, players advance only if they win (in contrast to double-elimination or round-robin tournaments). If we have 16 entrants, we might be interested in the number of paths a particular player can take to victory, where a path is taken to mean a sequence of opponents. ||

Counting problems, in general, sound complicated, and often we must do our counting subject to many restrictions. The way to solve such problems is to break them down into a series of simple tasks that are easy to count, and employ known rules of combining tasks. The following theorem is a first step in such a process, and is sometimes known as the Fundamental Theorem of Counting.

THEOREM 1.2.4: If a job consists of k separate tasks, the ith of which can be done in n_i ways, $i = 1, \ldots, k$, then the entire job can be done in $n_1 \times n_2 \times \cdots \times n_k$ ways.

Proof: It suffices to prove the theorem for $k = 2$ (see Exercise 1.14). The proof is just a matter of careful counting. The first task can be done in n_1 ways, and for each of these ways we have n_2 choices for the second task. Thus, we can do the job in

$$\underbrace{(1 \times n_2) + (1 \times n_2) + \cdots + (1 \times n_2)}_{n_1 \text{ terms}} = n_1 \times n_2$$

ways, establishing the theorem for $k = 2$. □

Example 1.2.8: Although the Fundamental Theorem of Counting is a reasonable place to start, in applications there are usually more aspects of a problem to consider. For example, in the New York state lottery the first number can be chosen in 44 ways, and the second number in 43 ways, making a total of $44 \times 43 = 1,892$ ways of choosing the first two numbers. However, if a person is allowed to choose the same number twice, then the first two numbers can be chosen in $44 \times 44 = 1,936$ ways. ‖

The distinction being made in Example 1.2.8 is between counting *with replacement* and counting *without replacement*. There is a second crucial element in any counting problem, whether or not the ordering of the tasks is important. To illustrate with the lottery example, suppose the winning numbers are selected in the order 12, 37, 35, 9, 13, 22. Does a person who selected 9, 12, 13, 22, 35, 37 qualify as a winner? In other words, does the order in which the task is performed actually matter? Taking all of these considerations into account, we can construct a 2×2 table of possibilities:

Possible methods of counting

	Without replacement	With replacement
Ordered		
Unordered		

Before we begin to count, the following definition gives us some extremely helpful notation.

DEFINITION 1.2.3: For a positive integer n, $n!$ (read n factorial) is the product of all of the positive integers less than or equal to n. That is,

$$n! = n \times (n-1) \times (n-2) \times \cdots \times 3 \times 2 \times 1.$$

Furthermore, we define $0! = 1$. □

Let us now consider counting all of the possible lottery tickets under each of these four cases.

1. *Ordered, without replacement* From the Fundamental Theorem of Counting, the first number can be selected in 44 ways, the second in 43 ways, etc. So there are

$$44 \times 43 \times 42 \times 41 \times 40 \times 39 = \frac{44!}{38!} = 5,082,517,440$$

possible tickets.

2. *Ordered, with replacement* Since each number can now be selected in 44 ways (because the chosen number is replaced) there are

$$44 \times 44 \times 44 \times 44 \times 44 \times 44 = 44^6 = 7,256,313,856$$

possible tickets.

3. *Unordered, without replacement* We know the number of possible tickets when the ordering must be accounted for, so what we must do is divide out the redundant orderings. Again using the Fundamental Theorem, six numbers can be arranged in $6 \times 5 \times 4 \times 3 \times 2 \times 1$ ways, so the total number of unordered tickets is

$$\frac{44 \times 43 \times 42 \times 41 \times 40 \times 39}{6 \times 5 \times 4 \times 3 \times 2 \times 1} = \frac{44!}{6!38!} = 7,059,052.$$

This form of counting plays a central role in much of statistics—so much, in fact, that it has earned its own notation.

DEFINITION 1.2.4: For nonnegative integers n and r, $n \geq r$, we define the symbol $\binom{n}{r}$, read n *choose* r, as

$$\binom{n}{r} = \frac{n!}{r!(n-r)!}.$$

In our lottery example, the number of possible tickets (unordered, without replacement) is $\binom{44}{6}$. These numbers are also referred to as *binomial coefficients*, for reasons that will become clear in Chapter 3.

4. *Unordered, with replacement* This is the most difficult case to count. You might first guess that the answer is $44^6/(6 \times 5 \times 4 \times 3 \times 2 \times 1)$, but this is not correct (it is too small).

To count in this case, it is easiest to think of placing 6 markers on the 44 numbers. In fact, we can think of the 44 numbers defining bins in which we can place the six markers, M, as shown, for example, in the accompanying figure.

M		MM	M		...	M	M		
1	2	3	4	5	...	41	42	43	44

The number of possible tickets is then equal to the number of ways that we can put the 6 markers into the 44 bins. But this can be further reduced by noting that all we need to keep track of is the arrangement of the markers and the walls of the bins. Note further that the two outermost walls play no part. Thus, we have to count all of the arrangements of 43 walls (44 bins yield 45 walls, but we disregard the two end walls) and 6 markers. We therefore have 43 + 6 = 49 objects, which can be arranged in 49! ways. However, to eliminate the redundant orderings we must divide by both 6! and 43!, so the total number of arrangements is

$$\frac{49!}{6!43!} = 13,983,816.$$

Although all of the preceding derivations were done in terms of an example, it should be easy to see that they hold in general. For completeness, we can summarize these situations in the following table:

	Number of possible arrangements of size r from n objects	
	Without replacement	With replacement
Ordered	$\dfrac{n!}{(n-r)!}$	n^r
Unordered	$\dbinom{n}{r}$	$\dbinom{n+r-1}{r}$

1.2.4 Equally Likely Outcomes

The counting techniques of the previous section are useful when the sample space S is a finite set and all the outcomes in S are equally likely. Then probabilities of events can be calculated by simply counting the number of outcomes in the event. To see this, suppose that $S = \{s_1, \ldots, s_N\}$ is a finite sample space. Saying that all the outcomes are equally likely means that $P(\{s_i\}) = 1/N$ for every outcome s_i. Then, using (3) from the definition of a probability function, we have, for any event A,

$$P(A) = \sum_{s_i \in A} P(\{s_i\}) = \sum_{s_i \in A} \frac{1}{N} = \frac{\text{\# of elements in } A}{\text{\# of elements in } S}.$$

For large sample spaces, the counting techniques might be used to calculate both the numerator and denominator of this expression.

Example 1.2.9: Consider choosing a five-card poker hand from a standard deck of 52 playing cards. Obviously, we are sampling without replacement from the deck. But to specify the possible outcomes (possible hands), we must decide whether we think of the hand as being dealt sequentially (ordered) or all at once (unordered). If we wish to calculate probabilities for events that depend on the order, such as the probability of an ace in the first two cards, then we must use the ordered outcomes. But if our events do not depend on the order, we can use the unordered outcomes. For this example we use the unordered outcomes, so the sample space consists of all the five-card hands that can be chosen from the 52-card deck. There are $\binom{52}{5} = 2,598,960$ possible hands. If the deck is well shuffled and the cards are randomly dealt, it is reasonable to assign probability 1/2,598,960 to each possible hand.

We now calculate some probabilities by counting outcomes in events. What is the probability of having four aces? How many different hands are there with four aces? Having specified that four of the cards are aces, there are 48 different ways of specifying the fifth card. Thus,

$$P(\text{four aces}) = \frac{48}{2,598,960},$$

less than 1 chance in 50,000. Only slightly more complicated counting, using Theorem 1.2.4, allows us to calculate the probability of having four of a kind. There are 13 ways to specify which denomination there will be four of. Having specified these four cards, there are 48 ways of specifying the fifth. Thus, the total number of hands with four of a kind is (13)(48) and

$$P(\text{four of a kind}) = \frac{(13)(48)}{2,598,960} = \frac{624}{2,598,960}.$$

To calculate the probability of exactly one pair (not two pair, no three of a kind, etc.) we combine some of the counting techniques. The number of hands with exactly one pair is

$$13\binom{4}{2}\binom{12}{3}4^3 = 1,098,240. \qquad (1.2.11)$$

Expression (1.2.11) comes from Theorem 1.2.4 because

$13 = \#$ of ways to specify the denomination for the pair,

$\binom{4}{2} = \#$ of ways to specify the two cards from that denomination,

$\binom{12}{3} = \#$ of ways of specifying the other three denominations,

$4^3 = \#$ of ways of specifying the other three cards from those denominations.

Thus,

$$P(\text{exactly one pair}) = \frac{1,098,240}{2,598,960}. \qquad \|$$

1.3 Conditional Probability and Independence

All of the probabilities that we have dealt with thus far have been unconditional probabilities. A sample space was defined and all probabilities were calculated with respect to that sample space. In many instances, however, we are in a position to update the sample space based on new information. In such cases, we want to be able to update probability calculations or to calculate *conditional probabilities.*

Example 1.3.1: Four cards are dealt from the top of a well-shuffled deck. What is the probability that they are the four aces? We can calculate this probability by the methods of the previous section. The number of distinct groups of four cards is

$$\binom{52}{4} = 270,725.$$

Only one of these groups consists of the four aces and every group is equally likely, so the probability of being dealt all four aces is 1/270,725.

We can also calculate this probability by an "updating" argument, as follows. The probability that the first card is an ace is 4/52. *Given that the first card is an ace,* the probability that the second card is an ace is 3/51 (there are 3 aces and 51 cards left). Continuing this argument, we get the desired probability as

$$\frac{4}{52} \times \frac{3}{51} \times \frac{2}{50} \times \frac{1}{49} = \frac{1}{270,725}. \qquad \|$$

In our second method of solving the problem, we updated the sample space after each draw of a card; we calculated conditional probabilities.

DEFINITION 1.3.1: If A and B are events in S, and $P(B) > 0$, then the *conditional probability of A given B*, written $P(A|B)$, is

$$P(A|B) = \frac{P(A \cap B)}{P(B)}. \qquad (1.3.1)$$

Note that what happens in the conditional probability calculation is that B becomes the sample space: $P(B|B) = 1$. The intuition is that our original sample space, S, has been updated to B. All further occurrences are then calibrated with respect to their relation to B. In particular, note what happens to conditional probabilities of disjoint sets. Suppose A and B are disjoint, so $P(A \cap B) = 0$. It then follows that $P(A|B) = P(B|A) = 0$.

