# Contents

1 Introduction ................................................................. 1
  1.1 An Overview .................................................................... 1

2 Probability ........................................................................ 5
  2.1 Introduction ...................................................................... 5
  2.2 Sample spaces and The Algebra of sets ............................... 8
  2.3 The Probability Function ................................................. 19
  2.4 Conditional Probability ................................................... 28
  2.5 Independence .................................................................... 45
  2.6 Combinatoric .................................................................... 60
  2.7 Combinatorial Probability ............................................... 86

3 Random Variables .................................................................. 92
  3.1 Introduction ...................................................................... 92
  3.2 Binomial and Hypergeometric Probabilities ......................... 96
  3.3 Discrete Random Variables .............................................. 106
  3.4 Continuous random Variables ......................................... 123
  3.5 Expected Values ............................................................ 138
  3.6 The Variance .................................................................... 151
  3.7 Joint Densities .................................................................. 160
  3.8 Transformation and Combining Random Variables ............... 179
  3.9 Further properties of Mean and Variance ............................ 193
  3.10 Order Statistics — skip .................................................. 201
  3.11 Conditional Densities ...................................................... 201
  3.12 Moment Generating Functions ......................................... 203
  3.13 Taking a second look at statistics(Interpreting Means) ......... 215
Chapter 1

Introduction

1.1 An Overview

The book covers two broad topics
  • Mathematics of Statistics
  • Practice of statistics

Mathematics of Statistics refers to the probability that supports and justifies the various method used to analyze data. 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 \textit{before}, \textit{during} and \textit{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

+----------------+-----------------+
| 0              | .5              | 1               |
+----------------+-----------------+
| event never    | event and "not event" always will occur | event are likely will occur |

Three basics methods on assigning probability to an event.

1) classical approach. Credited to Gerolamo Cardano. Require that (1) the number of possible outcomes is finite and (2) all outcomes are equally likely. The probability of an event consisting m outcomes is \( \frac{m}{n} \), where n is the total possible outcomes. Example: Tossing a fair six sided die give \( n = 6 \). The probability of getting either 2, 4 or 6 is \( \frac{m}{n} = \frac{3}{6} \). (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.
\[ P(head) = \frac{\text{number of head}}{1000} = \frac{\text{number times event happen}}{\text{number of experiments}} \]

\[ P(tail) = \frac{\text{number of tail}}{1000} \]

3) subjective
Just guest 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**: $S = \{ \text{HHH, HTH, HHT, THH, HTT, THT, TTH, TTT} \}$

**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**: $S = \{ \text{T, HT, HHT, HHHT, \cdots} \}$

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:**

\[ A = \{ 134, 135, 136, 234, 235, 236 \} \]

Practice: 2.2.12

**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\} 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**

\( A = \{1, 3, 5, 7\} \) \( B = \{2, 4, 6, 8\} \)

\( A \cap B = \emptyset = \emptyset \)

**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**

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

\( A^C = \{1, 3, 5, 7, 9\} \)

**Example 2.2.11**

Suppose the events \( A_1, A_2, \cdots, A_k \) 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) \( F \cap R \cap V = \{E_1, E_2\} \)
(b) \( F^C \cap R \cap V^C = \{S_1, S_2, T_1, T_2\} \)
(c) \( F \cap R^C \cap V = \{A, I\} \)

**More Practices 2.2.16, 2.2.28, 2.2.29**

**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) \( A \cup B \)
(d) \( A \cap B^C \)
(e) \( A^C \cup B \)
(f) \( A^C \cap B \)

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 represent 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,

$$N(S) = 100$$
N(L) = 15
N(H) = 30
N(L \cap H) = 10

Summarize these information in a venn diagram. Notice that

N(L \cup H) = \text{number of alumni suffering from at least one affliction}

= 5 + 10 + 20
= 35

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 Morgan's 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 intersection of their complements.

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

Practice

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 function that behaved in accordance with a prescribed set of axioms.
Denote $P(A)$ as a probability of $A$, where $P$ is the 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 straightforward 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 $P(B)$?

Answer:

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. $P(A_1 \cup A_2 \cup A_3) = P($at least one 6th appears$) = 1 - P($no 6th appear$) = 1 - \left(\frac{5}{6}\right)^3$.

Practice

2.3.6

2.3.10
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*.

Consider a fair die being tossed, with $A$ defined as the event 6 appears. Clearly, $P(A) = 1/6$. But suppose that the die has already been tossed by someone who refuses to tell us whether or not $A$ occurred but does enlighten us to the extent of confirming that $B$ occurred, where $B$ is the event Even number appears. What are the chances of $A$ now? Here, common sense can help us: There are three equally
likely even numbers making up the event Bone of which satisfies the event A, so the updated probability is $1/3$. Notice that the effect of additional information, such as the knowledge that B has occurred, is to revise indeed, to shrink the original sample space S to a new set of outcomes S’ . In this example, the original S contained six outcomes, the conditional sample space, three (see Figure 2.4.1).

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.

**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 2.4.2
Consider the set of families having two children. Assume that the four possible birth sequences (younger child is a boy, older child is a boy), (younger child is a boy, older child is a girl), and so on are equally likely. (sequences (b, b), (b, g), (g, b), and (g, g) has a 1/4 probability of occurring.) What is the probability that both children are boys given that at least one is a boy?

Let A be the event that both children are boys, and let B be the event that at least one child is a boy. Then

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

since A is a subset of B (so the overlap between A and B is just A). But A has one outcome \{(b, b)\} and B has three outcomes (b, g), (g, b), (b, b). Applying Definition 2.4.1, then, gives \[ P(A|B) = \frac{1/4}{3/4} = 1/3 \]

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
(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,

\[
P(A \cup B) = P(B \cap A^C) + P(A \cap B) + P(B \cap A^C)
\]
\[
P(A \cap B) = 0.4 - 0.2 - 0.1 = 0.1
\]
\[
P(B) = P(A \cap B) + P(B \cap A^C) = 0.1 + 0.1 = 0.2
\]
\[
P(A|B) = \frac{0.1}{0.2} = 0.5
\]

**Example 2.4.5**

Max and Muffy are two myopic deer hunters who shoot simultaneously at a nearby sheepdog that they have mistaken for a 10-point buck. Based on years of well documented ineptitude, it can be assumed that Max has a 20% chance of hitting a stationary target at close range, Muffy has a 30% chance, and the probability is 0.06 that they will
both be on target. Suppose that the sheepdog is hit and killed by exactly one bullet. What is the probability that Muffy fired the fatal shot? Let $A$ be the event that Max hit the dog, and let $B$ be the event that Muffy hit the dog. Then $P(A) = 0.2$, $P(B) = 0.3$, and $P(A \cap B) = 0.06$.

{myopic: not able to see clearly things that are far away
ineptitude: incompetence}

We are trying to find $P(B|(A^C \cap B) \cup (A \cap B^C))$ where the event $(A^C \cap B) \cup (A \cap B^C)$ is the union of $A$ and $B$ minus the intersection. that is, it represents the event that either $A$ or $B$ but not both occur

Notice, that the intersection of $B$ and $(A^C \cap B) \cup (A \cap B^C)$ is the event $A^C \cap B$. Therefore, from Definition 2.4.1,

$$P(B|(A^C \cap B) \cup (A \cap B^C)) = [P(A^C \cap B)]/[P(A^C \cap B) \cup (A \cap B^C)]$$
$$= [P(B) − P(A \cap B)]/[P(A \cup B) − P(A \cap B)]$$
$$= [0.3 − 0.06]/[0.2 + 0.3 − 0.06 − 0.06]$$
$$= 0.63$$
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.
\[ B = \{sum = 9, 10, 11, 12\} \]
\[ B = \{(4,5), (5,4), (3,6), (6,3), (5,5), (4,6), (6,4), (5,6), \ldots \} \]
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

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, $A_1, A_2, \ldots, A_n$ give a general case for:

$$P(A_1 \cap A_2 \cap \cdots \cap A_n)$$
$$= P(A_n|A_1 \cap A_2 \cap \cdots \cap A_{n-1})$$
$$\cdot P(A_{n-1}|A_1 \cap A_2 \cap \cdots \cap A_{n-2})$$
$$\cdots P(A_2|A_1)P(A_1) \quad (2.4.1)$$

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
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 \( S \). 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 \( P(B|A_1) = 4/5, \ P(B|A_2) = 3/5, \ P(A_1) = 2/6, \ P(A_2) = 4/6. \) Then \( P(B) = P(B|A_2)P(A_2) + P(B|A_1)P(A_1) = 4/5 \cdot 2/6 + 3/5 \cdot 4/6 = 2/3 \)

**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, Ashey 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 internship?

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,

$$P(B|A_1) = 0.70 \quad P(B|A_2) = 0.20$$

$$P(A_1) = 0.10 \quad P(A_2) = 1 - P(A_1) = 0.90$$

From the Total Probability Theorem,

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

$$= (0.70)(0.10) + (0.2)(0.90) = 0.25$$

**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**

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” cor-
responds 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)

$A_2$: coin come up tails (i.e., chip came from box II)

We are trying to find $P(A_2|B)$, where $P(B|A_1) = 3/7, P(B|A_2) = 6/9, P(A_1) = 2/3, P(A_2) = 1/3$

so

$$P(A_2|B) = \frac{P(B|A_2)P(A_2)}{P(B|A_1)P(A_1) + P(B|A_2)P(A_2)}$$

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

**Example 2.4.16**

According to the manufacturers specifications, your home burglar alarm has a 95% chance of going off if someone breaks into your house. During the two years you have lived there, the alarm has gone off on five different nights, each time for no apparent reason. Suppose the alarm goes off tomorrow night. What is the probability that someone is trying to break into your house? (Note: Police statistics show that the chances of any particular house in your
neighborhood being burglarized on any given night are two in ten thousand.) Let $B$ be the event Alarm goes off tomorrow night, and let $A_1$ and $A_2$ be the events House is being burglarized and House is not being burglarized, respectively.

Then

\[ P(B|A_1) = 0.95 \]
\[ P(B|A_2) = \frac{5}{730} \text{(i.e., five nights in two years)} \]
\[ P(A_1) = \frac{2}{10000} \]
\[ P(A_2) = 1 - P(A_1) = \frac{9998}{10000} \]

The probability in question is $P(A_1|B)$. Intuitively, it might seem that $P(A_1|B)$ should be close to 1 because the alarms performance probabilities look good $P(B|A_1)$ is close to 1 (as it should be) and $P(B|A_2)$ is close to 0 (as it should be). Nevertheless, $P(A_1|B)$ turns out to be surprisingly small:

\[
P(A_1|B) = \frac{P(B|A_1)P(A_1)}{P(B|A_1)P(A_1) + P(B|A_2)P(A_2)}
\]
\[= \frac{(0.95)(2/10,000)}{(0.95)(2/10000) + (5/730)(9998/10000)}
\]
\[= 0.027\]
That is, if you hear the alarm going off, the probability is only 0.027 that your house is being burglarized. Computationally, the reason $P(A_1|B)$ is so small is that $P(A_2)$ is so large. The latter makes the denominator of $P(A_1|B)$ large and, in effect, washes out the numerator. Even if $P(B|A_1)$ were substantially increased (by installing a more expensive alarm), $P(A_1|B)$ would remain largely unchanged (see Table 2.4.3).

<table>
<thead>
<tr>
<th>Table 2.4.3</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(B</td>
</tr>
<tr>
<td>0.95</td>
</tr>
<tr>
<td>0.97</td>
</tr>
<tr>
<td>0.99</td>
</tr>
<tr>
<td>0.999</td>
</tr>
<tr>
<td>$P(A_1</td>
</tr>
<tr>
<td>0.027</td>
</tr>
<tr>
<td>0.028</td>
</tr>
<tr>
<td>0.028</td>
</tr>
<tr>
<td>0.028</td>
</tr>
</tbody>
</table>

**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 drawn 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

\[
P(A_W|B) = \frac{P(B|A_W)P(A_W)}{P(B|A_W)P(A_W) + P(B|A_R)P(A_R)}
\]

\[
= \frac{(2/4)(2/3)}{(2/4)(2/3) + (3/4)(1/3)} = 4/7
\]

**Homework**
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,
\[ P(A \cap B) = P(A)P(B). \] 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 \( P(HH) = P(H \cap H) = 1/2 \cdot 1/2 = 1/4 \)

**Example 2.5.5**

Myra and Carlos are summer interns working as proofreaders for a local newspaper. Based on aptitude tests, Myra has a 50% chance of spotting a hyphenation error, while Carlos picks up on that same kind of mistake 80% of the time. Suppose the copy they are proofing contains a hyphenation error. What is the probability it goes undetected?

