An Introduction to Mathematical Statistics and Its Applications
Richard Larsen and Morris Marx
## Contents

### 1 Introduction
  1.1 Introduction ................................................................. 1

### 2 Probability
  2.1 Introduction ................................................................. 4
  2.2 Sample spaces and The Algebra of sets ................................... 7
  2.3 The Probability Function ..................................................... 18
  2.4 Conditional Probability ....................................................... 27
  2.5 Independence ................................................................. 39
  2.6 Combinatoric ................................................................. 48
  2.7 Combinatorial Probability ..................................................... 63

### 3 Random Variables
  3.1 3.1 Introduction ............................................................. 68
  3.2 Binomial and Hypergeometric Probabilities .................................. 71
  3.3 Discrete Random Variables .................................................... 81
  3.4 Continuous random Variables ................................................. 98
Chapter 1

Introduction

For the historical aspect of Statistic and Probability read pages 1-11.

1.1 Introduction

Why Statistical Techniques are needed?
Want to do some research like:
Research Questions:

• Do stock Market rise and Fall randomly?
• Can External forces such as the phases of the moon affect admission to mental hospital?
• What kind of relationship exists between exposure to radiation and cancer mortality?

Difficult or impossible to perform in the lab.
Can be answered by collecting data, making assumptions about the conditions that generated the
data and then drawing inferences about the assumption.

**Case Study 1.2.3 (4th Ed)**

In folklore, the full moon is often portrayed as something sinister, a kind of evil force possessing the power to control our behavior. Over the centuries, many prominent writers and philosophers have shared this belief. The possibility of lunar phases influencing human affairs is a theory not without supporters among the scientific community. Studies by reputable medical researchers have attempted to link the ”Transylvania effect,” as it has come to be known, with higher suicide rates, pyromania, and even epilepsy.

Note: Pyromania in more extreme circumstances can be an impulse control disorder to deliberately start fires to relieve tension or for gratification or relief. The term pyromania comes from the Greek word (’pyr’, fire).

The relationship between the admission rates to the emergency room of a Virginia mental health
clinic before, during and after the twelve full moons from August 1971 to July 1972.

Table 1.1: Admission Rates (Patients /Day )

<table>
<thead>
<tr>
<th>Year</th>
<th>Month</th>
<th>Before Full Moon</th>
<th>During Full Moon</th>
<th>After Full Moon</th>
</tr>
</thead>
<tbody>
<tr>
<td>1971</td>
<td>Aug.</td>
<td>6.4</td>
<td>5.0</td>
<td>5.8</td>
</tr>
<tr>
<td></td>
<td>Sept</td>
<td>7.1</td>
<td>13.0</td>
<td>9.2</td>
</tr>
<tr>
<td></td>
<td>Oct.</td>
<td>6.5</td>
<td>14.0</td>
<td>7.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1972</td>
<td>Jan.</td>
<td>10.4</td>
<td>9</td>
<td>13.5</td>
</tr>
<tr>
<td></td>
<td>Feb.</td>
<td>11.5</td>
<td>13.0</td>
<td>13.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>July</td>
<td>15.8</td>
<td>20.0</td>
<td>14.5</td>
</tr>
<tr>
<td></td>
<td>Averages</td>
<td>10.9</td>
<td>13.3</td>
<td>11.5</td>
</tr>
</tbody>
</table>

For these data, the average admission rate ”during” the full moon is higher than the ”before” and ”after” admission rates: 13.3 versus 10.9 and 11.5. Does that imply the ”Transylvania” effect is real? Not necessarily. The question that needs to be addressed is whether sample means of 13.3, 10.9 and 11.5 are significantly different or not. After doing a suitable statistical analysis, the conclusion is these three means are not statistically different which conclude that ”Transylvania” effect is not real. How do you make the decision? Based on some probability!

We will learn theory of probability in this class.
Chapter 2

Probability

2.1 Introduction

Read pages 22 through 23.

What is Probability?

Consider tossing a coin once. What will be the outcome? The outcome is uncertain. Head or tail? What is the probability it will land on its head? What is the probability that it will land on its tail?

A probability is a numerical measure of the likelihood of the event (head or tail). It is a number that we attach to an event.

A probability is a number from 0 to 1. If we assign a probability of 0 to an event, this indicates that this event never will occur. A probability of 1 attached to a particular event indicates that this event always will occur. What if we assign a probability of .5?
This means that it is just as likely for the event to occur as for the event to not occur.

### THE PROBABILITY SCALE

<table>
<thead>
<tr>
<th>0</th>
<th>.5</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>event never</td>
<td>event and &quot;not event&quot; always</td>
<td>always will occur</td>
</tr>
<tr>
<td>will occur</td>
<td>event are likely</td>
<td>will occur</td>
</tr>
<tr>
<td></td>
<td>to occur</td>
<td></td>
</tr>
</tbody>
</table>

Three basics methods on assigning probability to an event.

1) *classical approach*. Credited to Gerolamo Cardano. Require that (1) the number of possible outcomes \( m \) is finite and (2) all outcomes are equally likely. The probability of an event consisting \( m \) outcomes is \( m/n \), where \( n \) is the total possible outcomes. (Limited)

2) *empirical approach*. Credited to Richard von Misses. Needed identical experiments be repeated
many times let say n times. Can count the number of times event of interest occurs m. The probability of the event is the limit as n goes to infinity of m/n. In practice how to determine how large n is in order m/n to be good approximation of \( \lim_{n \to \infty} m/n \)

3) subjective - depend on situations.

Back to the coin tossed

1) Classical approach.
   \[ P(\text{head}) = \frac{\text{number of head}}{\text{the number of possible outcome}} = 1/2 \]
   \[ P(\text{tail}) = \frac{\text{number of tail}}{\text{the number of possible outcome}} = 1/2 \]
   This approach is based on assumption that the event head and tail are equally likely to occur.

2) Empirical approach.
   Toss the coin 1000 times. Count how may times it landed on the head or tail.
   \[ P(\text{head}) = \frac{\text{number of head}}{1000} = \frac{\text{number times event happen}}{\text{number of experiment}} \]
   \[ P(\text{tail}) = \frac{\text{number of tail}}{1000} \]

3) subjective
Just guess the probability of head or probability of a tail.

2.2 Sample spaces and The Algebra of sets

Just as statistics build on probability theory, probability in turns build upon set theory.

**Definition of key terms:**

*experiment*: Procedure that can be repeated, theoretically an infinite number of time and has well defined set of possible outcomes.

*sample outcome s*: each potential eventualities of an experiment

*sample space S*: the totalities of sample outcomes

*event*: collection of sample outcomes

**Example 2.2.1**

Consider the experiment tossing a coin three times.

*experiment*: tossing a coin three times

*sample outcome s*: HTH

*sample space*: $S = \{HHH, HTH, HHT, THH, HTT, THT, TTH, TTT\}$
**event:** Let A represent outcomes having 2 head so
\[ A = \{ \text{HTH, HHT, THH} \} \]

**Example 2.2.4**

A coin is tossed until the first tail appears

**sample outcome s:** HT

**sample space:** \( S = \{ T, \text{HT, HHT, HHHT, } \cdots \} \)

Note in example 2.2.4 sample outcomes are infinite.

**Practice:**

2.2.1 A graduating engineer has signed up for three jobs interview. She intends to categorize each one as being either being a ”success” (1) or a ”failure” (0) depending on whether it leads to a plant trip. Write out the appropriate sample space. What outcomes are in the event A:Second success occurs on third interview? In B: First success never occur? Hint: Notice the similarity between this situation and the coin tossing experiment in Example 2.2.1.

**Answer:** \( S = \{ 111, 110, 101, 011, 001, 010, 100, 000 \} \)

\[ A = \{ 101, 011 \} \]

\[ B = \{ 000 \} \]
2.2.2 Three dice are tossed, one red, one blue, and one green. What outcomes make up the event A that the sum of the three faces showing equals five?

Answer: \( A = \{113, 122, 131, 212, 221, 311\} \)

Practice

2.2.3 An urn contains six chips numbered 1 through 6. Three are drawn out. What outcomes are in the event A ”Second smallest chip is a 3”? Assume that the order of the chips is irrelevant.

Answer: \( A = \{134, 135, 136, 234, 235, 236,\} \)

Unions, Intersections and Complements

Operations performed among events defined on the sample space is referred to as algebra of set

Definition 2.2.1. Let \( A \) and \( B \) be any two events defined over the same sample space \( S \). Then

a. The intersection of \( A \) and \( B \), written as \( A \cap B \), is the event whose outcomes belong to both \( A \) and \( B \).

b. The union of \( A \) and \( B \), written as \( A \cup B \), is the event whose outcomes belong to either \( A \) or \( B \) of both.
Example
\[ A = \{1, 2, 3, 4, 5, 6, 7, 8\} \quad B=\{2,4,6,8\} \]
\[ A \cap B = \{2, 4, 6, 8\} \]
\[ A \cup B = \{1, 2, 3, 4, 5, 6, 7, 8\} \]

Example 2.2.7
Let \( A \) be the set of \( x \) for which \( x^2 + 2x = 8 \); let \( B \) be the set for which \( x^2 + x = 6 \). Find \( A \cap B \) and \( A \cup B \).

**Answer:** Since the first equation factor into \((x + 4)(x - 2) = 0\), its solution set is \( A = \{-4, 2\}\). Similarly, the second equation can be written \((x + 3)(x - 2) = 0\), making \( B = \{-3, 2\}\), Therefore
\[ A \cap B = \{2\} \]
\[ A \cup B = \{-4, -3, 2\} \]

Definition 2.2.2.
Events \( A \) and \( B \) defined over the same sample space \( S \) are said to be mutually exclusive if they have no outcomes in common - that is, if \( A \cap B = \emptyset \), where \( \emptyset \) is the null set.

Example
\[ A = \{1, 3, 5, 7\} \quad B=\{2,4,6,8\} \]
\[ A \cap B = \emptyset = \emptyset \]

**Definition 2.2.3.**

Let A be any event defined on a sample space S. The *complement* of A, written \( A^C \), is the event consisting of all the outcomes in S other than those contained in A.

**Example**

\( S = \{ 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 \} \)
\( A = \{ 2, 4, 6, 8, 10 \} \)
\( A^C = \{ \} \)

**Example 2.2.11**

Suppose the events \( A_1, A_2, \ldots, A_k \) are intervals of real numbers such that \( A_i = \{ x : 0 \leq x \leq 1/i \} \). Describe the set \( A_1 \cup A_2 \cup \ldots, \cup A_k = \bigcup_{i=1}^{k} A_i \) and \( A_1 \cap A_2 \cap \ldots, \cap A_k = \bigcap_{i=1}^{k} A_i \).

Notice that \( A_i 's \) are telescoping sets. That is, \( A_1 \) is the interval \( 0 \leq x \leq 1 \), \( A_2 \) is the interval \( 0 \leq x \leq 1/2 \) and so on. It follows, then that the union of the \( k \) \( A_i 's \) is simply \( A_1 \) while the intersection of the \( A_i 's \) (that is their overlap) is \( A_k \).

**Practice:**

Let \( A \) be the set of \( x \) for which \( x^2 + 2x - 8 \leq 0 \); let \( B \) be the set for which \( x^2 + x - 6 \leq 0 \). Find \( A \cap B \)
and $A \cup B$.