Example 1.3.1 (Continued): Although the probability of getting all four aces is quite small, let us see how the conditional probabilities change given that some aces have already been obtained. Four cards will again be dealt from a well-shuffled deck, and we now calculate

$$P(\text{4 aces in 4 cards}|i \text{ aces in } i \text{ cards}), \quad i = 1,2,3.$$

The event {4 aces in 4 cards} is a subset of the event {i aces in i cards}. Thus, from the definition of conditional probability, (1.3.1), we know that

$$\begin{aligned} P(\text{4 aces in 4 cards}|i \text{ aces in } i \text{ cards}) &= \frac{P(\{\text{4 aces in 4 cards}\} \cap \{i \text{ aces in } i \text{ cards}\})}{P(i \text{ aces in } i \text{ cards})} \\ &= \frac{P(\text{4 aces in 4 cards})}{P(i \text{ aces in } i \text{ cards})}. \end{aligned}$$

The numerator has already been calculated, and the denominator can be calculated with a similar argument. The number of distinct groups of i cards is $\binom{52}{i}$, and

$$P(i \text{ aces in } i \text{ cards}) = \frac{\binom{4}{i}}{\binom{52}{i}}.$$

Therefore, the conditional probability is given by

$$P(\text{4 aces in 4 cards}|i \text{ aces in } i \text{ cards}) = \frac{\binom{52}{i}}{\binom{52}{4}\binom{4}{i}} = \frac{(4-i)!48!}{(52-i)!} = \frac{1}{\binom{52-i}{4-i}}.$$

For $i = 1$, 2, and 3, the conditional probabilities are .00005, .00082, and .02041, respectively. $\|$

For any B for which $P(B) > 0$, it is straightforward to verify that the probability function $P(\cdot|B)$ satisfies Kolmogorov's Axioms (see Exercise 1.41). You may suspect that requiring $P(B) > 0$ is redundant. Who would want to condition on an event of probability zero? Interestingly, sometimes this is a particularly useful way of thinking of things. However, we will defer these considerations until Chapter 4.

Conditional probabilities can be particularly slippery entities and sometimes require careful thought. Consider the following often-told tale.

Example 1.3.2: Three prisoners, A, B, and C are on death row. The governor decides to pardon one of the three and chooses at random the prisoner to pardon. He informs the warden of his choice but requests that the name be kept secret for a few days.

The next day, A tries to get the warden to tell him who had been pardoned. The warden refuses. A then asks which of B or C will be executed. The warden thinks for a while, then tells A that B is to be executed.

Warden's reasoning: Each prisoner has a $\frac{1}{3}$ chance of being pardoned. Clearly, either B or C must be executed, so I have given A no information about whether A will be pardoned.

A's reasoning: Given that B will be executed, then either A or C will be pardoned. My chance of being pardoned has risen to $\frac{1}{2}$.

It should be clear that the warden's reasoning is correct, but let us see why. Let A, B, and C denote the events that A, B, and C is pardoned, respectively. We know that $P(A) = P(B) = P(C) = \frac{1}{3}$. Let W denote the event that the warden says B will die. Using (1.3.1), A can update his probability of being pardoned to

$$P(A|W) = \frac{P(A \cap W)}{P(W)}.$$

What is happening can be summarized in the following table:

Prisoner pardoned	Warden tells A	
A	B dies	each with equal probability
A	C dies	
B	C dies	
C	B dies	

Using this table, we can calculate

$$\begin{aligned} P(W) &= P(\text{warden says B dies}) \\ &= P(\text{warden says B dies and A pardoned}) \\ &\quad + P(\text{warden says B dies and C pardoned}) \\ &\quad + P(\text{warden says B dies and B pardoned}) \\ &= \frac{1}{6} + \frac{1}{3} + 0 = \frac{1}{2}. \end{aligned}$$

Thus, using the warden's reasoning,

$$P(A|W) = \frac{P(A \cap W)}{P(W)}$$

$$(1.3.2) \qquad = \frac{P(\text{warden says B dies and A pardoned})}{P(\text{warden says B dies})} = \frac{1/6}{1/2} = \frac{1}{3}.$$

However, A falsely interprets the event W as equal to the event B^c and calculates

$$P(A|B^c) = \frac{P(A \cap B^c)}{P(B^c)} = \frac{1/3}{2/3} = \frac{1}{2}.$$

We see that conditional probabilities can be quite slippery and require careful interpretation. For some other variations of this problem, see Exercise 1.47. ||

Re-expressing (1.3.1) gives a useful form for calculating intersection probabilities,

$$(1.3.3) \qquad P(A \cap B) = P(A|B)P(B),$$

which is essentially the formula that was used in Example 1.3.1. We can take advantage of the symmetry of (1.3.3) and also write

$$(1.3.4) \qquad P(A \cap B) = P(B|A)P(A).$$

When faced with seemingly difficult calculations, we can break up our calculations according to (1.3.3) or (1.3.4), whichever is easier. Furthermore, we can equate the right-hand sides of these equations to obtain (after rearrangement)

$$(1.3.5) \qquad P(A|B) = P(B|A)\frac{P(A)}{P(B)},$$

which gives us a formula for "turning around" conditional probabilities. Equation (1.3.5) is often called Bayes' Rule for its discoverer, Sir Thomas Bayes (although see Stigler, 1983).

Bayes' Rule has a more general form than (1.3.5), one that applies to partitions of a sample space. We therefore take the following as the definition of Bayes' Rule.

BAYES' RULE: Let $A_1, A_2, \ldots$ be a partition of the sample space, and let B be any set. Then, for each $i = 1, 2, \ldots$,

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{j=1}^{\infty} P(B|A_j)P(A_j)}. \qquad \square$$

Example 1.3.3: When coded messages are sent, there are sometimes errors in transmission. In particular, Morse code uses "dots" and "dashes," which are known to occur in the proportion of 3:4. This means that for any given symbol

$$P(\text{dot sent}) = \frac{3}{7} \quad \text{and} \quad P(\text{dash sent}) = \frac{4}{7}.$$

Suppose there is interference on the transmission line, and with probability $\frac{1}{8}$ a dot is mistakenly received as a dash, and vice versa. If we receive a dot, can we be sure that a dot was sent? Using Bayes' Rule, we can write

$$P(\text{dot sent} \mid \text{dot received}) = P(\text{dot received} \mid \text{dot sent})\frac{P(\text{dot sent})}{P(\text{dot received})}.$$

Now, from the information given, we know that $P(\text{dot sent}) = \frac{3}{7}$ and $P(\text{dot received} \mid \text{dot sent}) = \frac{7}{8}$. Furthermore, we can also write

$$\begin{aligned} P(\text{dot received}) &= P(\text{dot received} \cap \text{dot sent}) + P(\text{dot received} \cap \text{dash sent}) \\ &= P(\text{dot received} \mid \text{dot sent})P(\text{dot sent}) \\ &\quad + P(\text{dot received} \mid \text{dash sent})P(\text{dash sent}) \\ &= \frac{7}{8} \times \frac{3}{7} + \frac{1}{8} \times \frac{4}{7} = \frac{25}{56}. \end{aligned}$$

Combining these results, we have that the probability of correctly receiving a dot is

$$P(\text{dot sent} \mid \text{dot received}) = \frac{(7/8) \times (3/7)}{25/56} = \frac{21}{25}. \qquad \|$$

In some cases it may happen that the occurrence of a particular event, B, has no effect on the probability of another event, A. Symbolically, we are saying that

$$P(A|B) = P(A). \tag{1.3.6}$$

If this holds, then by Bayes' Rule (1.3.5) we have

$$\begin{aligned} P(B|A) &= P(A|B)\frac{P(B)}{P(A)} \\ &= P(A)\frac{P(B)}{P(A)} \qquad \text{(from(1.3.6))} \qquad (1.3.7) \\ &= P(B), \end{aligned}$$

so the occurrence of A has no effect on B. Moreover, since $P(B|A)P(A) = P(A \cap B)$, it then follows that

$$P(A \cap B) = P(A)P(B),$$

which we take as the definition of statistical independence.

DEFINITION 1.3.2: Two events, A and B, are *statistically independent* if

$$P(A \cap B) = P(A)P(B). \tag{1.3.8}$$

Note that independence could have been equivalently defined by either (1.3.6) or (1.3.7) (as long as either $P(A) > 0$ or $P(B) > 0$). The advantage of (1.3.8) is that it treats the events symmetrically and will be easier to generalize to more than two events.

Many gambling games provide models of independent events. The spins of a roulette wheel and the tosses of a pair of dice are both series of independent events.

Example 1.3.4: The gambler introduced at the start of the chapter, the Chevalier de Meré, was particularly interested in the event that he could throw at least 1 six in 4 rolls of a die. We have

$$\begin{aligned} P(\text{at least 1 six in 4 rolls}) &= 1 - P(\text{no six in 4 rolls}) \\ &= 1 - \prod_{i=1}^{4} P(\text{no six on roll } i), \end{aligned}$$

where the last equality follows by independence of the rolls. On any roll, the probability of *not* rolling a six is $\frac{5}{6}$, so

$$P(\text{at least 1 six in 4 rolls}) = 1 - \left(\frac{5}{6}\right)^4 = .518. \qquad \|$$

Independence of A and B implies independence of the complements also. In fact, we have the following theorem.

THEOREM 1.3.1: If A and B are independent events, then the following pairs are also independent:

a. A and B^c
b. A^c and B
c. A^c and B^c

Proof: We will prove only (a), leaving the rest as Exercise 1.45. To prove (a) we must show

$$P(A \cap B^c) = P(A)P(B^c).$$

From the definition of conditional probability we have

$$\begin{aligned} P(A \cap B^c) &= P(B^c|A)P(A) \\ &= [1 - P(B|A)]P(A) \quad \text{(since } P(\cdot|A) \text{ is a probability function)} \\ &= [1 - P(B)]P(A) \quad (A \text{ and } B \text{ are independent}) \\ &= P(B^c)P(A). \end{aligned}$$

□

Independence of more than two events can be defined in a manner similar to (1.3.8), but we must be careful. For example, we might think that we could say $A, B,$ and C are independent if $P(A \cap B \cap C) = P(A)P(B)P(C)$. However, this is not the correct condition.

Example 1.3.5: Let an experiment consist of tossing two dice. For this experiment the sample space is

$$S = \{(1,1),(1,2),\ldots,(1,6),(2,1),\ldots,(2,6),\ldots,(6,1),\ldots,(6,6)\},$$

that is, S consists of the 36 ordered pairs formed from the numbers 1 to 6. Define the following events:

$$\begin{aligned} A &= \{\text{doubles appear}\} = \{(1,1),(2,2),(3,3),(4,4),(5,5),(6,6)\}, \\ B &= \{\text{the sum is between 7 and 10}\}, \\ C &= \{\text{the sum is 2 or 7 or 8}\}. \end{aligned}$$

The probabilities can be calculated by counting among the 36 possible outcomes. We have

$$P(A) = \frac{1}{6}, \quad P(B) = \frac{1}{2}, \quad \text{and} \quad P(C) = \frac{1}{3}.$$

Furthermore,

$$\begin{aligned} P(A \cap B \cap C) &= P(\text{the sum is 8, composed of double fours}) \\ &= \frac{1}{36} \\ &= \frac{1}{6} \times \frac{1}{2} \times \frac{1}{3} \\ &= P(A)P(B)P(C). \end{aligned}$$

However,

$$\begin{aligned} P(B \cap C) &= P(\text{sum equals 7 or 8}) \\ &= \frac{11}{36} \\ &\neq P(B)P(C). \end{aligned}$$

Similarly, it can be shown that $P(A \cap B) \neq P(A)P(B)$; therefore, the requirement $P(A \cap B \cap C) = P(A)P(B)P(C)$ is not a strong enough condition to guarantee pairwise independence. ||

A second attempt at a general definition of independence, in light of the previous example, might be to define $A, B,$ and C to be independent if all the pairs are independent. Alas, this condition also fails.

