1

Probability

The mathematical foundation upon which mathematical statistics and likelihood inference are built is probability theory. Modern probability theory is primarily due to the foundational work of the Russian mathematician Andrei Nikolaevich Kolmogorov (1903–1987). Kolmogorov published his treatise on probability in 1933 [1], which framed probability theory in a rigorous mathematical framework. Kolmogorov's work provided probability theory with an axiomatic mathematical structure that produces a consistent and coherent theory of probability. Specifically, Kolmogorov's structure is based on measure theory, which deals with assigning numerical values to sets (i.e. measuring a set) and the theory of integration and differentiation.

1.1 Sample Spaces, Events, and σ-Algebras

The structure under which probabilities are relevant and can be assigned in a consistent and coherent fashion requires a probability model consisting of a chance experiment, the collection of all possible outcomes of the chance experiment, and a function that assigns probabilities to collections of outcomes of the chance experiment.

Definition 1.1 A chance experiment is any task for which the outcome of the task is unknown until the task is actually performed.

Experiments where the outcome is known before the experiment is actually performed are called *deterministic experiments* and are not interesting with regard to probability. Probability theory, probability assignments, and statistics only apply to chance experiments. The set of possible outcomes of a chance experiment is called the *sample space*, and the sample space defines one component of a probability model.

Mathematical Statistics: An Introduction to Likelihood Based Inference, First Edition. Richard J. Rossi.

Definition 1.2 The sample space associated with a chance experiment is the set of all possible outcomes of a chance experiment. The sample space will be denoted by Ω.

The sample space consists of the outcomes that are considered feasible and interesting. Probabilities can only be assigned to outcomes or subsets of outcomes in the sample space.

Example 1.1 Suppose that a chance experiment consists of flipping a two-sided coin with heads on one side and tails on the other side. The most commonly used sample space is $\Omega = \{\text{Heads, Tails}\}$; however, another possible sample space that could be used is $\Omega' = \{\text{heads, tails, edge}\}$; these two sample spaces produce two different probability models for the same chance experiment. As long as Kolmogorov's measure theoretic approach is used, both probability models will produce consistent and coherent probability assignments.

Chance experiments cover a wide range of everyday tasks such as dealing a hand of cards, forecasting weather, driving in excess of the speed limit at the risk of getting a speeding ticket, and buying a lottery ticket. In each of these cases there is a chance experiment where the outcome is unknown until the experiment is actually completed.

Example 1.2 Suppose that a chance experiment consists of weighing a brown trout randomly selected from the Big Hole River in Montana. A reasonable sample space for this chance experiment is $\Omega = (0, 50]$ since the largest known brown trout to come from the river is less than 50 lb. If the upper limit on the weight of a Big Hole brown trout is unknown, it would also be reasonable to use $\Omega = (0, \infty)$ for the sample space. Choosing the probability assignment takes care of probabilities for likely and unlikely values of the weight of a Big Hole River brown trout.

Note that in many chance experiments, the limits of the sample space will be unknown, and in this case, the sample space can be taken to be an infinite length subset of $\mathbb{R}$. The sample space is only the list of possible outcomes, while the choice of the function used to make the probability assignments controls the probabilities of the values in the sample space. The three components required of a probability model are the sample space, a collection of subsets of the sample space for which probabilities will be assigned, and the function used to assign the probabilities to subsets of the sample space.

Under Kolmogorov's probability structure, not all subsets of the sample space can be assigned probabilities. The collection of subsets of the sample space that can be assigned probabilities must have a particular structure so that the probability assignments are coherent and consistent. In particular, the collection of subsets of Ω that can be assigned probabilities must be a *σ-algebra*.

Definition 1.3 Let $\mathcal{A}$ be a collection of events of Ω. $\mathcal{A}$ is said to be a σ-algebra of events if and only if

i) $\Omega \in \mathcal{A}$.
ii) $A^c \in \mathcal{A}$ whenever $A \in \mathcal{A}$.
iii) $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$ whenever $A_i \in \mathcal{A},\ \forall i$.

A subset of Ω that is in a σ-algebra associated with Ω is called an *event*.

Definition 1.4 An event A of a sample space Ω is any subset in a σ-algebra associated with Ω. An event A is said to have occurred when the chance experiment results in an outcome in A.

A σ-algebra associated with a sample space Ω contains the only events that probabilities can be assigned to. There are many σ-algebras of subsets associated with a sample space (see Example 1.3); however, the appropriate σ-algebra must be chosen so that it is large enough to contain all of the relevant events to be considered. It is important to note that in order to have a consistent and coherent probability assignment, not all events of Ω can be assigned probabilities.

Example 1.3 Examples of σ-algebras associated with a sample space Ω include the following:

1) The trivial σ-algebra $\mathcal{A}_0 = \{\emptyset, \Omega\}$. This is the smallest σ-algebra possible and not very useful for a probability model.
2) $\mathcal{A}_1 = \{\emptyset, A, A^c, \Omega\}$, where A is a subset of Ω. This is the smallest σ-algebra that includes the event A.
3) The Borel σ-algebra, which is the smallest σ-algebra containing all of the open intervals of $\mathbb{R}$. The Borel σ-algebra can only be used when the elements of Ω are real numbers, and in this case, it is a commonly used σ-algebra.

The σ-algebra of events associated with Ω will also include all of the compound events that can constructed using the basic set operations intersection, union, and complementation. The definitions for the compound events are given in Definitions 1.5–1.7.

Definition 1.5 Let A and B be events of Ω. The event formed by the intersection of the events A and B is denoted by $A \cap B$ and is defined to be $A \cap B = \{\omega \in \Omega : \omega \in A \text{ and } \omega \in B\}$.

Definition 1.6 Let A and B be events of Ω. The event formed by the union of the events A and B is denoted by $A \cup B$ and is defined to be $A \cup B = \{\omega \in \Omega : \omega \in A \text{ or } \omega \in B\}$.

Definition 1.7 Let A be an event of Ω. The event that is the complement of the event A is denoted by A^c and is defined to be $A^c = \{\omega \in \Omega : \omega \notin A\}$.

Note that union and intersection are commutative operations. That is, $A \cup B = B \cup A$ and $A \cap B = B \cap A$. Also, with complementation, $(A^c)^c = A$. Another set operation that is used to create a compound event is the *set difference*. The set difference between two sets A and B consists of the elements of A that are not elements of B.

Definition 1.8 Let A and B be events of Ω. The set difference $A - B$ is defined to be $A - B = \{\omega \in \Omega : \omega \in A \text{ and } \omega \notin B\}$.

Set difference is not a commutative operation, and $A - B$ can also be written as $A \cap B^c$. The following example illustrates how compound events can be created using the set operations union, intersection, complementation, and set difference.

Example 1.4 Suppose that a card will be drawn from a standard deck of 52 playing cards. Then, the sample space is

$$\Omega = \{AH, \dots, KH, AD, \dots, KD, AC, \dots, KC, AS, \dots, KS\},$$

where in the outcome XY, X is the denomination of the card $(A, 2, \dots, K)$ and Y is the suit of the card (H, D, C, S). Let A be the event that a heart is selected, and let B be the event that an ace is selected. Then, $A = \{AH, \dots, KH\}$, $B = \{AH, AD, AC, AS\}$, and

$$\begin{aligned}
A \cap B &= \{AH\},\\
A \cup B &= \{AH, \dots, KH, AD, AC, AS\},\\
A^c &= \{AD, \dots, KD, AC, \dots, KC, AS, \dots, KS\},\\
B^c &= \{2H, \dots, KH, 2D, \dots, KD, 2C, \dots, KC, 2S, \dots, KS\},\\
A - B &= \{2H, \dots, KH\},\\
B - A &= \{AD, AC, AS\}.
\end{aligned}$$

Events that share no common elements are called *disjoint events* or *mutually exclusive events*.

Definition 1.9 Two events A and B of Ω are said to be disjoint when $A \cap B = \emptyset$.

When two events A and B are disjoint, the chance experiment cannot result in an outcome where both the events A and B occur. The events A and A^c are always disjoint events as are $A - B$ and $B - A$.

Compound events can also be constructed using the set operations intersection, union, and complementation on a family of sets, say $\mathcal{F} = \{A_i : i \in \Delta\}$, where the index set Δ is a finite or countably infinite set. In most cases, Δ will be taken to be a subset of $\mathbb{N}$, and the compound events created using intersection and union are

$$\omega \in \bigcap_{i\in\Delta} A_i, \quad \text{if and only if } \omega \in A_i, \ \forall i \in \Delta,$$
$$\omega \in \bigcup_{i\in\Delta} A_i, \quad \text{if and only if } \omega \in A_i \text{ for some } i \in \Delta.$$

Example 1.5 Suppose that a two-sided coin will be flipped until the first head appears. Let A_i be the event that the first head appears on the ith flip, B be the event that it takes at least two flips of the coin to observe the first head, and let C be the event that it takes less than 10 flips to observe the first head. Then,

$$B = \bigcup_{i=2}^{\infty} A_i$$

and

$$C = \bigcup_{i=1}^{9} A_i.$$

The set laws given in Theorems 1.1 and 1.2 can often be used to simplify the computation of the probability of a compound event.

Theorem 1.1 (De Morgan's Laws) *If $\{A_i : i \in \Delta\}$ is a family of events of Ω, Δ is a subset of $\mathbb{N}$, and $\mathcal{D} \subset \Delta$, then*

i) $\left(\bigcup_{i\in\mathcal{D}} A_i\right)^c = \bigcap_{i\in\mathcal{D}} A_i^c.$

ii) $\left(\bigcap_{i\in\mathcal{D}} A_i\right)^c = \bigcup_{i\in\mathcal{D}} A_i^c.$

Corollary 1.1 *If A and B are events of Ω, then*

i) $(A \cup B)^c = A^c \cap B^c$.
ii) $(A \cap B)^c = A^c \cup B^c$.

Note that the complement of a union is the intersection of the complements, and the complement of an intersection is the union of the complements.

Theorem 1.2 (Distributive Laws) *If $\{A_i : i \in \Delta\}$ is a family of events of Ω, Δ is a subset of $\mathbb{N}$, and $\mathcal{D} \subset \Delta$, then*

i) for any event B of Ω, $B \cap \left(\bigcup_{i\in\mathcal{D}} A_i\right) = \bigcup_{i\in\mathcal{D}} (B \cap A_i)$.

ii) for any event B of Ω, $B \cup \left(\bigcap_{i\in\mathcal{D}} A_i\right) = \bigcap_{i\in\mathcal{D}} (B \cup A_i)$.

In particular, Theorem 1.1 holds for $\mathcal{D} = \{1, 2, \dots, n\}$ or $\mathcal{D} = \mathbb{N}$. That is,

$$\left(\bigcup_{i=1}^{n} A_i\right)^{\mathrm{c}} = \bigcup_{i=1}^{n} A_i^{\mathrm{c}} \quad \text{and} \quad \left(\bigcup_{i=1}^{\infty} A_i\right)^{\mathrm{c}} = \bigcup_{i=1}^{\infty} A_i^{\mathrm{c}}$$

and

$$\left(\bigcap_{i=1}^{n} A_i\right)^{\mathrm{c}} = \bigcap_{i=1}^{n} A_i^{\mathrm{c}} \quad \text{and} \quad \left(\bigcap_{i=1}^{\infty} A_i\right)^{\mathrm{c}} = \bigcap_{i=1}^{\infty} A_i^{\mathrm{c}}$$

Similarly, Theorem 1.2 also holds for the finite set $\mathcal{D} = \{1, 2, \dots, n\}$ and the infinite set $\mathcal{D} = \mathbb{N}$. The simplest version of the *Distributive Laws* is given in Corollary 1.2

Corollary 1.2 *If A, B, and C are events of Ω, then*

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

A family $\mathcal{P}$ of disjoint events of a sample space Ω whose union is Ω is called a *partition*. Partitions also can be used to simplify the computation of the probability of an event.