**Answer:** The solution set for the first inequality is $[-4, 2]$, then $A = \{x : -4 \leq x \leq 2\}$. Similarly, the second inequality has a solution $[-3, 2]$, making $B = \{x : -3 \leq x \leq 2\}$, Therefore

\[
A \cap B = \{x : -3 \leq x \leq 2\}
\]

\[
A \cup B = \{x : -4 \leq x \leq 2\}
\]

**2.2.22** Suppose that each of the twelve letters in the word \{T E S S E L L A T I O N\} is written on a chip. Define the events F, R, and C as follows:

F: letters in first half of alphabet
R: letters that are repeated
V: letters that are vowels

Which chips make up the following events:

(a) $F \cap R \cap V$
(b) $F^C \cap R \cap V^C$
(c) $F \cap R^C \cap V$

**Answer:**

(a) $F \cap R \cap V = \{E1, E2\}$
(b) \( F^C \cap R \cap V^C = \{ S1, S2, T1, T2 \} \)

(c) \( F \cap R^C \cap V = \{ A, I \} \)

**Homework** Due Feb 20th

**2.2.16** Sketch the regions in the xy-plane corresponding to \( A \cup B \) and \( A \cap B \) if

\[
A = \{ (x, y) : 0 < x < 3, 0 < y < 3 \}
\]
\[
B = \{ (x, y) : 2 < x < 4, 2 < y < 4 \}
\]

**2.2.28.** Let events \( A \) and \( B \) and sample space \( S \) be defined as the following intervals:

\[
S = \{ x : 0 \leq x \leq 10 \}
\]
\[
A = \{ x : 0 < x < 5 \}
\]
\[
B = \{ x : 3 \leq x \leq 7 \}
\]

Characterize the following events:

(a) \( A^C \)

(b) \( A \cap B \)

(c) \( A \cup B \)
2.2.29 A coin is tossed four times and the resulting sequence of Heads and/or Tails is recorded. Defined the events A, B, and C as follows:

A: Exactly two heads appear
B: Heads and tails alternate
C: First two tosses are heads

(a) Which events, if any, are mutually exclusive?
(b) Which events, if any, are subsets of other sets?

Expressing Events Graphically: Venn Diagram

Read Textbook Notes on page 25 through 26

Relationships based on two or more events can sometimes be difficult to express using only equations or verbal descriptions. An alternative approach that can be used highly effective is to repre-
sent the underlying events graphically in a format known as a Venn diagrams.

**Example 2.2.13 (4th Ed)**

When swampwater Tech’s class of ’64 held its fortieth reunion, one hundredth grads attended. fifteen of those alumni were lawyers and rumor had it that thirty of the one hundredth were psychopaths. If ten alumni were both lawyers and psychopath, how many suffered from neither afflictions?

Let L be the set of lawyers and H, the set of psychopaths. If the symbol N(Q) is defined to be the number of members in set Q, then,

\[
\begin{align*}
N(S) &= 100 \\
N(L) &= 15 \\
N(H) &= 30 \\
N(L \cap H) &= 10 \\
\end{align*}
\]

Summarize these information in a venn diagram. Notice that

\[N(L \cup H) = \text{number of alumni suffering from at least one affliction}\]
\[ \begin{align*}
&= 5 + 10 + 20 \\
&= 35
\end{align*} \]

Therefore alumni who were neither lawyers of psychopaths is \(100 - 35 = 65\).

We can also see that \(N(L \cup H) = N(L) + N(H) - N(L \cap H)\)

**Practice**

2.2.31 During orientation week, the latest Spiderman movie was shown twice at State University. Among the entering class of 6000 freshmen, 850 went to see it the first time, 690 the second time, while 4700 failed to see it either time. How many saw it twice?

Answer: \(850 + 690 - 1300 = 240\).

2.2.32. *De Morgans laws* Let A and B be any two events. Use Venn diagrams to show that

(a) the complement of their intersection is the union of their complement.

\[(A \cap B)^C = A^C \cup B^C\]

(b) the complement of their union is the intersec-
\[ (A \cup B)^C = A^C \cap B^C \]

**Homework**

**2.2.36** Use Venn diagrams to suggest an equivalent way of representing the following events:

(a) \((A \cap B^C)^C\)
(b) \(B \cup (A \cap B)^C\)
(c) \(A \cap (A \cap B)^C\)

**2.2.37** A total of twelve hundredth graduates of State Tech have gotten into medical school in the past several years. Of that number, one thousand earned scores of twenty seven or higher on the Medical College Admission Test (MCAT) and four hundred had GPA that were 3.5 or higher. Moreover, three hundred had MCATs that were twenty seven or higher and GPA that were 3.5 or higher. What proportion of those twelve hundred graduates got into medical school with an MCAT lower than twenty seven and a GPA below 3.5?

**2.2.38**
2.2.40 For two events A and B defined on a sample space \( S \), \( N(A \cap B^C) = 15 \), \( N(A^C \cap B) = 50 \), and \( N(A \cap B) = 2 \). Given that \( N(S) = 120 \), how many outcomes belong to neither \( A \) nor \( B \)?

2.3 The Probability Function

The following definition of probability was entirely a product of the twentieth century. Modern mathematicians have shown a keen interest in developing subjects axiomatically. It was to be expected, then, that probability would come under such scrutiny and be defined not as a ratio (classical approach) or as the limit of a ratio (empirical approach) but simply as a \textit{function} that behaved in accordance with a prescribed set of axioms.

Denote \( P(A) \) as a probability of \( A \), where \( P \) is the \textit{probability function}. It is a mapping from a set \( A \) (event) in a sample space \( S \) to a number.

The major breakthrough on this front came in 1993 when Andrey Kolmogorov published Foundation of the Theory of Probability. Kolmogorovs work was a masterpiece of mathematical elegance-
it reduced the behavior of the probability function to a set of just three or four simple postulates, three if the sample space is finite and four if $S$ is infinite.

Three Axioms (Kolmogorov) that are necessary and sufficient for characterizing the probability function $P$:

**Axiom 1** Let $A$ be any event defined over $S$. Then $P(A) \geq 0$

**Axiom 2** $P(S)=1$

**Axiom 3** Let $A$ and $B$ be any two mutually exclusive events defined over $S$. Then $P(A \cup B) = P(A) + P(B)$. (Additivity or finite additivity)

When $S$ has an infinite number of members, a fourth axiom is needed:

**Axiom 4.** Let $A_1, A_2, \ldots$, be events defined over
S. If $A_i \cap A_j = \emptyset$ for each $i \neq j$, then

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

Note that from Axiom 4 follows Axiom 3, but in general the inverse does not hold.

Some basic properties of $P$ that are the consequence form the Kolgomorov Axiom are:

**Theorem 2.3.1.** $P(A^C) = 1 - P(A)$

**Proof** By Axiom 2 and Definition 2.2.3 (complement of an event : $S = A \cup A^C$)

$$P(S) = 1 = P(A \cup A^C),$$

but $A$ and $A^C$ are mutually exclusive, so by axiom 2, $P(A \cup A^C) = P(A) + P(A^C)$ and the result follows.

**Theorem 2.3.2.** $P(\emptyset) = 0$

**Proof** Since $\emptyset = S^C$, $P(S^C) = 1 - P(S) = 0$
**Theorem 2.3.3.** If $A \subset B$, then $P(A) \leq P(B)$

**Proof** Note that event $B$ may be written in the form

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

where $A$ and $(B \cap A^C)$ are mutually exclusive. Therefore $P(B) = P(A) + P(B \cap A^C)$ which implies that $P(B) \geq P(A)$ since $P(B \cap A^C) > 0$

---

**Theorem 2.3.4.** For any event $A$, $P(A) \leq 1$.

**Proof** The proof follows immediately from Theorem 2.3.3 because $A \subset S$ and $P(S) = 1$

---

**Theorem 2.3.5.** Let $A_1, A_2, \cdots, A_n$ be defined over $S$. If $A_i \cap A_j = \emptyset$ for $i \neq j$ then

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

**Proof** The proof is a straight forward induction argument with axiom 3 being the starting point.
Theorem 2.3.6. \( P(A \cup B) = P(A) + P(B) - P(A \cap B) \)

**Proof** The Venn diagram for \( A \cup B \) suggests that the statement of the theorem is true. Formally, we have from Axiom 3 that

\[
P(A) = P(A \cap B^C) + P(A \cap B)
\]

and

\[
P(B) = P(B \cap A^C) + P(A \cap B)
\]

adding these two equations gives

\[
P(A) + P(B) = [P(A \cap B^C) + P(B \cap A^C) + P(A \cap B)] + P(A \cap B).
\]

By Theorem 2.3.5, the sum in the brackets is \( P(A \cup B) \). If we subtract \( P(A \cap B) \) from both sides of the equations, the result follows.

**Example 2.3.1**

Let \( A \) and \( B \) be two events defined on the sample space \( S \) such that \( P(A) = 0.3, P(B) = 0.5 \) and \( P(A \cup B) = 0.7 \). Find (a) \( P(A \cap B) \), (b) \( P(A^C \cup B^C) \), and (c) \( P(A^C \cap B) \).
(a) From Theorem 2.3.6 we have

\[ P(A \cap B) = P(A) + P(B) - P(A \cup B) \]
\[ = 0.3 + 0.5 - 0.7 = 0.1 \]

(b) From De Morgan’s laws \( A^C \cup B^C = (A \cap B)^C \), so \( P(A^C \cup B^C) = P(A \cap B)^C = 1 - (A \cap B) = 1 - 0.1 = 0.9 \)

(c) The event \( A^C \cap B \) can be shown by venn diagram. From the diagram

\[ P(A^C \cap B) = P(B) - (A \cap B) = 0.5 - 0.1 = 0.4 \]

**Example 2.3.2**

Show that \( P(A \cap B) \geq 1 - P(A^C) - P(B^C) \) for any two events \( A \) and \( B \) defined on a sample space \( S \).

From Example 2.3.1 (a) and Theorem 2.3.1

\[ P(A \cap B) = P(A) + P(B) - P(A \cup B) \]
\[ = 1 - P(A^C) + 1 - P(B^C) - P(A \cup B). \]
But $P(A \cup B) \leq 1$ from Theorem 2.3.4, so $P(A \cap B) \geq 1 - P(A^C) - P(B^C)$

**Read Example 2.3.4**

**Example 2.3.5** Having endured (and survived) the mental trauma that comes from taking two years of Chemistry, a year of Physics and a year of Biology, Biff decides to test the medical school waters and sent his MCATs to two colleges, $X$ and $Y$. Based on how his friends have fared, he estimates that his probability of being accepted at $X$ is 0.7, and at $Y$ is 0.4. He also suspects there is a 75% chance that at least one of his application will be rejected. What is the probability that he gets at least one acceptance?

Answer: Let $A$ be the event ”school $X$ accept him” and $B$ the event ”school $Y$ accept him”. We are given that $P(A) = 0.7$, $P(B) = 0.4$, $P(A^C \cup B^C) = 0.75$. the question is $P(A \cup B)$. From Theorem 2.3.6,

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$
From de Morgan’s law
\[ A^C \cup B^C = (A \cap B)^C \]
so \[ P(A \cap B) = 1 - P(A \cap B)^C = 1 - 0.75 = 0.25 \]

It follows that Biff’s prospect is not that bleak - he has an 85% chance of getting in somewhere:
\[ P(A \cup B) = P(A) + P(B) - P(A \cap B) \]
\[ = 0.7 + 0.4 - 0.25 = 0.85 \]

Practice

2.3.2 Let \( A \) and \( B \) be two events defined on \( S \). Suppose that \( P(A) = 0.4, P(B) = 0.5 \), and \( P(A \cap B) = 0.1 \), What is the probability that \( A \) or \( B \) but not both occur?

Answer:
\[ P(A \text{ or } B \text{ but not } A \text{ and } B) = P(A \cup B) - P(A \cap B) \]
\[ = P(A) + P(B) - P(A \cap B) - P(A \cap B) = 0.7 \]

2.3.4 Let \( A \) and \( B \) be two events defined on \( S \). If the probability that at least one of them occurs is 0.3 and the probability that \( A \) occurs but \( B \) does not occur is 0.1, what is \( P(B) \)?
Given \( P(A \cup B) = 0.3 \)
\[
P(A \cap B^c) = P(A) - P(A \cap B) = 0.1
\]
What is \( P(B) \)

Since \( P(A \cup B) = P(A) + P(B) - P(A \cap B) \)
Then \( P(B) = 0.3 - 0.1 = 0.2 \)

**2.3.5** Suppose that three fair dice are tossed. Let \( A_i \) be the event that a 6 shows on the \( i \)th die, \( i = 1, 2, 3 \) Does \( P(A_1 \cup A_2 \cup A_3) = \frac{1}{2} \)? Explain.
Answer: No. \( P(A_1 \cup A_2 \cup A_3) = P(\text{at least one 6th appears}) = 1 - P(\text{no 6th appear}) = 1 - \left(\frac{5}{6}\right)^3 \).