Example 1.3.6: Let the sample space S consist of the 3! permutations of the letters a, b, and c along with the three triples of each letter. Thus,

$$S = \left\{ \begin{matrix} \text{aaa} & \text{bbb} & \text{ccc} \\ \text{abc} & \text{bca} & \text{cba} \\ \text{acb} & \text{bac} & \text{cab} \end{matrix} \right\}.$$

Furthermore, let each element of S have probability $\frac{1}{9}$. Define

$$A_i = \{i\text{th place in the triple is occupied by a}\}.$$

It is then easy to count that

$$P(A_i) = \frac{1}{3}, \quad i = 1, 2, 3,$$

and

$$P(A_1 \cap A_2) = P(A_1 \cap A_3) = P(A_2 \cap A_3) = \frac{1}{9},$$

so the A_is are pairwise independent. But

$$P(A_1 \cap A_2 \cap A_3) = \frac{1}{9} \neq P(A_1)P(A_2)P(A_3),$$

so the A_is do not satisfy the probability requirement. ||

The preceding two examples show that simultaneous (or mutual) independence of a collection of events requires an extremely strong definition. The following definition works.

DEFINITION 1.3.3: A collection of events $A_1, \ldots, A_n$ are mutually independent if for any subcollection $A_{i_1}, \ldots, A_{i_k}$, we have

$$P\left(\bigcap_{j=1}^{k} A_{i_j}\right) = \prod_{j=1}^{k} P(A_{i_j}).$$

Example 1.3.7: Consider the experiment of tossing a coin three times. A sample point for this experiment must indicate the result of each toss. For example, HHT could indicate that two heads, then a tail were observed. The sample space for this experiment has eight points, namely,

$$\{\text{HHH}, \text{HHT}, \text{HTH}, \text{THH}, \text{TTH}, \text{THT}, \text{HTT}, \text{TTT}\}.$$

Let H_i, $i = 1, 2, 3$, denote the event that the ith toss is a head. For example,

$$H_1 = \{\text{HHH}, \text{HHT}, \text{HTH}, \text{HTT}\}. \tag{1.3.9}$$

If we assign probability $\frac{1}{8}$ to each sample point, then using enumerations such as (1.3.9), we see that $P(H_1) = P(H_2) = P(H_3) = \frac{1}{2}$. This says that the coin is fair and has an equal probability of landing heads or tails on each toss.

Under this probability model, the events H_1, H_2, and H_3 are also mutually independent. To verify this we note that

$$P(H_1 \cap H_2 \cap H_3) = P(\{\text{HHH}\}) = \frac{1}{8} = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = P(H_1)P(H_2)P(H_3).$$

To verify the condition in Definition 1.3.3, we also must check each pair. For example,

$$P(H_1 \cap H_2) = P(\{\text{HHH}, \text{HHT}\}) = \frac{2}{8} = \frac{1}{2} \cdot \frac{1}{2} = P(H_1)P(H_2).$$

The equality is also true for the other two pairs. Thus, H_1, H_2, and H_3 are mutually independent. That is, the occurrence of a head on any toss has no effect on any of the other tosses.

It can be verified that the assignment of probability $\frac{1}{8}$ to each sample point is the only probability model that has $P(H_1) = P(H_2) = P(H_3) = \frac{1}{2}$ and H_1, H_2, and H_3 mutually independent. ||

1.4 Random Variables

In many experiments it is easier to deal with a summary variable than with the original probability structure. For example, in an opinion poll, we might decide to ask 50 people whether they agree or disagree with a certain issue. If we record a "1" for agree and "0" for disagree, the sample space for this experiment has 2^{50} elements, each an ordered string of 1s and 0s of length 50. We should be able to reduce this to a reasonable size! It may be that the only quantity of interest is the number of people who agree (equivalently, disagree) out of 50 and, if we define a variable X = number of 1s recorded out of 50, we have captured the essence of the problem. Note that the sample space for X is the set of integers $\{0, 1, 2, \ldots, 50\}$ and is much easier to deal with than the original sample space.

In defining the quantity X, we have defined a mapping (a function) from the original sample space to a new sample space, usually a set of real numbers. In general, we have the following definition.

DEFINITION 1.4.1: A *random variable* is a function from a sample space S into the real numbers.

Example 1.4.1: In most experiments random variables are implicitly used. Here are some examples.

Experiment	Random variable
Toss two dice	X = sum of the numbers
Toss a coin 25 times	X = number of heads in 25 tosses
Apply different levels of fertilizer to corn plants	X = yield/acre

||

In defining a random variable, we have also defined a new sample space (the range of the random variable). We must now check formally that our probability function, which is defined on the original sample space, can be used for the random variable.

Suppose we have a sample space

$$S = \{s_1, \ldots, s_n\}$$

with a probability function P and we define a random variable X with range $\mathcal{X} = \{x_1, \ldots, x_m\}$. We can define a probability function P_X on $\mathcal{X}$ in the following way. We will observe $X = x_i$ if and only if the outcome of the random experiment is an $s_j \in S$ such that $X(s_j) = x_i$. Thus,

$$P_X(X = x_i) = P\left(\{s_j \in S : X(s_j) = x_i\}\right). \tag{1.4.1}$$

Note that the left-hand side of (1.4.1), the function P_X, is an *induced* probability function on $\mathcal{X}$, defined in terms of the original function P. Equation (1.4.1) formally defines a probability function, P_X, for the random variable X. Of course, we have to verify that P_X satisfies the Kolmogorov Axioms, but that is not a very difficult job (see Exercise 1.52). Because of the equivalence in (1.4.1), we will simply write $P(X = x_i)$ rather than $P_X(X = x_i)$.

A note on notation: Random variables will always be denoted with uppercase letters and the realized values of the variable (or its range) will be denoted by the corresponding lowercase letter. Thus, the random variable X can take the value x.

Example 1.4.2: Consider again the experiment of tossing a fair coin three times from Example 1.3.7. Define the random variable X to be the number of heads obtained in the three tosses. A complete enumeration of the value of X for each point in the sample space is given at the top of the next page.

s	$X(s)$
HHH	3
HHT	2
HTH	2
THH	2
HTT	1
THT	1
TTH	1
TTT	0

The range for the random variable X is $\mathcal{X} = \{0,1,2,3\}$. Assuming that all eight points in S have probability $\frac{1}{8}$, by simply counting in the above display we see that the induced probability function on $\mathcal{X}$ is this:

x	$P_X(X=x)$
0	$\frac{1}{8}$
1	$\frac{3}{8}$
2	$\frac{3}{8}$
3	$\frac{1}{8}$

For example, $P_X(X=1) = P(\{\text{HTT}, \text{THT}, \text{TTH}\}) = \frac{3}{8}$. ||

Example 1.4.3: It may be possible to determine P_X even if a complete listing, as in Example 1.4.2, is not possible. Let S be the 2^{50} strings of 50 zeros and ones, X = number of 1s, and $\mathcal{X} = \{0,1,2,\ldots,50\}$, as mentioned at the beginning of this section. Suppose that each of the 2^{50} strings is equally likely. The probability that $X = 27$ can be obtained by counting all of the strings with 27 ones in the original sample space. Since each string is equally likely, it follows that

$$P_X(X=27) = \frac{\#\text{ strings with 27 ones}}{\#\text{ strings}} = \frac{\binom{50}{27}}{2^{50}}.$$

In general, for any $i \in \mathcal{X}$,

$$P_X(X=i) = \frac{\binom{50}{i}}{2^{50}}.$$ ||

The previous illustrations had both a finite S and finite $\mathcal{X}$, and definition of P_X was straightforward. Such is also the case if $\mathcal{X}$ is countable. If $\mathcal{X}$ is uncountable, we define the induced probability function, P_X, in a manner similar to (1.4.1). For any set $A \subset \mathcal{X}$,

$$P_X(X \in A) = P(\{s \in S : X(s) \in A\}). \tag{1.4.2}$$

This does define a legitimate probability function for which the Kolmogorov Axioms can be verified. (To be precise, we use (1.4.2) to define probabilities only for a certain Borel field of subsets of $\mathcal{X}$. But we will not concern ourselves with these technicalities.)

1.5 Distribution Functions

With every random variable X, we associate a function called the cumulative distribution function of X.

DEFINITION 1.5.1: The *cumulative distribution function* or *cdf* of a random variable X, denoted by $F_X(x)$, is defined by

$$F_X(x) = P_X(X \le x), \quad \text{for all } x.$$

Example 1.5.1: Consider the experiment of tossing three fair coins, and let X = number of heads observed. The cdf of X is

x	$F_X(x)$
$(-\infty, 0)$	0
$[0,1)$	$\frac{1}{8}$
$[1,2)$	$\frac{1}{2}$
$[2,3)$	$\frac{7}{8}$
$[3,\infty)$	1

The step function $F_X(x)$ is graphed in Figure 1.5.1.

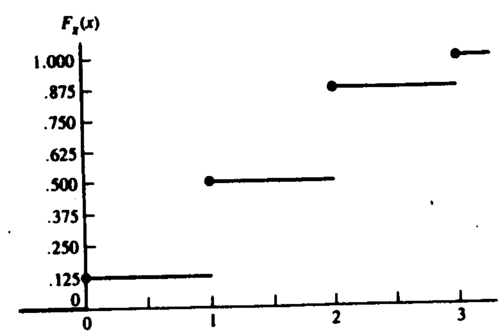


FIGURE 1.5.1 Cdf of Example 1.5.1

There are several points to note from Figure 1.5.1. F_X is defined for all values of x, not just those in $\mathcal{X} = \{0, 1, 2, 3\}$. Thus, for example,

$$F_X(2.5) = P(X \leq 2.5) = P(X = 0, 1, \text{ or } 2) = \frac{7}{8}.$$

Note that F_X has jumps at the values of $x_i \in \mathcal{X}$ and the size of the jump at x_i is equal to $P(X = x_i)$. Also, $F_X(x) = 0$ for $x < 0$ since X cannot be negative and $F_X(x) = 1$ for $x \geq 3$ since x is certain to be less than such a value. ||

As is apparent from Figure 1.5.1, F_X can be discontinuous, with jumps at certain values of x. By the way in which F_X is defined, however, at the jump points F_X takes the value at the top of the jump. (Note the open and closed endpoints on the intervals in Example 1.5.1.) This is known as *right-continuity*—the function is continuous when a point is approached from the right. The property of right-continuity is a consequence of the definition of the cdf. In contrast, if we had defined $F_X(x) = P_X(X < x)$ (note strict inequality), F_X would then be *left-continuous*. The size of the jump at any point x is equal to $P(X = x)$.

Every cdf satisfies certain properties, some of which are obvious when we think of the definition of $F_X(x)$ in terms of probabilities.

THEOREM 1.5.1: The function $F(x)$ is a cdf if and only if the following three conditions hold:

a. $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to\infty} F(x) = 1$.
b. $F(x)$ is a nondecreasing function of x.
c. $F(x)$ is right-continuous. That is, for every number x_0, $\lim_{x\downarrow x_0} F(x) = F(x_0)$.