Let \( A \) and \( B \) be the events that Myra and Carlos, respectively, catch the mistake. By assumption, \( P(A) = 0.50 \) and \( P(B) = 0.80 \). What we are looking for is the probability of the complement of a union. That is,
\[ P(\text{Error goes undetected}) \]
\[ = 1 - P(\text{Error is detected}) \]
\[ = 1 - P(\text{Myra or Carlos or both see the mistake}) \]
\[ = 1 - P(A \cup B) \]
\[ = 1 - \{ P(A) + P(B) - P(A \cap B) \} \]
\[ \text{(from Theorem 2.3.6)} \]

Since proofreaders invariably work by themselves, events A and B are necessarily independent, so \( P(A \cap B) \) would reduce to the product \( P(A).P(B) \). It follows that such an error would go unnoticed 10% of the time:

\[ P(\text{Error goes undetected}) \]
\[ = 1 - \{ 0.50 + 0.80 - (0.50)(0.80) \} \]
\[ = 1 - 0.90 \]
\[ = 0.10 \]

**Example 2.5.7** Emma and Josh have just gotten engaged. What is the probability that they have
different blood types? Assume that blood types for both men and women are distributed in the general population according to the following proportions:

<table>
<thead>
<tr>
<th>Blood Type</th>
<th>Proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>40%</td>
</tr>
<tr>
<td>B</td>
<td>10%</td>
</tr>
<tr>
<td>AB</td>
<td>5%</td>
</tr>
<tr>
<td>O</td>
<td>45%</td>
</tr>
</tbody>
</table>

First, note that the event Emma and Josh have different blood types includes more possibilities than does the event Emma and Josh have the same blood type. That being the case, the complement will be easier to work with than the question originally posed. We can start, then, by writing

\[
P(\text{Emma and Josh have different blood types}) = 1 - P(\text{Emma and Josh have the same blood type})
\]

Now, if we let \( E_X \) and \( J_X \) represent the events that Emma and Josh, respectively, have blood type \( X \), then the event Emma and Josh have the same blood type is a union of intersections, and we can write
\[ P(\text{Emma and Josh have the same blood type}) = P\{(E_A \cap J_A) \cup (E_B \cap J_B) \cup (E_{AB} \cap J_{AB}) \cup (E_O \cap J_O)\} \]

Since the four intersections here are mutually exclusive, the probability of their union becomes the sum of their probabilities. Moreover, ” blood type ” is not a factor in the selection of a spouse, so \( E_X \) and \( J_X \) are independent events and \( P(E_X \cap J_X) = P(E_X)P(J_X) \). It follows, then, that Emma and Josh have a 62.5\% chance of having different blood types:

\[
P(\text{Emma and Josh have different blood types}) = 1 - \left\{ P(E_A)P(J_A) + P(E_B)P(J_B) + P(E_{AB})P(J_{AB}) + P(E_O)P(J_O) \right\}
\]
\[
= 1 - \left\{ (0.40)(0.40) + (0.10)(0.10) + (0.05)(0.05) + (0.45)(0.45) \right\}
\]
\[
= 0.625
\]
**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.**
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

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

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

\[ P(A \cap B) = P(A) \cdot P(B) \] (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 $k, k = 1 \cdots n$,

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

Example 2.5.8

An insurance company plans to assess its future liabilities by sampling the records of its current policyholders. A pilot study has turned up three clients, one living in Alaska, one in Missouri, and one in Vermont, whose estimated chances of surviving to the year 2015 are 0.7, 0.9, and 0.3, respectively. What is the probability that by the end of 2014 the company will have had to pay death benefits to exactly one of the three?

Let $A_1$ be the event Alaska client survives through 2014. Define $A_2$ and $A_3$ analogously for the Missouri client and Vermont client, respectively. Then the event $E$: Exactly one dies can be written as the union of three intersections:
\[
E = (A_1 \cap A_2 \cap A_3^C) \cup (A_1 \cap A_2^C \cap A_3) \\
\cup (A_1^C \cap A_2 \cap A_3)
\]

Since each of the intersections is mutually exclusive of the other two,

\[
P(E) = P(A_1 \cap A_2 \cap A_3^C) + P(A_1 \cap A_2^C \cap A_3) \\
+ P(A_1^C \cap A_2 \cap A_3)
\]

Furthermore, there is no reason to believe that for all practical purposes the fates of the three are not independent. That being the case, each of the intersection probabilities reduces to a product, and we can write

\[
P(E) = P(A_1) \cdot P(A_2) \cdot P(A_3^C) \\
+ P(A_1) \cdot P(A_2^C) \cdot P(A_3) \\
+ P(A_1^C) \cdot P(A_2) \cdot P(A_3)
\]

\[
= (0.7)(0.9)(0.7) + (0.7)(0.1)(0.3) \\
+ (0.3)(0.9)(0.3)
\]

\[
= 0.543
\]
Comment Declaring events independent for reasons other than those prescribed in Definition 2.5.2 is a necessarily subjective endeavor. Here we might feel fairly certain that a random person dying in Alaska will not affect the survival chances of a random person residing in Missouri (or Vermont). But there may be special circumstances that invalidate that sort of argument. For example, what if the three individuals in question were mercenaries fighting in an African border war and were all crew members assigned to the same helicopter? In practice, all we can do is look at each situation on an individual basis and try to make a reasonable judgment as to whether the occurrence of one event is likely to influence the outcome of another event.

**Practice**

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 listing the sample outcomes, it can be shown that

\[ P(A) = \frac{1}{3}, \quad P(B) = \frac{1}{2}, \quad P(C) = \frac{1}{6} \]
\[ P(A \cap B) = \frac{1}{6}, \quad P(A \cap C) = \frac{1}{18}; \]
\[ P(B \cap C) = \frac{1}{18} \]

and \[ P(A \cap B \cap C) = \frac{1}{36} \] Note that equation 2.5.1 is satisfied,

\[ P(A \cap B \cap C) = \frac{1}{36} = P(A) \cdot P(B) \cdot P(C) \]
\[ = \frac{1}{3} \cdot \frac{1}{2} \cdot \frac{1}{6}. \]

But Equation 2.5.2 is not satisfied since \[ P(B \cap C) = \frac{1}{18} \neq P(B) \cdot P(C). \]

**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 \( P(A \cap B) = 0.2, \quad P(A) = 0.6, \) 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 Events**

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. They are also referred to as Repeated Independent 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 \( P(\text{at least one double six in } n \text{ throws}) \) exceeds 0.50 is \( n=25 \)

**Homework** 2.5.23, 2.5.26, 2.5.27

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 $A_i, i = 1, 2, \cdots, k$ can be performed in $n_i$ ways, $i = 1, 2, \cdots, k$, 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 (A), fourteen entrees (E), six desserts (D), and five beverages (B). How many different meals are possible if a diner intends to order only three courses? (Consider the beverage to be a "course.")

\[ 4 \cdot 14 \cdot 6 + 4 \cdot 6 \cdot 5 + 14 \cdot 6 \cdot 5 + 4 \cdot 14 \cdot 5 = 1156 \]

AED + ADB + EDB + AEB

Example 2.6.1

The combination lock on a briefcase has two dials, each marked off with sixteen notches (see Figure 2.6.2). To open the case, a person first turns the left dial in a certain direction for two revolutions and then stops on a particular mark. The right dial is set in a similar fashion, after having been turned in a certain direction for two revolutions. How many different settings are possible?

In the terminology of the multiplication rule, opening the briefcase corresponds to the four-step sequence \((A_1, A_2, A_3, A_4)\) detailed in Table 2.6.1. Applying the previous corollary, we see that 1024 dif-
ferent settings are possible:

\[
\text{number of different settings} = n_1 \cdot n_2 \cdot n_3 \cdot n_4 \\
= 2 \cdot 16 \cdot 2 \cdot 16 \\
= 1024
\]

**Example 2.6.3**

In 1824 Louis Braille invented what would eventually become the standard alphabet for the blind. Based on an earlier form of night writing used by the French army for reading battlefield communiqués in the dark, Braille’s system replaced each written character with a six-dot matrix:

\[
\ldots \\
\ldots \\
\ldots \\
\ldots
\]

where certain dots were raised, the choice depending on the character being transcribed. The letter e, for example, has two raised dots and is written
Punctuation marks, common words, suffixes, and so on, also have specified dot patterns. In all, how many different characters can be enciphered in Braille? See Figure 2.6.3

Think of the dots as six distinct operations, numbered 1 to 6 (see Figure 2.6.3). In forming a Braille letter, we have two options for each dot: We can raise it or not raise it. The letter e, for example, corresponds to the six-step sequence (raise, do not raise, do not raise, do not raise, raise, do not raise). The number of such sequences, with \( k = 6 \) and \( n_1 = n_2 = \cdots = n_6 = 2 \), is \( 2^6 \), or 64. One of those sixty-four configurations, though, has no raised dots, making it of no use to a blind person. Figure 2.6.4 shows the entire sixty-three-character Braille alphabet.

**Problem-Solving Hints** (Doing combinatorial problems)
Combinatorial questions sometimes call for problem-solving techniques that are not routinely used in other areas of mathematics. The three listed below are especially helpful.

1. Draw a diagram that shows the structure of the outcomes that are being counted. Be sure to include (or indicate) all relevant variations. A case in point is Figure 2.6.3. Almost invariably, diagrams such as these will suggest the formula, or combination of formulas, that should be applied.

2. Use enumerations to test the appropriateness of a formula. Typically, the answer to a combinatorial problem—that is, the number of ways to do something—will be so large that listing all possible outcomes is not feasible. It often is feasible, though, to construct a simple, but analogous, problem for which the entire set of outcomes can be identified (and counted). If the proposed formula does not agree with the simple-case enumeration, we know that our analysis of the original question is incorrect.

3. If the outcomes to be counted fall into struc-
turally different categories, the total number of outcomes will be the sum (not the product) of the number of outcomes in each category. Recall (from Question 2.6.9)

Suggested Practice (NOT COLLECTED)
2.6.1, 2.6.3, 2.6.4, 2.6.14, 2.6.16

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 exam-
ple, 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: $\frac{n}{1} \frac{n-1}{2} \cdots \frac{n-(k-2)}{k-1} \frac{n-(k-1)}{k}$

Position in sequence

See Figure 2.6.6, a tree diagram

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

Example 2.6.7 How many permutations of length $k = 3$ can be formed from the set of $n = 4$ distinct elements, $A, B, C,$ 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.

Example 2.6.12

A new horror movie, Friday the 13th, Part X, will star Jasons great-grandson (also named Jason) as a psychotic trying to dispatch (as gruesomely as possible) eight camp counselors, four men and four women. (a) How many scenarios (i.e., victim orders) can the screen writers devise, assuming they want Jason to do away with all the men before going after any of the women? (b) How many scripts are
possible if the only restriction imposed on Jason is that he save Muffy for last?

a. Suppose the male counselors are denoted A, B, C, and D and the female counselors, W, X, Y, and Z. Among the admissible plots would be the sequence pictured in Figure 2.6.11, where B is done in first, then D, and so on. The men, if they are to be restricted to the first four positions, can still be permuted in $4P_4 = 4!$ ways. The same number of arrangements can be found for the women. Furthermore, the plot in its entirety can be thought of as a two-step sequence: first the men are eliminated, then the women. Since $4!$ ways are available to do the former and $4!$ the latter, the total number of different scripts, by the multiplication rule, is $4!4!$, or 576.

<table>
<thead>
<tr>
<th>Men</th>
<th>Women</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>Y</td>
</tr>
<tr>
<td>D</td>
<td>Z</td>
</tr>
<tr>
<td>A</td>
<td>W</td>
</tr>
<tr>
<td>C</td>
<td>X</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>4</td>
<td>8</td>
</tr>
</tbody>
</table>

Order of killing

Figure 2.6.11
b. If the only condition to be met is that Muffy be dealt with last, the number of admissible scripts is simply \(7P7 = 7!\), that being the number of ways to permute the other seven counselors (see Figure 2.6.12).

B W Z C Y A D Muffy
1 2 3 4 5 6 7 8
Order of killing
Figure 2.6.12

**Example 2.6.13**

Consider the set of nine-digit numbers that can be formed by rearranging without repetition the integers 1 through 9. For how many of those permutations will the 1 and the 2 precede the 3 and the 4? That is, we want to count sequences like 7 2 5 1 3 6 9 4 8 but not like 6 8 1 5 4 2 7 3 9.

At first glance, this seems to be a problem well beyond the scope of Theorem 2.6.1. With the help of a symmetry argument, though, its solution is surprisingly simple.

Think of just the digits 1 through 4. By the corollary on p. 74, those four numbers give rise to
4!(=24) permutations. Of those twenty-four, only four \((1, 2, 3, 4), (2, 1, 3, 4), (1, 2, 4, 3),\) and \((2, 1, 4, 3)\) have the property that the 1 and the 2 come before the 3 and the 4. It follows that \(\frac{4}{24}\) of the total number of nine-digit permutations should satisfy the condition being imposed on 1, 2, 3, and 4. Therefore, number of permutations where 1 and 2 precede 3 and 4 = \(\frac{4}{24} \cdot 9! = 60,480\)

**Comment** Computing \(n!\) can be quite cumbersome, even for \(n\)'s that are fairly small: We saw in Example 2.6.9, for instance, that 16! is already in the trillions. Fortunately, an easy-to-use approximation is available. According to Stirling’s formula, \(n! = \sqrt{2\pi n} n^{n+1/2} e^{-n}\)

In practice, we apply Stirling’s formula by writing \(\log_{10}(n!) = \log_{10}(\sqrt{2\pi}) + (n + \frac{1}{2}) + n \log_{10}(e)\)

**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?

\(4P_4\) and \(2P_2 \cdot 2P_2\)

More Practice : 2.6.22, 2.6.23, 2.6.27

Counting Permutations (when the objects are not all distinct)

Permutation is an ordered arrangement of the numbers, terms, etc., of a set into specified groups

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 \(C:\)

\(ABC\)
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:

- AAB
- ABA
- BAA

**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

\[
\frac{n!}{n_1!n_2!n_3!}
\]

As we will see, there are many real-world applications where the \( n \) objects to be permuted belong to \( r \) different categories, each category containing one or more identical objects.