**Homework** Due Feb 20th

2.3.6
2.3.10
2.3.11
2.3.12
2.3.14
2.3.16
2.4 Conditional Probability

In the previous section we were given two separate probability of events $A$ and $B$. Knowing $P(A \cap B)$ we can find the probability of $A \cup B$. Sometimes knowing that certain event $A$ has happened can change the probability of $B$ happen compare with two individual probability of events $A$ and $B$. This is called conditional probability.

The symbol $P(A|B)$ read the probability of $A$ given $B$ is used to denote a conditional probability. Specifically, $P(A|B)$ refers to the probability that $A$ will occur given that $B$ has already occurred.

\textbf{Definition 2.4.1}

Let $A$ and $B$ any two events defined on $S$ such that $P(B) > 0$. The conditional probability of $A$, assuming that $B$ has already occurred, is written $P(A|B)$ and, is given by

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$
Comment: From definition 2.4.1,

\[ P(A \cap B) = P(A|B)P(B) \]

**Example 2.4.3**

Two events \( A \) and \( B \) are defined such that (1) the probability that \( A \) occurs but \( B \) does not occur is 0.2, (2) the probability that \( B \) occurs but \( A \) does not occur is 0.1, and (3) the probability that neither occurs is 0.6. What is \( P(A|B) \)?

Answer: Given (1) \( P(A \cap B^C) = 0.2 \). (2) \( P(B \cap A^C) = 0.1 \). (3) \( P(A \cup B)^C = 0.6 \). We want \( P(A|B) = \frac{P(A \cap B)}{P(B)} \). Therefore we need \( P(B) \) and \( P(A \cap B) \). From (3) we have \( P(A \cup B) = 0.4 \). Draw the venn diagram. From the venn diagram,

\[
\begin{align*}
P(A \cup B) &= P(B \cap A^C) + P(A \cap B) + P(B \cap A^C) \\
P(A \cap B) &= 0.4 - 0.2 - 0.1 = 0.1 \\
P(B) &= P(A \cap B) + P(B \cap A^C) = 0.1 + 0.1 = 0.2 \\
P(A|B) &= 0.1/0.2 = 0.5
\end{align*}
\]

**Practice**
2.4.1 Suppose that two fair dice are tossed. What is probability that the sum equals 10 given that it exceeds 8?

Answer: Let $A$: event sum of the two faces is 10 $A = \{(5, 5), (4, 6), (6, 4)\}$. Let $B$: event sum of the two faces exceed 8. $B = \{sum = 9, 10, 11, 12\}$ $B = \{(4, 5), (5, 4), (3, 6), (6, 3), (5, 5), (4, 6), (6, 4), (5, 6), (6, 5)\}$

Note that the number of elements in the sample space is 36 $Q$: $P(A|B) = P(A \cap B)/P(B)$?

2.4.2 Find $P(A \cap B)$ if $P(A) = 0.2$, $P(B) = 0.4$, and $P(A|B) + P(B|A) = 0.75$.

Answer:

$0.75 = P(A|B) + P(B|A) = P(A \cap B)/P(B) + P(A \cap B)/P(A)$

$\rightarrow P(A \cap B) = 0.1$

Homework Due 22nd February.

2.4.6, 2.4.7, 2.4.10, 2.4.11, 2.4.12, 2.4.16

Applying Conditional Probability to Higher-Order Intersection.

What is the formula for $P(A \cap B \cap C)$? If we let
\( A \cap B \) as \( D \) then
\[
P(A \cap B \cap C) = P(D \cap C) \\
= P(C|D)P(D) \\
= P(C|A \cap B)P(A \cap B) \\
= P(C|A \cap B)P(B|A)P(A)
\]
Repeating this same argument for \( n \) events, \( A_1, A_2, \cdots, A_n \) give a general case for:
\[
P(A_1 \cap A_2 \cap \cdots \cap A_n) \\
= P(A_n|A_1 \cap A_2 \cap \cdots \cap A_{n-1}) \\
\cdot P(A_{n-1}|A_1 \cap A_2 \cap \cdots \cap A_{n-2}) \\
\cdots P(A_2|A_1)P(A_1)
\] (2.4.1)

**Example 2.4.7**
A box contains 5 white chips, 4 black, and 3 red chips. Four chips are drawn sequentially and without replacement. What is the probability of obtaining the sequence (W,R,W,B) ? Define following four events:

A: white chip is drawn on 1st selection;
B: red chip is drawn on 2nd selection;
C: white chip is drawn on 3rd selection;  
D: black chip is drawn on 4th selection

Our objective is to find \( P(A \cap B \cap C \cap D) \), So \( P(A \cap B \cap C \cap D) = P(D|A \cap B \cap C)P(C|A \cap B)P(B|A)P(A) \). From the diagram, \( P(D|A \cap B \cap C) = \frac{4}{9}, P(C|A \cap B) = \frac{4}{10}, P(B|A) = \frac{3}{11}, P(A) = \frac{5}{12} \)

Therefore \[
P(A \cap B \cap C \cap D) = \frac{4}{9} \cdot \frac{4}{10} \cdot \frac{3}{11} \cdot \frac{5}{12} = \frac{240}{11880} = 0.02
\]

**Homework** Due 22 Feb.

2.4.21, 2.4.24

**Calculating ”Unconditional” Probability**

**Also called Total Probability**

Let’s partition \( S \) into mutually exclusive partitions namely: \( A_1, A_2, \cdots, A_n \), where the union is
Let $B$ denote an event defined on $S$. See venn diagram in Figure 2.4.7 in the text. the next theorem gives formula for the "unconditional" probability of $B$.

**Theorem 2.4.1 (Total Probability Theorem)** Let $\{A_i\}_{i=1}^n$ be a set of events defined over $S$ such that $S = \bigcup_{i=1}^n A_i$, $A_i \cap A_j = \emptyset$ for $i \neq j$, and $P(A_i) > 0$ for $1 = 1, 2, \cdots, n$. For any event $B$

$$P(B) = \sum_{i=1}^n P(B|A_i)P(A_i)$$

**Proof.** By the conditioned imposed on the $A_i$'s

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

and

$$P(B) = P(B \cap A_1) + P(B \cap A_2) + \cdots + P(B \cap A_n).$$

But each $P(B \cap A_i)$ can be written as the product of $P(B|A_i)P(A_i)$ and the result follows.

**Example 2.4.8**
Box I contains two red chips and four white chips; box II, three red and one white. A chip is drawn at random from box I and transferred to box II. Then a chip is drawn from box II. What is the probability that the chip drawn from box II is red?

Let $B$ the event Chip drawn from urn II is red; let $A_1$ and $A_2$ be the events Chip transferred from urn I is red and Chip transferred from urn I is white, respectively. Then $P(B|A_1) = \frac{4}{5}$, $P(B|A_2) = \frac{3}{5}$, $P(A_1) = \frac{2}{6}$, $P(A_2) = \frac{4}{6}$. Then $P(B) = P(B|A_2)P(A_2) + P(B|A_1)P(A_1) = \frac{4}{5} \cdot \frac{2}{6} + \frac{3}{5} \cdot \frac{4}{6} = \frac{2}{3}$

**Example 2.4.10**

Ashley is hoping to land a summer internship with a public relation firm. If her interview goes well, she has a 70% chance of getting a offer. If the interview is a bust, though, her chance of getting the position drop to 20%. Unfortunately, Ash ey tends to babble incoherently when she is under stress, so the likelihood of the interview going well is only 0.10. What is the probability that Ashley gets the
Let $B$ be the event ”Ashley is offered internship,” let $A_1$ be the event ”interview goes well” and $A_2$ be the event ”interview does not go well”. By the assumption,

\begin{align*}
P(B|A_1) &= 0.70 \quad P(B|A_2) = 0.20 \\
P(A_1) &= 0.10 \quad P(A_2) = 1 - P(A_1) = 0.90
\end{align*}

From the Total Probability Theorem,

\begin{align*}
P(B) &= P(B|A_2)P(A_2) + P(B|A_1)P(A_1) \\
&= (0.70)(0.10) + (0.2)(0.90) = 0.25
\end{align*}

**Practice**

2.4.25

A toy manufacturer buys ball bearings from three different suppliers - 50 % of her total order comes from supplier 1, 30 % from supplier 2, and the rest from supplier 3. Past experience has shown that the quality control standards of the three suppliers are not all the same. Two percent of the ball bearings produced by supplier 1 are defective, while suppliers 2 and 3 produce defective 3 % and 4 % of the time,
respectively. What proportion of the ball bearings in the toy manufacturer’s inventory are defective?

Let $B$ be the event that ball bearings are defective. $A_1$ be the event ball bearing from manufacturer 1, $A_2$ be the event ball bearing from manufacturer 2, $A_3$ be the event ball bearing from manufacturer 3. From the information $P(B|A_1) = 0.02$, $P(B|A_2) = 0.03$, $P(B|A_3) = 0.04$. $P(A_1) = 0.5$, $P(A_2) = 0.3$, $P(A_3) = 0.2$,

$$P(B) = P(B|A_3)P(A_3) + P(B|A_2)P(A_2) + P(B|A_1)P(A_1)$$

$$= (0.04)(0.2) + (0.03)(0.3) + (0.02)(0.5) = 0.027$$

**Homework** Due 22 Feb

2.4.26, 2.4.28, 2.4.30

**Bayes theorem**

If we know $P(B|A_i)$ for all $i$, the theorem enables us to compute conditional probabilities in the other direction that is we can use the $P(B|A_i)$s to find $P(A_i|B)$. It is like a certain kind of ”inverse” probability.

**Theorem 2.4.2** (Bayes) Let $\{A_i\}_{i=1}^n$ be a set
of \( n \) events, each with positive probability, that partition \( S \) in such a way that \( \bigcup_{i=1}^{n} A_i = S \) and \( A_i \cap A_j = \emptyset \) for \( i \neq j \). For any event \( B \) (also defined on \( S \), where \( P(B) > 0 \),

\[
P(A_j | B) = \frac{P(B | A_j)P(A_j)}{\sum_{i=1}^{n} P(B | A_i)P(A_i)}
\]

for any \( 1 \leq j \leq n \)

**Proof.**

From definition 2.4.1,

\[
P(A_j | B) = \frac{P(A_j \cap B)}{P(B)} = \frac{P(B | A_j)P(A_j)}{P(B)}.
\]

But Theorem 2.4.1 allows the denominator to be written as \( \sum_{i=1}^{n} P(B | A_i)P(A_i) \) and the result follows.

**PROBLEM SOLVING HINT (Working with partitioned sample Spaces)**

Learning to identify which part of the ”given” corresponds to \( B \) and which part correspond to the \( A_i \)’s. The following hints may help.

1) Pay attention to the last one or two sentences. Is
the question asking for the unconditional probability or conditional probability

2) If unconditional probability denote \( B \) as the event whose probability we are trying to find. If Conditional probability denote \( B \) as the event that already happened.

3) Ones \( B \) has been identified, reread the beginning of the question and assign the \( A_i \)'s.

**Example 2.4.13**

A biased coin, twice as likely to come up heads as tails, is tossed once. If it shows heads, a chip is drawn from box I, which contains three white and four red chips; if it shows tails, a chip is drawn from box II, which contains six white and three red chips. Given that a white chip was drawn, what is the probability that the coin came up tails? (Figure 2.4.10 shows the situation).