Outline of proof: To prove necessity, the three properties can be verified by writing F in terms of the probability function (Exercise 1.55). To prove sufficiency, that if a function F satisfies the three conditions of the theorem then it is a cdf for some random variable, is much harder. It must be established that there exists a sample space S, a probability function P on S, and a random variable X defined on S such that F is the cdf of X. □

Example 1.5.2: Suppose we do an experiment that consists of tossing a coin until a head appears. Let p = probability of a head on any given toss and define a random variable X = number of tosses required to get a head. Then, for any $x = 1, 2, \ldots$,

$$P(X = x) = (1-p)^{x-1}p, \tag{1.5.1}$$

since we must get $x - 1$ tails followed by a head for the event to occur and all trials are independent. From (1.5.1) we calculate, for any positive integer x,

$$P(X \leq x) = \sum_{i=1}^{x} P(X = i) \tag{1.5.2}$$

$$= \sum_{i=1}^{x} (1-p)^{i-1}p, \quad x = 1, 2, \ldots.$$

The partial sum of the geometric series is

$$\sum_{k=1}^{n} t^{k-1} = \frac{1-t^n}{1-t}, \quad t \neq 1, \tag{1.5.3}$$

a fact that can be established by induction (see Exercise 1.57). Applying (1.5.3) to our probability, we find that the cdf of the random variable X is

$$\begin{aligned} F_X(x) &= P(X \leq x) \\ &= \frac{1-(1-p)^x}{1-(1-p)}p \\ &= 1-(1-p)^x, \quad x = 1, 2, \ldots. \end{aligned}$$

The cdf $F_X(x)$ is flat between the nonnegative integers, as in Example 1.5.1.

It is easy to show that if $0 < p < 1$, then $F_X(x)$ satisfies the conditions of Theorem 1.5.1. First,

$$\lim_{x\to-\infty} F_X(x) = 0$$

since $F_X(x) = 0$ for all $x < 0$, and

$$\lim_{x\to\infty} F_X(x) = \lim_{x\to\infty} 1-(1-p)^x = 1,$$

where x goes only through integer values when this limit is taken. To verify property (b), we simply note that the sum in (1.5.2) contains more *positive* terms as x increases. Finally, to verify (c), note that, for any x, $F_X(x+\epsilon) = F_X(x)$ if $\epsilon > 0$ is sufficiently small. Hence,

$$\lim_{\epsilon\downarrow 0} F_X(x+\epsilon) = F_X(x),$$

so $F_X(x)$ is right-continuous. $F_X(x)$ is the cdf of a distribution called the *geometric distribution* (after the series) and is pictured in Figure 1.5.2 at the top of page 32. ||

Example 1.5.3: An example of a continuous cdf is the function

$$F_X(x) = \frac{1}{1+e^{-x}}, \tag{1.5.4}$$

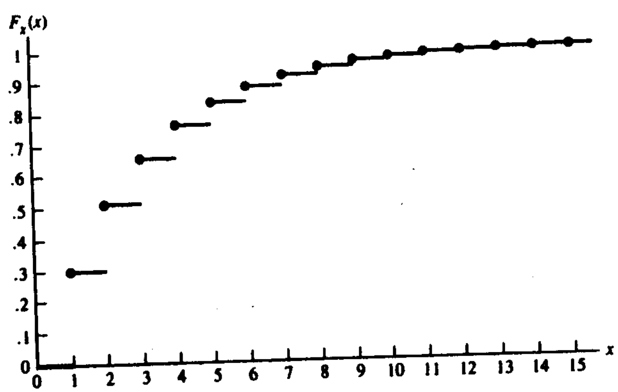


FIGURE 1.5.2 Geometric cdf, $p = .3$

which satisfies the conditions of Theorem 1.5.1. For example,

$$\lim_{x\to-\infty} F_X(x) = 0 \quad \text{since} \quad \lim_{x\to-\infty} e^{-x} = \infty$$

and

$$\lim_{x\to\infty} F_X(x) = 1 \quad \text{since} \quad \lim_{x\to\infty} e^{-x} = 0.$$

Differentiating $F_X(x)$ gives

$$\frac{d}{dx}F_X(x) = \frac{e^{-x}}{(1+e^{-x})^2} > 0,$$

showing that $F_X(x)$ is increasing. F_X is not only right-continuous, but also continuous. This is a special case of the *logistic distribution*. ||

Example 1.5.4: If F_X is not a continuous function of x, it is possible for it to be a mixture of continuous pieces and jumps. For example, if we modify $F_X(x)$ of (1.5.4) to be, for some $\epsilon, 1 > \epsilon > 0$,

$$F_Y(y) = \begin{cases} \dfrac{1-\epsilon}{1+e^{-y}} & \text{if } y < 0 \\ \epsilon + \dfrac{(1-\epsilon)}{1+e^{-y}} & \text{if } y \geq 0 \end{cases} \qquad (1.5.5)$$

then $F_Y(y)$ is the cdf of a random variable Y (see Exercise 1.54). The function F_Y has a jump of height ϵ at $y = 0$, and otherwise is continuous. This model might be appropriate if we were observing the reading from a gauge, a reading that could (theoretically) be anywhere between $-\infty$ and ∞. This particular gauge, however, sometimes sticks at 0. We could then model our observations with F_Y where ϵ is the probability that the gauge sticks. ||

Whether a cdf is continuous or has jumps corresponds to the associated random variable being continuous or not. In fact, the association is such that it is convenient to define continuous random variables in this way.

DEFINITION 1.5.2: A random variable X is *continuous* if $F_X(x)$ is a continuous function of x. A random variable X is *discrete* if $F_X(x)$ is a step function of x.

We close this section with a theorem formally stating that F_X completely determines the probability distribution of a random variable X. We first need the notion of two random variables being identically distributed.

DEFINITION 1.5.3: The random variables X and Y are *identically distributed* if for every set A, $P(X \in A) = P(Y \in A)$.

Note that two random variables that are identically distributed are not necessarily equal. That is, Definition 1.5.3 does not say that $X = Y$.

Example 1.5.5: Consider the experiment of tossing a fair coin three times as in Example 1.4.2. Define the random variables X and Y by

$$X = \text{number of heads observed}$$

and

$$Y = \text{number of tails observed.}$$

The distribution of X is given in Example 1.4.2 and it is easily verified that the distribution of Y is exactly the same. That is, for each $k = 0,1,2,3$, we have $P(X = k) = P(Y = k)$. So X and Y are identically distributed. However, for no sample points do we have $X(s) = Y(s)$. ||

THEOREM 1.5.2: The following two statements are equivalent:

a. The random variables X and Y are identically distributed.
b. $F_X(x) = F_Y(x)$ for every x.

Proof: To show equivalence we must show that each statement implies the other. We first show that (a) $\Rightarrow$ (b).

Since X and Y are identically distributed, we have for any set A,

$$P(X \in A) = P(Y \in A).$$

In particular, for a set $(-\infty, x]$ we have

$$P(X \in (-\infty, x]) = P(Y \in (-\infty, x]) \quad \text{for all } x.$$

But this last equality is

$$P(X \le x) = P(Y \le x), \quad \text{for all } x,$$

or that $F_X(x) = F_Y(x)$, for all x.

The converse implication, that (b) $\Rightarrow$ (a), is much more difficult to prove. The above argument showed that if the X and Y probabilities agreed on all sets, then they agreed on intervals. We now must prove the opposite, that is, if the X and Y probabilities agree on all intervals, then they agree on all sets. To show this requires heavy use of Borel fields; we will not go into these details here. Suffice it to say that it is necessary to prove only that the two probability functions agree on all intervals (Chung, 1974). □

1.6 Density and Mass Functions

Associated with a random variable X and its cdf F_X is another function, called either the probability density function (pdf) or probability mass function (pmf). The terms pdf and pmf refer, respectively, to the continuous and discrete cases. Both pdfs and pmfs are concerned with "point probabilities" of random variables.

DEFINITION 1.6.1: The *probability mass function (pmf)* of a discrete random variable X is given by

$$f_X(x) = P(X = x) \quad \text{for all } x.$$

Example 1.6.1: For the geometric distribution of Example 1.5.2, we have the pmf

$$f_X(x) = P(X = x) = \begin{cases} (1-p)^{x-1}p & \text{for } x = 1, 2, \ldots \\ 0 & \text{otherwise} \end{cases}$$

Recall that $P(X = x)$ or, equivalently, $f_X(x)$ is the size of the jump in the cdf at x. We can use the pmf to calculate probabilities. Since we can now measure the probability of a single point, we need only sum over all of the points in the appropriate event. Hence, for positive integers a and b, with $a \le b$, we have

$$\begin{aligned} P(a \le X \le b) &= \sum_{k=a}^{b} f_X(k) \\ &= \sum_{k=a}^{b} (1-p)^{k-1} p. \end{aligned}$$

As a special case of this we get

$$P(X \le b) = \sum_{k=1}^{b} f_X(b) = F_X(b). \tag{1.6.1}$$

||

A widely accepted convention, which we will adopt, is to use an uppercase letter for the cdf and the corresponding lowercase letter for the pmf or pdf.

We must be a little more careful in our definition of a pdf in the continuous case. If we naively try to calculate $P(X = x)$ for a continuous random variable, we get the following. Since $\{X = x\} \subset \{x - \epsilon < X \le x\}$ for any $\epsilon > 0$, we have from Theorem 1.2.2(c) that

$$P(X = x) \le P(x - \epsilon < X \le x) = F_X(x) - F_X(x - \epsilon),$$

for any $\epsilon > 0$. Therefore,

$$0 \le P(X = x) \le \lim_{\epsilon \downarrow 0} [F_X(x) - F_X(x - \epsilon)] = 0,$$

by the continuity of F_X. However, if we understand the purpose of the pdf, its definition will become clear.

From Example 1.6.1, we see that a pmf gives us "point probabilities." In the discrete case, we can sum over values of the pmf to get the cdf (as in (1.6.1)). The analogous procedure in the continuous case is to substitute integrals for sums, and we get

$$P(X \le x) = F_X(x) = \int_{-\infty}^{x} f_X(t)dt.$$

Using the Fundamental Theorem of Calculus, if $f_X(x)$ is continuous, we have the further relationship

$$\frac{d}{dx} F_X(x) = f_X(x). \tag{1.6.2}$$

Note that the analogy with the discrete case is almost exact. We "add up" the "point probabilities" $f_X(x)$ to obtain interval probabilities.

DEFINITION 1.6.2: The *probability density function* or *pdf*, $f_X(x)$, of a continuous random variable X is the function that satisfies

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt \quad \text{for all } x. \tag{1.6.3}$$

A note on notation: The expression "X has a distribution given by $F_X(x)$" is abbreviated symbolically by "$X \sim F_X(x)$," where we read the symbol "$\sim$" as "is distributed as." We can similarly write $X \sim f_X(x)$ or, if X and Y have the same distribution, $X \sim Y$.

In the continuous case we can be somewhat cavalier about the specification of interval probabilities. Since $P(X = x) = 0$ if X is a continuous random variable,

$$P(a < X < b) = P(a < X \leq b) = P(a \leq X < b) = P(a \leq X \leq b).$$

It should be clear that the pdf (or pmf) contains the same information as the cdf. This being the case, we can use either one to solve problems and should try to choose the simpler one.