**Theorem 2.6.1** The number of ways to arrange \( n \) objects, \( n_1 \) being of one kind, \( n_2 \) of a second kind, ..., and \( n_r \), of an \( r \)th kind, is

\[
\frac{n!}{n_1!n_2!...n_r!} \quad \text{where} \quad \sum_{i=1}^{r} n_i = n.
\]
Comment  Ratios like $\frac{n!}{n_1! \cdot n_2! \cdots n_r!}$ are all called multinomial coefficients because the general term in the expansion of $(x_1 + x_2 + \ldots + x_r)^n$ is
\[
\frac{n!}{n_1! \cdot n_2! \cdots n_r!} x_1^{n_1} x_2^{n_2} \cdots x_r^{n_r}
\]

**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?

Quarter  Dime  Dime  Quarter  Nickel  Dime
1        2        3        4        5        6
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 2.6.16**

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?

**MORE PRACTICE**

2.6.36, 2.6.40, 2.6.41, 2.6.42

**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 \textit{binomial coefficients}.

**EXAMPLE 2.6.20**

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 2.6.21**
A chemist is trying to synthesize a part of a straight-chain aliphatic hydrocarbon polymer that consists of twenty-one radicals ten ethyls (E), six methyls (M), and five propyls (P). Assuming all arrangements of radicals are physically possible, how many different polymers can be formed if no two of the methyl radicals are to be adjacent? Imagine arranging the Es and the Ps without the Ms. Figure 2.6.15 shows one such possibility. Consider the sixteen spaces between and outside the Es and Ps as indicated by the arrows in Figure 2.6.15. In order for the Ms to be nonadjacent, they must occupy any six of these locations. But those six spaces can be chosen in \(\binom{16}{6}\) ways. And for each of the positionings of the Ms, the Es and Ps can be permuted in \(\frac{15!}{10!5!}\) ways (Theorem 2.6.2).
So, by the multiplication rule, the total number of polymers having nonadjacent methyl radicals is 24,048,024:

\[
\left(\begin{array}{c}16 \\ 6 \end{array}\right) \cdot \frac{15!}{10!5!} = \frac{16!}{10!6!} \cdot \frac{15!}{10!5!} = (8008)(3003) = 24,048,024
\]

**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 expansion 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:

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

\(k\) will successively take on the values 0 through 5.

Therefore the binomial theorem is

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

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

**FIGURE 2.6.16**

**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 \( \binom{5}{2} = 10 \)

**Example 2.6.23** The answers to combinatorial questions can sometimes be obtained using quite different approaches. What invariably distinguishes one solution from another is the way in which outcomes are characterized. For example, suppose you have just ordered a roast beef sub at a sandwich shop, and now you need to decide which, if any,
of the available toppings (lettuce, tomato, onions, etc.) to add. If the shop has eight extras to choose from, how many different subs can you order? One way to answer this question is to think of each sub as an ordered sequence of length eight, where each position in the sequence corresponds to one of the toppings. At each of those positions, you have two choices—add or do not add that particular topping. Pictured in Figure 2.6.17 is the sequence corresponding to the sub that has lettuce, tomato, and onion but no other toppings. Since two choices (add or do not add) are available for each of the eight toppings, the multiplication rule tells us that the number of different roast beef subs that could be requested is $2^8$, or 256.

An ordered sequence of length eight, though, is not the only model capable of characterizing a roast beef sandwich. We can also distinguish one roast beef sub from another by the particular combination of toppings that each one has. For example, there are

$$\binom{8}{4} = 70$$

different subs having exactly four toppings. It follows that the total number of different
sandwiches is the total number of different combinations of size \( k \), where \( k \) ranges from 0 to 8. Reassuringly, that sum agrees with the ordered sequence answer:

\[
\text{total number of different roast beef subs} = \binom{8}{0} + \binom{8}{1} + \binom{8}{2} + \cdots + \binom{8}{8} \\
= 1 + 8 + 28 + \cdots + 1 \\
= 256
\]

What we have just illustrated here is another property of binomial coefficients namely, that

\[
\sum_{k=0}^{n} \binom{n}{k} = 2^n \quad (2.6.2)
\]

The proof of Equation 2.6.2 is a direct consequence of Newtons binomial expansion (see the second comment following Theorem 2.6.3).

**More practice**

2.6.51, 2.6.53, 2.6.54, 2.6.55
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 revelent.

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 $\binom{3n}{2}$. Moreover,

$$P(A) = P(2 \text{ red } \cup 2 \text{ whites } \cup 2 \text{ blues})$$
$$= P(2 \text{ red }) + P(2 \text{ whites }) + P(2 \text{ blues})$$
$$= \frac{3\binom{n}{2}}{\binom{3n}{2}}$$

and

$$P(B) = P( \text{ two } 1 \text{'s } \cup \text{ two } 2 \text{'s } \cup \cdots \cup \text{ two } n \text{'s})$$
$$= n\left(\frac{3}{2}\right) / \left(\frac{3n}{2}\right).$$
Therefore

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

**Example 2.7.3**

Twelve fair dice are rolled. What is the probability that

a. the first six dice all show one face and the last six dice all show a second face?

b. not all the faces are the same?

c. each face appears exactly twice?

a. The sample space that corresponds to the experiment of rolling twelve dice is the set of ordered sequences of length twelve, where the outcome at every position in the sequence is one of the integers 1 through 6. If the dice are fair, all \(6^{12}\) such sequences are equally likely.

Let \(A\) be the set of rolls where the first six dice show one face and the second six show another face. Figure 2.7.3 shows one of the sequences in the event
A. Clearly, the face that appears for the first half of the sequence could be any of the six integers from 1 through 6.

<table>
<thead>
<tr>
<th>Faces</th>
<th>2</th>
<th>2</th>
<th>2</th>
<th>2</th>
<th>2</th>
<th>4</th>
<th>4</th>
<th>4</th>
<th>4</th>
<th>4</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
</tr>
</tbody>
</table>

Position in sequence

Figure 2.7.3

Five choices would be available for the last half of the sequence (since the two faces cannot be the same). The number of sequences in the event A, then, is \( _6P_2 = 6 \cdot 5 = 30 \). Applying the \( m/n \) rule gives

\[
P(A) = \frac{30}{6^{12}} = 1.4 \times 10^{-8}
\]

b. Let B be the event that not all the faces are the same. Then \( P(B) = 1 - P(B^C) = 1 - (6/6^{12}) \) since there are six sequences

\[
(1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,)
\]

\[
, \cdots , (6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6,)
\]

where the twelve faces are all the same.

c. Let C be the event that each face appears exactly twice. From Theorem 2.6.2, the number of ways each face can appear exactly twice is

\[
\frac{12!}{(2! \cdot 2! \cdot 2! \cdot 2! \cdot 2! \cdot 2!)}
\]

Therefore,
\[ P(C) = \frac{12!/(2! \cdot 2! \cdot 2! \cdot 2! \cdot 2! \cdot 2!)}{6^{12}} = 0.0034 \]

**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}, \( n(S) = \binom{10}{4} \)

\[
P(A \cap B) = \frac{n(A \cap B)}{n(S)} = \frac{\binom{7}{2} \binom{3}{2}}{\binom{10}{4}}
\]

**Practice:**

2.7.2, 2.7.3, 2.7.7, 2.7.14,
Chapter 3

Random Variables

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).

i.e. Roll a die: $S=\{1,2,3,4,5,6\}$ ; $N(S)=6=n$
P(getting a 1) = P(getting a 2) = \cdots = P(getting a 6) = \frac{1}{6} = \frac{1}{n}

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 Random Variables 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

(\text{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, no, no, yes, no, no, yes, no).

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 \textit{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 \textit{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 \textit{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>
<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>

Table 3.2.1

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 $k = 0, 1, 2, \text{ or } 3$.

<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,
\[ P(\text{k heads}) = \text{ (number of ways to arrange k heads and n - k tails)} \]
\[ \cdot \text{(probability of any particular sequence having k heads and n - k tails)} \]
\[ = \text{ (number of ways to arrange k heads and n - k tails)} \]
\[ \cdot p^k(1 - p)^{n-k} \]

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

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

**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 one 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 \( \left(\frac{99}{100,000}\right)^{10,000} = 0.905 \). Thus, the probability that \( k \geq 1 \) is \( 1 - 0.905 = 0.095 \).

Practice: 3.2.2, 3.2.4, 3.2.5, 3.2.8

Practice 3.2.11
If a family has four children, is it more likely they will have two boys and two girls or three of one sex and one of the other? Assume that the probability of a child being a boy is 1/2 and that the births are independent events.

0.375 and 0.5

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:

\[
P(\text{k reds drawn}) = \binom{n}{k}(r/N)^k(1 - r/N)^{n-k},
\]

\( k = 0, 1, 2, \ldots, n \)

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

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

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 sub-
sets. 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?

$$P(2A \cap 2A^C) = \frac{\binom{4}{2}\binom{8}{3}}{\binom{12}{5}} = \frac{14}{33}$$
3.3 Discrete Random Variables

The binomial and hypergeometric distributions described in Section 3.2 are special cases of some important general concepts that we want to explore more fully in this section. Previously in Chapter 2, we studied in depth the situation where every point in a sample space is equally likely to occur (recall Section 2.6).

How to assign probabilities to outcomes that are not binomial or hypergeometric is one of the major questions investigated in this chapter.

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 \textit{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 appears on the $s$th 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, $p(s) = (1/2)^s, s = 1, 2, \cdots$

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
\[ p(s) = \frac{1}{1+\lambda} \left( \frac{\lambda}{1+\lambda} \right)^s \], \( s = 0, 1, 2, \ldots ; \lambda > 0 \)
a discrete probability function? Why or why not?
A simple inspection \( p(s) \geq 0 \) for all \( s \)
\[
\sum_{s \in S} p(s) = \sum_{s=0}^{\infty} \frac{1}{1+\lambda} \left( \frac{\lambda}{1+\lambda} \right)^s
\]

\[
= \frac{1}{1+\lambda} \left( \frac{1}{1 - \frac{\lambda}{1+\lambda}} \right)
\]

\[
= \frac{1}{1+\lambda} \cdot \frac{1+\lambda}{1} = 1
\]

**Defining ”New” sample Spaces**

We have seen how the function \( p(s) \) associates a probability with each outcome, \( s \), in a sample space. Related is the key idea that outcomes can often be grouped or reconfigured in ways that may facilitate problem solving.

The function that replaces the outcome \( s \) with the numerical value is called a random variable.

**Definition 3.3.2** A function whose domain is a sample space \( S \) and whose values form a finite or countably infinite set of real numbers is called dis-
crete 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

$$p_X(k) = P\left(\{s \in S|X(s) = k\}\right)$$

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
\begin{align*}
p_X(k) = P(X = k) &= \binom{n}{k} p^k (1 - p)^{n-k}, \\
&\quad k = 0, 1, 2 \cdots, n
\end{align*}

**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 \( X(s, f, s) = 2(10000) - 1(2000) = \)
Moreover, the R.V $X$ equals $18,000 whenever 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>
<thead>
<tr>
<th>Table 3.3.4</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. Defective</td>
</tr>
<tr>
<td>0</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</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

\[ P(Y = y) = P(2X + 1 = y) = P(X = \frac{y - 1}{2}) \]
\[ = p_X \left( \frac{y - 1}{2} \right) \]
\[ = \binom{n}{\frac{y - 1}{2}} p^{\frac{(y-1)}{2}} (1 - p)^{n-\frac{(y-1)}{2}} \]
\[ = \binom{4}{\frac{y - 1}{2}} \left( \frac{2}{3} \right)^{\frac{y-1}{2}} \left( 1 - \left( \frac{2}{3} \right) \right)^{4-\frac{(y-1)}{2}} \]
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 $X = \text{larger no. drawn}$ 1, \\

<table>
<thead>
<tr>
<th>Outcome</th>
<th>$X = \text{larger no. drawn}$</th>
</tr>
</thead>
<tbody>
<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>

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>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Outcome</th>
<th>$X = \text{larger no. drawn}$</th>
<th>sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,2</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>1,3</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>1,4</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>1,5</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>2,3</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>2,4</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>2,5</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>3,4</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>3,5</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>4,5</td>
<td>5</td>
<td>9</td>
</tr>
</tbody>
</table>
3.4 Continuous random Variables

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.

Rolling a pair of dice and recording the faces that appear is an experiment with a discrete sample space; choosing a number at random from the interval $[0, 1]$ would have a continuous sample space.

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

Homework: 3.3.4, 3.3.5, 3.3.13, 3.3.14
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).

If a random variable X is defined on the sample space, the probabilities associated with its outcomes are assigned by the probability density function $p_X(X = k)$.

This will not work when S is continuous.

The fact that a continuous sample space has an uncountably infinite number of outcomes eliminates the option of assigning a probability to each point as we did in the discrete case with the function p(s).

We begin this section with a particular pdf de-
fined on a discrete sample space that suggests how we might define probabilities, in general, on a continuous sample space.

Suppose an electronic surveillance monitor is turned on briefly at the beginning of every hour and has a 0.905 probability of working properly, regardless of how long it has remained in service.

If we let the random variable $X$ denote the hour at which the monitor first fails, then $p_X(k)$ is the product of $k$ individual probabilities:

$$p_X(k) = P(X = k) = P(\text{Monitor fails for the first time at the } k\text{th hour}) = P(\text{Monitor functions properly for first } k - 1 \text{ hours} \cap \text{Monitor fails at the } k\text{th hour}) = (0.905)^{k-1}(0.095), \ k = 1, 2, 3, \ldots$$

Figure 3.4.1 shows a probability histogram of $p_X(k)$
for \( k \) values ranging from 1 to 21. Here the height of the \( k \)th bar is \( p_X(k) \), and since the width of each bar is 1, the area of the \( k \)th bar is also \( p_X(k) \).