Since \( P(H) = 2P(T) \), \( P(H) = 2/3 \) and \( P(T) = 1/3 \). Define the events \( B \): white chip is drawn, \( A_1 \): coin come up heads (i.e., chip came from box I)
A₂: coin came up tails (i.e., chip came from box II)

We are trying to find \( P(A₂|B) \), where

\[
P(B|A₁) = \frac{3}{7}, \quad P(B|A₂) = \frac{6}{9}, \quad P(A₁) = \frac{2}{3}, \quad P(A₂) = \frac{1}{3}
\]

so

\[
P(A₂|B) = \frac{P(B|A₂)P(A₂)}{P(B|A₁)P(A₁) + P(B|A₂)P(A₂)}
\]

\[
= \frac{(6/9)(1/3)}{(3/7)(2/3) + (6/9)(1/3)} = \frac{7}{16}
\]

**Practice**

2.4.40 Box I contains two white chips and one red chip; box II has one white chip and two red chips. One chip is draw at random from box I and transferred to box II. Then one chip is drawn from box II. Suppose that a red chip is selected from box II. What is the probability that the chip transferred was white?

\( A_R \): transferred red is chip

\( A_W \): transferred chip is white;

Let B denote the event that the chip drawn from
box II is red; Then

$$P(A_W|B) = \frac{P(B|A_W)P(A_W)}{P(B|A_W)P(A_W) + P(B|A_R)P(A_R)} \left(\frac{2/4}{2/4(2/3) + (3/4)(1/3)}\right) = 4/7$$

**Homework** Due Monday Feb 28

2.4.44, 2.4.49, 2.4.52

2.5 Independence

In section 2.4 we introduced conditional probability. It often is the case, though, that the probability of the given event remains unchanged, regardless of the outcome of the second event—that is, $P(A|B) = P(A)$

**Definition 2.5.1.** Two events $A$ and $B$ are said to be independent if $P(A \cap B) = P(A)P(B)$

Comment: When two events are independent, then
\[ P(A \cap B) = P(A)P(B). \] We have

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

Therefore, when event A and event B are independent \( P(A|B) = P(A) \)

**Example 2.5.2.** Suppose that A and B are independent, Does it follow that \( A^C \) and \( B^C \) are also independent?

Answer: Yes!!

We need to show that \( P(A^C \cap B^C) = P(A^C)P(B^C) \)

We start with

\[ P(A^C \cup B^C) = P(A^C) + P(B^C) - P(A^C \cap B^C) \]

But

\[ P(A^C \cup B^C) = P(A \cap B)^C = 1 - P(A \cap B). \]

Therefore

\[ 1 - P(A \cap B) = 1 - P(A) + 1 - P(B) - P(A^C \cap B^C) \]

Since A and B are independent,
\[ P(A \cap B) = P(A)P(B). \text{ So} \]

\[
P(A^C \cap B^C)
\]

\[
= P(A^C) + P(B^C) - P(A^C \cup B^C)
\]

\[
= P(A^C) + P(B^C) - P[(A \cap B)^C]
\]

\[
= 1 - P(A) + 1 - P(B) - [1 - P(A)P(B)]
\]

\[
= [1 - P(A)][1 - P(B)] = P(A^C)P(B^C)
\]

**Example 2.5.4**

Suppose that two events, A and B each having nonzero probability are mutually exclusive. Are they also independent?

No. If A and B are mutually exclusive then \( P(A \cap B) = 0 \), But \( P(A) \cdot P(B) > 0 \) (By assumption)

**Deducing Independence.**

Sometimes the physical circumstances surrounding two events makes it obvious that the occurrence (or nonoccurrence) of one has absolutely no influence or effect on the occurrence (or nonoccurrence) of the other. If it should be the case, then the two events will necessarily be independent in the
sense of definition 2.5.1. Example is tossing a coin twice. Clearly what happen in the first toss will not influence what happen at the second toss. So
\[ P(HH) = P(H \cap H) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4} \]

**Practice 2.5.1**
Suppose \( P(A \cap B) = 0.2 \), \( P(A) = 0.6 \), and \( P(B) = 0.5 \)
a. Are A and B mutually exclusive?
b. Are A and B independent?
c. Find \( P(A^C \cup B^C) \).
   a) No, because \( P(A \cap B) > 0 \).
   b) No, because \( P(A \cap B) = 0.2 \) while \( P(A)P(B) = (0.6)(0.5) = 0.3 \)
   c) \( P(A^C \cup B^C) = P((A \cap B)^C) = 1 - P(A \cap B) = 1 - 0.2 = 0.8 \).

**Practice 2.5.2.** Spike is not a terribly bright student. His chances of passing chemistry are 0.35, mathematics, 0.40, and both 0.12. Are the event ”Spike passes chemistry” and ”Spike passes mathematics” independent? What is the probability that
he fails both.

Answer: not independent, 0.37

**Homework.** Due Monday Feb 28

2.5.4, 2.5.7, 2.5.9

**Defining the Independence of More Than Two Events**

It is not immediately obvious how to extend definition of independence, say, three events. To call A, B, and C independent, should we require

\[
P(A \cap B \cap C) = P(A) \cdot P(B) \cdot P(C) \quad (2.5.1)
\]

or should we impose the definition we already have on the three pairs of events (pairwise independent)

\[
P(A \cap B) = P(A) \cdot P(B) \quad (2.5.2)
\]

\[
P(B \cap C) = P(B) \cdot P(C)
\]

\[
P(A \cap C) = P(A) \cdot P(C)
\]

Neither condition by itself is sufficient.

If the three events satisfy 2.5.1 and 2.5.2, we call them independent (or mutually independent) Generally 2.5.1 does not imply (2.5.2), nor does 2.5.2 imply 2.5.1
More generally, the independence of $n$ events requires that the probabilities of all possible intersections equal the products of all the corresponding individual probabilities.

**Definition 2.5.2** Events $A_1, A_2, \cdots, A_n$ are said to be independent if all $k, k = 1 \cdots n$,

$$P(\bigcap_{i=1}^{k} A_i) = P(A_1)P(A_2)\cdots P(A_k)$$

**Example**

2.5.11 Suppose that two fair dice (one red and one green) are thrown, with events $A, B,$ and $C$ defined

- $A$: a 1 or a 2 shows on the red die
- $B$: a 3, 4, or 5 shows on the green die
- $C$: the dice total is 4, 11, or 12. Show the these events satisfy Equation 2.5.1 but not 2.5.2. By list-
ing the sample outcomes, it can be shown that

\[ P(A) = 1/3, \ P(B) = 1/2, \ P(C) = 1/6 \]

\[ P(A \cap B) = 1/6, \ P(A \cap C) = 1/18; \]

\[ P(B \cap C) = 1/18 \]

and \( P(A \cap B \cap C) = 1/36 \) Note that equation 2.5.1 is satisfied,

\[ P(A \cap B \cap C') = 1/36 = P(A) \cdot P(B) \cdot P(C') \]

\[ = 1/3 \cdot 1/2 \cdot 1/6. \]

But Equation 2.5.2 is not satisfied since \( P(B \cap C') = 1/18 \neq P(B) \cdot P(C'). \)

**Practice** 1) Suppose that a fair coin is flipped three times. Let \( A_1 \) be the event of a head on the first flip; \( A_2 \), a tail on the second flip; and \( A_3 \), a head on the third flip. Are \( A_1, A_2, \) and \( A_3 \) independent?

2) Suppose that two events A and B, each having nonzero probability, are mutually exclusive. Are they also independent?

3) Suppose that \( P(A \cap B) = 0.2, \ P(A) = 0.6, \) and \( P(B) = 0.5 \) a) Are A and B mutually exclusive? b)
Are A and B independent? C) Find $P(A^C \cup B^C)$

**Repeated Independent Trials**

It is not uncommon for an experiment to be the composite of a finite or countably infinite number of subexperiments, each of the latter being performed under essentially the same conditions. (tossing a coin three times) In general, the subexperiments comprising an experiment are referred to as trials. We will restrict our attention here to problems where the trials are independent- that is, for all $j$, the probability of any given outcome occurring on the $j$th trial is unaffected by what happened on the preceding $j-1$ trials.

**Example 2.5.10** Suppose a string decoration light you just bought has twenty-four bulbs wired in series. If each bulb has a 99.9% chance of ”working” the first time current is applied. What is the probability that the string itself will not work? (Note that if one or more bulb fails the string will not work)
Let $A_i$ be the event that $i$th bulb fails, $i = 1, 2, \cdots, 24$. Then

\[
P(\text{String fails}) = P(\text{at least one bulb fails}) = P(A_1 \cup A_2 \cup \cdots \cup A_{24})
\]

\[
= 1 - P(\text{String works}) = 1 - P(\text{all twenty four work})
\]

\[
= 1 - P(A_1^C \cap A_2^C \cap \cdots \cap A_{24}^C)
\]

If we assume that bulbs are presumably manufactured the same way, $P(A_i^C)$ is the same for all $i$, so

\[
P(\text{String fails}) = 1 - \{P(A_i^C)\}^{24}
\]

\[
= 1 - \{0.99\}^{24}
\]

\[
= 1 - 0.98 = 0.02
\]

Therefore the chances are one in fifty, in other words, that the string will not work the first time current is applied.

**Practice 2.5.25**

If two fair dice are tossed, what is the smallest
number of throws, \( n \), for which the probability of getting at least one double 6 exceed 0.5? (Note that this was one of the first problems that de Mere communicated to Pascal in 1654)

\[
P( \text{ at least one double six in } n \text{ throws } ) = 1 - P( \text{ no double sixes in } n \text{ throws } )
= 1 - \left( \frac{35}{36} \right)^n.
\]

By trial and error, the smallest \( n \) for which
\[
P( \text{ at least one double six in } n \text{ throws } ) \text{ exceeds } 0.50
\]

is \( n=25 \)

**Homework** 2.5.23, 2.5.26, 2.5.27 Due Monday Feb 28

2.6 Combinatoric

Recall \( P(A) = \frac{\# \text{ element in the event } A}{\# \text{ element in the sample space } S} \)

How to count the number of element?

**Counting ordered sequences (The multiplication Rule)**

The multiplication Rule *If operation \( A \) can
be performed in m different ways and operation B in n different ways, the sequence (operation A, operation B) can be performed in m.n different ways.

Proof. Use tree diagram

**Corollary 2.6.1** If operation $A_i$, $i = 1, 2, \ldots, k$ can be performed in $n_i$ ways, $i = 1, 2, \ldots, k$, respectively, then the ordered sequence (operation $A_1$, operation $A_2$, $\cdots$, operation $A_k$) can be performed in $n_1, n_2, \cdots, n_k$ ways.

**Example:** How many different ways can parents have three children.

**Answer:** For each child we will assume there are only two possible outcomes (thus neglecting effects of extra X or Y chromosomes, or any other chromosomal/birth defects). The number of ways can be calculated: $2 \cdot 2 \cdot 2 = 8$. These can be listed: BBB, BBG, BGB, BGG, GBB, GBG, GGB, GGG where B=boy, G=girl.

**Example** (from Question 2.6.9) A restaurant
offers a choice of four appetizers, fourteen entrees, six desserts, and five beverages. How many different meals are possible if a diner intends to order only three courses? (Consider the beverage to be a "course.")

$$4 \cdot 14 \cdot 6 + 4 \cdot 6 \cdot 5 + 14 \cdot 6 \cdot 5 + 4 \cdot 14 \cdot 5 = 1156$$

**Counting Permutations (when the objects are all distinct)**

Ordered sequences arise in two fundamentally different ways. The first is the scenario addressed by the multiplication rule - a process is comprised of \( k \) operations, each allowing \( n_i \) options, \( i = 1, 2, \cdots, k \); choosing one version of each operation leads to \( n_1, n_2, \cdots, n_k \) possibilities.

The second occurs when an ordered arrangement of some specified length \( k \) is formed from a finite collection of objects. Any such arrangement is referred to as a *permutation of length* \( k \). For example, given the three objects \( A, B, \) and \( C \), there are six different permutations of length two that can be formed if the objects cannot be repeated: \( AB, AC, BC, BA, CA, \) and \( CB \).
**Theorem 2.6.1** The number of permutations of length $k$ that can be formed from a set of $n$ distinct elements, repetitions not allowed, is denoted by the symbol $nP_k$, where

$$nP_k = n(n-1)(n-2) \cdots (n-k+1) = n!/(n-k)!$$

**Proof** Any of the $n$ objects may occupy the first position in the arrangement, any of $n-1$ the second, and so on - the number of choices available for filling the $k$th position will be $n-k+1$ (see Figure 2.6.6). The theorem follows, then, from the multiplication rule: There will be $n(n-1)...(n-k+1)$ ordered arrangements.