Example 1.6.2: For the logistic distribution of Example 1.5.3 we have

$$F_X(x) = \frac{1}{1+e^{-x}}$$

and, hence,

$$f_X(x) = \frac{d}{dx}F_X(x) = \frac{e^{-x}}{(1+e^{-x})^2}.$$

The area under the curve $f_X(x)$ gives us interval probabilities (see Figure 1.6.1):

$$P(a < X < b) = F_X(b) - F_X(a) = \int_{-\infty}^{b} f_X(x)dx - \int_{-\infty}^{a} f_X(x)dx = \int_{a}^{b} f_X(x)dx.$$

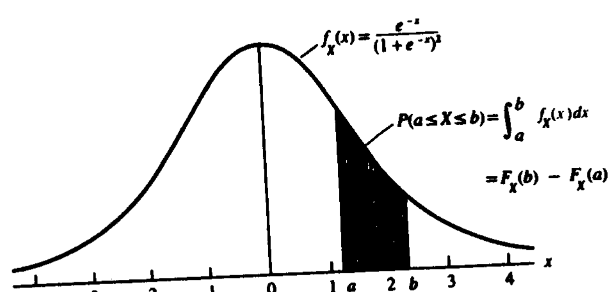


FIGURE 1.6.1 Area under logistic curve ||

There are really only two requirements for a pdf (or pmf), both of which are immediate consequences of the definition.

THEOREM 1.6.1: A function $f_X(x)$ is a pdf (or pmf) of a random variable X if and only if

a. $f_X(x) \geq 0$ for all x.

b. $\sum_x f_X(x) = 1$ (pmf) or $\int_{-\infty}^{\infty} f_X(x)dx = 1$ (pdf).

Proof: If $f_X(x)$ is a pdf (or pmf), then the two properties are immediate from the definitions. In particular, for a pdf, using (1.6.3) and Theorem 1.5.1, we have that

$$1 = \lim_{x\to\infty} F_X(x) = \int_{-\infty}^{\infty} f_X(t)dt.$$

The converse implication is equally easy to prove. Once we have $f_X(x)$ we can define $F_X(x)$ and appeal to Theorem 1.5.1. □

From a purely mathematical viewpoint, any nonnegative function with a finite positive integral (or sum) can be turned into a pdf or pmf. For example, if $h(x)$ is any nonnegative function that is positive on a set A, 0 elsewhere, and

$$\int_{\{x\in A\}} h(x)dx = K < \infty$$

for some constant $K > 0$, then the function $f_X(x) = h(x)/K$ is a pdf of a random variable X taking values in A.

Actually, the relationship (1.6.3) does not always hold because $F_X(x)$ may be continuous but not differentiable. In fact, there exist continuous random variables for which the integral relationship does not exist for *any* $f_X(x)$. These cases are rather pathological and we will ignore them. Thus, in this text, we will assume that (1.6.3) holds for any continuous random variable. In more advanced texts (for example, Chung (1974)) a random variable is called ***absolutely continuous*** if (1.6.3) holds.

EXERCISES

1.1 For each of the following experiments, describe the sample space.
  a. Toss a coin four times.
  b. Count the number of insect-damaged leaves on a plant.
  c. Measure the lifetime (in hours) of a particular brand of light bulb.
  d. Record the weights of 10-day-old rats.
  e. Observe the proportion of defectives in a shipment of electronic components.

1.2 Verify the following identities.
  a. $A\backslash B = A\backslash(A\cap B) = A\cap B^c$ b. $B = (B\cap A)\cup(B\cap A^c)$
  c. $B\backslash A = B\cap A^c$ d. $A\cup B = A\cup(B\cap A^c)$.

1.3 Finish the proof of Theorem 1.1.1. For any events A, B, and C defined on a sample space S, show that
  a. $A\cup B = B\cup A$ and $A\cap B = B\cap A$. (Commutativity)
  b. $A\cup(B\cup C) = (A\cup B)\cup C$ and $A\cap(B\cap C) = (A\cap B)\cap C$. (Associativity)
  c. $(A\cup B)^c = A^c\cap B^c$ and $(A\cap B)^c = A^c\cup B^c$. (DeMorgan's Laws)

1.4 Prove the general version of DeMorgan's Laws. Let $\{A_\alpha : \alpha \in \Gamma\}$ be a (possibly uncountable) collection of sets. Prove that
  a. $(\cup_\alpha A_\alpha)^c = \cap_\alpha A_\alpha^c$. b. $(\cap_\alpha A_\alpha)^c = \cup_\alpha A_\alpha^c$.

1.5 Formulate and prove a version of DeMorgan's Laws that applies to a finite collection of sets $A_1, \ldots, A_n$.

1.6 Let S be a sample space. Show that the collection $B = \{\emptyset, S\}$ is a Borel field.

1.7 Let S be a sample space, and let $B = \{$all subsets of S, including S itself$\}$. Show that B is a Borel field.

1.8 Show that the intersection of two Borel fields is a Borel field.

1.9 It was noted in Section 1.2.1 that statisticians who follow the De Finetti school do not accept the Axiom of Countable Additivity. However, this axiom can be derived from two others that are somewhat more self-evident. Suppose we have an infinite sequence of nested sets $A_1 \supset A_2 \supset \cdots \supset A_n \supset \cdots$ that are decreasing to the empty set, which we denote by $A_n \downarrow \emptyset$. Consider the following:

Axiom of Continuity: If $A_n \downarrow \emptyset$ then $P(A_n) \to 0$.

Prove that the Axiom of Continuity together with the Axiom of Finite Additivity imply the Axiom of Countable Additivity.

1.10 If $P(A) = \frac{1}{3}$ and $P(B^c) = \frac{1}{4}$, can A and B be disjoint? Explain.

1.11 Refer to the game of darts explained in Example 1.2.4, and to Figure 1.2.1.

a. Derive the general formula for the probability of scoring i points.

b. Show that $P(\text{scoring } i \text{ points})$ is a decreasing function of i, that is, as the points increase, the probability of scoring them decreases.

c. Show that $P(\text{scoring } i \text{ points})$ satisfies the Kolmogorov Axioms.

1.12 For events A and B, find formulas for the probabilities of the following events in terms of the quantities $P(A)$, $P(B)$, and $P(A \cap B)$.

a. either A or B or both
b. either A or B but not both
c. at least one of A or B
d. at most one of A or B

1.13 Suppose that a sample space S has n elements. Prove that the number of subsets that can be formed from the elements of S is 2^n.

1.14 Finish the proof of Theorem 1.2.4. Use the result established for $k = 2$ as the basis of an induction argument.

1.15 How many different sets of initials can be formed if every person has one surname and

a. exactly two given names?
b. either one or two given names?
c. either one or two or three given names?

1.16 In the game of dominos, each piece is marked with two numbers. The pieces are symmetrical so that the number pair is not ordered (so, for example, $(2,6) = (6,2)$). How many different pieces can be formed using the numbers $1, 2, \ldots, n$?

1.17 If n balls are placed at random into n cells, find the probability that exactly one cell remains empty.

1.18 There is interest in obtaining information about two species of fish that are known to inhabit a particular pond (other species will be ignored). Samples of size 5 will be drawn from the pond.

a. If there are M fish of species A and N of species B, describe the sample space of the experiment. In particular, how many elements are in the sample space?

b. Assuming that all samples are equally likely, what is the probability of getting at least three fish of species B in a sample?

c. A second experiment is now planned, in which there is interest in species A, B, C, and D. Samples of size 8 are to be taken. How many different samples are there?

1.19 If a multivariate function has continuous partial derivatives, the order in which the derivatives are calculated does not matter. Thus, for example, the function $f(x, y)$ of two variables has equal third partials

$$\frac{\partial^3}{\partial x^2 \partial y} f(x,y) = \frac{\partial^3}{\partial y \partial x^2} f(x,y).$$

a. How many fourth partial derivatives does a function of three variables have?

b. Prove that a function of n variables has $\binom{n+r-1}{r}$ rth partial derivatives.

1.20 A secretary types four letters to four people and addresses the four envelopes. If he inserts the letters at random, one in each envelope, what is the probability that exactly two letters will go into the correct envelopes? exactly three?

1.21 In a certain family, the four children take turns washing dishes. Out of a total of four breakages, three were caused by the youngest child, who thereafter was called clumsy. Was it justified to call the child clumsy, or could such an occurrence be reasonably attributed to chance?

1.22 A committee is to consist of four academicians and two industrialists, to be chosen from a larger group of eight academicians and five industrialists. How many ways can a committee be formed if

a. there are no additional restrictions?

b. two of the chosen academicians must be the two female members of the group of eight?

1.23 A way of approximating large factorials is through the use of *Stirling's Formula*:

$$n! \approx \sqrt{2\pi} n^{n+(1/2)} e^{-n},$$

a complete derivation of which is difficult. Instead, prove the easier fact

$$\lim_{n \to \infty} \frac{n!}{n^{n+(1/2)} e^{-n}} = \text{a constant}.$$

(*Hint:* Feller (1968) proceeds by using the monotonicity of the logarithm to establish that

$$\int_{k-1}^{k} \log x \, dx < \log k < \int_{k}^{k+1} \log x \, dx, \quad k = 1, \ldots, n,$$

and hence

$$\int_0^n \log x \, dx < \log n! < \int_1^{n+1} \log x \, dx.$$

Now compare $\log n!$ to the average of the two integrals. See Exercise 5.21 for another derivation.)

1.24 My telephone rings 12 times each week, the calls being randomly distributed among the 7 days. What is the probability that I get at least one call each day?

1.25 Verify the following identities for $n \geq 2$.

a. $\sum_{k=0}^{n} (-1)^k \binom{n}{k} = 0$
b. $\sum_{k=1}^{n} k \binom{n}{k} = n2^{n-1}$
c. $\sum_{k=1}^{n} (-1)^{k+1} k \binom{n}{k} = 0$

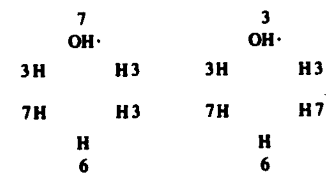
1.26 A man is given n keys, in random order, of which only one will unlock his door. He tries them successively (sampling without replacement). This procedure may require $1, 2, \ldots, n$ trials. Show that each of the n outcomes has probability $1/n$. (In other words, $P(\text{man succeeds on } k\text{th trial}) = 1/n, k = 1, 2, \ldots, n$.)

1.27 A closet contains n pairs of shoes. If $2r$ shoes are chosen at random ($2r < n$), what is the probability that there will be no matching pair in the sample?

1.28 In a draft lottery containing the 366 days of the year (including February 29), what is the probability that the first 180 days drawn (without replacement) are evenly distributed among the 12 months? What is the probability that the first 30 days drawn contain none from September?

1.29 Two people each toss a fair coin n times. Find the probability that they will score the same number of heads.

1.30 An electron spin resonance spectrometer (ESR) detects atoms in a molecule that have a free (unpaired) electron. A *hyperfine coupling constant* measures the magnetic attraction of the electron to such an atom. The output of an ESR gives a set of possible hyperfine coupling constants for the atoms in a molecule. Although each atom corresponds to one coupling constant from the set, different atoms can have the same constant. The exact matching of atoms to constants is immaterial because, for a given set of constants, all possible arrangements of those constants correspond to the same compound. For example, if the set of possible constants is $\{1, 2, \ldots, 10\}$, then the two assignments of constants $\{3, 3, 3, 6, 7, 7\}$ diagrammed below are considered equivalent. If there are n atoms in a molecule, and r distinct values in the set of possible coupling constants, how many different coupling assignments are there? (See Duling, Motten, and Mason (1988) for a more thorough description of this problem.)