Now, look at Figure 3.4.2, where the exponential curve \( y = 0.1e^{-0.1x} \) is superimposed on the graph of \( p_X(k) \). Notice how closely the area under the curve approximates the area of the bars. It follows that the probability that \( X \) lies in some given interval will be numerically similar to the integral of the exponential curve above that same interval.

For example, the probability that the monitor fails sometime during the first four hours would be the sum

\[
P(0 \leq X \leq 4) = \sum_{k=0}^{4} p_X(k)
\]

\[
= \sum_{k=0}^{4} (0.905)^{k-1}(0.095)
\]

\[
= 0.3297
\]

To four decimal places, the corresponding area
under the exponential curve is the same:

\[ \int_0^4 0.1e^{0.1x} \, dx = 0.3297 \]

Implicit in the similarity here between \( p_X(k) \) and the exponential curve \( y = 0.1e^{-0.1x} \) is our sought-after alternative to \( p(s) \) for continuous sample spaces.

Instead of defining probabilities for individual points, we will define probabilities for intervals of points, and those probabilities will be areas under the graph of some function (such as \( y = 0.1e^{-0.1x} \), where the shape of the function will reflect the desired probability measure to be associated with the sample space.

**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 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 $f(t)$

We have seen that the probability structure of any sample space with a finite or countably infinite number of outcomes is defined by the function $p(s) = P(\text{Outcome is } s)$. The function $f(t)$ serves an analogous purpose. Specifically, $f(t)$ defines the probability structure of $S$ in the sense that the probability of any interval in the sample space is the integral of $f(t)$.

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

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

See figure 3.4.3

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

Notice that the shape of $f(t)$ (see Figure 3.4.4) implies that outcomes close to 1 are more likely to occur than are outcomes close to 0. For example,
\[ P([0, 1/3]) = \int_0^{1/3} 3t^2 \, dt = t^3 \bigg|_0^{1/3} = 1 - 8/27 = 19/27 \]

while

\[ P([2/3, 1]) = \int_{2/3}^{1} 3t^2 \, dt = t^3 \bigg|_{2/3}^{1} = 1 - 8/27 = 19/27 \]

**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

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

**Fitting \( f(t) \) to Data: The Density-Scaled Histogram**

How to determine which \( f(t) \) to use for certain data set?
Create histogram for the data. Specifically probability density histogram.

Unlike frequency histogram (refer to figure 3.4.6 and figure 3.4.7.) to make sure that area under the histogram is 1.

**Continuous Probability Density Functions (pdf)**

**Definition 3.4.2**

Let $Y$ be a function from a sample space $S$ to the real numbers. The function $Y$ is called a continuous random variable if there exists a function $f_Y(y)$ such that for any real numbers $a$ and $b$, with $a < b$

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

The function $f_Y(y)$ is the probability density function (pdf) for $Y$. As in the discrete case, the cumulative distribution function (cdf) is defined by

$$F_Y(y) = \Pr(Y \leq y)$$

The cdf in the continuous case is just an integral of $f_Y(y)$, that is, $F_Y(y) = \int_0^y f_Y(t) \, dt$
Let \( f(y) \) be an arbitrary real-valued function defined on some subset \( S \) of the real numbers. If
1. \( f(y) \geq 0 \) for all \( y \) in \( S \) and
2. \( \int_S f_Y(y) \, dy = 1 \)

then \( f(y) = f_Y(y) \) for all \( y \), where the random variable \( Y \) is the identity mapping.

**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
\[
\int_0^1 6y(1-y) \, dy = 6\left[\frac{y^2}{2} - \frac{y^3}{3}\right]_0^1 = 1
\]

**Continuous Cumulative Distribution functions**

**Definition 3.4.3.** The cdf for a continuous
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**

**More Practices**
3.4.2, 3.4.4, 3.4.5, 3.4.6, 3.4.8, 3.4.11, 3.4.12, 3.4.16, 3.4.17

**Practice 3.4.5.**

The length of time, \( Y \), that a customer spends in line at a bank teller's window before being served is described by the exponential pdf

\[ f_Y(y) = 0.2e^{-0.2y}, \quad y > 0. \]

(a) What is the probability that a customer will wait more than ten minutes?
\[
P(Y > 10) = \int_{10}^{+\infty} 0.2e^{-0.2y} \, dy
\]
\[
= \frac{0.2e^{-0.2y}}{-0.2} \bigg|_{10}^{+\infty}
\]
\[
= \lim_{b \to +\infty} e^{-0.2y} \bigg|_{y=10}^{y=b}
\]
\[
= \lim_{b \to +\infty} \frac{1}{eb} - (-\frac{1}{e^5})
\]
\[
= \frac{1}{e^5} = 0.135
\]

(b) Suppose the customer will leave if the wait is more than ten minutes. Assume that the customer goes to the bank twice next month. Let the random variable \(X\) be the number of times the customer leaves without being served. Calculate \(p_X(1)\).

Let \(A = \text{Probability customers leaves on first trip}\) and \(B = \text{Probability customers leaves on second trip}\).

\[
P(A) = P(B) = 0.135
\]

\[
P_X(1) = P(A)P(B^C) + P(A^C)P(B)
\]
\[
= 2(0.135)(0.865) = 0.23355
\]
3.5 Expected Values

measure of central tendency - mean - Expected value 
\( \mu_x \) or \( \mu_Y \)

**Definition 3.5.1**

Let \( X \) be a discrete random variable with probability function \( p_X(k) \). The expected value of \( X \) is denoted as \( E(X) \) (or \( \mu \) or \( \mu_X \)) is given by

\[
E(X) = \mu = \mu_X = \sum_{all \ x} x \cdot p_X(x)
\]

Similarly if \( Y \) be a continuous random variable with probability function \( f_Y(y) \).

\[
E(Y) = \mu = \mu_Y = \int_{-\infty}^{\infty} y \cdot f_Y(y) \, dy
\]

**Comment:** We assume that both the sum and the integral in Definition 3.5.1 converges absolutely
\[
\sum_{all \ x} |x| p_X(x) < \infty, \quad \int_{-\infty}^{\infty} |y| \cdot f_Y(y) \, dy < \infty
\]

**Theorem 3.5.1** Suppose \( X \) is a binomial R.V. with parameter \( n \) and \( p \). then \( E(X) = np \).

Proof According to Definition 3.5.1, \( E(X) \) for a binomial random variable is the sum

\[
E(X) = \sum_{x=0}^{n} x \cdot p_X(x)
\]

\[
= \sum_{x=0}^{n} x \binom{n}{x} p^x (1 - p)^{n-x}
\]

\[
= \sum_{x=0}^{n} \frac{x \cdot n!}{x!(n-x)!} p^x (1 - p)^{n-x}
\]

\[
= \sum_{x=1}^{n} \frac{n!}{(x-1)!(n-x)!} p^x (1 - p)^{n-x}
\]

\[
= np \sum_{x=1}^{n} \frac{(n-1)!}{(x-1)!(n-x)!} p^{x-1}(1 - p)^{n-x}
\]
At this point, a trick is called for. If $E(X) = \sum_{\text{all } k} g(x)$ can be factored in such a way that $E(X) = h \sum_{\text{all } k} p_X^*(x)$, where $p_X^*(x)$ is the pdf for some random variable $X$, then $E(X) = h$, since the sum of a pdf over its entire range is 1. Here, suppose that $np$ is factored out of Equation 3.5.2. Then

$$E(X) = np \sum_{x=1}^{n} \binom{n-1}{x-1} p^{x-1} (1 - p)^{n-x}$$

Now, let $j = x - 1$. It follows that

$$E(X) = np \sum_{j=0}^{n-1} \binom{n-1}{j} p^{j} (1 - p)^{(n-1)-j}$$

Finally, letting $m = n - 1$ gives

$$E(X) = np \sum_{j=0}^{m} \binom{m}{j} p^{j} (1 - p)^{m-j}$$

and, since the value of the sum is 1 (why?), $E(X) = np$
Comment: The statement of Theorem 3.5.1 make sense since for example if a multiple choice test have 100 questions, each with five possible answer, we would expect to get twenty correct, just by guessing. \( E(X) = np = 100 \cdot \left( \frac{1}{5} \right) \)

**Theorem 3.5.2** suppose \( X \) is a hypergeometric R.V. with parameters \( r, w, \) and \( n \). That is suppose an urn contains \( r \) red balls and \( w \) white balls. A sample of size \( n \) is drawn simultaneously from the urn. Let \( X \) be the number of red balls in the sample. Then

\[
E(X) = n \cdot \frac{r}{r+w}
\]

**Example 3.5.6**

The distance, \( Y \), that a molecule in a gas travels before colliding with another molecule can be modeled by the exponential pdf

\[
f_Y(y) = \frac{1}{\mu} e^{-y/\mu} \quad y \geq 0
\]

where \( \mu \) is a positive constant known as mean free
path. Find $E(Y)$.

Since the R.V. here is continuous, its expected value is

$$E(Y) = \int_{0}^{\infty} y \frac{1}{\mu} \frac{-y}{\mu} dy$$

Integrating by parts give $E(Y) = \mu$

The above Equation shows that $\mu$ is aptly named it does, in fact, represent the average distance a molecule travels, free of any collisions. Nitrogen ($N_2$), for example, at room temperature and standard atmospheric pressure has $\mu = 0.00005$ cm. An $N_2$ molecule, then, travels that far before colliding with another $N_2$ molecule, on the average.

**Example 3.5.7**

One continuous distribution that has a number of interesting applications in Physics is the Raleigh distribution, where the pdf is given by

$$f_Y(y) = \frac{y}{a^2} e^{-\left(\frac{y^2}{2a^2}\right)} \quad a > 0; \quad 0 \leq y < \infty$$

Calculate $E(Y)$. 
\[ E(Y) = \int_{0}^{\infty} y \cdot \frac{y}{a^2} e^{-\left(\frac{y^2}{2a^2}\right)} dy. \]

Letting \( \nu = \frac{y}{\sqrt{2}a} \), \( \sqrt{2}a d\nu = dy \). Then

\[ E(Y) = 2\sqrt{2}a \int_{0}^{\infty} \nu^2 \cdot e^{-\left(\nu^2\right)} d\nu \]

\[ = 2\sqrt{2}a \frac{1}{4} \sqrt{\pi} = a \sqrt{\pi/2} \]

This is due to \( \int_{0}^{\infty} \nu^2 \cdot e^{-\left(\nu^2\right)} d\nu \) is a special case of the general form \( \nu^{2k} \cdot e^{-\left(\nu^2\right)} \). For \( k = 1 \)

\[ \int_{0}^{\infty} \nu^{2k} \cdot e^{-\left(\nu^2\right)} d\nu \]

\[ = \int_{0}^{\infty} \nu^2 \cdot e^{-\left(\nu^2\right)} d\nu \]

\[ = \frac{1}{4} \sqrt{\pi} \]

Let \( u = \nu^2, du = 2\nu d\nu \) we have
\[
\int_0^\infty ue^{-u} \frac{du}{2\sqrt{u}} = \frac{1}{2} \int_0^\infty u^{1/2} e^{-u} du
\]

Define Gamma function \( \Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt \)
\( \Gamma(1) = 1 \)
\( \Gamma(x + 1) = x\Gamma(x) \)
\( \Gamma(1/2 + 1) = 1/2\Gamma(1/2) = 1/2\sqrt{\pi} \)
therefore
\[
\frac{1}{2} \int_0^\infty u^{1/2} e^{-u} du = \frac{\sqrt{\pi}}{4}
\]

A second Measure of Central Tendency: The Median

**Definition 3.5.2.** If \( X \) is a discrete R.V, the median, \( m \), is that point for which \( P(X < m) = P(X > m) \). In the event that \( P(X \leq m) = 0.5 \) and \( P(X \geq m') = 0.5 \), the median is the arithmetic average, \( (m+m')/2 \). If \( y \) is a continuous R.V., its median is the solution to the integral equation \( \int_{-\infty}^{m'} f_Y(y) dy = 0.5 \).
Example 3.5.8 If a random variable’s pdf is symmetric, both $\mu$ and $m$ will be equal. Should $p_X(k)$ and $f_Y(y)$ not be symmetric, though, the difference between the expected value and the median can be considerable, especially if the symmetry takes the form of extreme skewness. The situation described here is a case in point.

Soft glow makes a 60-watt light bulb that is advertised to have an average life of one thousand hours. Assuming that the performance claim is valid, is it reasonable for consumer to conclude that soft glow bulb they bought will last for approximately one thousand hours?

No! If the average life of a bulb is one thousand hours, the pdf $f_Y(y)$, modeling the length of time $Y$, that it remains lit before burning out is likely to have the form

$$f_Y(y) = 0.001e^{-0.001y}, \quad y > 0. \quad (3.5.1)$$

Equation 3.5.1 is very skewed pdf. The median is to the left of the mean.
The median is the solution to equation
\[ \int_{0}^{m} 0.001 e^{-0.001y} = 0.5 \]
where \( m = 693 \)

So even though the average life of these bulb is one thousand hours, there is a 50% chance that the one that you buy will last 693 hours.

**Homework:** 3.5.8, 3.5.10, 3.5.12, 3.5.16, 3.5.27

**The Expected Value of a Function of a Random Variable**

**Theorem 3.5.3.** Suppose \( X \) is a discrete random variable with pdf \( p_X(x) \). Let \( g(X) \) be a function of \( X \). Then the expected value of the random variable \( g(X) \) is given by

\[
E[g(X)] = \sum_{\text{all } x} g(x) \cdot p_X(x)
\]

provided that \( \sum_{\text{all } x} |g(x)|p_X(x) < \infty \).