Definition 1.10 A collection of events $\mathcal{P} = \{A_i : i \in \mathbb{N}\}$ is said to be a partition of a sample space Ω if and only if

i) $\bigcup_{i=1}^{\infty} A_i = \Omega$.
ii) $A_i \cap A_j = \emptyset$, whenever $i \neq j$.

A partition may consist of a finite number of events. That is, if $\{A_1, A_2, \dots, A_n\}$ is a collection of disjoint events whose union is Ω, a partition $\mathcal{P} = \{A_j : j \in \mathbb{N}\}$ is formed by letting $A_j = \emptyset$ for $j > n$.

Example 1.6 Let A be an event in Ω. The simplest partition of a sample space Ω is $\mathcal{P} = \{A, A^{\mathrm{c}}\}$ since $A \cup A^{\mathrm{c}} = \Omega$ and $A \cap A^{\mathrm{c}} = \emptyset$.

Theorem 1.3 shows that a partition $\mathcal{P}$ can be used to partition an event B into disjoint events whose union is the event B.

Theorem 1.3 *If $\mathcal{P} = \{A_i : i \in \mathbb{N}\}$ is a partition of Ω, then for any event B of Ω.*

$$B = \bigcup_{i=1}^{\infty}(B \cap A_i).$$

Proof. Let $\{A_i : i \in \mathbb{N}\}$ be a partition of Ω and B an event of Ω. Then,

$$B = B \cap \Omega = B \cap \left(\bigcup_{i=1}^{\infty} A_i\right) = \bigcup_{i=1}^{\infty}(B \cap A_i).$$

∎

Corollary 1.3 shows that an arbitrary event A of Ω can be used to partition any other event B of Ω.

Corollary 1.3 *If A and B are events of Ω, then $B = (B \cap A) \cup (B \cap A^c)$.*

Problems

1.1.1 Determine a reasonable sample space when the chance experiment involves

a) selecting a student at random and recording their GPA.
b) selecting an adult male at random and measuring their weight.
c) selecting an adult female at random and measuring their weight.
d) selecting a student at random and recording the color of their hair.
e) rolling two standard six-sided dice and summing the outcomes on each die.
f) selecting a person at random and recording their birthday.
g) selecting a student at random and recording how many credits they are enrolled in.

1.1.2 Two numbers will be drawn at random and without replacement from the numbers 1, 2, 3, 4, 5. Let A be the event that at least one even number is drawn in the two draws, and let B be the event that the sum of the draws is equal to 4. Determine the outcomes in

a) sample space.
b) A.
c) B.
d) $A \cap B$.
e) $A \cup B$.
f) $B - A$.

1.1.3 Flip a two-sided coin four times. Let A be the event exactly two heads are flipped and let B be the event at least one tail is flipped. Determine the outcomes in

a) the sample space.
b) A.
c) A^c.
d) B.
e) $A \cap B$.
f) $A \cup B$.
g) $A - B$.
h) $B - A$.

1.1.4 Draw a card at random from a standard deck of 52 playing cards. Let A be the event that an ace is drawn, and let B be the event that a diamond is drawn. Determine the outcomes in

a) the sample space.
b) A.
c) B.
d) B^c.
e) $A \cap B$.
f) $A \cup B$.
g) $A - B$.
h) $B - A$.

1.1.5 Flip a two-sided coin until a head appears. Let A be the event the first head is flipped on the fourth flip and let B be the event the first head is flipped in less than five flips. Determine the outcomes in

a) A.
b) B.
c) $A \cap B$.
d) $A \cup B$.

1.1.6 Let A be an event of Ω. Show that $(A^c)^c = A$.

1.1.7 Let A and B be events of Ω. If $A \subset B$, show that $B = A \cup (B - A)$.

1.1.8 Let A and B be events of Ω. Show that $A \cup B = A \cup (B - A)$.

1.1.9 Let A and B be events of Ω. Show that $A - B$ and $B - A$ are disjoint events.

1.1.10 Let A and B be disjoint events of Ω, and let C be any other event of Ω. Show that $A \cap C$ and $B \cap C$ are disjoint events.

1.1.11 Let $\{A_i\}$ be a countable collection of events of Ω with $\bigcup_{i=1}^{\infty} A_i = \Omega$, and let $B_1 = A_1$ and $B_k = A_k - \left(\bigcup_{i=1}^{k-1} A_i\right)$ for $n \geq 2$. Show that $\{B_k\}$ is a partition of Ω.

1.1.12 Let $\Omega = \mathbb{R}$ and define $A_i = [-\frac{1}{n}, n)$ for $i \in \mathbb{N}$. Determine

a) $\bigcup_{i=1}^{10} A_i$.
b) $\bigcap_{i=1}^{10} A_i$.
c) $\bigcup_{i=1}^{\infty} A_i$.
d) $\bigcap_{i=1}^{\infty} A_i$.

1.1.13 Let $\Omega = \mathbb{R}^+$ and define $A_i = (\frac{1}{n}, 1 + \frac{1}{n})$ for $i \in \mathbb{N}$. Determine

a) $\bigcup_{i=1}^{20} A_i$.

b) $\bigcap_{i=1}^{20} A_i$.

c) $\bigcup_{i=1}^{\infty} A_i$.

d) $\bigcap_{i=1}^{\infty} A_i$.

1.1.14 Show that $\{\emptyset, A, A^c, \Omega\}$ is a σ-algebra.

1.1.15 Show that if $\mathcal{A}$ is a σ-algebra and $A, B \in \mathcal{A}$, then $A \cap B \in \mathcal{A}$.

1.1.16 Let $\mathcal{A}$ be a σ-algebra and let $A_i \in \mathcal{A}$ for $i \in \mathbb{N}$. Show that $\bigcap_{i=1}^{\infty} A_i \in \mathcal{A}$.

1.1.17 Determine the smallest σ-algebra that contains the events A and B.

1.2 Probability Axioms and Rules

The third component of a probability model is the *probability function*, which is a set function whose domain $\mathcal{A}$ is a σ-algebra of events of Ω. Kolmogorov's measure theoretic approach to probability requires a probability function satisfying the properties given in Definition 1.11.

Definition 1.11 (Kolmogorov's Probability Function) Let Ω be the sample space associated with a chance experiment, and let $\mathcal{A}$ be a σ-algebra of events of Ω. A set function P on $\mathcal{A}$ satisfying the following three properties is called a probability function or a probability measure.

A1: $P(\Omega) = 1$.
A2: $P(A) \geq 0$ for every event $A \in \mathcal{A}$.
A3: If $\{A_i : i \in \mathbb{N}\} \subset \mathcal{A}$ is a collection of disjoint events, then

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

The triple $(\Omega, \mathcal{A}, P)$ is called a probability space.

Conditions A1–A3 are known as *Kolmogorov's Axioms of Probability*, and the sets in the σ-algebra $\mathcal{A}$ are the measurable sets and the only sets that can be assigned probabilities. Theorem 1.4 reveals some of the basic consequences of Kolmogorov's Axioms.

Theorem 1.4 *Let* $(\Omega, \mathcal{A}, P)$ *be a probability space and let* $A, B \in \mathcal{A}$*.*

i) $P(\emptyset) = 0$.
ii) *If* $A \subset B$, *then* $P(A) \leq P(B)$.
iii) $0 \leq P(A) \leq 1$.
iv) $P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i)$ *when* $A_1, A_2, \ldots, A_n$ *are disjoint events in* $\mathcal{A}$*.*

Proof. Let $(\Omega, \mathcal{A}, P)$ be a probability space and let $A, B \in \mathcal{A}$.

i) Since $\Omega = \Omega \bigcup_{i=1}^{\infty} \emptyset$, let $A_1 = \Omega$ and $A_{i+1} = \emptyset$ for $i \in \mathbb{N}$. Then, $\{A_i\}$ is a collection of disjoint sets and

$$1 = P(\Omega) = P\left(\bigcup_{i=1}^{\infty} A_i\right) = \underbrace{\sum_{i=1}^{\infty} P(A_i)}_{\text{Axiom 3}}$$

$$= P(A_1) + \sum_{i=2}^{\infty} P(A_i) = 1 + \sum_{i=2}^{\infty} P(A_i).$$

Thus, $\sum_{i=2}^{\infty} P(A_i) = 0$ and since $P(A_i) \geq 0,\ \forall i \in \mathbb{N}$. Hence, it follows that $P(A_{i+1}) = P(\emptyset) = 0,\ \forall i \in \mathbb{N}$.

ii) Since $B = A \cup (B - A)$ and $A \cap (B - A) = \emptyset$, it follows that

$$P(B) = P(A \cup (B - A)) = P(A) + P(B - A) \underbrace{\geq}_{P(B-A)\geq 0} P(A).$$

iii) This follows directly from parts (i) and (ii), since $A \subset \Omega$ and $\emptyset \subset A$, it follows that $0 = P(\emptyset) \leq P(A) \leq P(\Omega) = 1$.

iv) The proof of part (iv) is left as an exercise. ∎

Further consequences of Kolmogorov's Axioms are given in Theorem 1.5. In particular, the results given in Theorem 1.5 provide several useful rules for computing probabilities.

Theorem 1.5 *Let* $(\Omega, \mathcal{A}, P)$ *be a probability space and let* $A, B \in \mathcal{A}$*. Then,*

i) $P(A^c) = 1 - P(A)$.
ii) $P(B) = P(B \cap A) + P(B \cap A^c)$.
iii) $P(A - B) = P(A) - P(A \cap B)$.

iv) *if* $A \subset B$, *then* $P(B - A) = P(B) - P(A)$.
v) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
vi) *if* A *is countable, say* $A = \{\omega_i : i \in \mathbb{N}\}$, *then* $P(A) = \sum_{i=1}^{\infty} P(\{\omega_i\})$.

Proof. Let $(\Omega, \mathcal{A}, P)$ be a probability space and let $A, B \in \mathcal{A}$.

i) Note that $\Omega = A \cup A^c$ and A and A^c are disjoint events. Thus,

$$1 = P(\Omega) = P(A \cup A^c) = \underbrace{P(A) + P(A^c)}_{\text{since disjoint}}.$$

Therefore, $P(A^c) = 1 - P(A)$.

ii) From Corollary 1.3, $B = (A \cap B) \cup (A^c \cap B)$ and the events $A \cap B$ and $A^c \cap B$ are disjoint. Thus,

$$P(B) = P[(A \cap B) \cup (A^c \cap B)] = P(A \cap B) + P(A^c \cap B).$$

iii) The proof of part (iii) is left as an exercise.
iv) The proof of part (iv) is left as an exercise.
v) Since $A \cup B$ can be written as the disjoint union of $A \cap B^c$, $A \cap B$, and $A^c \cap B$, it follows that

$$\begin{aligned} P(A \cup B) &= P(A \cap B^c) + P(A \cap B) + P(A^c \cap B) \\ &= P(A - B) + P(A \cap B) + P(B - A). \end{aligned}$$

Hence, by Theorem 1.5 part (iii), it follows that

$$\begin{aligned} P(A \cup B) &= P(A - B) + P(A \cap B) + P(B - A) \\ &= P(A) - P(A \cap B) + P(A \cap B) + P(B) - P(A \cap B) \\ &= P(A) + P(B) - P(A \cap B). \end{aligned}$$

vi) The proof of part (vi) is left as an exercise. ■

Examples 1.7 and 1.8 illustrate the use of Theorem 1.5 for computing probabilities.