Choices: $\begin{array}{c} n \quad n-1 \quad \cdots \quad n-(k-2) \quad n-(k-1) \\ 1 \quad 2 \quad \cdots \quad k-1 \quad k \end{array}$ Position in sequence

**Figure 2.6.6**

**Corollary 2.6.2** The number of ways to permute an entire set of $n$ distinct objects is $nP_n = n(n-1)(n-2)...1 = n!$

**Example 2.6.7** How many permutations of length $k = 3$ can be formed from the set of $n = 4$ distinct
elements, $A, B, C, \text{ and } D$?

According to Theorem 2.6.1, the number should be $24$:

$$\frac{n!}{(n-k)!} = \frac{4!}{(4-3)!} = \frac{4\cdot3\cdot2\cdot1}{1} = 24$$

Confirming that figure, Table 2.6.2 lists the entire set of $24$ permutations and illustrates the argument used in the proof of the theorem.

**Practice**

2.6.17. The board of a large corporation has six members willing to be nominated for office. How many different "president/vice president/treasurer" slates could be submitted to the stockholders?

Answer: $6 \cdot 5 \cdot 4 = 120$

2.6.18. How many ways can a set of four tires be put on a car if all the tires are interchangeable? How many ways are possible if two of the four are snow tires?

$4P_4 \text{ and } 2P_2 \cdot 2P_2$

**Homework: 2.6.22**
Counting Permutations (when the objects are not all distinct)

The corollary to Theorem 2.6.1 gives a formula for the number of ways an entire set of n objects can be permuted if the objects are all distinct. Fewer than $n!$ permutations are possible, though, if some of the objects are identical. For example, there are $3! = 6$ ways to permute the three distinct objects $A$, $B$, and $C$:

\begin{align*}
ABC \\
ACB \\
BAC \\
BCA \\
CAB \\
CBA \\
\end{align*}

If the three objects to permute, are $A$, $A$, and $B$ - that is, if two of the three are identical - the number of permutations decreases to three:

\begin{align*}
AAB \\
ABA \\
BAA \\
\end{align*}

Illustration 2 Suppose you want to order a
group of n objects where some of the objects are the same.

Think about the letters in the word EAR. How many different ways can we arrange the letters to form different three letter words? Easy, right. We have three letters we can write first, we have two letters next, and then the last letter. \( 3 \times 2 \times 1 = 6 \) different three letter words. EAR, ERA, ARE, AER, REA, RAE.

Now think about the letters in the word EYE. How many different ways can we arrange the letters to form different three letter words? Easy, right. Just like before. We have three letters we can write first, we have two letters next, and then the last letter. \( 3 \times 2 \times 1 = 6 \) different three letter words. EYE, EYE, YEE, YEE, EEY, EEY. But wait a second. Three are the same as another three. Actually there are only three distinguishable ways for the word EYE.

The number of distinguishable permutations of n objects where \( n_1 \) are one type, \( n_2 \) are of another type, and so on is \( \frac{n!}{n_1!n_2!n_3!} \).
Theorem 2.6.2 The number of ways to arrange \( n \) objects, \( n_1 \) being of one kind, \( n_2 \) of a second kind, \ldots, and \( n_r \) of an \( r \)th kind, is \[
\frac{n!}{n_1! \cdot n_2! \cdots n_r!}
\] where \( \sum r n_i = n \).

Comment Ratios like \( \frac{n!}{(n_1! \ n_2! \cdots n_r!)} \) are all called multinomial coefficients because the general term in the expansion of \((x_1 + x_2 + \ldots + x_r)^n\) is \[
\frac{n!}{n_1! \ n_2! \cdots n_r! x_1^{n_1} \ x_2^{n_2} \cdots x_r^{n_r}}
\]

Example 2.6.14 A pastry in a vending machine costs 85 cents. In how many ways can a customer put in two quarters, three dimes, and one nickel?

<table>
<thead>
<tr>
<th>Quarter</th>
<th>Dime</th>
<th>Dime</th>
<th>Quarter</th>
<th>Nickel</th>
<th>Dime</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

Order in which coins are deposited

Figure 2.6.13

If all coins of a given value are considered identical, then a typical deposit sequence, say, QDDQND (see Figure 2.6.13), can be thought of as a permutation of \( n = 6 \) objects belonging to \( r = 3 \) categories,
where
\[ n_1 = \text{number of nickels} = 1 \]
\[ n_2 = \text{number of dimes} = 3 \]
\[ n_3 = \text{number of quarters} = 2 \]

By Theorem 2.6.2, there are sixty such sequences:
\[ \frac{n!}{n_1! \cdot n_2! \cdot n_3!} = \frac{6!}{1! \cdot 3! \cdot 2!} = 60 \]

Of course, had we assumed the coins were distinct (having been minted at different places and different times), the number of distinct permutations would have been 6!, or 720.

**Example 2.6.15**

What is the coefficient of \(x^{23}\) in the expression of \((1 + x^5 + x^9)^{100}\)?

First consider \((a + b)^2 = (a + b)(a + b) = a^2 + 2ab + b^2\)

The coefficient \(ab\) is 2 come from two different multiplication of \(ab\) and \(ba\). Similarly for the coefficient of \(x^{23}\) in the expansion of \((1 + x^5 + x^9)^{100}\)

will the number of ways that one term from each of the one hundredth factors \((1 + x^5 + x^9)\) can be
multiplied together to form $x^{23}$.

$$x^{23} = x^9 \cdot x^9 \cdot x^5 \cdot 1 \cdot 1 \cdots 1$$

It follows that the coefficient $x^{23}$ is the number of to permute two $x^9$’s one $x^5$ and ninety seven 1’s. So the

$$\text{coefficient of} x^{23} = \frac{100!}{2!1!97!} = 485100$$

**Practice**

2.6.34 Which state name can generate more permutations, TENNESSEE or FLORIDA?

**Homework**

2.6.36, 2.6.40

**Counting Combination**

We call a collection of $k$ *unordered* elements a *combination of size $k$*. For example, given a set of $n = 4$ distinct elements - $A$, $B$, $C$, and $D$ - there are *six* ways to form combinations of size 2:

- $A$ and $B$
- $B$ and $C$
A and C  B and D  
A and D  C and D

A general formula for counting combinations can be derived quite easily from what we already know about counting permutations.

**Theorem 2.6.3.** The number of ways to form combinations of size \( k \) from a set of \( n \) distinct objects, repetitions not allowed, is denoted by the symbols \( \binom{n}{k} \) or \( nC_k \), where

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

**Proof.** Let the symbol \( \binom{n}{k} \) denote the number of combinations satisfying the conditions of the theorem. Since each of those combinations can be ordered in \( k! \) ways, the product \( k! \binom{n}{k} \) must equal the number of permutations of length \( k \) that can be formed from \( n \) distinct elements. But \( n \) distinct elements can be formed into permutations of length \( k \) in \( n(n-1) \cdots (n-k+1) = \frac{n!}{(n-k)!} \) ways. Therefore,

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

Solving for \( \binom{n}{k} \) gives the result.
Comment. It often helps to think of combinations in the context of drawing objects out of an urn. If an urn contains \( n \) chips labeled 1 through \( n \), the number of ways we can reach in and draw out different samples of size \( k \) is \( \binom{n}{k} \). With respect to this sampling interpretation for the formation of combinations, \( \binom{n}{k} \) is usually read ”\( n \) things taken \( k \) at a time” or ”\( n \) choose \( k \)”.

Comment. The symbol \( \binom{n}{k} \) appears in the statement of a familiar theorem from algebra,

\[
(x + y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}
\]

Since the expression being raised to a power involves two terms, \( x \) and \( y \), the constants \( \binom{n}{k}, k = 0, 1, \ldots, n \), are commonly referred to as binomial coefficients.

**EXAMPLE 2.6.18**

Eight politicians meet at a fund-raising dinner. How many greetings can be exchanged if each politician shakes hands with every other politician exactly once?
Imagine the politicians to be eight chips - 1 through 8 - in an urn. A handshake corresponds to an unordered sample of size 2 chosen from that urn. Since repetitions are not allowed (even the most obsequious and overzealous of campaigners would not shake hands with himself!), Theorem 2.6.3 applies, and the total number of handshakes is
\[
\binom{n}{k} = \frac{8!}{2!6!}
\]
or 28.

**EXAMPLE 2.6.22**

Consider Binomial expansion \((a + b)^n = (a + b)(a + b) \cdots (a + b)\)

When \(n=4\), \((a+b)^4 = a^4 + 4a^3b + 6a^2b^2 + 4ab^3 + b^4\)

Notice: The literal factors are all the combinations of a and b where the sum of the exponents is 4: \(a^4, a^3b, a^2b^2, ab^3, b^4\). The degree of each term is 4. In the expansion of \((a + b)^4\), the **binomial coefficients** are 1 4 6 4 1; The coefficients from left to right are the same right to left.

The answer to the question, ”What are the binomial coefficients?” is called the binomial theorem. It shows how to calculate the coefficients in the ex-
pansion of \((a + b)^n\).

The symbol for a binomial coefficient is \(\binom{n}{k}\). The upper index \(n\) is the exponent of the expansion; the lower index \(k\) indicates which term.

For example, when \(n = 5\), each term in the expansion of \((a + b)^5\) will look like this:

\[
\binom{n}{k} a^{5-k} b^k
\]

\(k\) will successively take on the values 0 through 5. Therefore the binomial theorem is

\[
(a + b)^n = \sum_{k=0}^{n} \binom{n}{k} a^{5-k} b^k
\]

http://www.themathpage.com/aprecalc/binomial-theorem.htm

Binomial coefficients have many interesting properties. Perhaps the most famous is Pascal’s triangle, a numerical array where each entry is equal to the sum of the two numbers appearing diagonally above it (see Figure 2.6.16). Note that each entry in Pascal’s triangle can be expressed as a binomial
coefficient, and the relationship just described appears to reduce to a simple equation involving those coefficients:
\[
\binom{n+1}{k} = \binom{n}{k} + \binom{n}{k-1}
\]  
Equation 2.6.1

Prove that Equation 2.6.1 holds for all positive integers \(n\) and \(k\).

**FIGURE 2.6.16**

Consider a set of \(n + 1\) distinct objects \(A_1, A_2, \ldots, A_{n+1}\).

We can obviously draw samples of size \(k\) from that set in \(\binom{n+1}{k}\) different ways. Now, consider any particular object - for example, \(A_1\). Relative to \(A_1\), each of those \(\binom{n+1}{k}\) samples belongs to one of two categories: those containing \(A_1\) and those not containing \(A_1\). To form samples containing \(A_1\), we need to select \(k - 1\) additional objects from the remaining \(n\). This can be done in \(\binom{n}{k-1}\) ways. Similarly, there are \(\binom{n}{k}\) ways to form samples not containing \(A_1\). Therefore, \(\binom{n+1}{k}\) must equal \(\binom{n}{k} + \binom{n}{k-1}\).

**Practice**
2.6.50. How many straight lines can be drawn between five points (A, B, C, D, and E), no three of which are collinear?

Since every (unordered) set of two letters describes a different line, the number of possible lines is \( \binom{5}{2} = 10 \)

Homework
2.6.54

2.7 Combinatorial Probability

In Section 2.6 our concern focused on counting the number of ways a given operation, or sequence of operations, could be performed. In Section 2.7 we want to couple those enumeration results with the notion of probability. Putting the two together makes a lot of sense – there are many combinatorial problems where an enumeration, by itself, is not particularly relevant.

In a combinatorial setting, making the transition from an enumeration to a probability is easy. If there are \( n \) ways to perform a certain operation and
a total of \( m \) of those satisfy some stated condition—call it \( A \)—then \( P(A) \) is defined to be the ratio \( m/n \). This assumes, of course, that all possible outcomes are equally likely. Historically, the "\( m \) over \( n \)" idea is what motivated the early work of Pascal, Fermat, and Huygens (recall section 1.3). Today we recognize that not all probabilities are so easily characterized. Nevertheless, the \( m/n \) model—the so-called classical definition of probability—is entirely appropriate for describing a wide variety of phenomena.