If \( Y \) is a continuous R.V. with pdf \( f_Y(y) \), and if \( g(Y) \) is a continuous function, the expected value
of the R.V $g(Y)$ is

$$E[g(Y)] = \int_{-\infty}^{\infty} g(y) \cdot f_Y(y) dy,$$

provided that $\int_{-\infty}^{\infty} |g(y)| \cdot f_Y(y) dy < \infty$.

**Corollary.** For any R.V $W$, $E(aW + b) = aE(W) + b$ where $a$ and $b$ are constants.

Proof. Suppose $W$ is continuous; the proof for the discrete case is similar. By theorem 3.5.3, $E(aW + b) = \int_{-\infty}^{\infty} (aw + b) f_W(w) dw$, but the latter can be written as $a \int_{-\infty}^{\infty} w \cdot f_W(w) dw + b \int_{-\infty}^{\infty} f_W(w) dw = aE(W) + b \cdot 1 = aE(W) + b$

**Example 3.5.10**

suppose the amount of propellant, $Y$, put into a can of spray paint is a random variable with pdf

$$f_Y(y) = 3y^2, \quad 0 < y < 1$$

Experience have shown that the largest surface area that can be painted by a can having $Y$ amount of propellant is twenty times the area of a circle generated by a radius of $Y$ ft. If the Purple Dominoes, a
newly formed urban group, have just bought their first can of spray paint, can they expect to have enough to cover a $5' \times 8'$ subway panel?

No. By assumption, the maximum area (in $ft^2$) that can be covered by a can of paint is described by the function

$$\text{area} = g(Y) = 20\pi Y^2$$

According to the second statement in Theorem 3.5.3, though, the average value for $g(Y)$ is slightly less than the desired $40 ft^2$:

$$E[g(Y)] = \int_0^1 20\pi y^2 \cdot 3y^2 dy = \frac{60\pi y^5}{5} \mid_0^1 = 12\pi = 37.7 ft^2$$

**Example 3.5.11**

A fair coin is tossed until a head appears. You will be given $\left(\frac{1}{2}\right)^k$ dollars if that first head occurs on the $k^{th}$ toss. How much money can you expect to be paid? Let the random variable $X$ denote the
toss at which the first head appears. Then

\[ p_X(x) = P(X = x) \]

\[ = P(1 \text{st } k-1 \text{ tosses are tails and } k\text{th toss is a head}) \]

\[ = \left(\frac{1}{2}\right)^{x-1} \frac{1}{2} \]

\[ = \left(\frac{1}{2}\right)^x \]
More over,

\[ E(\text{ Amount won}) = E \left[ \left( \frac{1}{2} \right)^X \right] \]

\[ = E[g(X)] \]

\[ = \sum_{\text{all } x} g(x) \cdot p_X(x) \]

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

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

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

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

\[ = \frac{1}{1 - \frac{1}{4}} - 1 = \$0.33 \]

Practice 3.5.28
3.6 The Variance

Dispersion of the distribution.

**Definition 3.6.1** The variance of a random variable is the expected value of its squared deviations from $\mu$. If $X$ is discrete with pdf $p_X(x)$,

$$\text{Var}(X) = \sigma^2 = E[(X-\mu)^2] = \sum_{\text{all } x} (x-\mu)^2 \cdot p_X(x).$$

If $Y$ is continuous with pdf $f_Y(y)$,

$$\text{Var}(Y) = \sigma^2 = E[(Y-\mu)^2] = \int_{-\infty}^{\infty} (y-\mu)^2 \cdot f_Y(y) \, dy.$$  

If $E(X^2)$ or $E(Y^2)$ is not finite, the variance is not defined.

**Comment:** The unit of the variance is the square of the unit of the R.V. For application the square
root of the variance namely standard deviation can be used to measure the dispersion. It has the same unit as the R.V.

\[
\sigma = \begin{cases} 
\sqrt{\sum_{\text{all } x} (x - \mu)^2 \cdot p_X(x)} & \text{if } X \text{ is discrete} \\
\sqrt{\int_{-\infty}^{\infty} (y - \mu)^2 \cdot f_Y(y) \, dy} & \text{if } Y \text{ is continuous}
\end{cases}
\]

**Theorem 3.6.1**

Let \( W \) be any R.V, discrete or continuous, having mean \( \mu \) and for which \( E(W^2) \) is finite. The

\[
\text{Var}(W) = \sigma^2 = E(W^2) - \mu^2
\]

**Proof.** We will proof the theorem for the continuous case. The argument for discrete \( W \) is similar. In theorem 3.5.3, let \( g(W) = (W - \mu)^2 \). Then

\[
\text{Var}(W) = E[(W - \mu)^2] = \int_{-\infty}^{\infty} g(w) f_W(w) \, dw
\]

\[
= \int_{-\infty}^{\infty} (w - \mu)^2 f_W(w) \, dw
\]

Squaring out the term \((w - \mu)^2\) that appear in the integrand and using the additive property of inte-
grals give
\[
\int_{-\infty}^{\infty} g(w) f_W(w) \, dw = E(W^2) - \mu^2
\]

**Read Example 3.6.1 for Variance of hypergeometric distribution**

**Theorem 3.6.2**

Let $W$ be any R.V, discrete or continuous, having mean $\mu$ and for which $E(W^2)$ is finite. Then $\text{Var}[(aW + b)] = a^2 \text{Var}(W)$.

Proof. Using the same approach taken in the proof of theorem 3.6.1, it can be shown that $E[(aW + b)^2] = a^2 E(W^2) + 2ab\mu + b^2$. We also know from corrollary to Theorem 3.5.3 that $E(aW + b) = a\mu + b$. Using Theorem 3.6.1, then we can write

\[
\text{Var}(aW + b) = E[(aW + b)^2] - [E(aW + b)]^2
\]

\[
= [a^2 E(W^2) + 2ab\mu + b^2] - [a\mu + b]^2
\]

\[
= [a^2 E(W^2) + 2ab\mu + b^2] - [a^2\mu^2 + 2ab\mu + b^2]
\]

\[
= a^2[E(W^2) - \mu^2] = a^2 \text{Var}(W)
\]
Example 3.6.2
A R.V $Y$ is described by the pdf
\[ f_Y(y) = 2y, \quad 0 < y < 1 \]
What is the standard deviation of $3Y + 2$?
First, we need to find the variance of $Y$.

\[ E(Y) = \int_0^1 y \cdot 2ydy = \frac{2}{3} \]
and
\[ E(Y^2) = \int_0^1 y^2 \cdot 2ydy = \frac{1}{2} \]
So
\[ \text{Var}(Y) = E(Y^2) - \mu^2 = \frac{1}{2} - \left( \frac{2}{3} \right)^2 = \frac{1}{18} \]
Then by Theorem 3.6.2
\[ \text{Var}(3Y + 2) = (3)^2 \cdot \text{Var}(Y) = 9 \cdot \frac{1}{8} = \frac{1}{2} \]
which makes the standard deviation of $3Y + 2$ equals to $\sqrt{\frac{1}{2}}$ or 0.71.

Practice: 3.6.4: Compute the variance for a uniform random variable defined on the unit interval.
\( f_Y(y) = 1 \quad 0 \leq y \leq 1 \)

\[
\mu = \int_0^1 y \cdot 1 \, dy = 1/2
\]

\[
E(Y^2) = \int_0^1 y^2 \cdot 1 \, dy = 1/3.
\]

\[
\text{Var}(Y) = 1/3 - (1/2)^2 = 1/12
\]

Practice: 3.6.2, 3.6.6, 3.6.8, 3.6.11, 3.6.14

**Higher Moments.**

\( E(W) \) (measure of central tendency - location measure) is the first moment about the origin

\[
\sigma^2 = E(W^2) - [E(W)]^2 \quad \text{(measure of dispersion)}
\]

the second moment about the mean

How about skewness of the distribution? Can use third moment.

**Definition 3.6.2.** Let \( W \) be any random vari-
able with pdf $f_W(w)$. For any positive integer $r$,

1. The $r$th moment of $W$ about the origin, $\mu_r$, is given by $\mu_r = E(W^r)$, provided that $\int_{-\infty}^{\infty} |w|^r \cdot f_W(w)dw < \infty$ or provided that the analogous on the summation holds if $W$ is discrete. When $r = 1$, we usually drop the subscript and write $E(W) = \mu$ rather than $\mu_1$.

2. The $r$th moment of $W$ about the mean, $\mu'_r$, is given by $\mu'_r = E[(W - \mu)^r]$, provided the finiteness conditions of part 1 hold.

Comment. We can express $\mu'_r$ in terms of $\mu_j$, $j = 1, 2, \ldots, r$ by simply writing out the binomial expansion of $(W - \mu)^r$:

Recall binomial expansion

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

then

$$(W - \mu)^r = (W + (-\mu))^r = \sum_{j=0}^{r} \binom{r}{j} W^j (-\mu)^{r-j}$$

and
\[ E(W - \mu)^r = E(W + (-\mu))^r \]
\[ = E \left[ \sum_{j=0}^{r} \binom{r}{j} W^j (-\mu)^{r-j} \right] \]
\[ \mu'_r = E[(W - \mu)^r] = \sum_{j=0}^{r} \binom{r}{j} E(W^j) (-\mu)^{r-j} \]

Thus,
\[ \mu'_2 = E[(W - \mu)^2] = \sigma^2 = \mu_2 - \mu_1^2 \]
\[ \mu'_3 = E[(W - \mu)^3] = \mu_3 - 3\mu_1\mu_2 + 2\mu_1^3 \]
\[ \mu'_4 = E[(W - \mu)^4] = \mu_4 - 4\mu_1\mu_3 + 6\mu_1^2\mu_2 - 3\mu_1^4 \]

**Example 3.6.3**

a) Coefficient of skewness = \( \gamma_1 = \frac{E[(W-\mu)^3]}{\sigma^3} \)

Note that the division by \( \sigma^3 \) makes \( \gamma_1 \) dimensionless. Note also that when the pdf is symmetry, \( E[(W - \mu)^3] \) will be zero.
When \( \gamma_1 > 0 \) distribution is skew right
When \( \gamma_1 < 0 \) distribution is skew left
b) Coefficient of kurtosis $\gamma_2$ (measure the flatness or peakedness of a pdf is a shape parameter relative to the bell shape pdf)

$$\gamma_2 = \frac{E[(W - \mu)^4]}{\sigma^4} - 3$$

Low kurtosis - flat
High kurtosis - peaked

For certain pdf's $\gamma_2$ is useful measure of peakedness: a relatively flat pdf's are said to be platykurtic; more peaked pdf's are called leptokurtic.

Sometimes $E(W^j)$ is not finite. How to determine?

**Theorem 3.6.3.** If the $k$th moment of a random variables exists, all moments of order less that $k$ exist (finite).

**Example 3.6.4** (fourth edition)
The pdf for a Student t random variable is given by

$$f_Y(y) = \frac{c(n)}{(1 + \frac{y^2}{n})^{(n+1)/2}}, \quad -\infty < y < \infty, \quad n \geq 1$$
where $n$ is referred to as the "degree of freedom" of the distribution, and $c(n)$ is a constant. By definition, the $2k$ moment is the integral

$$E(Y^{2k}) = c(n) \cdot \int_{-\infty}^{\infty} \frac{y^{2k}}{(1 + \frac{y^2}{n})^{(n+1)/2}}$$

Is $E(Y^{2k})$ finite?

Not necessarily! Recall from calculus that an integral of the form

$$\int_{-\infty}^{\infty} \frac{1}{y^\alpha} dy$$

will converge only if $\alpha > 1$.

The convergence of $\frac{y^{2k}}{(1 + \frac{y^2}{n})^{(n+1)/2}}$ are the same as those for

$$\frac{y^{2k}}{y^{2(n+1)/2}} = \frac{1}{y^{n+1-2k}}.$$ 

Therefore, if $E(Y^{2k})$ to be finite, we must have

$$n + 1 - 2k > 1$$
or equivalently, $2k < n$. thus for a Student $t$ random variable with, say $n = 9$ degree of freedom has $E(X^8) < \infty$, but no moment of order higher than eight exists.

**Practices: 3.6.19, 3.6.20, 3.6.22**

### 3.7 Joint Densities

This section introduce the concepts, definition and mathematical techniques associated with distribution based on two (or more) random variables.