Example 1.7 Let $(\Omega, \mathcal{A}, P)$ be a probability space and let $A, B \in \mathcal{A}$. Suppose that A and B are events of Ω with $P(A) = 0.6$, $P(B) = 0.75$, and $P(A \cap B) = 0.55$. Then,

1) $P(A^c) = 1 - P(A) = 1 - 0.6 = 0.4$.
2) $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.6 + 0.75 - 0.55 = 0.8$.
3) $P(A - B) = P(A) - P(A \cap B) = 0.6 - 0.55 = 0.05$.
4) $P(A^c \cap B^c) = P[(A \cup B)^c] = 1 - P(A \cup B) = 1 - 0.8 = 0.2$.
5) $P(A^c \cup B) = P(A^c) + P(B) - P(A^c \cap B) = 0.4 + 0.75 - (0.75 - 0.55) = 0.95$.

Example 1.8 Suppose that 70% of the fishermen in Montana use fly fishing gear, 40% use spin fishing gear, and 10% use both fly and spin fishing gear. The probability that a fisherman in Montana uses fly fishing gear (F) but not spin fishing gear (S) is

$$P(F - S) = P(F) - P(F \cap S) = 0.70 - 0.10 = 0.60,$$

and the probability that a fisherman in Montana uses spin fishing gear but not fly fishing gear is

$$P(S - F) = P(S) - P(S \cap F) = 0.40 - 0.10 = 0.30.$$

Theorem 1.5 part (ii) is known as the *Law of Total Probability*. The Law of Total Probability can be used to solve many probability problems, and in particular, it is useful when there are two or more cases that must be considered when computing the probability of an event. A generalized version of the Law of Total Probability is given in Theorem 1.6.

Theorem 1.6 (General Law of Total Probability) *Let* $(\Omega, \mathcal{A}, P)$ *be a probability space and let* $A \in \mathcal{A}$. *If* $B_i \in \mathcal{A},\ \forall i \in \mathbb{N}$ *and* $\{B_i\}$ *is a partition of* Ω, *then,*

$$P(A) = \sum_{i=1}^{\infty} P(A \cap B_i).$$

Proof. The proof of Theorem 1.6 follows directly from Theorem 1.3 and axiom A3. ∎

Note that Theorem 1.6 also works with a finite partition $\{B_1, B_2, \dots, B_n\}$ in which case $P(A) = \sum_{i=1}^{n} P(A \cap B_i)$.

Example 1.9 Suppose that a game is played where two dice are rolled until either a total of 6 or 7 is rolled. Let A be the event that a 6 is rolled before a 7. Note that $A = \bigcup_{i=1}^{\infty} (A \cap B_i)$, where B_i = the game terminates on the ith roll.

Thus, the probability of event A can be computed using the General Law of Total Probability with

$$P(A) = \sum_{i=1}^{\infty} P(A \cap B_i).$$

This example will be revisited in a later section after further development of the rules of probability.

Example 1.10 Suppose that two marbles are drawn at random and without replacement from an urn containing 5 white marbles, 9 red marbles, and

11 black marbles. Let A be the event that two marbles of the same color are selected, W be the event that the first marble selected is white, R be the event that the first marble selected is red, and B be the event that the first marble selected is black. Then, W, R, and B partition the sample space, and the probability of drawing two marbles of the same color is

$$P(A) = P(A \cap W) + P(A \cap R) + P(A \cap B).$$

Theorem 1.7 (Boole's Inequality) *If $\{A_i\}$ is a collection of events in $\mathcal{A}$, then* $P\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} P(A_i)$.

Proof. Let $\{A_i\}$ be a collection of events in $\mathcal{A}$ and let

$$B_i = A_i - \left(\bigcup_{j=1}^{i-1} A_i\right)$$

Then, $\bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{\infty} B_i$, $B_i \cap B_j = \emptyset$ for $i \neq j$, and $B_i \subset A_i$, $\forall i$. Hence,

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = P\left(\bigcup_{i=1}^{\infty} B_i\right) = \underbrace{\sum_{i=1}^{\infty} P(B_i) \leq \sum_{i=1}^{\infty} P(A_i)}_{\text{since } B_i \subset A_i}.$$

■

Boole's Inequality is also often referred to as *Bonferroni's Inequality* and is sometimes used when performing multiple comparisons in a hypothesis testing scenario.

Example 1.11 Suppose that five hypothesis tests will be carried out with each test having a probability of false rejection equal to 0.01. Let A_i be the event that the ith test makes a false rejection. Then, by Boole's Inequality, the probability of making at least one false rejection in the five hypothesis tests is

$$P\left(\bigcup_{i=1}^{5} A_i\right) \leq \sum_{i=1}^{5} P(A_i) = 0.05.$$

Using Boole's Inequality is one way to protect against making false rejections in a multiple comparison setting and can be generalized to handle any number of hypothesis tests. For example, if the goal is to have the overall probability of making at least one false rejection less than α when n hypothesis tests are performed, then by taking $P(A_i) = \frac{\alpha}{n}$, the overall chance of making one or more false rejections is less than α. This procedure is referred to as the *Bonferroni Multiple Comparison Procedure.*

Problems

1.2.1 Suppose that $P(A) = 0.6$, $P(B) = 0.75$, and $P(A \cap B) = 0.55$. Determine

a) $P(A^c)$. b) $P(B^c)$.
c) $P(A \cup B)$. d) $P(A - B)$.
e) $P(A^c \cap B^c)$. f) $P(A^c \cup B^c)$.
g) $P(A^c \cup B)$. h) $P[(A - B) \cup (B - A)]$.

1.2.2 Suppose that $A \subset B$, $P(A) = 0.4$, and $P(B) = 0.6$. Determine

a) $P(A^c)$. b) $P(B^c)$.
c) $P(A \cap B)$. d) $P(A \cup B)$.
e) $P(B - A)$. f) $P(A - B)$.

1.2.3 Suppose that A and B are disjoint events with $P(A) = 0.5$ and $P(B) = 0.35$. Determine

a) $P(A \cap B)$. b) $P(A \cup B)$.
c) $P(B - A)$. d) $P(A - B)$.
e) $P[(A \cup B)^c]$. f) $P(A^c \cap B^c)$.

1.2.4 Suppose that $P(B) = 0.6$, $P(A \cap B) = 0.45$, and $P(A - B) = 0.25$. Determine

a) $P(B - A)$. b) $P(A)$.
c) $P(A \cup B)$. d) $P(A^c \cap B^c)$.

1.2.5 Suppose that $\Omega = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$ and $P(\{i\}) = \frac{i}{55}$ for $i \in \Omega$. If a number is drawn at random, determine the probability that

a) an even number is drawn.
b) a multiple of 3 is drawn.
c) a number less than 5 is drawn.
d) a prime number is drawn.

1.2.6 A large computer sales company reports that 80% of their computers are sold with a DVD drive, 95% with a CD drive, and 75% with both. Determine the probability that

a) a computer without a DVD drive is sold.
b) a computer with a DVD drive or a CD drive is sold.
c) a computer with a DVD drive but not a CD drive is sold.
d) a computer with a DVD drive or a CD drive but not both is sold.

1.2.7 Mr. Jones watches the 6 p.m. news 50% of the time, he watches the 11 p.m. news 75% of the time, and watches both the 6 p.m. and 11 p.m. news 28% of the time. Determine the probability that

a) Mr. Jones watches either the 6 p.m. or 11 p.m. news.
b) Mr. Jones watches either the 6 p.m. or 11 p.m. news, but not both.
c) Mr. Jones watches neither the 6 p.m. nor 11 p.m. news.
d) Mr. Jones watches the 6 p.m. but not the 11 p.m. news.
e) Mr. Jones watches the 11 p.m. but not the 6 p.m. news.

1.2.8 Suppose that the probability of a two-child family having two male children is 0.23, having a male child first followed by a female child is 0.25, having a female child first followed by a male child is 0.25, and having two female children is 0.27. Determine the probability that a two-child family has

a) at least one male child.
b) at least one female child.
c) a male first child.
d) a female second child.

1.2.9 Airlines A and B have 9 a.m. flights from San Francisco to Seattle. Suppose that the probability that airline A's flight is fully booked is 0.80, the probability that airline B's flight is fully booked is 0.75, and the probability that both airlines 9 a.m. flights to Seattle are fully booked is 0.68. Determine the probability that

a) airline A or airline B has a fully booked flight.
b) neither airline A nor airline B has a fully booked flight.
c) airline A has a fully booked flight but airline B does not.

1.2.10 Suppose that A, B, and C are disjoint events with $P(A) = 0.1$, $P(B) = 0.25$, and $P(C) = 0.6$. Determine

a) $P(A \cap B \cap C)$.
b) $P(A \cup B \cup C)$.

1.2.11 Suppose that A, B, and C are events with $A \subset B \subset C$. If $P(A) = 0.1$, $P(B) = 0.25$, and $P(C) = 0.6$, determine

a) $P(A \cap B \cap C)$.
b) $P(A \cup B \cup C)$.
c) $P(C - A)$.
d) $P(C - B)$.

1.2.12 Suppose that $P(A) = 0.25$, $P(B - A) = 0.3$, and $P[C - (A \cup B)] = 0.1$. Determine

a) $P(A \cup B)$.
b) $P(A \cup B \cup C)$.

1.2.13 Prove: If A and B are events, then
a) the probability that event A or event B occurs, but not both, is $P(A \cup B) - P(A \cap B)$.
b) the probability that exactly one of the events A or B occurs is $P(A) + P(B) - 2P(A \cap B)$.
c) $P(A \cap B) \leq P(A \cup B) \leq P(A) + P(B)$.

1.2.14 Prove: If A, B, and C are events, then
a) $P(A \cup B \cup C) \leq P(A) + P(B) + P(C)$.
b) $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$.

1.2.15 Prove Theorem 1.4 part (iv).

1.2.16 Prove
a) Theorem 1.5 part (iii).
b) Theorem 1.5 part (iv).
c) Theorem 1.5 part (vi).

1.2.17 Suppose that $\Omega = \mathbb{N}$ and for $i \in \mathbb{N}$, $P(\{i\}) = \frac{k}{3^i}$, where k is an unknown constant. Determine the value of
a) k.
b) $P(\{\omega \in \mathbb{N} : \omega \leq 5\})$.

1.2.18 Let $\Omega = \mathbb{N}$ and for $i \in \mathbb{N}$, let $A_i = \{\omega \in \mathbb{N} : \omega \leq i\}$. If $P(A_i) = 1 - \left(\frac{1}{2}\right)^i$, determine
a) $P(A_i - A_j)$ for $i, j \in \mathbb{N}$ and $i > j$.
b) $P(\{i\})$ for $i \in \mathbb{N}$.

1.2.19 Let $\Omega = (0, \infty)$ and for $t \in (0, \infty)$, let $A_t = \{\omega \in \mathbb{N} : \omega \geq t\}$. If $P(A_i) = \mathrm{e}^{-t}$, determine $P(A_s - A_t)$ for $t, s \in (0, \infty)$ and $t > s$.

1.3 Probability with Equally Likely Outcomes

When Ω is a finite set and each of the outcomes in Ω has the same chance of occurring, then the outcomes in Ω are said to be *equally likely outcomes*. When the outcomes of the chance experiment are equally likely, computing the probability of any event A is often simple. In particular, if $N(A)$ is the number of outcomes in event A and $N(\Omega) = N$, then $P(A) = \frac{N(A)}{N}$.