1.31 An employer is about to hire one new employee from a group of N candidates, whose future potential can be rated on a scale from 1 to N. The employer proceeds according to the following rules:

i. Each candidate is seen in succession (in random order) and a decision is made whether to hire the candidate.
ii. Having rejected $m-1$ candidates ($m > 1$), the employer can hire the mth candidate only if the mth candidate is better than the previous $m-1$.

Suppose a candidate is hired on the ith trial. What is the probability that the best candidate was hired?

1.32 Approximately one-third of all human twins are identical (one-egg) and two-thirds are fraternal (two-egg) twins. Identical twins are necessarily the same sex, with male and female being equally likely. Among fraternal twins, approximately one-fourth are both female, one-fourth are both male, and half are one male and one female. Finally, among all U.S. births, approximately 1 in 90 is a twin birth. Define the following events:

$$A = \{\text{a U.S. birth results in twin females}\}$$

$$B = \{\text{a U.S. birth results in identical twins}\}$$

$$C = \{\text{a U.S. birth results in twins}\}$$

a. State, in words, the event $A \cap B \cap C$.
b. Find $P(A \cap B \cap C)$.

1.33 Two pennies, one with $P(\text{head}) = u$ and one with $P(\text{head}) = w$, are to be tossed together independently. Define

$$p_0 = P(0 \text{ heads occur}),$$

$$p_1 = P(1 \text{ head occurs}),$$

$$p_2 = P(2 \text{ heads occur}).$$

Can u and w be chosen such that $p_0 = p_1 = p_2$? Prove your answer.

1.34 In a town of $n+1$ inhabitants, a person tells a rumor to a second person, who in turn repeats it to a third person, etc. At each step the recipient of the rumor is chosen at random from the n people. Find the probability that the rumor will be told exactly r times
a. before returning to the originator. b. without being repeated to any person.

1.35 An airport bus deposits 25 passengers at 7 stops. Each passenger is as likely to get off at any stop as at any other, and the passengers act independently of one another. The bus makes a stop only if someone wants to get off. What is the probability that nobody gets off at the third stop?

1.36 Two players, A and B, alternately and independently flip a coin and the first player to obtain a head wins. Assume player A flips first.
a. If the coin is fair, what is the probability that A wins?
b. Suppose that $P(\text{head}) = p$, not necessarily $\frac{1}{2}$. What is the probability that A wins?
c. Show that for all $p, 0 < p < 1, P(\text{A wins}) > \frac{1}{2}$. (*Hint:* Try to write $P(\text{A wins})$ in terms of the events $E_1, E_2, \ldots$, where $E_i = \{\text{head first appears on } i\text{th toss}\}$.)

1.37 Suppose that 5% of men and .25% of women are color-blind. A person is chosen at random and that person is color-blind. What is the probability that the person is male? (Assume males and females to be in equal numbers.)

1.38 An insurance company has three types of customers—high risk, medium risk, and low risk. Twenty percent of its customers are high risk, 30% are medium risk, and 50% are low risk. Also, the probability that a customer has at least one accident in the current year is .25 for high risk, .16 for medium risk, and .10 for low risk.
a. Find the probability that a customer chosen at random will have at least one accident in the current year.
b. Find the probability that a customer is high risk, given that the person has had at least one accident during the current year.

1.39 In a certain population, 1% of the people are color-blind. A subset is to be randomly chosen. How large must the subset be if the probability of its containing at least one color-blind person is to be .95 or more? (Assume that the population is large enough to be considered infinite, so the selection can be considered to be with replacement.)

1.40 Two litters of a particular rodent species have been born, one with two brown-haired and one gray-haired (litter 1), and the other with three brown-haired and two gray-haired (litter 2). We select a litter at random, and then select an offspring at random from the selected litter.
a. What is the probability that the animal chosen is brown-haired?
b. Given that a brown-haired offspring was selected, what is the probability that the sampling was from litter 1?

1.41 Prove that if $P(\cdot)$ is a legitimate probability function and B is a set with $P(B) > 0$, then $P(\cdot|B)$ also satisfies Kolmogorov's Axioms.

1.42 Refer to the dart game of Example 1.2.4. Suppose we do not assume that the probability of hitting the dart board is 1, but rather is proportional to the area of the dart board. Assume that the dart board is mounted on a wall that is hit with probability 1, and the wall has area A.
a. Using the fact that the probability of hitting a region is proportional to area, construct

a probability function for $P(\text{scoring } i \text{ points})$, $i = 0, \ldots, 5$. (No points are scored if the dart board is not hit.)

b. Show that the conditional probability distribution $P(\text{scoring } i \text{ points}|\text{board is hit})$ is exactly the probability distribution of Example 1.2.4.

1.43 Prove each of the following statements. (Assume that any conditioning event has positive probability.)

a. If $P(B) = 1$ then $P(A|B) = P(A)$ for any A.

b. If $A \subset B$ then $P(B|A) = 1$ and $P(A|B) = P(A)/P(B)$.

c. If A and B are mutually exclusive, then

$$P(A|A \cup B) = \frac{P(A)}{P(A) + P(B)}.$$

d. $P(A \cap B \cap C) = P(A|B \cap C)P(B|C)P(C)$.

1.44 A pair of events A and B cannot be simultaneously *mutually exclusive* and *independent*. Prove that if $P(A) > 0$ and $P(B) > 0$, then

a. If A and B are mutually exclusive they cannot be independent.

b. If A and B are independent they cannot be mutually exclusive.

1.45 Finish the proof of Theorem 1.3.1. If A and B are independent events, show that the following pairs are also independent.

a. A^c and B b. A^c and B^c

1.46 If the probability of hitting a target is $\frac{1}{5}$, and ten shots are fired independently, what is the probability of the target being hit at least twice? What is the conditional probability that the target is hit at least twice, given that it is hit at least once?

1.47 Here we will look at some variations of Example 1.3.2. A similar, but somewhat more complicated, problem is discussed by Selvin (1975).

a. In the warden's calculation of Example 1.3.2 it was assumed that if A were to be pardoned, then with equal probability the warden would tell A that either B or C would die. However, this need not be the case. The warden can assign probabilities γ and $1 - \gamma$ to these events, as shown here:

Prisoner pardoned	Warden tells A	
A	B dies	with probability γ
A	C dies	with probability $1 - \gamma$
B	C dies	
C	B dies	

Calculate $P(A|W)$ as a function of γ. For what values of γ is $P(A|W)$ less than, equal to, or greater than $\frac{1}{3}$?

b. Suppose again that $\gamma = \frac{1}{2}$, as in the example. After the warden tells A that B will die, A thinks for a while and realizes that his original calculation was false. However, A then gets a bright idea. A asks the warden if he can swap fates with C. The warden, thinking that no information has been passed, agrees to this. Prove that A's reasoning is now correct and that his probability of survival has jumped to $\frac{2}{3}$!

1.48 A fair die is cast until a 6 appears. What is the probability that it must be cast more than five times?

1.49 The Smiths have two children. At least one of them is a boy. What is the probability that both children are boys? (See Gardner (1961) for a complete discussion of this problem.)

1.50 As in Example 1.3.3, consider telegraph signals "dot" and "dash" sent in the proportion 3:4, where erratic transmissions cause a dot to become a dash with probability $\frac{1}{4}$, and a dash to become a dot with probability $\frac{1}{3}$.

a. If a dash is received, what is the probability that a dash has been sent?

b. Assuming independence between signals, if the message dot–dot was received, what is the probability distribution of the four possible messages that could have been sent?

1.51 Standardized tests provide an interesting application of probability theory. Suppose first that a test consists of 20 multiple-choice questions, each with 4 possible answers. If the student guesses on each question, then the taking of the exam can be modeled as a sequence of 20 independent events. Find the probability that the student gets at least 10 questions correct, given that he is guessing.

1.52 Show that the induced probability function defined in (1.4.1) defines a legitimate probability function in that it satisfies the Kolmogorov Axioms.

1.53 Seven balls are distributed randomly into seven cells. Let X_i = the number of cells containing exactly i balls. What is the probability distribution of X_3? (That is, find $P(X_3 = x)$ for every possible x.)

1.54 Prove that the following functions are cdfs.

a. $\frac{1}{2} + \frac{1}{\pi}\tan^{-1}(x), x \in (-\infty, \infty)$

b. $(1 + e^{-x})^{-1}, x \in (-\infty, \infty)$

c. $e^{-e^{-x}}, x \in (-\infty, \infty)$

d. $1 - e^{-x}, x \in (0, \infty)$

e. the function defined in (1.5.5)

1.55 Prove the necessity part of Theorem 1.5.1.

1.56 A cdf F_X is *stochastically greater* than a cdf F_Y if $F_X(t) \leq F_Y(t)$ for all t and $F_X(t) < F_Y(t)$ for some t. Prove that if $X \sim F_X$ and $Y \sim F_Y$, then

$$P(X > t) \geq P(Y > t) \quad \text{for every } t$$

and

$$P(X > t) > P(Y > t) \quad \text{for some } t,$$

that is, X tends to be bigger than Y.

1.57 Verify formula (1.5.3), the formula for the partial sum of the geometric series.

1.58 An appliance store receives a shipment of 30 microwave ovens, 5 of which are (unknown to the manager) defective. The store manager selects 4 ovens at random, without replacement, and tests to see if they are defective. Let X = number of defectives found. Calculate the pmf and cdf of X and plot the cdf.

1.59 Let X be a continuous random variable with pdf $f(x)$ and cdf $F(x)$. For a fixed number x_0, define the function

$$g(x) = \begin{cases} f(x)/[1 - F(x_0)] & x \geq x_0 \\ 0 & x < x_0. \end{cases}$$

Prove that $g(x)$ is a pdf. (Assume that $F(x_0) < 1$.)

1.60 A certain river floods every year. Suppose that the low-water mark is set at 1 and the high-water mark Y has distribution function

(d) Reading off the appropriate probabilities from the tree diagram we find that the answer to (d) is

$$
\begin{aligned}
&P(\text{the laboratory will report the presence of Chemical X in a sample})\\
&= P(A_+B_+C_+)\\
&= P(A_+B_+C_+ \mid X_+)P(X_+) + P(A_+B_+C_+ \mid X_-)P(X_-)\\
&= P(X_+)P(A_+ \mid X_+)P(B_+ \mid A_+X_+)P(C_+ \mid A_+B_+X_+)\\
&\quad + P(X_-)P(A_+ \mid X_-)P(B_+ \mid A_+X_-)P(C_+ \mid A_+B_+X_-)\\
&= (.10)(.99)(.98)(.97) + (.90)(.05)(.10)(.15) = .09478
\end{aligned}
$$

In other words, the laboratory will report the presence of Chemical X in approximately 9.5% of the samples.

(e) Reading off the appropriate probabilities from the tree diagram we find that the answer to (e) is

$$
\begin{aligned}
&P(\text{the laboratory will report the absence of Chemical X in a sample})\\
&= P(A_- \cup A_+B_- \cup A_+B_+C_-)\\
&= P(X_+)P(A_- \cup A_+B_- \cup A_+B_+C_- \mid X_+)\\
&\quad + P(X_-)P(A_- \cup A_+B_- \cup A_+B_+C_- \mid X_-)\\
&= (.10)\cdot[(.01) + (.99)(.02) + (.99)(.98)(.03)]\\
&\quad + (.90)\cdot[(.95) + (.05)(.90) + (.05)(.10)(.85)]\\
&= .9052
\end{aligned}
$$

In other words, the laboratory will report the absence of Chemical X in approximately 90.5% of the samples. This answer may, of course, be also obtained by referring to (d).