**Discrete Joint Pdfs**

**Definition 3.7.1.** Suppose $S$ is a discrete sample space on which two r.v. $X$ and $Y$ are defined. The joint probability density function of $X$ and $Y$ (or joint pdf) is denoted
\[ p_{X,Y}(x, y) = P(\{s|X(s) = x \text{ and } Y(s) = y\}) \]
\[ = P(X = x, Y = y) \]

**Example 3.7.1** A supermarket has two express lanes. Let \( X \) and \( Y \) denote the number of customers in the first and second, respectively, at any given time. During nonrush hours, the joint pdf of \( X \) and \( Y \) is summarized by the following table:

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.1</td>
<td>0.2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0.2</td>
<td>0.25</td>
<td>0.05</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0.05</td>
<td>0.05</td>
<td>0.025</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0.025</td>
<td>0.05</td>
</tr>
</tbody>
</table>

Find \( P(|X - Y| = 1) \), the probability that \( X \) and \( Y \) differ by exactly 1.
By definition

\[ P(|X - Y| = 1) \]
\[ = \sum \sum p_{X,Y}(x, y) \quad |x-y|=1 \]
\[ = p_{X,Y}(0, 1) + p_{X,Y}(1, 0) + p_{X,Y}(1, 2) \]
\[ + p_{X,Y}(2, 1) + p_{X,Y}(2, 3) + p_{X,Y}(3, 2) \]
\[ = 0.2 + 0.2 + 0.05 + 0.05 + 0.25 + 0.25 = 0.55 \]

**Example 3.7.2**

Suppose two fair dice are rolled. Let \( X \) be the sum of the numbers showing, and let \( Y \) be the larger of the two. So, for example,

\[ p_{X,Y}(2, 3) = P(X = 2, Y = 3) = P(\emptyset) = 0 \]
\[ p_{X,Y}(4, 3) = P(X = 4, Y = 3) = P(\{(1, 3), (3, 1)\}) \]
\[ = \frac{2}{36} \]

and

\[ p_{X,Y}(6, 3) = P(X = 6, Y = 3) = P(\{(3, 3)\}) = \frac{1}{36}. \]

The entire joint pdf is given in table 3.7.1.
<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>Row</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1/36</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1/36</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>2/36</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2/36</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>1/36</td>
<td>2/36</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3/36</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>2/36</td>
<td>2/36</td>
<td>0</td>
<td>0</td>
<td>4/36</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0</td>
<td>1/36</td>
<td>2/36</td>
<td>2/36</td>
<td>0</td>
<td>5/36</td>
</tr>
<tr>
<td>x</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2/36</td>
<td>2/36</td>
<td>2/36</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1/36</td>
<td>2/36</td>
<td>2/36</td>
<td>5/36</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2/36</td>
<td>2/36</td>
<td>4/36</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1/36</td>
<td>2/36</td>
<td>3/36</td>
</tr>
<tr>
<td>11</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2/36</td>
<td>2/36</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1/36</td>
<td>1/36</td>
</tr>
<tr>
<td>Col</td>
<td>1/36</td>
<td>3/36</td>
<td>5/36</td>
<td>7/36</td>
<td>9/36</td>
<td>11/36</td>
<td></td>
</tr>
</tbody>
</table>

Notice that the row totals in the right hand margin of the table give the pdf for $X$. Similarly, the column totals along the bottom detail the pdf for $Y$. Those are not coincidences. Theorem 3.7.1 gives a formal statement of the relationship between the joint pdf and the individual pdfs.

**Theorem 3.7.1** Suppose that $p_{X,Y}(x,y)$ is the
joint pdf of the discrete random variables $X$ and $Y$. Then

$$p_X(x) = \sum_{all \ y} p_{X,Y}(x, y) \quad \text{and}$$

$$p_Y(y) = \sum_{all \ x} p_{X,Y}(x, y)$$

**Definition 3.7.2.** An individual pdf obtained by summing a joint pdf over all values of the other random variable is called a *marginal pdf*.

**Continuous Joint Pdfs**

If $X$ and $Y$ are both continuous random variables, Definition 3.7.1 does not apply because $P(X = x, Y = y)$ will be identically 0 for all $(x, y)$. As was the case in single-variable situation, the joint pdf for two continuous variables will be defined as a function, in which when integrated yields the probability that $(X, Y)$ lied in a specified region of $xy$—plane.

**Definition 3.7.3.** Two random variables defined on the same set of real numbers are jointly
continuous if there exists a function $f_{X,Y}(x, y)$ such that for any region $R$ in the $xy-$plane

$$P((X, Y) \in R) = \int \int_R f_{X,Y}(x, y) \, dx \, dy.$$ 

the function $f_{X,Y}(x, y)$ is the joint pdf of $X$ and $Y$.

Note: Any function $f_{X,Y}(x, y)$ for which

1. $f_{X,Y}(x, y) \geq 0$ for all $x$ and $y$

2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x, y) \, dx \, dy = 1$

qualifies as a joint pdf.

**Example 3.7.3**

Suppose that the variation in two continuous random variables, $X$ and $Y$, can be modeled by the joint pdf $f_{X,Y}(x, y) = cxy$ for $0 < y < x < 1$. Find $c$.

By inspection, $f_{X,Y}(x, y)$ will be nonnegative as long as $c \geq 0$. The particular $c$ that qualifies $f_{X,Y}(x, y)$ as a joint pdf, though, is the one that makes the volume under $f_{X,Y}(x, y)$ equal to 1. But
\[
\int \int_S cxy \ dydx = 1
\]
\[
= c \int_0^1 \left[ \int_0^x (xy) \ dy \right] \ dx
\]
\[
= c \int_0^1 x \left( \frac{y^2}{2} \bigg|_0^x \right) \ dx
\]
\[
= c \int_0^1 \left( \frac{x^3}{2} \right) \ dx
\]
\[
= \frac{x^4}{8} \bigg|_0^1
\]
\[
= \left( \frac{1}{8} \right) c
\]

Therefore, \( c = 8 \).

**Example 3.7.4**

A study claim that the daily number of hours, \( X \), a teenager watches television and the daily number of hours, \( Y \), he works on his homework are approximated by the joint pdf,

\[
f_{X,Y}(x, y) = x \ y \ \ e^{-(x+y)}, \quad x > 0, \ y > 0
\]
What is the probability that the amount of time the student chosen at random watch TV is at least twice the amount of time working on his homework?

We wanted $P(X > 2Y)$. Let $R$ be the region such that $X > 2Y$ when $x > 0, y > 0$. So $P(X > 2Y)$ is the volume under $f_{X,Y}(x, y)$ above the region $R$

$$P(X > 2Y) = \int_0^\infty \int_0^{x/2} xye^{-(x+y)} \, dy \, dx$$

Separating variables, we can write

$$P(X > 2Y) = \int_0^\infty xe^{-x} \int_0^{x/2} ye^{-(y)} \, dy \, dx.$$ 

And the double integral reduces to $\frac{7}{27}$:

$$P(X > 2Y) = \int_0^\infty xe^{-x} \left[ 1 - \left( \frac{x}{2} + 1 \right) e^{-x/2} \right] \, dx.$$ 

$$= \int_0^\infty xe^{-x} \, dx - \int_0^\infty \frac{x^2}{2} e^{-3x/2} \, dx - \int_0^\infty xe^{-3x/2} \, dx$$ 

$$= 1 - \frac{16}{54} - \frac{4}{9} = \frac{7}{27}.$$
Geometric Probability

One particularly important special case of definition 3.7.3 is the joint uniform pdf, which is represented by a surface having a constant height everywhere above a specified rectangular xy-plane. That is:

\[ f_{X,Y}(x, y) = \frac{1}{(b - a)(d - c)}, \quad a \leq x \leq b, \quad c \leq y \leq d \]

If \( R \) is some region in the rectangle where \( X \) and \( Y \) are defined, \( P[(X, Y) \in R] \) reduces to simple ratios of areas

\[ P[(X, Y) \in R] = \frac{\text{area of } R}{(b - a)(d - c)} \]

**practice 3.7.1.** If \( p_{X,Y}(x, y) = cxy \) at points (1, 1), (2, 1), (2, 2), and (3, 1) and equals to 0 elsewhere. Find \( c \).

\[ 1 = \sum_{x,y} p(x, y) = \sum_{x,y} cxy \]

**Practices:** 3.7.2, 3.7.4, 3.7.8, 3.7.10, 3.7.11, 3.7.13

Marginal Pdfs for Continuous Random
Theorem 3.7.2. Suppose $X$ and $Y$ are jointly continuous with joint pdf $f_{X,Y}(x, y)$. Then the marginal pdfs, $f_X(x)$ and $f_Y(y)$ are given by

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy$$
and

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dx$$

Example 3.7.7
Suppose $f_{X,Y}(x, y) = \frac{1}{6}$, $0 \leq x \leq 3$, $0 \leq y \leq 2$

Applying Theorem 3.7.2 gives

$$f_X(x) = \int_{0}^{2} f_{X,Y}(x, y) dy = \int_{0}^{2} \frac{1}{6} dy = \frac{1}{3},$$

$$0 \leq x \leq 3$$

Example 3.7.8
Consider the case where $X$ and $Y$ are two continuous random variables, jointly distributed over the first quadrant of the xy-plane according to the joint
pdf,

\[ f_{X,Y}(x, y) = \begin{cases} 
  y^2 e^{-y(x+1)} & x \geq 0, \ y \geq 0 \\
  0 & \text{elsewhere}
\end{cases} \]

Find the two marginal cdf’s.

a) \[ f_X(x) = \int_0^\infty y^2 e^{-y(x+1)} dy \]

using substitution and integrating by parts twice: \( w = y(x + 1) \) making \( du = (x + 1)dy \). This gives

\[ f_X(x) = \frac{1}{x + 1} \int_0^\infty \frac{w^2}{(x + 1)^2} e^{-w} dw \]

\[ = \frac{1}{(x + 1)^3} \int_0^\infty w^2 e^{-w} dw \]

After applying by part twice to get \( \int_0^\infty w^2 e^{-w} dw \),
we get

\[ f_X(x) = \frac{1}{(x+1)^3} \left[ -w^2 e^{-w} - 2we^{-w} - 2e^{-w} \right] \bigg|_0^\infty \]

\[ = \frac{1}{(x+1)^3} \left[ 2 - \lim_{w \to \infty} \left( \frac{w^2}{e^w} + \frac{2w}{e^w} + \frac{2}{e^w} \right) \right] \]

\[ = \frac{2}{(x+1)^3} \quad x \geq 0 \]

Finding \( f_Y(y) \) is a bit easier.

\[ f_Y(y) = \int_0^\infty y^2 e^{-y(x+1)} \, dx \]

\[ = y^2 e^{-y} \int_0^\infty e^{-yx} \, dx = y^2 e^{-y} \left( \frac{1}{y} \right) \left( -e^{-yx} \bigg|_0^\infty \right) \]

\[ = ye^{-y}, \quad y \geq 0 \]

**Homework:** 3.7.19, 3.7.20, 3.7.21, 3.7.22

**Joint Cdf**
**Definition 3.7.4.**

Let $X$ and $Y$ be any two random variables. The joint cumulative distribution function of $X$ and $Y$ (or joint cdf) is denoted $F_{X,Y}(u,v)$, where

$$F_{X,Y}(u,v) = P(X \leq u \text{ and } Y \leq v)$$

**Example 3.7.9**

Find the joint cdf, $F_{X,Y}(u,v)$ for two random variables $X$ and $Y$ whose joint pdf is $f_{X,Y}(x,y) = \frac{4}{3}(x + xy), 0 \leq x \leq 1, 0 \leq y \leq 1$
\[
F_{X,Y}(u, v) = P(X \leq u \text{ and } Y \leq v)
= \frac{4}{3} \int_0^v \int_0^u (x + xy) \, dx \, dy
= \frac{4}{3} \int_0^v \left( \int_0^u (x + xy) \, dx \right) \, dy
= \frac{4}{3} \int_0^v \left( \frac{x^2}{2} (1 + y) \right|_0^u \right) \, dy
= \frac{4}{3} \int_0^v \frac{u^2}{2} (1 + y) \, dy
= \frac{4u^2}{3} \left( y + \frac{y^2}{2} \right) \bigg|_0^v
= \frac{4u^2}{3} \left( v + \frac{v^2}{2} \right)
\]

For what values of \( u \) and \( v \) if \( F_{X,Y}(u, v) \) defined?

**Theorem 3.7.3**

Let \( F_{X,Y}(u, v) \) be the joint cdf associated with the continuous random variables \( X \) and \( Y \). Then the joint pdf of \( X \) and \( Y \), \( f_{X,Y}(x, y) \) is a second partial derivative of the joint cdf - that is \( f_{X,Y}(x, y) = \)
\frac{\partial^2}{\partial x \partial y} F_{X,Y}(x, y) \text{ provided that } F_{X,Y}(x, y) \text{ has continuous second derivatives.}

**Example 3.7.10**

What is the joint pdf of the random variables $X$ and $Y$ whose joint cdf is $F_{X,Y}(x, y) = \frac{1}{3}x^2(2y+y^2)$

$$f_{X,Y}(x, y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x, y)$$

$$= \frac{\partial^2}{\partial x \partial y} \frac{1}{3}x^2(2y+y^2)$$

$$= \frac{\partial}{\partial y} \frac{2}{3}x(2y+y^3)$$

$$= \frac{2}{3}x(2+2y) = \frac{4}{3}(x+xy)$$

Compare with example 3.7.9

Read the format for Multivariate Densities.

**Homework: 3.7.25, 3.7.27, 3.7.29, 3.7.30**

**Independence of two random variables**

**Definition 3.7.5.** Two random variables $X$ and $Y$ are said to be independent if for every interval
A and every interval $B$, $P(X \in A \text{ and } Y \in B) = P(X \in A)P(Y \in B)$

**Theorem 3.7.4.** The random variables $X$ and $Y$ are independent if and only if there are functions $g(x)$ and $h(y)$ such that

$$f_{X,Y}(x, y) = g(x)h(y) \quad (3.7.1)$$

If equation 3.7.1 holds, there is a constant $k$ such that $f_X(x) = kg(x)$ and $f_Y(y) = (1/k)h(y)$.

Notes: If $X$ and $Y$ are independent then $f_{X,Y}(x, y) = f_X(x)f_Y(y)$

**Example 3.7.11.**

Suppose that the probabilistic behavior of two random variables $X$ and $Y$ described by the joint pdf $f_{X,Y}(x, y) = 12xy(1 - y), 0 \leq x \leq 1, 0 \leq y \leq 1$. Are $X$ and $Y$ independent? If they are find $f_X(x)$ and $f_Y(y)$.

From theorem 3.7.4, if $f_{X,Y}(x, y) = g(x)h(y)$, then $X$ and $Y$ are independent. Let $g(x) = 12x$ and $g(y) = y(1 - y)$.