Theorem 1.8 *If Ω is a finite sample space with N possible outcomes, say $\Omega = \{\omega_1, \omega_2, \dots, \omega_N\}$, and the outcomes in Ω are equally likely to occur, then the probability of an event A is $P(A) = \frac{N(A)}{N}$.*

Proof. Follows directly from Theorem 1.5 part (vi) with each outcome having equal probability $\frac{1}{N(\Omega)}$. ∎

Equally likely outcomes often arise when sampling at random from a finite population. For example, when drawing a card at random from a standard deck of 52 playing cards, each of the 52 cards has an equal chance of being selected. Other common chance experiments in which equally likely outcomes arise include flipping a fair coin n times, rolling n fair dice, selecting n balls at random from an urn, randomly drawing lottery numbers, dealing cards from a well shuffled deck, and choosing names at random out of a hat.

Examples 1.12–1.14 illustrate probability computations where the sample space contains a finite number of equally likely outcomes.

Example 1.12 A card will be drawn at random from a standard deck of 52 playing cards. The sample space associated with this chance experiment is

$$\Omega = \{\underbrace{AH, 2H, \ldots, KH, AD, 2D, \ldots, KD}_{\text{red cards}}, \underbrace{AS, 2S, \ldots, KS, AC, 2C, \ldots, KC}_{\text{black cards}}\}.$$

Thus, $N(\Omega) = 52$. Let A be the event that an ace is drawn and B be the event that a black card is drawn. Then, $P(A) = \frac{4}{52}$, $P(B) = \frac{26}{52}$, $P(A \cap B) = \frac{2}{52}$, and $P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{4}{52} + \frac{26}{52} - \frac{2}{52} = \frac{28}{52}$.

Theorem 1.9 *If a chance experiment consists of repeating a task that has N possible outcomes n times, then $N(\Omega) = N^n$.*

Thus, when a chance experiment consists of drawing twice, at random and with replacement, from a collection of N objects, there are N^2 possible outcomes in Ω. For example, if two cards are drawn at random and with replacement from a standard deck of 52 playing cards, then there are $52^2 = 2704$ possible outcomes in Ω.

Example 1.13 If a chance experiment consists of flipping a fair coin 10 times, then there are $2^{10} = 1024$ possible outcomes in Ω. If A is the event that at least one of the flips is a head, then

$$P(A) = 1 - P(A^c) = 1 - P(0 \text{ heads are flipped}) = 1 - \frac{1}{1024} = \frac{1023}{1024}.$$

Theorem 1.10 *When a chance experiment consists of drawing at random and without replacement twice from a collection of N objects, there are $N(N-1)$ outcomes in Ω.*

For example, if two cards are to be drawn at random and without replacement from a standard deck of 52 playing cards, then $\mathbb{N}(\Omega) = 52 \times 51 = 2652$. Theorem 1.10 can be generalized to a chance experiment consisting of drawing n times, at random and without replacement, from a collection of N objects.

In this case, $N(\Omega) = N(N-1)\cdots(N-n+1)$. For example, when three objects are drawn at random and without replacement from a collection of 10 objects, the sample space for this chance experiment will contain $10 \times 9 \times 8 = 720$ outcomes.

Example 1.14 Draw two numbers at random and without replacement from the numbers 1, 2, 3, 4, and 5. The sample space associated with this chance experiment contains 20 equally likely outcomes. The event that at least one number greater than or equal to 4 is chosen comprised the outcomes (4,1), (1,4), (4,2), (2,4), (4,3), (3,4),(4,5), (5,4), (5,3), (3,5), (5,2), (2,5), (5,1), and (1,5). Thus,

$$P(\text{at least one number greater than or equal to 4 is chosen}) = \frac{14}{20}.$$

Probabilities concerning chance experiments with equally likely outcomes will be considered in more detail in Section 1.6.

Problems

1.3.1 Draw a card at random from a standard deck of 52 playing cards. Let A be the event that an ace is drawn, B be the event that a black card is drawn, C be the event that a club is drawn, and K be the event that a king is drawn. Determine

a) $P(A \cap B)$.
b) $P(A \cup B)$.
c) $P(A \cap C)$.
d) $P(A \cup K)$.
e) $P(A \cap (B \cup C))$.
f) $P(A \cup B \cup K)$.

1.3.2 A fair six-sided die is to be rolled twice. Determine
a) the sample space associated with this chance experiment.
b) the probability that the total is 5.
c) the probability that the absolute value of the difference between the two rolls is 2.
d) the probability that the sum of the two rolls is even.

1.3.3 A fair die will be rolled and then a fair coin will be flipped twice. Determine the
a) probability that an even number is rolled and two heads are flipped.
b) probability that a 1 is rolled and two tails are flipped.
c) probability that a 1 is rolled and at least one tail is flipped.

1.3.4 Three fair dice will be rolled. Determine the
a) number of outcomes in Ω.
b) probability that exactly two sixes will be rolled.
c) probability that no sixes will be rolled.

1.3.5 A fair coin will be flipped five times. Determine the
a) number of outcomes in Ω.
b) probability that exactly two heads will be flipped.
c) probability that fewer than two heads will be flipped.
d) probability that at least two heads will be flipped.

1.3.6 Two marbles will be drawn at random and without replacement from an urn containing three white marbles, one blue marble, and one red marble. Determine the
a) number of outcomes in Ω.
b) probability that only white marbles are drawn.
c) probability that no white marbles are drawn.
d) probability that the blue marble is drawn.

1.3.7 Determine the number of outcomes in Ω when the chance experiment consists of drawing
a) three objects at random and without replacement from a collection of 12 objects.
b) three objects at random and with replacement from a collection of 12 objects.

1.3.8 Determine the number of outcomes in Ω when the chance experiment consists of drawing
a) four cards at random and with replacement from a standard deck of 52 playing cards.
b) four cards at random and without replacement from a standard deck of 52 playing cards.

1.3.9 Let Ω contain N equally likely outcomes. Show that
a) $P(A^c) = \frac{N-N(A)}{N}$. b) $P(A-B) = \frac{N(A)-N(A\cap B)}{N}$.

1.4 Conditional Probability

When probability theory is applied to real-world problems, it is often the case that an event A being studied is dependent on several other factors. For example, when a two-sided coin is flipped, the probability of flipping a head depends on whether or not the coin is a fair coin or a coin biased in favor of heads or tails. Without the knowledge of the type of coin being flipped, the probability of heads must be computed unconditionally. On the other hand, when the type of coin being flipped is known, the probability of flipping a head can be conditioned on this knowledge.

In general, a statistical model is built to explain how a set of explanatory variables $X_1, X_2, \dots, X_p$ are related to a response variable Y, and conditional probability models form the foundation of statistical modeling. For example, a researcher studying the incidence of lung cancer may be interested in the incidence of lung cancer among individuals who are long-term smokers or have been exposed to asbestos, rather than the unconditional rate of lung cancer among the general population, because the incidence of lung cancer would be expected to be higher for long-term smokers than it would be for the general population.

A probability computed utilizing known information about a chance experiment is called a *conditional probability*.

Definition 1.12 Let $(\Omega, \mathcal{A}, P)$ be a probability space, and let A and B be events in $\mathcal{A}$. The conditional probability of the event A given the event B, denoted by $P(A|B)$, is defined to be

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

provided $P(B) > 0$. The conditional probability $P(A|B)$ is said to be undefined when $P(B) = 0$.

$P(A|B)$ is stated as "the probability of the event A, given the event B has or will occur" or simply as "the conditional probability of event A given the event B." The event B is the known condition upon which the probability of A is computed, and the event B serves as the conditional sample space. That is, since B is assumed to have or will occur, only the outcomes in B are relevant to the chance experiment and the conditional probability that event A occurs. Thus, given the event B, the event A occurs only when the chance experiment results in an outcome in $A \cap B$.

Example 1.15 Suppose that a fair coin is flipped twice with $\Omega = \{HH, HT, TH, TT\}$ and each of the outcomes in Ω is equally likely. Let B be the event that at least one head has been flipped, and let A be the event that exactly two heads are flipped. Consider $P(A|B) = \frac{P(A \cap B)}{P(B)}$.

Since there is only one outcome in $A \cap B$, namely HH, and B contains the outcomes HH, HT, and TH, it follows that

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{1}{4}}{\frac{3}{4}} = \frac{1}{3}.$$

Note that $P(A)$ unconditionally is $\frac{1}{4}$; however, given the knowledge that at least one head was flipped, there is a larger chance that A will occur.

The probability rules for conditional probability are similar probability rules given in Section 1.2. Several conditional probability rules are given in Theorem 1.11.

Theorem 1.11 *Let $(\Omega, \mathcal{A}, P)$ be a probability space and let $A, B, C \in \mathcal{A}$. Then,*

i) $0 \leq P(A|B) \leq 1$.
ii) $P(B|B) = 1$.
iii) $P(A^c|B) = 1 - P(A|B)$.
iv) $P(A \cup B|C) = P(A|C) + P(B|C) - P(A \cap B|C)$.
v) $P(A \cap B) = P(A)P(B|A) = P(B)P(A|B)$.

Proof. Let $(\Omega, \mathcal{A}, P)$ be a probability space and let $A, B, C \in \mathcal{A}$.

i) First, $P(A|B) = \dfrac{P(A \cap B)}{P(B)}$ and $A \cap B \subset B$. Thus, it follows that $0 \leq P(A \cap B) \leq P(B)$, and hence, $0 \leq P(A|B) \leq 1$.
ii) $P(B|B) = \dfrac{P(B \cap B)}{P(B)} = \dfrac{P(B)}{P(B)} = 1$.
iii) Note that,

$$\begin{aligned} P(A^c|B) &= \frac{P(A^c \cap B)}{P(B)} = \frac{P(B) - P(A \cap B)}{P(B)} \\ &= \frac{P(B)}{P(B)} - \frac{P(A \cap B)}{P(B)} = 1 - P(A|B). \end{aligned}$$

iv) The proof of (iv) is left as an exercise.
v) The proof of (v) is left as an exercise.

■

The result given in Theorem 1.11(v) is known as the *Multiplication Law* and can be generalized to conditional probabilities involving than more than two events. For example, with three events A, B, and C, the Multiplication Law is $P(A \cap B \cap C) = P(A)P(B|A)P(C|A \cap B)$.

The Multiplication Law is often used in computing probabilities in chance experiments involving sampling without replacement from a collection of N objects, such as the chance experiment in Example 1.16.

Example 1.16 Suppose that an urn contains 8 red, 10 blue, and 7 green marbles. Three marbles will be drawn from the urn at random and without replacement. The probability of selecting three red marbles consists of selecting a red marble on the first draw, say event $R1$; a red marble on the second draw, say event $R2$; and selecting a red marble on the third draw, say event $R3$. Thus, by the Multiplication Law,

$$\begin{aligned} P(R1 \cap R2 \cap R3) &= P(R1) \times P(R2|R1) \times P(R2|R1 \cap R2) \\ &= \frac{8}{25} \times \frac{7}{24} \times \frac{6}{23} = \frac{336}{13\,800}. \end{aligned}$$

The Multiplication Law also provides an alternative form for the Law of Total Probability, which is given in Theorem 1.12.