**Example 2.7.1**

A box contains eight chips, numbered 1 through 8. A sample of three is drawn without replacement. What is the probability that the largest chip in the sample is 5? Let \( A \) be the event "Largest chip in sample is a 5" Figure 2.7.1 shows what must happen in order for \( A \) to occur: (1) the 5 chip must be selected, and (2) two chips must be drawn from the subpopulation of chips numbered 1 through 4. By the multiplication rule, the number of samples satisfying event \( A \) is the product \( \binom{1}{1} \cdot \binom{4}{2} \). The sample space \( S \) for the experiment of drawing three chips
from the box contains \(\binom{8}{3}\) outcomes, all equally likely. In this situation, then \(m = \binom{1}{1} \cdot \binom{4}{2}, n = \binom{8}{3}\), and

\[
P(A) = \frac{\binom{1}{1} \cdot \binom{4}{2}}{\binom{8}{3}} = 0.11
\]

**Example 2.7.2**

A box contains \(n\) red chips numbered 1 through \(n\), \(n\) white chips numbered 1 through \(n\), and \(n\) blue chips numbered 1 through \(n\). Two chips drawn are drawn at random and without replacement. What is the probability that the two drawn are either the same color or the same number?

Let \(A\) be the event that the two chips drawn are the same color, let \(B\) be the event that they have the same number. We are looking for \(P(A \cup B)\). Since \(A\) and \(B\) here are mutually exclusive,

\[
P(A \cup B) = P(A) + P(B).
\]

With \(3n\) chips in the box, the total number of ways draw an unordered sample of size 2 is \(\binom{3n}{2}\). More-
over,
\[ P(A) = P(2 \text{ red } \cup 2 \text{ whites } \cup 2 \text{ blues}) \]
\[ = P(2 \text{ red}) + P(2 \text{ whites}) + P(2 \text{ blues}) \]
\[ = \frac{3\binom{n}{2}}{\binom{3n}{2}} \]
and
\[ P(B) = P(\text{ two 1 's } \cup \text{ two 2 's } \cup \cdots \cup \text{ two n 's}) \]
\[ = n\binom{3}{2}/\binom{3n}{2}. \]

Therefore
\[ P(A \cup B) = \frac{3\binom{n}{2} + n\binom{3}{2}}{\binom{3n}{2}} \]
\[ = \frac{n + 1}{3n - 1} \]

**Practice**

2.7.1 Ten equally qualified marketing assistants are candidates for promotion to associate buyer; seven are men and three are women. If the company intends to promote four of the ten at random, what is the probability that exactly two of the four are women?
Let A: \{ Exactly two of the four are women \} 
B: \{ two of the four are men \}, \ n(S) = \binom{10}{4}

\[ P(A \cap B) = \frac{n(A \cap B)}{n(S)} \]

\[ = \frac{\binom{7}{2} \binom{3}{2}}{\binom{10}{4}} \]

Homework:
2.7.2 and 2.7.7
Chapter 3

Random Variables

3.1 3.1 Introduction

Throughout Chapter 2, probabilities were assigned to events - that is, to sets of sample outcomes. The events we dealt with were composed of either a finite or a countably infinite number of sample outcomes, in which case the event’s probability was simply the sum of the probabilities assigned to its outcomes. One particular probability function that came up over and over again in Chapter was the assignment of $1/n$ as the probability associated with each of the $n$ points in a finite sample space. This is the model that typically describes games of chance (and all of our combinatorial probability problems in Chapter 2).

The first objective of this chapter is to look at several other useful ways for assigning probabilities to
sample outcomes. In so doing, we confront the desirability of ”redefining” sample spaces using functions known as *random variables*. How and why these are used - and what their mathematical properties are - become the focus of virtually everything covered in Chapter 3.

As a case in point, suppose a medical researcher is testing eight elderly adults for their allergic reaction (yes or no) to a new drug for controlling blood pressure. One of the $2^8 = 256$ possible sample points would be the sequence (yes, no, no, yes, no, no, yes, no), signifying that the first subject had an allergic reaction, the second did not, the third did not, and so on. Typically, in studies of this sort, the particular subjects experiencing reactions is of little interest: what does matter is the *number* who show a reaction. If that were true here, the outcome’s relevant information (i.e., the number of allergic reactions) could be summarized by the number 3.

Suppose $X$ denotes the number of allergic reactions among a set of eight adults. Then $X$ is said to be a *random variable* and the number 3 is the *value* of the random variable for the outcome (yes,
In general, random variables are functions that associate numbers with some attribute of a sample outcome that is deemed to be especially important. If $X$ denotes the random variable and $s$ denotes a sample outcome, then $X(s) = t$, where $t$ is a real number. For the allergy example, $s = (\text{yes, no, no, yes, no, no, yes, no})$ and $t = 3$.

Random variables can often create a dramatically simpler sample space. That certainly is the case here - the original sample space has $256$ ($= 2^8$) outcomes, each being an ordered sequence of length eight. The random variable $X$, on the other hand, has only nine possible values, the integers from 0 to 8, inclusive.

In terms of their fundamental structure, all random variables fall into one of two broad categories, the distinction resting on the number of possible values the random variable can equal. If the latter is finite or countably infinite (which would be the case with the allergic reaction example), the random variable is said to be discrete; if the outcomes
can be any real number in a given interval, the number of possibilities is uncountably infinite, and the random variable is said to be continuous. The difference between the two is critically important, as we will learn in the next several sections.

The purpose of Chapter 3 is to introduce the important definitions, concepts and computational techniques associated with random variables, both discrete and continuous. Taken together these ideas form the bedrock of modern probability and statistics.

3.2 Binomial and Hypergeometric Probabilities

This section looks at two specific probability scenarios that are especially important, both for their theoretical implications as well as for their ability to describe real-world problems. What we learn in developing these two models will help us understand random variables in general, the formal discussion of which beings in Section 3.3.

The Binomial Probability Distribution
Binomial probabilities apply to situations involving a series of independent and identical trials (Bernoulli), where each trial can have only one of two possible outcomes. Imagine three distinguishable coins being tossed, each having a probability $p$ of coming up heads. The set of possible outcomes are the eight listed in Table 3.2.1. If the probability of any of the coins coming up heads is $p$, then the probability of the sequence $(H, H, H)$ is $p^3$, since the coin tosses qualify as independent trials. Similarly the probability of $(T, H, H)$ is $(1−p)p^2$. The fourth column of Table 3.2.1 shows the probabilities associated with each of the three-coin sequences.
Table 3.2.1

<table>
<thead>
<tr>
<th>1st</th>
<th>2nd</th>
<th>3rd</th>
<th>Prob.</th>
<th>Number of Heads</th>
</tr>
</thead>
<tbody>
<tr>
<td>H</td>
<td>H</td>
<td>H</td>
<td>$p^3$</td>
<td>3</td>
</tr>
<tr>
<td>H</td>
<td>H</td>
<td>T</td>
<td>$p^2(1 - p)$</td>
<td>2</td>
</tr>
<tr>
<td>H</td>
<td>T</td>
<td>H</td>
<td>$p^2(1 - p)$</td>
<td>2</td>
</tr>
<tr>
<td>T</td>
<td>H</td>
<td>H</td>
<td>$p^2(1 - p)$</td>
<td>2</td>
</tr>
<tr>
<td>H</td>
<td>T</td>
<td>T</td>
<td>$p(1 - p)^2$</td>
<td>1</td>
</tr>
<tr>
<td>T</td>
<td>H</td>
<td>T</td>
<td>$p(1 - p)^2$</td>
<td>1</td>
</tr>
<tr>
<td>T</td>
<td>T</td>
<td>H</td>
<td>$p(1 - p)^2$</td>
<td>1</td>
</tr>
<tr>
<td>T</td>
<td>T</td>
<td>T</td>
<td>$(1 - p)^3$</td>
<td>0</td>
</tr>
</tbody>
</table>

Suppose our main interest in the coin tosses is the number of heads that occur. Whether the actual sequence is, say (H, H, T) of (H, T, H) is immaterial, since each outcome contains exactly two heads. The last column of Table 3.2.1 shows the number of heads in each of the eight possible outcomes. Notice that there are three outcomes with exactly two heads, each having an individual probability of $p^2(1 - p)$. The probability, then, of the event ”two heads” is the sum of those three individual probabilities - that is, $3p^2(1 - p)$. Table 3.2.2
lists the probabilities of tossing \( k \) heads, where \( k = 0, 1, 2, \) or 3.

<table>
<thead>
<tr>
<th>Number of Heads</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>((1 - p)^3)</td>
</tr>
<tr>
<td>1</td>
<td>(3p(1 - p)^2)</td>
</tr>
<tr>
<td>2</td>
<td>(3p^2(1 - p))</td>
</tr>
<tr>
<td>3</td>
<td>(p^3)</td>
</tr>
</tbody>
</table>

Now, more generally, suppose that \( n \) coins are tossed, in which case the number of heads can equal any integer from 0 through \( n \). By analogy,
\[ P(k \text{ heads }) = \text{( number of ways to arrange } k \text{ heads and } n - k \text{ tails)} \cdot \text{( probability of any particular sequence having } k \text{ heads and } n - k \text{ tails)} \]
\[ = \text{( number of ways to arrange } k \text{ heads and } n - k \text{ tails)} \cdot p^k (1 - p)^{n-k} \]

The number of ways to arrange \( k \) H’s and \( n - k \) T’s, though is \( \frac{n!}{k!(n-k)!} \), or \( \binom{n}{k} \) (recall Theorem 2.6.2).

**Theorem 3.2.1** Consider a series of \( n \) independent trials, each resulting in one of two possible outcomes, ”success” or ”failure.” Let \( p = P(\text{success occurs at any given trial}) \) and assume that \( p \) remains constant from trial to trial. Then

\[ P(k \text{ successes}) = \binom{n}{k} p^k (1 - p)^{n-k}, \quad k = 0, 1, \cdots n \]
\[ P(x \text{ successes}) = \binom{n}{x} p^x (1 - p)^{n-x}, \quad x = 0, 1, \ldots n \]

**Comment** The probability assignment given by the equation in Theorem 3.2.1 is known as the *binomial distribution*.

**Example 3.2.1** An information technology center uses nine aging disk drives for storage. The probability that any one of them is out of service is 0.06. For the center to function properly, at least seven of the drives must be available. What is the probability that the computing center can get its work done?

The probability that a drive is available is \( p = 1 - 0.06 = 0.94 \). Assuming the devices operate independently, the number of disk drives available has a binomial distribution with \( n = 9 \) and \( p = 0.94 \). The probability that at least seven disk drives work is a reassuring 0.986:

\[ \binom{9}{7} (0.94)^7 (0.06)^2 + \binom{9}{8} (0.94)^8 (0.06)^1 \]
\[ + \binom{9}{9} (0.94)^9 (0.06)^0 = 0.986 \]

Practice:
Suppose that since the early 1950’s some ten-thousand independent UFO sightings have been reported to civil authorities. If the probability that any sighting is genuine on the order of one in one hundred thousand, what is the probability that at least on of the ten thousand was genuine?

The probability of \( k \) sightings is given by the binomial probability model with \( n = 10,000 \) and \( p = 1/100,000 \). The probability of at least one genuine sighting is the probability that \( k \geq 1 \). The probability of the complementary event, \( k = 0 \), is \( (99,999/100,000)^{10,000} = 0.905 \). Thus, the probability that \( k \geq 1 \) is \( 1 - 0.905 = 0.095 \).

Homework: 3.2.2, 3.2.4, 3.2.5, 3.2.8
The Hypergeometric Distribution

The second ”special” distribution that we want to look at formalizes the urn problems that frequented Chapter 2. Our solutions to those earlier problems tended to be enumerations in which we listed the entire set of possible samples, and then counted the ones that satisfied the event in question. The inefficiency and redundancy of that approach should now be painfully obvious. What we are seeking here is a general formula that can be applied to any and all such problems, much like the expression in Theorem 3.2.1 can handle the full range of question arising from the binomial model.