From Theorem 3.7.4 $f_X(x) = kg(x)$, therefore
\[
\int_0^1 kg(x)\,dx = 1 = \int_0^1 k12x\,dx = \left. \frac{12k}{2}x^2 \right|_0^1
\]

So \( k = 1/6 \), therefore \( f_X(x) = 2x, 0 \leq x \leq 1 \) and \( f_Y(y) = 6y(1 - y), 0 \leq y \leq 1 \)

**Practice 3.7.43.** Suppose that random variables \( X \) and \( Y \) are independent with marginal pdf \( f_X(x) = 2x, 0 \leq x \leq 1 \) and \( f_Y(y) = 3y^2, 0 \leq y \leq 1 \). find \( P(Y < X) \)?

\[
P(Y < X) = \int_0^1 \int_0^x f_{X,Y}(x, y)\,dy\,dx
\]

\[
= \int_0^1 \int_0^x (2x)(3y^2)\,dy\,dx
\]

\[
= \int_0^1 2x^4\,dx = 2/5
\]

**Independence of \( n(> 2) \) random variables.**

**Definition 3.7.6** The \( n \) random variables \( X_1, X_2, \ldots , X_n \) are said to be independent if there
are functions
\[ g_1(x_1), g_2(x_2), \cdots, g_n(x_n) \]
such that for every \( x_1, x_2, \cdots, x_n \)
\[
f_{X_1,X_2,\cdots,X_n}(x_1, x_2, \cdots, x_n) = \\
g_1(x_1) \cdot g_2(x_2) \cdots g_n(x_n)
\]

Example 3.7.12 Consider \( k \) urns, each holding \( n \) chips numbered 1 through \( n \). A chip is to be drawn at random from each urn. What is the probability that all \( k \) chips will bear the same number? If \( X_1, X_2, \cdots, X_k \) denote the numbers on the 1st, 2nd, \( \ldots \), and \( k^{th} \) chips, respectively, we are looking for the probability that \( X_1 = X_2 = \cdots = X_k \).

In terms of the joint pdf,

\[
P(X_1 = X_2 = \cdots = X_k) = \\
\sum_{x_1=x_2=\cdots=x_k} p_{X_1,X_2,\cdots,X_k}(x_1, x_2, \cdots, x_k)
\]

Each of the selections here is obviously independent of all the others, so the joint pdf factors according to Definition 3.7.6, and we can write
\[ P(X_1 = X_2 = \cdots = X_k) \]
\[ = \sum_{x_1=x_2=\cdots=x_k}^{n} p_{X_1,X_2,\cdots,X_k}(x_1, x_2, \cdots, x_k) \]
\[ = \sum_{i=1}^{n} p_{X_1}(x_i) \cdot p_{X_2}(x_i) \cdots p_{X_k}(x_i) \]
\[ = n \cdot \left( \frac{1}{n} \cdot \frac{1}{n} \cdots \frac{1}{n} \right) \]
\[ = \frac{1}{n^{k-1}} \]

**Random Samples**

**Definition 3.7.6** addresses the question of independence as it applies to \( n \) random variables having marginal pdfs, say,

\( f_1(x_1), f_2(x_2), \cdots, f_n(x_n) \) that might be quite different. A special case of that definition occurs for virtually every set of data collected for statistical analysis.

Suppose an experimenter takes a set of \( n \) measurements, \( x_1, x_2, \cdots, x_n \) under the same conditions. Those \( X_i \)'s, then, qualify as a set of indepen-
dent random variables moreover, each represents the same pdf. The special- but familiar-notation for that scenario is given in Definition 3.7.7. We will encounter it often in the chapters ahead.

**Definition 3.7.7.** Let $X_1, X_2, \cdots, X_n$ be a set of $n$ independent random variables, all having the same pdf. Then $X_1, X_2, \cdots, X_n$ are said to be a random sample of size $n$.

**Homework:** 3.7.39, 3.7.42, 3.7.45, 3.7.46

3.8 Transformation and Combining Random Variables

In Section 3.4 we can find the pdf for $Y = aX + b$ given the pdf for $X$. In this section we can find the pdf for a random variable that is a function of $X$ and $Y$.

**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

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

$$= pX \left(\frac{y - 1}{2}\right)$$

$$= \left(\frac{n}{(y-1)}\right)p^{\frac{y-1}{2}}(1 - p)^{n-\frac{(y-1)}{2}}$$

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

**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

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

Finding the pdf for the sum of r.v.
Theorem 3.8.1. Suppose that $X$ and $Y$ are independent random variables. Let $W = X + Y$. Then

1. If $X$ and $Y$ are discrete r.v with pdf’s $p_X(x)$ $p_Y(y)$, respectively, then

$$p_W(w) = \sum_{\text{all } x} p_X(x)p_Y(w - x)$$

2. If $X$ and $Y$ are continuous r.v with pdf’s $f_X(x)$ $f_Y(y)$, respectively, then

$$f_W(w) = \int_{-\infty}^{\infty} f_X(x)f_Y(w - x)dx$$

Proof. 1.

$$p_W(w) = P(W = w) = P(X + Y = w)$$

$$= P\left(\bigcup_{\text{all } x}(X = x, Y = w - x)\right)$$

$$= \sum_{\text{all } x} P(X = x, Y = w - x)$$

$$= \sum_{\text{all } x} P(X = x)P(Y = w - x)$$

$$= \sum_{\text{all } x} p_X(x)p_Y(w - x)$$
where the next-to-last equality derives from the independence of $X$ and $Y$

2. Since $X$ and $Y$ are continuous random variables, we can find $f_W(w)$ by differentiating the corresponding cdf, $F_W(w)$. Here, $F_W(w) = P(X + Y) \leq w$ is found by integrating $f_{X,Y}(x, y) = f_X(x) \cdot f_Y(y)$ over the shaded region $R$ in figure 3.8.1.

By inspection

\[
F_W(w) = P(W < w) = P(X + Y < w)
= \int_{-\infty}^{\infty} \int_{-\infty}^{w-x} f_X(x) f_Y(y) \, dy \, dx
= \int_{-\infty}^{\infty} f_X(x) \left( \int_{-\infty}^{w-x} f_Y(y) \, dy \right) \, dx
= \int_{-\infty}^{\infty} f_X(x) F_Y(w - x) \, dx
\]

Assume that the integrand in the above equation is sufficiently smooth so that differentiation and in-
Integration can be interchanged. Then we can write

\[ f_W(w) = \frac{d}{dw} F_W(w) \]

\[ = \frac{d}{dw} \int_{-\infty}^{\infty} f_X(x) F_Y(w - x) dx \]

\[ = \int_{-\infty}^{\infty} f_X(x) \left( \frac{d}{dw} F_Y(w - x) \right) dx \]

\[ = \int_{-\infty}^{\infty} f_X(x) f_Y(w - x) dx \]

and the theorem is proved.

Comment. The integral in part (2) above is referred to as the convolution of the functions \( f_x \) and \( f_y \). Besides their frequent appearances in random-variable problems, convolutions turn up in many areas of mathematics and engineering.

Example 3.8.2 Suppose that \( X \) and \( Y \) are two independent binomial random variables, each with the same success probability but defined on \( m \) and \( n \) trials, respectively. Specifically,
\[ p_X(x) = \binom{m}{x} p^x (1 - p)^{m-x}, \quad x = 0, 1, \ldots, m \]

\[ p_Y(y) = \binom{n}{y} p^y (1 - p)^{n-y}, \quad y = 0, 1, \ldots, n \]

Find \( p_W(w) \), where \( W = X + Y \)

By Theorem 3.8.3 \( p_W(w) = \sum_{\text{all } x} p_X(x) p_Y(w-x) \), but the summation over all \( x \) needs to be interpreted as the set of values for \( x \) and \( w-x \) such that \( p_X(x) \) and \( p_Y(w-x) \), respectively are both nonzero. But that will be true for all integers \( x \) from 0 to \( w \), i.e.

\[ p_Y(w-x) = \binom{n}{w-x} p^{(w-x)} (1-p)^{n-(w-x)}, \quad (w-x) = 0, 1, \ldots, n \]

So \( w-x \geq 0 \), therefore \( x \leq w \)
Therefore

\[ p_W(w) = \sum_{\text{all } x} p_X(x)p_Y(w - x) \]

\[ = \sum_{x=0}^{w} \binom{m}{x} p^x (1 - p)^{m-x} \binom{n}{w-x} p^{w-x} (1 - p)^{n-(w-x)} \]

\[ = \sum_{x=0}^{w} \binom{m}{x} \binom{n}{w-x} p^w (1 - p)^{m+n-w} \]

**EXAMPLE 3.8.3** Suppose a radiation monitor relies on an electronic sensor, whose lifetime \( X \) is modeled by the exponential pdf \( f_X(x) = \lambda e^{-\lambda x}, x > 0 \). To improve the reliability of the monitor, the manufacturer has included an identical second sensor that is activated only in the event the first sensor malfunctions. (This is called *cold redundancy.*) Let the random variable \( Y \) denote the operating lifetime of the second sensor, in which case the lifetime of the monitor can be written as the sum \( W = X + Y \). Find \( f_W(w) \).
Since X and Y are both continuous random variables,
\[
f_W(w) = \int_{-\infty}^{\infty} f_X(x)f_Y(w-x)\,dx \quad (3.8.2)
\]
Notice that \( f_X(x) > 0 \) only if \( x > 0 \) and \( f_Y(w-x) > 0 \) only if \( x < w \). Therefore, the integral in (3.8.2) that goes from \(-\infty\) to \(\infty\) reduces to an integral from 0 to \(w\), and we can write

\[
f_W(w) = \int_{0}^{w} f_X(x)f_Y(w-x)\,dx
\]

\[
= \int_{0}^{w} \lambda e^{-\lambda x} \lambda e^{-\lambda(w-x)}\,dx
\]

\[
= \lambda^2 \int_{0}^{w} e^{-\lambda x} e^{-\lambda(w-x)}\,dx
\]

\[
= \lambda^2 e^{-\lambda w} \int_{0}^{w} \,dx = \lambda^2 we^{-\lambda w}, \ w \geq 0
\]

**Finding the Pdfs of Quotients and Products**

Interested in finding the pdf for

1) \( W = Y/X \)

2) \( W = XY \)
Theorem 3.8.4
Let $X$ and $Y$ be independent continuous R.V, with pdfs $f_X(x)$ and $f_Y(y)$, respectively. Assume that $X$ is zero for at most a set of isolated points. Let $W = Y/X$. Then

$$f_W(w) = \int_{-\infty}^{\infty} |x| f_X(x) f_Y(wx) dx$$

Example 3.8.4
Let $X$ and $Y$ be independent continuous R.V, with pdfs

$$f_X(x) = \lambda e^{-\lambda x}, \ x > 0 \text{ and } f_Y(y) = \lambda e^{-\lambda y}, \ y > 0,$$

respectively. Define $W = Y/X$. Find $f_W(w)$
\( f_W(w) = \int_0^\infty x \lambda e^{-\lambda x} \lambda e^{-\lambda wx} \, dx \)

\[
= \lambda^2 \int_0^\infty x e^{-\lambda(1+w)x} \, dx
\]

\[
= \frac{\lambda^2}{\lambda(1+w)} \int_0^\infty x(1+w)e^{-\lambda(1+w)x} \, dx
\]

\[
= \frac{\lambda^2}{\lambda(1+w)\lambda(1+w)} \frac{1}{1+w}, \quad w > 0
\]

**Theorem 3.8.5**

Let \( X \) and \( Y \) be independent continuous R.V, with pdfs \( f_X(x) \) and \( f_Y(y) \), respectively. Assume that \( X \) is zero for at most a set of isolated points. Let \( W = XY \). Then

\[
f_W(w) = \int_{-\infty}^{\infty} \frac{1}{|x|} f_X(x) f_Y(w/x) \, dx
\]

\[
= \int_{-\infty}^{\infty} \frac{1}{|y|} f_X(w/y) f_Y(y) \, dy
\]
Example 3.8.5
Let $X$ and $Y$ be independent continuous R.V. with pdfs
$f_X(x) = 1, \ 0 \leq x \leq 1$ and $f_Y(y) = 2y, \ 0 \leq y \leq 1$, respectively. Define $W = XY$. Find $f_W(w)$.

From Theorem 3.8.5
$f_W(w) = \int_{-\infty}^{\infty} \frac{1}{|x|} f_X(x) f_Y(w/x) dx$

The region of integration, though, needs to be restricted to values of $x$ such that the integrand is positive. But $f_Y(w/x)$ is positive only if $0 \leq w/x \leq 1$, which implies that $x \geq w$. Moreover, for $f_X(x)$ to be positive requires $0 \leq x \leq 1$. Any $x$ the from $w$ to 1 will yield a positive integrand. Therefore,

$$f_W(w) = \int_{w}^{1} \frac{1}{x}(1)(2w/x) dx$$

$$= 2w \int_{w}^{1} \frac{1}{x^2} dx$$

$$= 2 - 2w, \quad 0 \leq w \leq 1$$
3.9 Further properties of Mean and Variance

**Theorem 3.9.1**

1. Suppose $X$ and $Y$ are discrete r.v with joint pdf $p_{X,Y}(x, y)$ and let $g(X, Y)$ be a function of $X$ and $Y$. Then the expected value of the random variable $g(X, Y)$ is given by

$$E[g(X, Y)] = \sum_{\text{all } x} \sum_{\text{all } y} g(x, y) \cdot p_{X,Y}(x, y)$$

provided $\sum_{\text{all } x} \sum_{\text{all } y} |g(x, y)| \cdot p_{X,Y}(x, y)$

2. Suppose $X$ and $Y$ are continuous r.v with joint pdf $f_{X,Y}(x, y)$ and let $g(X, Y)$ be a function of $X$ and $Y$. Then the expected value of the random variable $g(X, Y)$ is given by

$$E[g(X, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) \cdot f_{X,Y}(x, y)$$

provided $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |g(x, y)| \cdot f_{X,Y}(x, y)$
Theorem 3.9.2 Let $X$ and $Y$ be any two random variables (discrete or continuous dependent or independent), and let $a$ and $b$ be any two constants. Then

$$E(aX + bY) = aE(X) + bE(Y)$$

provided $E(X)$ and $E(Y)$ are both finite.