Theorem 1.12 (Law of Total Probability) *Let A and B be events. Then,*

$$P(B) = P(B|A)P(A) + P(B|A^c)P(A^c).$$

Proof. Let A and B be events. From Theorem 1.4,

$$P(B) = P(B \cap A) + P(B \cap A^c),$$

and by the Multiplication Law, $P(B \cap A) = P(B|A)P(A)$ and $P(B \cap A^c) = P(B|A^c)P(A^c)$. Thus,

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

Example 1.17 Draw two cards at random and without replacement from a standard deck of 52 playing cards. Because the second draw is dependent on the outcome of the first draw, the Law of Total Probability will be used to determine the probability that an ace is drawn on the second draw, say event $A2$. Now, let $A1$ be the event that an ace was selected on the first draw. Then, $A1$ and $A1^c$ partition Ω, and by conditioning on the outcome of the first draw, it follows that $P(A2|A1) = \frac{3}{51}$ and $P(A2|A1^c) = \frac{4}{51}$. Thus, by Theorem 1.12,

$$\begin{aligned} P(A2) &= P(A1)P(A2|A1) + P(A1^c)P(A2|A1^c) \\ &= \frac{4}{52} \times \frac{3}{51} + \frac{48}{52} \times \frac{4}{51} = \frac{4}{52}. \end{aligned}$$

Theorem 1.13 generalizes the conditional form of the Law of Total Probability to partitions consisting of more than two events.

Theorem 1.13 (Generalized Law of Total Probability) *Let* $\{B_k : k \in \mathbb{N}\}$ *be a partition of* Ω *and let* $A \in \mathcal{A}$. *Then,* $P(A) = \sum_{k=1}^{\infty} P(A|B_k)P(B_k)$.

Proof. The proof follows directly from Theorem 1.6 and the Multiplication Law. ■

Note that the Generalized Law of Total Probability also applies to a finite partition, say $\{B_1, B_2, \dots, B_n\}$, $\forall n \in \mathbb{N}$, and in this case, $P(A) = \sum_{k=1}^{n} P(A|B_k)P(B_k)$.

Example 1.18 Suppose that a town T is in a region of high tornado activity. Let A be the event that a tornado hits town T, and let B_i be the event that there are $i \in \mathbb{W}$ tornadoes in town T's region during a tornado season. Suppose that $P(B_i) = \frac{e^{-3}3^i}{i!}$ and $P(A|B_i) = 1 - \frac{1}{2^i}$. The probability that a tornado hits town T is

$$
\begin{aligned}
P(A) &= \sum_{i=0}^{\infty} P(A|B_i)P(B_i) = \sum_{i=0}^{\infty} \left(1 - \frac{1}{2^i}\right) \times \frac{e^{-3}3^i}{i!} \\
&= \sum_{i=0}^{\infty} \frac{e^{-3}3^i}{i!} - \sum_{i=0}^{\infty} \frac{e^{-3}3^i}{2^i i!} = 1 - e^{-3} \sum_{i=1}^{\infty} \frac{\left(\frac{3}{2}\right)^i}{i!} \\
&= 1 - e^{-3}e^{\frac{3}{2}} = 1 - e^{-\frac{3}{2}} = 0.777.
\end{aligned}
$$

The Generalized Law of Total Probability shows that $P(A)$ can be computed when a partition $\{B_i\}$ is available and $P(B_i)$ and $P(A|B_i)$ are known for each of the events in the partition. Theorem 1.14, *Bayes' Theorem*, provides the solution to the inverse problem that concerns the probability of $B_i|A$.

Theorem 1.14 (Bayes' Theorem) *Let $\{B_k : k \in \mathbb{N}\}$ be a partition of Ω and let $A \in \mathcal{A}$, then for $i \in \mathbb{N}$*

$$
P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_{k=1}^{\infty} P(A|B_k)P(B_k)}.
$$

Proof. Let $\{B_k\}$ be a partition of Ω and let $A \in \mathcal{A}$. Let $i \in \mathbb{N}$ be arbitrary but fixed, then

$$
\begin{aligned}
P(B_i|A) &= \frac{P(B_i \cap A)}{P(A)} = \frac{\overbrace{P(A|B_i)P(B_i)}^{\text{by the Multiplication Law}}}{P(A)} \\
&= \underbrace{\frac{P(A|B_i)P(B_i)}{\sum_{k=1}^{\infty} P(A|B_k)P(B_k)}}_{\text{by the } \textit{Generalized Law of Total Probability}} .
\end{aligned}
$$

■

Note that the Generalized Law of Total Probability and Bayes' Theorem also work with finite partitions. For example, using B and B^c to partition Ω, Bayes' Theorem becomes $P(B|A) = \frac{P(A|B)P(B)}{P(A|B)P(A)+P(A|B^c)P(B^c)}$.

Example 1.19 Suppose that a new and rare infectious disease has been diagnosed. It is known that this new disease is contracted with probability $P(D) = 0.0001$, which is referred to as the *prevalence probability*. Suppose that a diagnostic test has been developed to diagnose this disease, and given that a person has the disease, the probability that the test is positive is $P(+|D) = 0.99$; $P(+|D)$ is called the *sensitivity* of the test and measures the ability of the test to correctly diagnose that an individual does have the disease. Also, suppose that given a person does not have the disease, the probability that the test is positive is $P(+|D^c) = 0.02$; the *specificity* of the test is $P(-|D^c) = 1 - P(+|D^c)$ and measures the ability of the test to correctly diagnose that an individual does not have the disease.

The probability that a person has the disease given a positive test result is

$$P(D|+) = \frac{P(+|D)P(D)}{P(+|D)P(D) + P(+|D^c)P(D^c)}$$
$$= \frac{0.0001 \times 0.99}{0.0001 \times 0.99 + 0.9999 \times 0.02} = 0.004\,926.$$

Example 1.20 Suppose that three silicon wafer plants produce blank DVDs with plant A manufacturing 45%, plant B manufacturing 30%, and plant C manufacturing 25%. The probability of a defective DVD given plant A produced the DVD is 0.01, the probability of a defective DVD given plant B produced the DVD is 0.02, and the probability of a defective DVD given plant C produced the DVD is 0.05. If a defective DVD was found, the probability that it was manufactured by plant A is

$$P(A|D) = \frac{P(D|A)P(A)}{P(D|A)P(A) + P(D|B)P(B) + P(D|C)P(C)}$$
$$= \frac{0.01 \times 0.45}{0.01 \times 0.45 + 0.02 \times 0.30 + 0.05 \times 0.25} = 0.196.$$

Similarly, if a defective DVD was found, the probability that it was manufactured by plant B is

$$P(B|D) = \frac{P(D|B)P(B)}{P(D|A)P(A) + P(D|B)P(B) + P(D|C)P(C)}$$
$$= \frac{0.02 \times 0.30}{0.01 \times 0.45 + 0.02 \times 0.30 + 0.05 \times 0.25} = 0.261.$$

Finally, for manufacturer C, the probability is

$$P(C|D) = \frac{P(D|C)P(C)}{P(D|A)P(A) + P(D|B)P(B) + P(D|C)P(C)}$$
$$= \frac{0.05 \times 0.25}{0.01 \times 0.45 + 0.02 \times 0.30 + 0.05 \times 0.25} = 0.543.$$

Thus, given a defective DVD is found, it was most likely produced by manufacturer C.

Example 1.21 Suppose that in Example 1.18, town T was hit by a tornado. The probability that there were three tornadoes in town T's region during a tornado season given that town T was hit by a tornado is

$$P(B_3|A) = \frac{P(B_3)P(A|B_3)}{\sum_{i=0}^{\infty} P(A|B_i)P(B_i)} = \frac{\left(1 - \frac{1}{2^3}\right) \frac{e^{-3}3^3}{3!}}{\sum_{i=0}^{\infty} \left(1 - \frac{1}{2^i}\right) \times \frac{e^{-3}3^i}{i!}}$$
$$= \frac{0.196}{0.777} = 0.252.$$

Problems

1.4.1 Suppose that an urn contains 10 red marbles and 5 black marbles. If two marbles are drawn from the urn at random and without replacement, determine the probability that

a) a red marble is drawn on the second draw given a black marble was drawn on the first draw.
b) a red marble is drawn on the second draw and a black marble was drawn on the first draw.
c) a red marble is drawn on the second draw.
d) two marbles of different colors are selected.
e) two marbles of the same color are selected.

1.4.2 Suppose that an urn contains 12 red marbles, 8 white marbles, and 5 black marbles. If three marbles are drawn from the urn at random and without replacement, determine the probability that

a) a white marble is drawn on the first draw, a red marble is drawn on the second draw, and a white marble is drawn on the third draw.
b) three marbles of the same color are selected.
c) a white marble is selected on the second draw.
d) a white marble is selected on the third draw.

1.4.3 Two marbles will be drawn at random and with replacement from an urn having 8 red marbles and 12 black marbles. Determine

a) the number of possible outcomes in Ω.
b) the probability of drawing marbles of different colors.
c) the probability of drawing marbles of the same color.

1.4.4 Two marbles will be drawn at random and with replacement from an urn having 8 red marbles, 4 black marbles, and 5 white marbles. Determine

a) the number of possible outcomes in Ω.
b) the probability of drawing two white marbles.
c) the probability of drawing one white marble.
d) the probability of drawing marbles of different colors.
e) the probability of drawing marbles of the same color.

1.4.5 Suppose that two cards are drawn at random and without replacement from a standard deck of 52 playing cards, determine the probability that

a) an ace is drawn on the second draw given a king was selected on the first draw.

b) an ace is drawn on the second draw and a king was selected on the first draw.
c) a club is drawn on the second draw.
d) cards of different colors are selected.

1.4.6 Suppose that urn A contains 5 red marbles and 10 black marbles and urn B contains 8 red marbles and 4 black marbles. If a marble is drawn at random from urn A and placed in urn B and then a marble is drawn from urn B at random, determine the probability that
a) a black marble is drawn from urn A and a red marble is drawn from urn B.
b) a red marble is drawn from urn B.
c) marbles of the same color are drawn from urns A and B.
d) marbles of the opposite colors are drawn from urns A and B.

1.4.7 Using the information in Example 1.19 with $P(+|D^c) = p$. Determine the value of p so that $P(D|+) = 0.75$.

1.4.8 A computer company has two assembly locations with 84% of their computers assembled in location A and 16% assembled in location B. Given that a computer is assembled in location A, the probability that a computer works perfectly is 0.98; given that a computer is assembled in location B, the probability that the computer works perfectly is 0.92. Determine the probability that
a) one of the computers supplied by this company works perfectly.
b) one of the computers supplied by this company was assembled in location A, given that it works perfectly.

1.4.9 A university club holds dances at three different bars. The club uses bar A 25% of the time, bar B 60% of the time, and bar C 15% of the time. Suppose that the probability that a fight breaks out given that bar A was used is 0.30, 0.10 given that bar B was used, and 0.50 given that bar C was used. Determine
a) the probability that a fight breaks out a club dance.
b) the probability bar B was used given that a fight broke out.
c) the probability bar C was used given that a fight broke out.
d) the bar most likely to have been used given that a fight broke out.

1.4.10 Suppose that two cards are drawn at random and without replacement from a standard deck of 52 playing cards, determine the probability that
a) a king, queen, or a jack is selected on the second draw.
b) a king was selected on the first draw given a king, queen, or jack was selected on the second draw.

1.4.11 A fishing fleet has three different captains it uses with a particular ship during salmon season. The fleet uses Captain I 65% of the time, Captain II 25% of the time, and Captain III 10% of the time. Suppose that the probability that the ship reaches its quota given that Captain I was used is 0.72, 0.43 given that Captain II was used, and 0.83 given that Captain III was used. Determine

a) the probability that the ship fulfills its quota.
b) the probability that Captain I was used given that the ship fulfills its quota.
c) the probability Captain III was used given that the ship fulfills its quota.
d) the captain most likely to have been used given the ship fulfills its quota.