Suppose an urn contains r red chips and w white chips, where r + w = N. Imagine drawing n chips from the urn one at a time without replacing any of the chips selected. At each drawing we record the color of the chip removed. The question is, what is the probability that exactly k red chips are included among the n that are removed?

Notice that the experiment just described is similar in some respects to the binomial model, but the method of sampling creates a critical distinction. If
each chip drawn was replaced prior to making another selection, then each drawing would be an independent trial, the chances of drawing a red in any given trial would be a constant $r/N$, and the probability that exactly $k$ red chips would ultimately be included in the $n$ selections would be a direct application of Theorem 3.2.1:

$$P(\text{ k reds drawn } ) = \binom{n}{k} \left(\frac{r}{N}\right)^k \left(1 - \frac{r}{N}\right)^{n-k},$$

$$k = 0, 1, 2, \ldots, n$$

However, if the chips drawn are not replaced, then the probability of drawing a red on any given attempt is not necessarily $r/N$: Its value would depend on the colors of the chips selected earlier. Since $p = P(\text{Red is drawn}) = P(\text{success})$ does not remain constant from drawing to drawing, the binomial model of Theorem 3.2.1 does not apply. Instead, probabilities that arise from the "no replacement" scenario just described are said to follow the hypergeometric distribution.
Theorem 3.2.2 Suppose an urn contains $r$ red chips and $w$ white chips, where $r + w = N$. If $n$ chips are drawn out at random, without replacement, and if $k$ denotes the number of red chips selected, then

$$P(k \text{ red chips are chosen}) = \frac{\binom{r}{k}\binom{w}{n-k}}{\binom{N}{n}} \quad (3.2.1)$$

where $k$ varies over all the integers for which $\binom{r}{k}$ and $\binom{w}{n-k}$ are defined. The probabilities appearing on the right-hand side of Equation 3.2.1 are known as the hypergeometric distribution.

Comment The appearance of binomial coefficients suggests a model of selecting unordered subsets. Indeed one can consider the model of selecting a subset of size $n$ simultaneously, where order doesn’t matter. In that case, the question remains: What is the probability of getting $k$ red chips and $n-k$ white chips? A moment’s reflection will show that the hypergeometric probabilities given in the statement of the theorem also answer that question. So, if our interest is simply counting the number of
red and white chips in the sample, the probabilities are the same whether the drawing of the sample is simultaneous or the chips are drawn in order without repetition.

**Example (from 3.2.20)**

A corporate board contains twelve members. The board decides to create a five-person Committee to hide Corporation Debt. Suppose four members of the boards are accountants. What is the probability that the committee will contain two accountants and three nonaccountants?

\[ P(2A \cap 2A^C) = \binom{4}{2} \binom{8}{3} \left(\frac{12}{5}\right) = \frac{14}{33} \]

**Homework: 3.2.22, 3.2.23, 3.2.25**

3.3 Discrete Random Variables

The purpose of this section is to (1) outline the general conditions under which probabilities can be
assigned to sample spaces and (2) explore the ways and means of redefining sample spaces through the use of random variables. The notation introduced in this section is especially important and will be used throughout the remainder of the book.

**Assigning Probabilities: The Discrete Case**

We begin with the general problem of assigning probabilities to sample outcomes, the simplest version of which occurs when the number of points in \( S \) is either finite or countably infinite. The probability function, \( p(s) \), that we are looking for in those cases satisfy the conditions in Definition 3.3.1.

**Definition 3.3.1** Suppose that \( S \) is finite or countably infinite sample space. Let \( p \) be a real valued function defined for each element of \( S \) such that

a) \( 0 \leq p(s) \) for each \( s \in S \)

b) \( \sum_{\text{all } s \in S} p(s) = 1 \)

Then \( p \) is said to be a **discrete probability function**.
Comment Once \( p(s) \) is defined for all \( s \), it follows that the probability of any event \( A \) - that is \( P(A) \) - is the sum of the probabilities of the outcomes comprising \( A \):

\[
P(A) = \sum_{\text{all } s \in A} p(s)
\]

Defined in this way, the function \( P(A) \) satisfies the probability axioms given in Section 2.3. The next several examples illustrate some of the specific form that \( p(s) \) can have and how \( P(A) \) is calculated.

Example 3.3.2 Suppose a fair coin is tossed until a head comes up for the first time. What are the chances of that happening on an odd-numbered toss?

Note that the sample space here is countably infinite and so is the set of outcomes making up the event whose probability we are trying to find. The \( P(A) \) that we are looking for, then, will be the sum of an infinite number of terms.

Let \( p(s) \) be the probability that the first head ap-
pears on the sth toss. Since the coin is presumed to be fair, \( p(1) = 1/2 \). Furthermore, we would expect that half the time, when a tail appears, the next toss would be a head, so \( p(2) = 1/2 \) times \( 1/2 = 1/4 \). In general, \( p(s) = (1/2)^s, s = 1, 2, \ldots \)

Does \( p(s) \) satisfy the conditions stated in Definition 3.3.1? Yes. Clearly, \( p(s) \) greater than or equal to 0 for all \( s \). To see that the sum of the probabilities is 1, recall the formula for the sum of a geometric series: If \( 0 < r < 1 \),

\[
\sum_{s=0}^{\infty} r^s = \frac{1}{1-r}
\]

Applying Equation 3.3.2 to the sample space here confirms that \( P(S) = 1 \):

\[
P(S) = \sum_{s=1}^{\infty} p(s) = \sum_{s=1}^{\infty} \left(\frac{1}{2}\right)^s
\]

\[
= \sum_{s=0}^{\infty} \left(\frac{1}{2}\right)^s - \left(\frac{1}{2}\right)^0 = \frac{1}{1-\frac{1}{2}} - 1 = 1
\]

Now, let \( A \) be the event that the first head appears on an odd-numbered toss. Then \( P(A) = p(1) + p(3) + p(5) + \cdots \)
\[
\sum_{s=0}^{\infty} p(2s + 1) = \sum_{s=0}^{\infty} \left(\frac{1}{2}\right)^{2s+1}
\]

\[
= \sum_{s=0}^{\infty} \left(\frac{1}{2}\right)^{2s+1}
\]

\[
= \frac{1}{2} \sum_{s=0}^{\infty} \left(\frac{1}{4}\right)^s
\]

\[
= \frac{1}{2} \left[ \frac{1}{1 - \frac{1}{4}} \right] = \frac{2}{3}
\]

**Example 3.3.4**

Is

\[
p(s) = \frac{1}{1 + \lambda} \left(\frac{\lambda}{1 + \lambda}\right), \ s = 0, 1, 2, \cdots; \lambda > 0
\]

a discrete probability function? Why or why not?

A simple inspection \( p(s) \geq 0 \) for all \( s \)
\[
\sum_{\text{all } s \in S} p(s) = \sum_{s=0}^{\infty} \frac{1}{1+\lambda} \left( \frac{\lambda}{1+\lambda} \right) \\
= \frac{1}{1+\lambda} \left( \frac{1}{1 - \frac{\lambda}{1+\lambda}} \right) \\
= \frac{1}{1+\lambda} \frac{1+\lambda}{1} = 1
\]

**Definition 3.3.2** A function whose domain is a sample space \( S \) and whose values form a finite of countably infinite set of real numbers is called discrete random variable. We denote random variable by upper case \( X \) or \( Y \).

**Example 3.3.5.**

Consider tossing two dice, an experiment for which the sample space is a set of ordered pairs, \( S = \{(i, j)|i = 1, 2, \cdots, 6; j = 1, 2, \cdots, 6\} \). For a variety of games the sum showing is what matters on a given turn. That being the case, the original sample \( S \) of thirty six ordered pairs would not
provide a particularly convenient backdrop for discussing the rules of those games. It would be better to work directly with sums. Of course the eleven possible sums (from two to twelve) are simply the different values of the random variable $X$ where $X(i, j) = i + j$.

Comment: In the above example, suppose we define a random variable $X_1$ that gives the result on the first die and a random variable $X_2$ that gives the result on the second die. Then $X = X_1 + X_2$. Note how easily we could extend this idea to the toss of three dice or ten dice. The ability to conveniently express complex events in the terms of simple ones is an advantage of the random variable concept that we will see playing out over and over again.

**The probability density Function**

**Definition 3.3.3.** Associated with every discrete random variable $X$ is a probability density function (or pdf), denoted $p_X(k)$ where
\[ p_X(k) = P\{s \in S| X(s) = k\} \]

Note that \( p_X(k) = 0 \) for any \( k \) not in the range of \( X \). For notational simplicity, we will usually delete all references to \( s \) and \( S \) and write \( p_X(k) = P(X = k) \)

**Comment.** We have already discussed at length two examples of the function \( p_X(k) \). Recall the binomial distribution derived in Section 3.2. If we let the random variable \( X \) denote the number of successes in \( n \) independent trials, then Theorem 3.2.1 states that

\[ p_X(k) = P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}, \]
\[ k = 0, 1, 2 \cdots , n \]

**EXAMPLE 3.3.6**

Consider rolling two dice as described in Example 3.3.5. Let \( i \) and \( j \) denote the faces showing on the
first and the second die respectively, and define the 
R.V. $X$ to be the sum of the two faces: $X(i, j) = i + j$. Find $p_X(k)$.

<table>
<thead>
<tr>
<th>$k$</th>
<th>$p_X(k)$</th>
<th>$k$</th>
<th>$p_X(k)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1/36</td>
<td>8</td>
<td>5/36</td>
</tr>
<tr>
<td>3</td>
<td>2/36</td>
<td>9</td>
<td>4/36</td>
</tr>
<tr>
<td>4</td>
<td>3/36</td>
<td>10</td>
<td>3/36</td>
</tr>
<tr>
<td>5</td>
<td>4/36</td>
<td>11</td>
<td>2/36</td>
</tr>
<tr>
<td>6</td>
<td>5/36</td>
<td>12</td>
<td>1/36</td>
</tr>
<tr>
<td>7</td>
<td>6/36</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**EXAMPLE 3.3.7**

Acme Industries typically produces three electric power generators per day: some pass the company’s quality control inspection on their first try and are ready to be shipped: others need to be retooled. The probability of a generator needing further work is 0.05 If a generator is ready to be shipped, the firm earns a profit of $10,000. If it needs to be retooled, it ultimately cost the firm $2000. Let $X$ be the random variable quantifying the company’s daily
profit. Find $p_X(k)$.

the underlying sample space here is a set of $n = 3$ independent trials, where

\[ p = P(\text{generator passes inspection}) = 0.95. \]

If the random variable $X$ is to measure the company’s daily profit, then

\[
X = \$10000 \times (\text{no. of generators passing inspection}) - \$2000 \times (\text{no. of generators needing retooling})
\]

What are the possible profit? k? It depend on the number of defective(non defective). The pdf of the profit will correspond to the pdf of number of defective which is distributed as binomial.