Exercises

1. Applicants for employment at a certain large firm are initially interviewed by any one of three personnel officers. If an interview is favorable, the applicant is given a written test. If the applicant's test result is favorable, he or she is interviewed by a technical staff member. A favorable report from the technical staff member places the applicant on a waiting list for the next job opening. The following statistics are known concerning the application process at this firm: personnel officers A, B, and C interview 40%, 35%, and 25%, respectively, of the applicants. These personnel officers give favorable reports to 60%, 75%, and 80%, respectively, of the applicants. Only 10% of the applicants who take the written test receive favorable results, and of these applicants only 30% receive favorable reports from the interview with a technical staff member.

   (a) Draw a tree diagram of the empirical probabilities defined by these statistics.
   (b) What percentage of all applicants are eventually placed on the waiting list for employment?
   (c) What percentage of applicants who receive favorable results on their initial interviews with the personnel officer are eventually placed on the waiting list for employment?

2. Alice and Mary are engaged in a chess tournament. The first person to win either 2 games in a row or a total of 3 games will be declared the winner.
   (a) If Mary has a probability of .6 of winning each of the individual games, what is the probability that Mary will win the tournament?
   (b) What is the probability that Mary will win the tournament after fewer than 4 games? (*Hint:* draw a tree diagram of the pertinent probabilities. This problem was considered in Example 2.12 of Chapter 2.)

3. Players A, B, C are engaged in a series of games of chance. The series begins with A and B as opponents. The winner of each game continues the series by playing against the player who was not a participant in the most recent game. The series continues until one player has won 3 games.
   (a) Draw a tree diagram of the possible outcomes with the associated probabilities.
   (b) What is the probability that player A will win the series of games?
   (c) What is the probability that player B will win the series of games? (*Note:* this problem was considered in Exercise 5 of Section 2.5 of Chapter 2.)

Supplementary Exercises

1. Suppose that 5% of males and 0.25% of females are color blind. If a color-blind person is chosen at random, what is the probability that the person is female? Assume that there are an equal number of males and females.

2. Perhaps not all politicians are liars, but candidates A, B, and C are—at least some of the time. Suppose that candidates A, B, and C lie 5%, 10%, and 15% of the time, respectively; tell half-truths 40%, 45%, and 50% of the time, respectively; and are truthful 55%, 45%, and 35% of the time, respectively. Furthermore, suppose that candidate A issues twice as many statements as candidate B and three times as many statements as candidate C.
   (a) If a lie was issued by one of these candidates, what is the probability that candidate A was responsible?
   (b) What percent of the combined statements issued by these candidates are half-truths?

3. Suppose that each child born to a couple is equally likely to be a boy or a girl, independently of the sex of the other children in the family. If a couple has four children, then what are the probabilities of

the following events:
(a) All four children are girls.
(b) All four children are of the same sex.
(c) Exactly two are boys.
(d) The eldest is a girl.
(e) There is at least one boy.

4. Players A and B take turns rolling a die. The first player to get a 6 wins the game. If player A has the first turn, what is the probability that he will be the winner. (*Hint:* see Example 3.16.)

5. Suppose that in Exercise 4 there are three players.
(a) What is the probability that player A will be the winner if he has the first turn?
(b) What is the probability that player B will be the winner if he has the second turn?

6. Twenty items, 6 of which are defective and 14 not defective, are inspected one after the other. If these items are chosen at random, what is the probability that:
(a) The first three items inspected are defective?
(b) The first three items inspected are not defective?
(c) Among the first three items inspected, there is one defective and two not defective?

7. In a nail factory, machines A, B, and C manufacture 20%, 30%, and 50% of the total output, respectively. Of their outputs, 4%, 5%, and 1%, respectively, are defective nails. A nail is chosen at random and found to be defective. What is the probability that the nail came from machine A? B? C?

8. Suppose that in the production of a certain item, defects of one type occur with probability .05 and defects of a second type occur with probability .01. If the occurrence or nonoccurrence of one type of defect does not affect the occurrence or nonoccurrence of the other, what is the probability that, with respect to these types of defects,
(a) an item of this type is defective?
(b) an item of this type has both kinds of defects?
(c) an item of this type has only one type of defect, given that it is defective?

9. A system consists of components A and B. The probability that component A fails is .01, that component B fails is .02, and that both components fail is .001.
(a) Are the events of failure of components A and B independent events?
(b) What is the conditional probability of a failure in component A, given that there is a failure in component B?
(c) What is the conditional probability of a failure in component B, given that there is a failure in component A?

(d) Does the event of a failure in one component increase or decrease the likelihood of failure in the other component?
(e) What is the probability of a failure in at least one of the components?
(f) What is the probability of a failure in exactly one of the components?

10. Let S be a sample space. Let C be an event in S with $P(C) > 0$. Prove
(a) $0 \le P(A \mid C) \le 1$ for all events A in S;
(b) $P(S \mid C) = 1$;
(c) $P(A \cup B \mid C) = P(A \mid C) + P(B \mid C)$ for mutually exclusive events A and B in S (i.e., prove that if P_C is a set function that is defined by the relation $P_C(A) = P(A \mid C)$, then P_C is a probability set function.)

gives us a clue to the meaning of λ. As we saw in Section 4.5.5.2, if X is a binomial random variable with parameters p and n, and if n is large, the distribution of X becomes essentially Poisson with parameter $\lambda = np$. Thus it seems reasonable to conclude that, if the numerical value of an experiment is a Poisson random variable X and if the experiment is repeated a number of times, we would expect the average value of X to be λ. For example, if the number of misprints on a page is a Poisson random variable X with parameter $\lambda = 3$, on the average 3 misprints per page can be expected.

Exercises

1. Show that the Bernoulli distribution is the binomial distribution with $n = 1$.
2. Prove that $b(x; n, p) = b(n - x; n, 1 - p)$.
3. Use Table 1 in the Appendix to find (a) $b(7; 10, .3)$; (b) $b(6; 16, .4)$; (c) $b(10; 15, .6)$; (d) $b(3; 7, .75)$. (*Hint:* refer to Exercise 2 for parts (c) and (d).)
4. Assume that 1% of model airplanes are sold with missing parts. If a child receives five model airplanes on her birthday, what is the probability that one of them has missing parts?
5. If the probability is .05 that a baseball game will go into an extra inning, find the probabilities that among 15 baseball games
   (a) exactly one will go into an extra inning;
   (b) at most 2 will go into extra innings;
   (c) at least 2 will go into extra innings.
6. If the probability that the Internal Revenue Service will audit an income tax return reporting gross income over $100,000 is .65, find the probabilities that among 20 such returns
   (a) exactly 10 will be audited;
   (b) at least 10 will be audited;
   (c) between 8 and 12 (inclusive) will be audited.
7. If one-third of the employees of a certain company bring their lunches, find the probabilities that among 4 such employees
   (a) exactly two bring their lunches;
   (b) none of them bring their lunches;
   (c) more than one of them bring their lunches.
8. An editor wants to check whether her new proofreader meets the standards of missing typographical errors in no more than 1% of the galley proofs of a book. She decides to proofread 10 randomly selected galley proofs of a large book that has already been checked by her new proofreader. If all 10 pages are free of typographical errors, she will assume that the proofreader meets the standards. If more than 2 pages have uncorrected typographical errors, she will fire the proofreader. If one or two pages have typographical errors, she will read another 10 pages before passing judgment. Assuming that the editor is a flawless proofreader, what is the probability that after reading 10 pages she will
   (a) fire the proofreader even though only 1% of the pages have typographical errors;
   (b) assume that the proofreader meets the standards when 5% of the pages have typographical errors;
   (c) burden herself with another 10 pages of proofreading when only 1% of the pages have typographical errors?

9. A quality-control inspector examines 10 out of every shipment of 1,000 packages of frozen chicken for spoilage. The shipment is approved if and only if all 10 packages pass inspection. Find the probability of a shipment being approved even though 10% of the packages are badly spoiled.
10. Ms. Jones drives into town once a week to see her psychiatrist. In the past she parked her car at a lot for $5, but she decided that in the next 5 weeks she will park at the fire hydrant and risk getting tickets with fines of $25 per offense. If the probability of getting a ticket is .1, what is the probability that Ms. Jones will pay more in fines in 5 weeks than she would pay in parking fees if she had opted not to break the law?
11. A quality-control inspector examines a random selection of 3 rifles from each case of 24 rifles that is ready for shipment to retailers. If such a case contains 5 rifles with imperfections, find the probabilities that the inspector's sample of 3 will
   (a) contain none of the imperfect rifles;
   (b) contain exactly one rifle with imperfections;
   (c) contain at least one rifle with imperfections.
12. If 4 of 20 recently acquired works of art are frauds, what is the probability that an art inspector who randomly selects 5 of these works for inspection will have a sample that
   (a) contains none of the frauds;
   (b) contains all of the frauds;
   (c) contains only one fraud.
13. An employer asks a random sampling of 3 of his 100 employees whether they would prefer a 9–5 or an 8–4 workday.
   (a) If 50% of his employees prefer the 9–5 schedule, what is the probability that the sample will contain two who prefer the 9–5 schedule?
   (b) Use the binomial distribution to approximate the answer to (a).
14. A random sampling of 4 members of a 150-member club has shown that 3 prefer no smoking in the clubhouse dining room. What is the probability that this will occur if in fact only 20% of members prefer no smoking in the dining room? Find this probability assuming that the sample was obtained under
   (a) sampling without replacement, and

(b) sampling with replacement.
(c) Compare the two answers.

15. Find the following binomial probabilities and compare with the Poisson approximations.
(a) $b(3; 20, .01)$ (b) $b(5; 100, .05)$

16. Let X be the number of telephone calls arriving within a fixed time interval at a switchboard. Explain why it may be reasonable to assume that X is a Poisson random variable.

17. Suppose that the number of letters lost in the mail in a certain city on a given day is a Poisson random variable X with parameter $\lambda = 5$. What is the probability that in a given day
(a) exactly 5 letters will be lost in the mail;
(b) at least one letter will be lost in the mail.

18. Use the Poisson approximation of the binomial distribution to determine the probabilities that
(a) there will be 15 heads in 30 tosses of a coin;
(b) there will be fewer than four 1's in 60 throws of a die.

19. A large shipment of disposable flashlights contains 1% that are defective. Use the Poisson approximation of the binomial distribution to find the probability that among 200 flashlights randomly selected from the shipment
(a) exactly 3 will be defective;
(b) at most 2 will be defective;
(c) at least 3 will be defective.

20. The number of public buses that need repair in any one day in a certain city is a Poisson random variable X with parameter $\lambda = 10$. What is the probability that on a given day more than 15 buses in this city will need repair?

21. Show that the function $p(x; \lambda) = (\lambda^x e^{-\lambda})/x!$, for $x = 0, 1, 2, \ldots$, satisfies conditions (a) and (b) of Theorem 4.1 and is, therefore, truly a probability function. (*Hint:* $\sum_{x=0}^{\infty} (\lambda^x/x!) = e^{\lambda}$.)