1.4.12 A gold mining company uses four different remediation techniques, say R1, R2, R3, and R4. The company uses remediation technique R1 20% of the time, remediation technique R2 40% of the time, remediation technique R3 30% of the time, and remediation technique R4 10% of the time. Suppose that the probability of a successful remediation given that technique R1 was used is 0.5, 0.3 given that R2 was used 0.2, given that technique R3 was used, and 0.25 given technique R4 was used. Determine the probability

a) of a successful remediation.
b) remediation technique R2 was used given that the remediation was successful.
c) remediation technique R1 or technique R3 was used given that the remediation was successful.

1.4.13 In a small town, three lawyers (A, B, and C) serve as public defenders. Court records indicate that lawyer A handles 40% of the cases, lawyer B 30% of the cases, and lawyer C 30% of the cases. Furthermore, the probability that a defendant is acquitted given that lawyer A handled the case is 0.25, 0.20 given that lawyer B handled the case, and 0.30 given that lawyer C handled the case. Determine the probability that

a) a defendant using a public defender is acquitted.
b) lawyer B handled the case given that a defendant using a public defender is acquitted.
c) lawyer C handled the case given that a defendant using a public defender is acquitted.
d) lawyer A handled the case given that a defendant using a public defender is acquitted.

1.4.14 In Example 1.21, given a tornado hit town T, determine the most likely number of tornadoes there were in town T's region. *Hint:* Determine the largest value of $P(B_i|A)$.

1.4.15 *Prove*: If $P(A|B) > P(A)$, then $P(B|A) > P(B)$.

1.4.16 Prove
a) Theorem 1.11 part (iv).
b) Theorem 1.11 part (v).

1.5 Independence

In some cases, the conditional probability of an event A given that the event B is the same as the unconditional probability of the event A. In this case, knowledge of the event B occurring does not provide any information about the probability that event A occurs, and the events A and B are said to be *independent events.*

Definition 1.13 Let $(\Omega, \mathcal{A}, P)$ be a probability space and let $A, B \in \mathcal{A}$. The events A and B are said to be independent events if and only if any one of the following three conditions is satisfied.

i) $P(A \cap B) = P(A)P(B)$.
ii) $P(A|B) = P(A)$, provided $P(B) > 0$.
iii) $P(B|A) = P(B)$, provided $P(A) > 0$.

Example 1.22 Suppose that a fair coin is flipped twice. Let A be the event that a head is flipped on the first flip, and let B be the event that a tail is flipped on the second flip. There are four equally likely outcomes in Ω and the event $A = \{HT, HH\}$ and the event $B = \{TT, HT\}$. Thus, $P(A) = \frac{1}{2}$, $P(B) = \frac{1}{2}$, and since $P(A \cap B) = P(\{HT\}) = \frac{1}{4}$, it follows that A and B are independent events.

Note that in a chance experiment where n objects are drawn with replacement, the outcome of each draw is independent of the outcome of any other draw; however, this is not the case when the draws are made without replacement.

Theorem 1.15 shows that when A and B are independent events, then so are the pairs of events A and B^c, A^c and B, and A^c and B^c.

Theorem 1.15 *If A and B are independent events, then*

i) A^c *and* B *are independent.*
ii) A *and* B^c *are independent.*
iii) A^c *and* B^c *are independent.*

Proof. Let A and B be independent events.

i) Consider $P(A^c \cap B)$.

$$P(A^c \cap B) = P(B) - P(A \cap B) = P(B) - P(A)P(B)$$
$$= P(B) \times [1 - P(A)] = P(B)P(A^c).$$

Thus, A^c and B are independent whenever A and B are independent.

ii) The proof of part (ii) is left as an exercise.

iii) The proof of part (iii) is left as an exercise.

■

The definition of independence can be generalized to a family of events, say $\{A_i : i \in \mathbb{N}\}$. For example, if every pair of events in $\mathcal{A}$ is a pair of independent events, then the events in $\{A_i : i \in \mathbb{N}\}$ are called a *pairwise independent* events. Definition 1.14 gives the definition of a family of *mutually independent* events.

Definition 1.14 A family of events $\{A_i : i \in \mathbb{N}\}$ is said to be a family of mutually independent events if and only if

$$P\left(\bigcap_{i \in J} A_i\right) = \prod_{i \in J} P(A_i)$$

for all subsets J of $\mathbb{N}$.

Note that, in a family of mutually independent events, the probability of the intersection of any subcollection of the family of events is the product on the events in the subcollection. For example, three events A, B, and C are mutually independent only when $P(A \cap B) = P(A)P(B), P(A \cap C) = P(A)P(C)$, $P(B \cap C) = P(B)P(C)$, and $P(A \cap B \cap C) = P(A)P(B)P(C)$. Furthermore, it is possible for a family to be a family of pairwise independent events but not a family of mutually independent events as illustrated in Example 1.23.

Example 1.23 Flip a fair coin twice so that the four outcomes in Ω are equally likely. Let $A = \{HH, HT\}, B = \{HT, TH\}$, and $C = \{HH, TH\}$. Then, $P(A) = P(B) = P(C) = \frac{1}{2}$ and

$$P(A \cap B) = P(\{HT\}) = \frac{1}{4} = P(A)P(B)$$
$$P(A \cap C) = P(\{HH\}) = \frac{1}{4} = P(A)P(C)$$
$$P(B \cap C) = P(\{TH\}) = \frac{1}{4} = P(B)P(C)$$

but

$$P(A \cap B \cap C) = P(\emptyset) = 0 \neq \frac{1}{8} = P(A)P(B)P(C).$$

Thus, A, B, and C are pairwise independent events but not mutually independent events.

In chance experiments consisting of n trials where the outcome of each trial is independent of the outcomes of the other trials, the outcomes of the trials are mutually independent events. Moreover, the probability of a particular outcome of the chance experiment is simply the product of the probabilities of the corresponding trial outcomes.

Example 1.24 A basketball practice will end when a particular player makes a shot from half court. If the outcome of each shot, hit (H) or miss (M), is independent, and the probability that a half court shot will be made is 0.05, then the probability that it will take five shots to end the practice is

$$\begin{aligned} P(\text{M and M and M and M and H}) &= P(\text{M})P(\text{M})P(\text{M})P(\text{M})P(\text{H}) \\ &= 0.95^4 \times 0.05 = 0.041. \end{aligned}$$

In general, the probability that it takes k shots to end the practice is $0.95^{k-1} \times 0.05$.

Example 1.25 Suppose that two fair dice will be rolled repeatedly and independently until a total of 6 or a total of 7 appears. To determine the probability that a total of 6 is rolled before a total of 7 is rolled, let A_i be the event that the first total of 6 is rolled before a total of 7 is rolled occurs on the ith roll. Then, the collection of events $\{A_i\}$ is a partition of the event that a 6 is rolled before a 7 and

$$P(6 \text{ before } 7) = P\left(\bigcup_{i=1}^{\infty} A_i\right) \overbrace{=}^{\text{disjoint}} \sum_{i=1}^{\infty} P(A_i).$$

Now, the event A_i occurs when the first $i-1$ rolls result in totals that are not 6 or 7, and the ith roll is a total of 6. Thus, since the rolls of the dice are independent,

$$P(A_i) = P\left(\bigcap_{i=1}^{i-1} (6 \cup 7)^c \cap 6\right) \overbrace{=}^{\text{ind.}} \left(\frac{25}{36}\right)^{i-1} \times \frac{5}{36}.$$

Hence,

$$\begin{aligned} P(6 \text{ before } 7) &= \sum_{i=1}^{\infty} P(A_i) = \sum_{i=1}^{\infty} \left(\frac{25}{36}\right)^{i-1} \times \frac{5}{36} \\ &= \frac{5}{36} \underbrace{\sum_{i=1}^{\infty} \left(\frac{25}{36}\right)^{i-1}}_{\text{geometric series}} = \frac{5}{36} \times \frac{1}{1 - \frac{11}{36}} = \frac{5}{11}. \end{aligned}$$

Problems

1.5.1 Suppose that $P(A) = 0.3$, $P(B) = 0.4$, and A and B are independent events. Determine

a) $P(A \cup B)$.
b) $P(A - B)$.
c) $P(A^c \cup B^c)$.
d) $P((A - B) \cup (B - A))$.

1.5.2 Two cards will be drawn at random and with replacement from a standard deck of playing cards. Determine the probability of drawing

a) two aces.
b) two cards of different suits.
c) two cards of the same suit.

1.5.3 Three geological exploration companies are searching independently for new shale deposits in Eastern Montana. Suppose that the probability that company A finds new shale in Eastern Montana is 0.4, the probability that company B finds new shale in Eastern Montana is 0.6, and the probability that company C finds new shale in Eastern Montana is 0.5. Determine the probability that

a) exactly two companies find new shale deposits in Eastern Montana.
b) at least two of the companies find new shale deposits in Eastern Montana.
c) at least one of the companies finds a new shale deposit in Eastern Montana.

1.5.4 The goal of a military operation is to destroy a strategic target by firing missiles at the target. The target will be destroyed when two missiles have hit the target. Suppose that each missile is fired at the target independently, and the probability that a missile hits the target is 0.6. If missiles will be fired until the second missile hits the target, determine

a) the probability that only two missiles will have to be fired in order to destroy the target.
b) the probability that four missiles will have to be fired in order to destroy the target.
c) the probability that at most four missiles will have to be fired in order to destroy the target.

1.5.5 Suppose that the probability of winning any money at all on a single play of the Lucky Dollar poker machine is 0.14. If each play of the Lucky Dollar poker machine is independent, determine the probability that

a) a player wins on five consecutive plays.
b) a player wins at least once in five consecutive plays.
c) a player's first win occurs on the kth play.
d) a player's first win occurs in 10 or fewer plays.

1.5.6 A basketball practice will end when a player hits a shot from half court. Suppose that each shot is independent of the others, and the probability that a half court shot will be made is 0.05. Determine the probability that it takes more than 20 shots to end practice.

1.5.7 Let A, B, C be mutually independent events. Show that
a) A and $B \cap C$ are independent events.
b) A and $B \cup C$ are independent events.
c) A and $B - C$ are independent events.

1.5.8 Let $\Omega = \{ABB, BAB, BBA, AAA\}$ and suppose that the outcomes in Ω are equally likely. Let A_i be the event that A occurs in the ith position for $i = 1, 2, 3$. Show that A_1, A_2, and A_3 are pairwise independent but not mutually independent.

1.5.9 Suppose that $P(A) = 0.25$, $P(B) = 0.3$, and $P(C) = 0.16$. Determine $P(A \cup B \cup C)$ when A, B, and C are
a) disjoint events.
b) mutually independent events.

1.5.10 Suppose that a multiple-choice test has n questions and each question has four possible choices. A student will guess at random on each question making each guess independent of the others, and the probability of guessing the correct answer on any question is 0.25. Determine the number of questions (n) that would have to be guessed so that the probability of guessing the correct answer on at least one question is at least 0.98.

1.5.11 An academic integrity committee consists of three members, two faculty representatives and one student representative. Suppose that each of the committee members votes independently on each student appeal case, and at least two committee members must agree with the student appeal for the committee to rule in favor of the student. Furthermore, based on the past records, the faculty rule in favor of a student appeal with probability 0.10 and the student rules in favor with probability 0.3. Determine the probability that the committee rules in favor of a student appeal.

1.5.12 Two fair dice will be rolled repeatedly and independently until a total of 6 or a total of 5 appears. Determine the probability that a total of 6 is rolled before a total of 5 is rolled.

1.5.13 Players A and B will alternate rolling a fair die independently until a 6 appears. The first player to roll a 6 wins the game. Determine the probability that player A wins the game
a) when player A rolls first.
b) when player B rolls first.
c) when the players randomly chose who goes first.