For instance $X(s, f, s) = 2(\$10000) - 1(\$2000) = \$18000$. Moreover, the R.V $X$ equals $\$18,000$ whenver the day’s output consists of two successes and one failure. It follows that

\[
P(X = \$18000) = p_X(18000) = \binom{3}{2}(0.95)^2(0.05)^1 = 0.135375\]
Table 3.3.4

<table>
<thead>
<tr>
<th>No. Defective</th>
<th>k= Profit</th>
<th>$p_X(k)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>$30000$</td>
<td>0.857375</td>
</tr>
<tr>
<td>1</td>
<td>$18000$</td>
<td>0.135375</td>
</tr>
<tr>
<td>2</td>
<td>$6000$</td>
<td>0.007125</td>
</tr>
<tr>
<td>3</td>
<td>-$6000$</td>
<td>0.000125</td>
</tr>
</tbody>
</table>

Linear Transformation

**Theorem 3.3.1.**
Suppose $X$ is a discrete random variable. Let $Y = aX + b$, where $a$ and $b$ are constants. Then

$$p_Y(y) = p_X\left(\frac{y-b}{a}\right)$$

**Proof.**

$$p_Y(y) = P(Y = y) = P(aX + b = y)$$

$$= P(X = \frac{(y-b)}{a}) = p_X\left(\frac{(y-b)}{a}\right)$$

**Practice 3.3.11** Suppose $X$ is a binomial distribution with $n = 4$ and $p = \frac{2}{3}$. What is the pdf of $2X + 1$
Given

\[ p_X(x) = \binom{n}{x} p^x (1-p)^{n-x} \]

Let \( Y = 2X + 1 \), then

\[ P(Y = y) = P(2X + 1 = y) = P(X = \frac{(y - 1)}{2}) \]

\[ = p_X \left( \frac{(y - 1)}{2} \right) \]

\[ = \binom{n}{\frac{(y - 1)}{2}} p^{\frac{(y - 1)}{2}} (1-p)^{n-\frac{(y - 1)}{2}} \]

\[ = \left( \frac{4}{\frac{(y - 1)}{2}} \right) \left( \frac{2}{3} \right)^{\frac{(y - 1)}{2}} (1 - \left( \frac{2}{3} \right))^{4-\frac{(y - 1)}{2}} \]

**The Cumulative Distribution function**

In working with random variables, we frequently need to calculate the probability that the value of a random variable is somewhere between two numbers. For example, suppose we have an integer-valued random variable. We might want to calculate an expression like

\[ P(s \leq X \leq t) \]
If we know the pdf for $X$, then

$$P(s \leq X \leq t) = \sum_{k=s}^{t} p_X(k).$$

but depending on the nature of $p_X(k)$ and the number of terms that need to be added, calculating the sum of $p_X(k)$ from $k = s$ to $k = t$ may be quite difficult. An alternate strategy is to use the fact that

$$P(s \leq X \leq t) = P(X \leq t) - P(X \leq s - 1)$$

where the two probabilities on the right represent cumulative probabilities of the random variable $X$. If the latter were available (and they often are), then evaluating $P(s \leq X \leq t)$

by one simple subtraction would clearly be easier than doing all the calculations implicit in $\sum_{k=s}^{t} p_X(k)$.

**Definition 3.3.4.** Let $X$ be a discrete random variable. For any real number $t$, the probability that $X$ takes on a value $\leq t$ is the cumulative distribution function (cdf) of $X$ (written $F_X(t)$). In formal notation, $F_X(t) = P\{s \in S | X(s) \leq t \}$. 
As was the case with pdfs, references to $s$ and $S$ are typically deleted, and the cdf is written $F_X(t) = P(X \leq t)$

**EXAMPLE 3.3.10**

Suppose we wish to compute $P(21 \leq X \leq 40)$ for a binomial random variable $X$ with $n = 50$ and $p = 0.6$. From Theorem 3.2.1, we know the formula for $p_X(k)$, so $P(21 \leq X \leq 40)$ can be written as a simple, although computationally cumbersome, sum:

$$P(21 \leq X \leq 40) = \sum_{k=21}^{40} \binom{50}{k}(0.6)^k(0.4)^{50-k}$$

Equivalently, the probability we are looking for can be expressed as the difference between two cdfs:

$$P(21 \leq X \leq 40) = P(X \leq 40) - P(X \leq 20) = F_X(40) - F_X(20)$$

As it turns out, values of the cdf for a binomial random variable are widely available, both in books
and in computer software. Here, for example, $F_X(40) = 0.9992$ and $F_X(20) = 0.0034$, so

$$P(21 \leq X \leq 40) = 0.9992 - 0.0034 = 0.9958$$

**Practice**

3.3.1 An urn contains five balls numbered 1 through 5. Two balls are drawn simultaneously.

a) Let $X$ be the larger of the two numbers drawn. Find $p_X(k)$.

b) Let $V$ be the sum of the two numbers drawn. Find $p_V(k)$.

a) Each outcome has probability $1/10$ Outcome $X = $ larger no. drawn 1,
Counting the number of each value of the larger of the two and multiplying by $1/10$ gives the pdf:

<table>
<thead>
<tr>
<th>$k$</th>
<th>$p_X(k)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1/10</td>
</tr>
<tr>
<td>3</td>
<td>2/10</td>
</tr>
<tr>
<td>4</td>
<td>3/10</td>
</tr>
<tr>
<td>5</td>
<td>4/10</td>
</tr>
<tr>
<td>Outcome</td>
<td>X = larger no. drawn</td>
</tr>
<tr>
<td>---------</td>
<td>----------------------</td>
</tr>
<tr>
<td>1, 2</td>
<td>2</td>
</tr>
<tr>
<td>1, 3</td>
<td>3</td>
</tr>
<tr>
<td>1, 4</td>
<td>4</td>
</tr>
<tr>
<td>1, 5</td>
<td>5</td>
</tr>
<tr>
<td>2, 3</td>
<td>3</td>
</tr>
<tr>
<td>2, 4</td>
<td>4</td>
</tr>
<tr>
<td>2, 5</td>
<td>5</td>
</tr>
<tr>
<td>3, 4</td>
<td>4</td>
</tr>
<tr>
<td>3, 5</td>
<td>5</td>
</tr>
<tr>
<td>4, 5</td>
<td>5</td>
</tr>
</tbody>
</table>

\[
\begin{array}{|c|c|}
\hline
k & p_X(k) \\
\hline
3 & 1/10 \\
4 & 1/10 \\
5 & 2/10 \\
6 & 2/10 \\
7 & 2/10 \\
8 & 1/10 \\
9 & 1/10 \\
\hline
\end{array}
\]

**Homework:** 3.3.4, 3.3.5, 3.3.13, 3.3.14
If a random variable $X$ is defined over a continuous sample space $S$ (contained uncountably infinite number of outcomes), then the r.v $X$ is said to be continuous r.v. Example of continuous r.v are time, temperature, weight and etc.

How do we assign a probability to this type of sample space?

When $S$ is discrete (countable), we can assign each outcome $s$ with $p(s)$.

This will not work for $S$ continuous.

**Definition 3.4.1** A probability function $P$ on a set of real numbers $S$ is called continuous if there exist a function $f(t)$ such that for any closed interval $[a, b] \subset S$, $P([a, b]) = \int_{a}^{b} f(t)dt$.

**Comment** If a probability function $P$ satisfies Definition 3.4.1, then $P(A) = \int_{A} f(t)dt$ for any set $A$ where the integral is defined.

Conversely, suppose a function $f(t)$ has two prop-
properties

a) \( f(t) \geq 0 \) for all \( t \).

b) \( \int_{-\infty}^{\infty} f(t)dt = 1 \).

If \( P(A) = \int_A f(t)dt \) for all \( A \), then \( P \) will satisfy the probability axioms given in section 2.3.

Choosing the Function \( f(t) \)

**Example 3.4.1.** The continuous equivalent of the equiprobable probability model on a discrete sample space is the function \( f(t) \) defined by \( f(t) = \frac{1}{b-a} \) for all \( t \) in the interval \([a, b]\) (and \( f(t) = 0 \) otherwise). This particular \( f(t) \) places equal probability weighting on every closed interval of the same length contained in the interval \([a, b]\). For example, suppose \( a = 0 \) and \( b = 10 \), and let \( A = [1, 3] \) and \( B = [6, 8] \), then \( f(t) = \frac{1}{10} \) and

\[
P(A) = \int_1^3 \left( \frac{1}{10} \right) dt = \frac{2}{10} = \int_6^8 \left( \frac{1}{10} \right) dt
\]

**Example 3.4.2**
Could $f(t) = 3t^2$, $0 \leq t \leq 1$ be used to defined the probability function for a continuous sample space whose outcomes consist of all the real numbers in the interval $[0, 1]$?

Yes because $(1) f(t) \geq 0$ for all $t$, and

$(2) \int_0^1 f(t)dt = \int_0^1 3t^2dt = 3t^3|_0^1 = 1$

**Example 3.4.3** By far the most important of all continuous probability function is the ”bell-shaped” curve, known more formally as the normal (Gaussian) distribution. The sample space for the normal distribution is the entire real line; its probability is given by

$$f(t) = \frac{1}{\sqrt{2\pi}\sigma}\exp \left[ -\frac{1}{2} \left( \frac{t - \mu}{\sigma} \right)^2 \right]$$

**Continuous Probability Density Functions**

**Definition 3.4.2**

A function $Y$ that maps a subset of the real numbers into the real numbers is called a *continuous random variable*. The *pdf* of $Y$ is the function
$f_Y(y)$ having the property that for any numbers $a$ and $b$,

$$P(a \leq Y \leq b) = \int_a^b f_Y(y)dy$$

Example 3.4.5

Suppose we would like a continuous random variable $Y$ to "select" a number between 0 and 1 in such a way that the intervals near the middle of the range would be more likely to be represented than intervals near 0 and 1. One pdf having that property is the function $f_Y(y) = 6y(1-y), 0 \leq y \leq 1$ (see figure 3.4.9). Do we know for certain that the function pictured is a legitimate pdf?

Yes because $f_Y(y) \geq 0$ for all $y$, and

$$\int_0^1 6y(1-y)dy = 6[y^2/2 - y^3/3]|_0^1 = 1$$

Continuous Cumulative Distribution functions

Definition 3.4.3. The cdf for a continuous ran-
Random variable $Y$ is the indefinite integral of its pdf:

$$F_Y(y) = \int_{-\infty}^{y} f_Y(r)dr$$

$$= P(\{s \in S|Y(s) \leq y\})$$

$$= P(Y \leq y)$$

**Theorem 3.4.1** Let $f_Y(y)$ be the pdf of a continuous R.V. with cdf $F_Y(y)$. Then

$$\frac{d}{dy} F_Y(y) = f_Y(y)$$

**Theorem 3.4.2** Let $Y$ be a continuous random variable with cdf $F_Y(y)$. Then

a) $P(Y > s) = 1 - F_Y(s)$

b) $P(r < Y \leq s) = F_Y(s) - F_Y(r)$

c) $\lim_{y \to \infty} F_Y(y) = 1$

d) $\lim_{y \to -\infty} F_Y(y) = 0$

**Transformation**

**Theorem 3.4.3** Suppose $X$ is a continuous R.V. Let $Y = aX + b$ where $a \neq 0$ and $b$ are
constant. Then
\[ f_Y(y) = \frac{1}{|a|} f_X \left( \frac{y - b}{a} \right) \]

**Proof.** We begin by writing an expression for the cdf of \( Y \):
\[ F_Y(y) = P(Y \leq y) = P(aX + b \leq y) = P(aX \leq y - b) \]

At this point we will consider two cases, First let \( a > 0 \). Then
\[ F_Y(y) = P(Y \leq y) = P(aX + b \leq y) \]
\[ = P(aX \leq y - b) = P \left( X \leq \frac{y - b}{a} \right) \]
and differentiating \( F_Y(y) \) yield \( f_Y(y) \).
\[ f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{d}{dy} F_X \left( \frac{y - b}{a} \right) \]
\[ = \frac{1}{a} f_X \left( \frac{y - b}{a} \right) \]
\[ = \frac{1}{|a|} f_X \left( \frac{y - b}{a} \right) \]

If \( a < 0 \)
\[ F_Y(y) = P(Y \leq y) = P(aX + b \leq y) \]
\[ = P(aX \leq y - b) = P \left( X \geq \frac{y - b}{a} \right) \]
\[ = 1 - P \left( X \leq \frac{y - b}{a} \right) \]

Differentiation yield

\[ f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{d}{dy} \left[ 1 - F_X \left( \frac{y - b}{a} \right) \right] \]
\[ = -\frac{1}{a} f_X \left( \frac{y - b}{a} \right) \]
\[ = \frac{1}{|a|} f_X \left( \frac{y - b}{a} \right) \]

**Practice 3.4.1**

**Homework**

3.4.2, 3.4.4, 3.4.6, 3.4.8, 3.4.10, 3.4.11, 3.4.12, 3.4.15, 3.4.16