22. Show that the function $b(x; n, p) = \binom{n}{x} p^x (1-p)^{n-x}$, for $x = 0, 1, \ldots, n$, satisfies conditions (a) and (b) of Theorem 4.1 and is, therefore, truly a probability function. (*Hint:* use the binomial expansion formula $(a+b)^n = \sum_{x=0}^{n} \binom{n}{x} a^x b^{n-x}$.)

23. The number of cars passing through a certain toll gate at a certain bridge during any time interval of length t (measured in minutes) is a Poisson process with parameter $\lambda = .3$.
(a) Find the probability of no car passing through this toll gate during a given 5-minute period.
(b) Find the probability of more than one car passing through during a given 5-minute period.
(c) Find the probability that between 15 and 25 cars pass through during a given one-hour period.

24. The number of customers arriving at a certain bank during any time interval of length t (measured in minutes) is a Poisson process with parameter $\lambda = 1$.
(a) Find the probability of 10 customers arriving at this bank during a given ten-minute period.
(b) Find the probability that customers will arrive at the rate of at least 1 per minute during a given period of 10 minutes.

4.6 Cumulative Distribution Functions of Discrete Random Variables

Often we are interested in estimating the chances that the values of a random phenomenon will remain at or below a certain fixed value. For example, what are the chances that the price of gold will remain at or below \$425 per troy ounce? What are the chances that a certain candidate will get no more than 30% of the votes? What are the chances that 20 tosses of a coin will produce no more than 10 heads? In each case, we are concerned with the probability that a given random variable X will take on values that are less than or equal to some fixed value x. In other words, we wish to find $P(X \leq x)$. This probability, considered as a function of x, is called the cumulative distribution function (cdf) or, more simply, the distribution function of the random variable X. A formal definition follows.

Definition 4.8 The function

$$F(x) = P(X \leq x) \qquad \text{for} \quad -\infty < x < \infty$$

is called the **cumulative distribution function** or the distribution function of the random variable X.

EXAMPLE 4.20 FINDING THE CUMULATIVE DISTRIBUTION FUNCTION

Let X be a discrete random variable whose only possible values are -1, 2, and 5. Suppose that the probability function of X is

$$f(x) = \begin{cases} 1/4 & \text{for } x = -1 \\ 1/2 & x = 2 \\ 1/4 & x = 5 \end{cases}$$

Find the cumulative distribution function of X.

Alternate Definition 4.8a The Cumulative Distribution Function of a Discrete Random Variable Let f be the probability function of the discrete random variable X. The function

$$F(x) = \sum_{t \le x} f(t) \quad \text{for} \quad -\infty < x < \infty$$

is called the cumulative distribution function or the distribution function of the random variable X. (The summation is over all values of $t \le x$ for which $f(t)$ is defined.)

EXAMPLE 4.21 FINDING PROBABILITIES WITH THE CUMULATIVE DISTRIBUTION FUNCTION

The following problem illustrates how we may use the cumulative distribution function to find probabilities and to find the probability function of a random variable.

Let X be a discrete random variable whose cumulative distribution function is

$$F(x) = \begin{cases} 0 & \text{for} \quad x < -3 \\ 1/6 & \quad -3 \le x < 6 \\ 1/2 & \quad 6 \le x < 10 \\ 1 & \quad 10 \le x \end{cases}$$

(a) Find $P(X \le 4)$, $P(-5 < X \le 4)$, $P(X = -3)$, $P(X = 4)$.
(b) Find the probability function of X.

SOLUTION

(a) $P(X \le 4) = F(4) = 1/6$, $P(-5 < X \le 4) = P(X \le 4) - P(X \le -5) = F(4) - F(-5) = 1/6 - 0 = 1/6$, $P(X = -3) = 1/6$ because the magnitude of the jump in the value of $F(x)$ at $x = -3$ is 1/6. (The alternate definition of the cumulative distribution function indicates that if $x_1, x_2, \dots$ are the possible values of X with positive probabilities, then at each point $x_1, x_2, \dots$ there is a jump in the value of $F(x)$. Furthermore, $f(x_i) = P(X = x_i)$ is the magnitude of the jump.) $P(X = 4) = 0$ because there is no jump in the value of $F(x)$ at $x = 4$.

(b) There are jumps in the value of $F(x)$ at $x = -3$, 6, and 10. The magnitudes of the jumps are 1/6, 2/6, and 3/6, respectively. Therefore, the probability function of X is

$$f(x) = \begin{cases} 1/6 & \text{for} \quad x = -3 \\ 2/6 & \quad x = 6 \\ 3/6 & \quad x = 10 \end{cases}$$

Exercises

1. Let X be a discrete random variable whose only possible values are -5, -1, 0, and 7. Find and graph the cumulative distribution function of X if the probability function of X is defined by the values $f(-5) = .3$, $f(-1) = .1$, $f(0) = .2$, and $f(7) = .4$.
2. Consider a throw of two dice. Let X be the sum of the values on the two dice. Find and graph the cumulative distribution function of X.
3. Let X be the number of heads in 5 flips of a fair coin. Find and graph the cumulative distribution function of X.
4. A box contains a penny, two nickels, and a dime. If two coins are selected at random from the box and X is the sum of values of the two coins, what is the cumulative distribution function of X? [*Hint:* the probability function was found in Exercise 6 of Section 4.3.]
5. Let X be a discrete random variable with cumulative distribution

$$F(x) = \begin{cases} 0 & \text{for} \quad x < 1 \\ .1 & \quad 1 \le x < 3 \\ .4 & \quad 3 \le x < 5 \\ .9 & \quad 5 \le x < 5.5 \\ 1.0 & \quad 5.5 \le x \end{cases}$$

   (a) Find $P(X \le 3)$, $P(X \le 4)$, $P(1.5 < X \le 5.2)$, $P(X = 3)$, $P(X = 3.5)$
   (b) Find the probability function of X.
6. Let $F(x)$ be the cumulative distribution function of a discrete random variable X. Explain why the following are true statements.
   (a) $\lim_{x \to -\infty} F(x) = 0$;
   (b) $\lim_{x \to \infty} F(x) = 1$;
   (c) If $a < b$ then $F(a) \le F(b)$.

Supplementary Exercises

1. How many times should you toss a coin if you want a probability of at least .9 of getting heads at least once?
2. If the probability that a certain book will get a favorable review from any reviewer is 1/3, how many reviewers should you ask to review the book if you want a probability of at least .95 of getting at least one favorable review?
3. Suppose that your dream is to be a partner in a prestigious law firm. You feel that your chances of being accepted as a partner in any one of a number of such law firms is .1. How many prestigious law firms should there be to make your dream come true with a probability of at least .99?

4. Based on his past experience in marketing, a manufacturer of typewriters feels that when his two new models are introduced, 65% of consumers will prefer design A to design B. He takes a straw vote of 20 randomly chosen company employees and finds that 17 of them prefer design A. If the manufacturer's subjective probabilities were correct, what is the probability that so many (i.e., 17 or more) would prefer design A? Should he revise his subjective probabilities on the basis of these results?

5. An advertisement for a certain tutoring service claims that 95% of its students show an improvement of 100 or more points in their SAT scores after completion of the course. A random sampling of 18 students showed that only 13 of them improved their scores by 100 or more points.
(a) If the claim in the advertisement is correct, what is the probability that so few (i.e., 13 or fewer from a sample of 18) would improve their scores by 100 or more points?
(b) Should we be suspicious of the claim in the advertisement?
(c) How should you answer part (b) if 16 of the 18 students showed an improvement of 100 or more points? Explain.

6. To qualify for a certain job, an applicant must demonstrate a minimum proficiency in mathematics by obtaining a passing grade of at least 10 correct answers in any one of a number of multiple-choice tests consisting of 20 questions with 4 alternative answers each. Applicants may take as many tests as they wish. Suppose that John feels that he can always answer 5 of the 20 questions correctly and he can guess on the others. He decides to keep taking the tests until he passes one of them.
(a) What is the probability that he will pass after taking fewer than 3 tests?
(b) How many tests should he take to have a probability of at least .99 of passing?
(c) Why should failing students press for makeup examinations, especially if the examinations are multiple-choice tests?

7. **The geometric distribution.** A random variable X has a *geometric distribution*, and is called a *geometric random variable*, if its probability function is

$$f(x) = g(x; p) = q^{x-1}p \qquad \text{for} \quad x = 1, 2, 3, \ldots$$

where p and q are constants such that $0 < p < 1$ and $p + q = 1$.
(a) Consider the experiment of rolling a die until a six occurs. Show that if X is the required number of rolls then X is a geometric random variable with $p = 1/6$.
(b) In the above experiment what is the probability that 4 throws of the die are required?
(c) What is the probability that at least 3 throws are required?

8. Consider a sequence of Bernoulli trials where the probability of success in each trial is p. Show that if X is the required number of trials for the occurrence of a success, then X is a geometric random variable.

9. Show that the function $g(x; p) = q^{x-1}p$ for $x = 1, 2, 3, \ldots$ satisfies conditions (a) and (b) of Theorem 4.1 and is, therefore, truly a probability function. (*Hint:* use the geometric series $\sum_{n=0}^{\infty} t^n = 1/(1-t)$ for $0 < t < 1$.)

10. Suppose that 10% of the watches manufactured by a certain company have defective parts.
(a) What is the probability that the eleventh watch inspected is the first one with defective parts?
(b) What is the probability that an inspector would have to inspect at least 10 but no more than 15 watches from this company to find one with defective parts?

11. Suppose that, for a certain couple, the probability of a male birth is .51 and a female birth is .49. If the couple decides to continue having children until they have one daughter, what is the probability that they will have 5 children?

12. **The negative binomial distribution.** A discrete random variable X is called a *negative binomial random variable*, and its probability function $f(x)$ is called a *negative binomial probability function*, if there is a positive integer n, and constants p and q, with $0 < p < 1$ and $q = 1 - p$, such that

$$f(x) = f(x; n, p) = \binom{x-1}{n-1} p^n q^{x-n} \qquad \text{for} \quad x = n, n+1, n+2, \ldots$$

Show that in a sequence of Bernoulli trials, if X is the number of the trial on which the nth success occurs, then X is a negative binomial random variable. (*Hint:* if the nth success occurs on the xth trial, then there are exactly $n - 1$ successes in the first $x - 1$ trials.)

13. One evening an editor of a certain journal was determined to continue reviewing manuscripts until she had read 3 that were acceptable for publication. If from past experience it is known that only 15% of manuscripts submitted to this editor will prove to be acceptable for publication, what is the probability that this editor will read 20 manuscripts that evening? (*Hint:* use the negative binomial distribution of Exercise 11.)

14. A certain firm is looking for 4 qualified engineers to add to its staff. If from past experience it is known that only 25% of engineers applying for a position with this firm is judged to be qualified, what is the probability that the firm will interview 30 applicants to fill the 4 positions?

References

1. Chung, Kai Lai. 1974. *Elementary probability theory with stochastic processes.* New York: Springer-Verlag.
2. Feller, William. 1950. *An introduction to probability theory and its applications,* Vol. I, 3d ed. New York: John Wiley & Sons, Inc.