1.5.14 If $\{A_i : i \in \mathbb{N}\}$ are mutually independent events, show that

$$P\left(\bigcup_{i=1}^{n} A_i\right) = 1 - \prod_{i=1}^{n}[1 - P(A_i)].$$

1.5.15 Let A and B be independent events. Show that
a) A and B^c are independent events.
b) A^c and B^c are independent events.

1.5.16 If $P(A) = 0$, show that A is independent of every other event B of Ω.

1.5.17 Let A and B be events and let A be a subset of B. Show that A and B are independent events when $P(A) = 0$ or $P(B) = 0$.

1.5.18 Let A and B be events, and let A be a nonempty subset of B with $P(A) > 0$. If $P(B) < 1$, show that A and B cannot be independent events.

1.6 Counting Methods

Probability computations often involve the enumeration of the possible outcomes of a chance experiment, enumeration of the outcomes in an event, or the enumeration of the ways an event can occur. The *Fundamental Principal of Counting*, given in Theorem 1.16, is one of the most important tools for enumeration and can be used on chance experiments with or without equally likely outcomes.

Theorem 1.16 (Fundamental Principle of Counting) *Let a chance experiment $\mathcal{E}$ be carried out by carrying out the subexperiments $\mathcal{T}_1, \mathcal{T}_2, \mathcal{T}_3, \ldots, \mathcal{T}_k$. If there are n_i possible outcomes to subexperiment $\mathcal{T}_i$, then the number of possible outcomes for the chance experiment $\mathcal{E}$ is* $N(\Omega) = \prod_{i=1}^{k} n_i$.

Example 1.26 Suppose that a pizza parlor offers pizzas based on three different types of crusts, two different sauces, and 15 different toppings. Making a three distinct topping pizza is based on five subexperiments, namely,

- $\mathcal{T}_1$ – choose the type of crust with $n_1 = 3$,
- $\mathcal{T}_2$ – choose the type of sauce with $n_2 = 2$,
- $\mathcal{T}_3$ – choose the first topping with $n_3 = 15$,
- $\mathcal{T}_4$ – choose the second topping with $n_4 = 14$ since one topping has already been chosen,
- and $\mathcal{T}_5$ – choose the third topping with $n_5 = 13$ since two toppings have already been chosen.

Thus, the number of possible pizzas having three different toppings is $N(\Omega) = 3 \times 2 \times 15 \times 14 \times 13 = 16\,380$.

Theorem 1.17 *If a chance experiment consists of drawing n times*

i) with replacement from N objects, then $N(\Omega) = N^n$.
ii) without replacement from N objects, then
$N(\Omega) = N(N-1) \times \cdots \times (N-n+1)$.

Proof. This theorem follows directly from the Fundamental Principle of Counting. ∎

Example 1.27 Suppose that a chance experiment involves rolling a fair six-sided die five times. Since there are six possible outcomes to each roll, $N(\Omega) = 6^5 = 7776$.

Example 1.28 Suppose that a chance experiment involves selecting three people at random and without replacement from a group of seven people to serve as club president, vice president, and secretary. Then, $N(\Omega) = 7 \times 6 \times 5 = 210$.

The Fundamental Principle of Counting can be very useful when computing probabilities associated with a chance experiment having equally likely outcomes. To use the Fundamental Principle of Counting to compute $P(A)$ when the chance experiment consists of several tasks and equally likely outcomes:

1) Define each of the k tasks involved in the chance experiment.
2) Determine the number of possible outcomes to task T_k, say n_k.
3) $N(\Omega) = \prod_{i=1}^{n} n_i$.
4) Determine the number of favorable outcomes to task T_k, say f_k.
5) $N(A) = \prod_{i=1}^{n} f_i$.
6) $P(A) = \frac{N(A)}{N(\Omega)}$.

Example 1.29 Suppose that a standard state license plate is made of three letters selected from $\{A, B, C, \dots, Z\}$ followed by three digits selected from $\{0, 1, 2, \dots, 9\}$. Then, the number of possible distinct license plates that can be made is $N(\Omega) = 26 \times 26 \times 26 \times 10 \times 10 \times 10 = 17\,576\,000$.

Now, if a license plate is selected at random, the probability that a license plate will have three different letters is

$$\begin{aligned} P(\text{three different letters}) &= \frac{N(\text{three different letters})}{N(\Omega)} \\ &= \frac{26 \times 25 \times 24 \times 10 \times 10 \times 10}{17\,576\,000} = \frac{15\,600\,000}{17\,576\,000} \\ &= 0.888. \end{aligned}$$

The probability that all three letters on a license plate match while none of the digits on a license plate match is

$$\frac{26 \times 1 \times 1 \times 10 \times 9 \times 8}{17\,576\,000} = \frac{18\,720}{17\,576\,000} = 0.0011.$$

The two most important considerations when computing probabilities in a chance experiment that involves sampling from N distinct objects are (i) whether or not the sampling is with replacement and (ii) whether or not the order in which the objects are sampled is important. For example, when dealing five cards without replacement to form a hand of cards, the order in which the five cards are dealt does not matter; that is, a hand of cards consisting of $AH, KS, 5D, 4C,$ and $2C$ is the same hand, no matter in which order the cards were dealt. On the other hand, the order in which the objects are selected may be important such as in the case where nine digits are chosen to form a social security number.

Definition 1.15 A permutation of n distinct objects is an ordered sequence of the n objects. A partial permutation of size r is an ordered sequence of r of the n objects.

For example, *abc* and *bca* are two different permutations of the letters *a*, *b*, and *c*, and *ab* and *cb* are different partial permutations of size 2 made from the letters *a*, *b*, and *c*.

Theorem 1.18 *The number of partial permutations of size r made from n distinct objects when sampling without replacement is* $\dfrac{n!}{(n-r)!}$.

Proof. This theorem follows directly from the Fundamental Principle of Counting. ∎

Example 1.30 There are $\dfrac{12!}{(12-4)!} = 11\,880$ partial permutations of size 4 when sampling with replacement from 12 distinct objects, and there are $12! = 479\,001\,600$ permutations of all 12 objects.

Example 1.31 Five couples are arranged in a row of 10 chairs. Let A be the event that five couples are seated in a fashion so that a husband is seated next to his wife. Now, the number of ways to seat 10 people in the 10 chairs is $N(\Omega) = 10! = 3\,628\,800$, and the event A comprised the following six tasks:

- $\mathcal{T}_1$ – Choose the order in which the 5 couples will be seated. $n_1 = 5! = 120$.
- $\mathcal{T}_2$ – Seat the first couple. $n_2 = 2$.
- $\mathcal{T}_3$ – Seat the second couple. $n_2 = 2$.
- $\mathcal{T}_4$ – Seat the third couple. $n_2 = 2$.
- $\mathcal{T}_5$ – Seat the fourth couple. $n_2 = 2$.
- $\mathcal{T}_6$ – Seat the fifth couple. $n_2 = 2$.

Thus, $N(A) = 5! \times 2^5 = 3840$. Hence, $P(A) = \frac{3840}{3\,628\,800} = 0.0011$.

When sampling without replacement, and the order the objects are selected is not important, the objects selected are called a *combination*.

Definition 1.16 A combination of size r is an unordered subset of r distinct objects selected from a set of n distinct objects.

For example, abc and cab are the same combination of the letters a, b, and c. The four combinations of size 3 for the letters a, b, c, and d are abc, abd, acd, bcd.

Theorem 1.19 *The number of combinations of size r that can be formed by selecting r objects without replacement from n distinct objects is* $\binom{n}{r} = \dfrac{n!}{r!(n-r)!}$.

Proof. The proof of Theorem 1.19 is left as an exercise. ■

Note that $\binom{n}{k}$ is the kth term in the nth row of Pascal's Triangle, and $\binom{n}{k}$ is also the kth coefficient in the *binomial expansion* of $(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$.

Example 1.32 The number of possible five-card poker hands that can be dealt from a standard deck of 52 playing cards is $\binom{52}{5} = 2\,598\,960$. Let A be the event a five-card hand contains four aces. Then, A is comprised of two tasks $\mathcal{T}_1$ – deal all four aces from the four aces in the deck to the hand and $\mathcal{T}_2$ – deal the fifth card from the remaining 48 cards in the deck to the hand. Thus,

$$P(A) = \frac{\binom{4}{4}\binom{48}{1}}{\binom{52}{5}} = \frac{1 \cdot 48}{2\,598\,960} = 0.000\,018.$$

The number of combinations of size r can also be used to count the number of ways n objects can be assigned to two cells with one cell receiving r objects and the other cell $n - r$ objects.

Example 1.33 The number of possible equally likely outcomes when a fair coin is flipped 20 times is $2^{20} = 1\,048\,576$. The number of outcomes resulting in 12 heads and eight tails is $\binom{20}{12} = 125\,970$, and thus, the probability of flipping 12 heads when a fair coin is flipped 20 times is $\frac{125\,970}{1\,048\,576} = 0.12$.

Example 1.34 A state lottery generally consists of selecting six numbers from 1 to n without replacement, and the order of selection is unimportant. A lottery player will select their own six numbers from 1 to n, and then the state will pick the six winning lottery numbers without replacement. Because order is unimportant, the number of possible lottery combinations is $\binom{n}{6}$. For example, when $n = 51$, there are $\binom{51}{6} = 18\,009\,460$ possible lottery combinations.

A player will usually win prize money for matching three, four, five, or six of the state's numbers. Now, for a player to match x of the winning numbers, x of the player's numbers must be match the six winning numbers and $6 - x$ of the player's numbers must match the 45 numbers that were not chosen. Thus, the probability that a player will match x of the state's six numbers is

$$P(x \text{ matched numbers}) = \frac{\binom{6}{x}\binom{n-6}{6-x}}{\binom{n}{6}}.$$

With $n = 51$ and $x = 3$, the probability that a player matches 3 of the winning numbers is $\frac{\binom{6}{3}\binom{45}{3}}{\binom{45}{6}} = 0.0158$.

Theorem 1.20 generalizes the number of ways n distinct objects can be assigned to more than two cells.

Theorem 1.20 *The number of ways n distinct objects can be distributed into k distinct cells is*

$$\frac{n!}{n_1!n_2!\cdots n_k!} = \binom{n}{n_1, n_2, n_3, \ldots, n_k},$$

where $n_i \geq 0$ *for* $i = 1, 2, 3, \ldots, k$ *and* $\sum_{i=1}^{k} n_i = n$.

Example 1.35 The number of ways that a standard deck of 52 playing cards can be dealt so that each player receives 13 cards is $\binom{52}{13,13,13,13} = \frac{52!}{13!^4}$. Let A be the event that one of the four players is dealt all four aces. Then, A is comprised the tasks $\mathcal{T}_1$ – choose one of the players to receive the four aces, $\mathcal{T}_2$ – distribute the four aces to the player, and $\mathcal{T}_3$ – distribute the remaining 48 cards so that each player has been dealt 13 cards. Thus,

$$P(A) = \frac{N(A)}{N(\Omega)} = \frac{\binom{4}{1}\binom{4}{4}\binom{48}{9,\ 13,\ 13,\ 13}}{\binom{52}{13,\ 13,\ 13,\ 13}} = 0.0106.$$

Counting methods can also be useful for enumerating the possible outcomes in chance experiments where the outcomes are not equally likely.

Example 1.36 Suppose that a multiple-choice test consists of 10 questions, with each question having four options of which only one option is the correct answer. If a student guesses at random on each question, then the probability that a student guesses correctly on any particular question is $\frac{1}{4}$. If a student guesses independently on each of the questions, the probability of answering seven questions correctly is

$$\binom{10}{7}\left(\frac{1}{4}\right)^7\left(\frac{3}{4}\right)^3 = 0.0031$$

since (1) the probability of any sequence of seven correct answers and three incorrect answers is $\left(\frac{1}{4}\right)^7\left(\frac{3}{4}\right)^3$ and (2) there are $\binom{10}{7}$ ways to assign the 10 questions to seven correct answers and three incorrect answers.

Problems

1.6.1 Phone numbers in a particular state begin with one of three area codes, followed by one of seven prefixes and finally four digits (i.e. xxx-xxx-xxxx). Determine how many possible distinct phone numbers are possible

a) without further restriction.
b) if a phone number must begin with the prefix 491.
c) if the last four digits cannot be 0000.
d) if the last three digits must be 000.

1.6.2 A particular state's license plates have the form LCCDDDD, where L is either an A for auto or a T for truck, CC is one of 56 county identifiers, and D is a digit. Determine how many distinct license plates there are

a) for automobiles.
b) for a particular county.
c) with no repeated digits.
d) where the product of the digits is even.

1.6.3 A pizza parlor makes pizzas with three different types of crusts, two types of sauces, 10 different toppings, and all pizzas come with cheese. Determine the number of distinct types of

a) one-topping pizzas that can be made.
b) two-topping pizzas that can be made.
c) pizzas that can be made.

1.6.4 Suppose that a club consists of six men and three women. From the nine club members, three members will be selected at random and without replacement to serve as club officers. If, the first member selected will serve as President, the second as Vice President, and the third as Treasurer, determine

a) the number of possible outcomes for selecting this club's officers.
b) the probability that all of the club's officers are female.
c) the probability that the club's president is female.

1.6.5 Three numbers are chosen at random and without replacement from a pool of 15 numbers consisting of 0, six positive numbers, and eight negative numbers. Determine the probability that

a) no negative numbers are chosen.
b) at least one negative number is chosen.
c) the product is positive.
d) the product is nonnegative.
e) the product is 0.
f) the product is negative.

1.6.6 A bridge hand consists of dealing 52 playing cards to four players such that each player receives 13 cards. Determine the probability that

a) each player has one ace.
b) one player has all four aces and all four kings.

1.6.7 From a standard deck of 52 playing cards, five cards are dealt at random and without replacement to four players so that each player receives five cards and 32 are left undealt.

a) How many possible ways are there to deal the four hands?
b) What is the probability that each player is dealt one ace?
c) What is the probability that one player receives all four aces?

1.6.8 A company has 25 trucks of which 20 are in working order and 5 are in the shop for repair. If four trucks are selected at random and without replacement, determine the probability that
a) all four of the selected trucks are in working order.
b) all four of the selected trucks are in the shop for repair.
c) at least one of the selected trucks is in working order.
d) at least two of the selected trucks are in working order.

1.6.9 A wardrobe consists of five pairs of pants of which three pairs are blue, 12 shirts of which 4 are white, seven pairs of socks, and black, brown, and tan shoes. Determine the number of distinct ensembles (i.e. pants, shirt, socks, and shoes) possible
a) with no restrictions.
b) with blue pants.
c) with blue pants, a white shirt, and black or brown shoes.

1.6.10 Suppose that four cards are drawn at random and without replacement from a standard deck of 52 playing cards. Determine the probability that
a) two aces are drawn.
b) two hearts are drawn.
c) at least one ace is drawn.
d) one ace and one king are drawn.

1.6.11 Suppose that four digits are to be selected at random with replacement. Determine the probability that
a) the product is even.
b) the product is odd.
c) the sum is even.
d) the product is positive.

1.6.12 Suppose that five male/female couples are to be seated in a row of chairs. Determine the probability that
a) the seating arrangement alternates MFMFMFMFMF.
b) the seating arrangement is MMMMMFFFFF.
c) no couple is seated together.

1.6.13 For the state lottery given in Example 1.34 with $n = 51$, compute the probability of matching

a) 0 numbers.
b) 1 number.
c) at least 3 numbers.
d) at most 1 number.

1.6.14 For the state lottery given in Example 1.34 with $n = 56$, compute the probability of matching

a) 3 numbers.
b) 4 numbers.
c) 5 numbers.
d) at least 3 numbers.

1.6.15 Prove Theorem 1.19.

1.6.16 Using the binomial expansion of $(1 + 1)^n$, show that $\sum_{k=0}^{n} \binom{n}{k} = 2^n$.

1.6.17 Show that

a) $\binom{n}{k} = \binom{n}{n-k}$.
b) $\binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1}$.

1.6.18 Show that $\binom{n}{n_1, n_2 \dots, n_k} = \binom{n}{n_1}\binom{n-n_1}{n_2} \cdots \binom{n-n_1-n_2-\cdots-n_{k-1}}{n_k}$ where $\sum_{i=1}^{k} n_k = n$.

1.6.19 Evaluate the following sums:

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

b) $$\sum_{k=0}^{n} \binom{n}{k} 2^k.$$

c) $$\sum_{k=0}^{n} \binom{n}{k} (\alpha - 1)^k.$$

1.6.20 Show that $\sum_{j=0}^{k} \binom{n}{j}\binom{m}{k-j} = \binom{n+m}{k}$ for $n, m, k \in \mathbb{N}$. *Hint:* Consider the binomial expansion of $(x + 1)^{n+m}$.

1.7 Case Study – The Birthday Problem

A famous probability problem posed sometime in the twentieth century is called the *Birthday Problem*. The Birthday Problem deals with the probability of people sharing common birthdays, which turns out to be a fairly common coincidence.

Statement of the Birthday Problem: What is the probability that two or more people in a room of n randomly assembled people share a common birthday?

The solution to the Birthday Problem depends on the assumptions being made on the assumed probability model, and different models will yield different solutions. The probability model being used to solve the Birthday Problem here is one of the simplest models possible; more complicated models may yield a more accurate probability for the solution to the Birthday Problem, but they are seldom used.

The Assumptions of the Probability Model:

- Because the people in the room are randomly assembled, assume that the birthdays of the individuals in the room are independent events.
- Because leap-year birthdays (i.e. February 29) are rare, assume that the birthdays all occur in nonleap years, leaving 365 possible birthdays.
- Assume that birthdays are uniformly distributed over the 365 possible birthdays so that each of the 365 birthdays is equally likely.

With these assumptions, the birthday problem is reduced to a repeated sampling problem where the chance experiment consists of choosing n birthdays with replacement from the 365 possible birthdays. Thus, under this model, $N(\Omega) = 365^n$.

Let A be the event that two or more people share the same birthday. While simple in statement, the event A is actually a fairly complex event. For example, a small sampling of the possible outcomes in A includes that only two people share the same birthday, three people share the same birthday, two people share the same birthday and another two share a different birthday, all but one person in the room share the same birthday, and all n people share the same birthday. To compute the probability of A directly, one would need to determine all of the possible ways at least two people in a room of n people share the same birthday and then sum the probabilities of these outcomes.

On the other hand, complementary event A^c is the event that no one in the room shares a common birthday. Because $P(A^c)$ is easier to compute, the probability of A will be found using the complement rule, $P(A) = 1 - P(A^c)$.

Now, the number of ways none of the individuals in the room share a common birthday can be modeled as sampling n birthdays without replacement from the 365 possible birthdays. In other words, the first person can have any of the 365 possible birthdays, the second person can have any of the 364 remaining birthdays, and so on, until the last person can have any of the $365 - n + 1$ remaining birthdays. Thus,

$$N(A^c) = 365 \times 364 \times 363 \times \cdots \times (365 - n + 1)$$

and

$$P(A^c) = \frac{365 \times 364 \times 363 \times \cdots \times (365 - n + 1)}{365^n}.$$

Finally, the probability that two or more people in a room of n randomly assembled people share a common birthday is

$$P(A) = 1 - P(A^c) = 1 - \frac{365 \times 364 \times 363 \times \cdots \times (365 - n + 1)}{365^n}$$
$$= 1 - \frac{\frac{365!}{(365-n)!}}{365^n}.$$

For example, for $n = 10$, the probability that two or more people in a room of 10 randomly assigned people share a common birthday is

$$P(A) = 1 - \frac{365 \times 364 \times 363 \times \cdots \times 356}{365^{10}} = 0.1169,$$

and for $n = 23$, $P(A) = 1 - \frac{365\times364\times363\times\cdots\times344}{365^{22}} = 0.5073$. Thus, there is better than a 50% chance that at least two people in a room of 23 randomly assembled people will share a common birthday.

The probability of two or more people in a room of n randomly assembled people sharing the same birthday plotted as a function of n is shown in Figure 1.1. Note that the probability at least two people in a room of n randomly assembled people will share a common birthday is greater than 0.90 when $n \geq 41$.

Finally, other probability models could have been used for modeling the probability that two or more people in a room of n randomly assembled people share a common birthday. For example, a probability model that includes February 29 as possible birthday could be considered; however, in this model, it would not

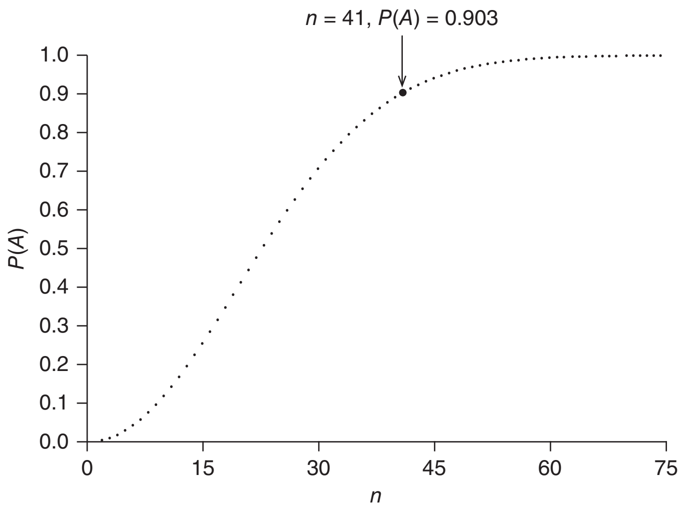

Figure 1.1 A plot of $P(A)$ for $n = 2, 3, \ldots, 75$.

be reasonable to assume that the 366 birthdays are equally likely. An alternative probability model could also be built using unequal empirical probabilities for each of the possible birthdays.

Problems

1.7.1 Determine the probability that in a room of 40 people, at least two people share the same birthday.

1.7.2 Determine the probability that in a room of 50 people, at least two people share the same birthday.

1.7.3 Using a probability model analogous to the one used in solving the Birthday Problem, determine the probability that two or more people in a room of n randomly assembled people have a birthday in the same week

a) assuming that there are 52 equally likely weeks in a year.
b) assuming that there are 52 equally likely weeks in a year and $n = 10$.
c) assuming that there are 52 equally likely weeks in a year and $n = 15$.
d) assuming that there are 52 equally likely weeks in a year and $n = 25$.

1.7.4 Using the setting of Problem 1.7.3, determine the value of n so that the probability that two or more people in a room of n randomly assembled people have a birthday in the same week is at least 0.50.

1.7.5 Suppose that the last two digits of a car license plate are numerical with possible values $00, 01, 02, \ldots, 99$. Determine the probability that at least two cars in a parking lot of

a) n randomly assembled cars have the same last two digits.
b) 10 randomly assembled cars have the same last two digits.
c) 20 randomly assembled cars have the same last two digits.

1.7.6 Using the setting of Problem 1.7.5, determine the value of n so that the probability that two or more cars in a parking lot of n randomly assembled cars have the same last two digits is at least 0.75.