Corollary. Let $W_1, W_2, \cdots, W_n$, be any r.v. for which $E(W_i) < \infty, i = 1, 2, \cdots, n$ and let $a_1, a_2, \cdots, a_n$ be any sets of constant. Then

$$E(a_1W_1 + a_2W_2 + \cdots + a_nW_n)$$

$$= a_1E(W_1) + a_2E(W_2) + \cdots + a_nE(W_n)$$

Example 3.9.3

Let $X$ be a binomial random variable defined on $n$ independent trials, each trial resulting in success with probability $p$. Find $E(X)$. Note first, that $X$ can be thought as a sum, $X = X_1 + X_2 + \cdots + X_n$, where $X_i$ represent the number of successes occurring at the $i$th trial:
\[ X_i = \begin{cases} 
1 & \text{if the } i\text{th trial produces a success} \\
0 & \text{if the } i\text{th trial produces a failure} 
\end{cases} \]

( Any \( X_i \) defined in this way on an individual trial is called \textit{Bernoulli} random variable. Every binomial r.v. then, can be thought of as the sum of \( n \) independent Bernoullis ). By assumption that \( p_{X_i}(1) = p \) and \( p_{X_i}(0) = 1 - p, i = 1, 2, \ldots, n \). Using the corollary,

\[ E(X) = E(X_1) + E(X_2) + \cdots + E(X_n). \]

But \( E(X_1) = 1 \cdot p + 0 \cdot (1 - p) = p \) so \( E(X) = np \)

\textbf{Example 3.9.5}

Ten fair dice are rolled. Calculate the expected value of the sum of the faces showing. If the random variable \( X \) denotes the sum of these faces showing on the ten dice, then

\[ X = X_1 + X_2 + \cdots + X_{10} \]

, where \( X_i \) is the number showing on the \( X_i \)th die, \( i = 1, 2, \cdots, 10 \). By assumption, \( p_X(k) = 1/6 \) for \( k = 1, 2, \cdots, 6 \), so
\[ E(X_i) = \sum_{k=1}^{6} \frac{1}{6} k = \frac{1}{6} \sum_{k=1}^{6} k = \frac{1}{6} \frac{6(7)}{2} \]

By theorem 3.9.2
\[ E(X) = E(X_1) + E(X_2) + \cdots + E(X_{10}) = 10(3.5) = 35 \]

**Expected values of products: A special Case**

**Theorem 3.9.3.** If X and Y are independent random variables,
\[ E(XY) = E(X) \cdot E(Y) \]
provided E(X) and E(Y) both exist.

**Proof.** Suppose X and Y are discrete random variable, and since X and Y are independent,
\[ p_{X,Y}(x, y) = p_X(x) \cdot p_Y(y), \]
then,
\[ E(XY) = \sum_{\text{all } x} \sum_{\text{all } y} xy \cdot p_{X,Y}(x, y) \]
\[ = \sum_{\text{all } x} \sum_{\text{all } y} xy \cdot p_X(x) \cdot p_Y(y) \]
\[ = \sum_{\text{all } x} x \cdot p_X(x) \sum_{\text{all } y} y \cdot p_Y(y) \]
\[ = E(X)E(Y) \]

**Practice 3.9.3**
Calculating The variance of a sum of Random Variable

When random variables are not independent the measure of relationship is called covariance.

**Definition 3.9.1** Given R.V. $X$ and $Y$ with finite variances, define the covariance of $X$ and $Y$, written $\text{Cov}(X, Y)$, as

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y)$$

**Theorem 3.9.4** If $X$ and $Y$ are independent, then $\text{Cov}(X, Y) = 0$.

**Proof:** If $X$ and $Y$ are independent then from theorem 3.9.3

$$E(XY) - E(X)E(Y),$$

therefore

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y) = E(X)E(Y) - E(X)E(Y)$$

Note: The converse of theorem 3.9.4 is NOT true. Just because $\text{Cov}(X, Y) = 0$, we cannot conclude that $X$ and $Y$ are independent. Read Example
3.9.7.

Theorem 3.9.5 Suppose R.V. $X$ and $Y$ with finite variances, and $a$ and $b$ are constants. Then

$$\text{Var}(aX + bY) = a^2\text{Var}(X) + b^2\text{Var}(Y) + 2ab\text{Cov}(X, Y)$$

Proof. For convenience, denote $E(X) = \mu_X$ and $E(Y) = \mu_Y$ and $E(aX + bY) = a\mu_X + b\mu_Y$

$$\text{Var}(aX + bY) = E[(aX + bY)^2] - [E(aX + bY)]^2$$
$$= E[(aX + bY)^2] - (a\mu_X + b\mu_Y)^2$$
$$= E(a^2X^2) + E(2abXY) + E(b^2Y^2)$$
$$- (a^2\mu_X^2 + 2ab\mu_X\mu_Y + b^2\mu_Y^2)$$
$$= E(a^2X^2) - a^2\mu_X^2 + E(b^2Y^2) - b^2\mu_Y^2$$
$$+ E(2abXY) - 2ab\mu_X\mu_Y$$
$$= a^2[E(X^2) - \mu_X^2] + b^2[E(Y^2) - \mu_Y^2]$$
$$+ 2ab[E(XY) - \mu_X\mu_Y]$$
$$= a^2\text{Var}(X) + b^2\text{Var}(Y) + 2ab\text{Cov}(X, Y)$$
Read Example 3.9.8.

**Corollary.** Suppose $W_1, W_2, \cdots, W_n$, are random variables with finite variances. Then

$$\text{Var}(\sum_{i=1}^{a} a_i W_i) = \sum_{i=1}^{a} a_i^2 \text{Var}(W_i) + 2 \sum_{i<j} a_i a_j \text{Cov}(W_i, W_j)$$

**Corollary.** Let $W_1, W_2, \cdots, W_n$, be a set of independent r.v. for which $E(W_i^2) < \infty, i = 1, 2, \cdots, n$. Let $a_1, a_2, \cdots, a_n$ be any sets of constant. Then

$$\text{Var}(a_1 W_1 + a_2 W_2 + \cdots + a_n W_n) = a_1^2 \text{Var}(W_1) + a_2^2 \text{Var}(W_2) + \cdots + a_n^2 \text{Var}(W_n)$$

**Example 3.9.9**

From Example 3.9.3,

Let $X$ be a binomial random variable defined on $n$ independent trials, each trial resulting in success with probability $p$. Find $E(X)$. Note first, that $X$ can be thought as a sum, $X = X_1 + X_2 + \cdots + X_n$, where $X_i$ represent the number of successes occurring at the $i$th trial:
\[ X_i = \begin{cases} 
1 & \text{if the } ith \text{ trial produces a success} \\
0 & \text{if the } ith \text{ trial produces a failure} 
\end{cases} \]

( Any \( X_i \) defined in this way on an individual trial is called \textit{Bernoulli} random variable. Every binomial r.v. then, can be thought of as the sum of \( n \) independent Bernoullis ). By assumption that \( p_{X_i}(1) = p \) and \( p_{X_i}(0) = 1 - p, i = 1, 2, \ldots, n \).

\[
E(X) = E(X_1) + E(X_2) + \cdots + E(X_n). \]

But

\[
E(X_1) = 1 \cdot p + 0 \cdot (1 - p) = p \quad \text{so} \quad E(X) = np \]

\[
E(X_i^2) = (1)^2 \cdot p + (0)^2 \cdot (1 - p) = p \]

\[
Var(X_i) = E(X_i^2) - [E(X_i)]^2 = p - p^2 = p(1 - p) \]

It follows, then, that the variance of a binomial random variable is \( np(1 - p) \):

\[
Var(X) = \sum_{i=1}^{n} Var(X_i) = np(1 - p) \]

\textbf{Example 3.9.11} In statistics, it is often necessary to draw inferences based on \( \bar{W} \), the average computed from a random sample of \( n \) observations. Two properties of \( \bar{W} \) are especially important. First, if the \( W_i \)'s come from a population where the mean is \( \mu \), the corollary from theorem
3.9.2 implies that $E(\bar{W} = \mu$. Second, if the $W_i$'s comes from the population whose variances is $\sigma^2$, then $\text{Var}(\bar{W}) = \frac{\sigma^2}{n}$

Proof.

$$
\bar{W} = \frac{1}{n} \sum_{i=1}^{n} W_i = \frac{1}{n} W_1 + \frac{1}{n} W_2 + \cdots + \frac{1}{n} W_n
$$

Practice: 3.9.14, 3.9.16, 3.9.17, 3.9.20

3.10 Order Statistics —- skip

3.11 Conditional Densities

Finding Conditional Pdfs for Discrete Random Variables

definition 3.11.1 let $X$ and $Y$ be discrete random variables. The conditional probability density function of $Y$ given $x$ - that is the probability that $Y$ takes on the value $y$ given that $X$ is equal $x$ - denoted as

$$
p_{Y|X}(y) = P(Y = y|X = x) = \frac{p_{X,Y}(x,y)}{p_X(x)}
$$
for $p_X(x) \neq 0$

Example 3.11.2 Assume that the probabilistic behavior of a pair of discrete random variables $X$ and $Y$ is described by the joint pdf

$$p_{X,Y}(x, y) = \frac{xy^2}{39}$$

define over the four points $(1, 2), (1, 3), (2, 2), (2, 3)$. Find the conditional probability that $X = 1$ given that $Y = 2$

$$p_{X|Y=2} = \frac{p_{X,Y}(1, 2)}{p_Y(2)} = \frac{1 \cdot 2^2/39}{1.2^2/39 + 2.2^2/39} = \frac{1}{3}$$

When r.v. $X$ and $Y$ are continuous then, 

$$f_{Y|X=x}(y) = f_{Y|x}(y) = \frac{f_{X,Y}(x, y)}{f_X(x)}$$

See Example 3.11.5
Practice: 3.11.2, 3.11.11, 3.11.16
Recall $E(X^k)$ is called $k^{th}$ central moment moment.

How to obtained them? BY the definition.

Moment generating function can be used to find the moment.

**Definition 3.12.1.** Let $W$ be a random variable. The moment generating function (mgf) for $W$ is denoted $M_W(t)$ and given by

$$M_W(t) =$$

$$E(e^{tW}) = \begin{cases} 
\sum_{k=1}^{\infty} e^{tk} p_W(k) & \text{if } W \text{ is discrete} \\
\int_{-\infty}^{\infty} e^{tw} f_W(w) dw & \text{if } W \text{ is continuous} 
\end{cases}$$

at all values of $t$ for which the expected value exists.

**Example 3.12.1** Suppose the r.v. $X$ has a geometric pdf,

$$p_X(k) = (1 - p)^{k-1}p, \ k = 1, 2, \cdots$$

[In practice, this is the pdf that models the occurrence of the first success in a series of independent trials, where each trial has a probability $p$ of
ending in success (recall Example 3.3.2)]

Find the $M_X(t)$, the moment generating function for $X$.

Since $X$ is discrete, the first part of Definition 3.12.1. applies, so

$$M_X(t) = E(e^{tX}) = \sum_{k=1}^{\infty} e^{tk} (1 - p)^{k-1} p$$

$$= \frac{p}{1 - p} \sum_{k=1}^{\infty} e^{tk} (1 - p)^k$$

$$= \frac{p}{1 - p} \sum_{k=1}^{\infty} [(1 - p)e^t]^k$$

The term $t$ in $M_X(t)$ can be any number in a neighborhood of zero, as long as $M_X(t) < \infty$. Here $M_X(t)$ is an infinite sum of the terms $[(1 - p)e^t]^k$ sum will be finite only if $(1 - p)e^t < 1$, or equivalently, if $t < ln\left[\frac{1}{1-p}\right]$. It is assume then in what follows that $t < ln\left[\frac{1}{1-p}\right]$.

Recall that

$$\sum_{k=0}^{\infty} r^k = \frac{1}{1 - r}$$
provided that \(0 < r < 1\). This formula can be used to solve the problem where \(r = (1 - p)e^t\) and \(0 < t < \ln\left[\frac{1}{1-p}\right]\). Specifically,

\[
M_X(t) = \frac{p}{1 - p} \left[ \sum_{k=0}^{\infty} [(1 - p)e^t]^k - [(1 - p)e^t]^0 \right] = \frac{p}{1 - p} \left[ \frac{1}{1 - (1 - p)e^t} - 1 \right] = \frac{pe^t}{1 - (1 - p)e^t}
\]

**Example 3.12.2.** Suppose that \(X\) is a binomial random variable with pdf

\[
p_X(x) = \binom{n}{x} p^x (1 - p)^{n-x}, \quad x = 0, 1, 2, \cdots, n
\]

Find \(M_X(t)\)
By definition 3.12.1.

\[ M_X(t) = E(e^{tX}) = \sum_{x=0}^{n} e^{tx} \binom{n}{x} p^x (1 - p)^{n-x} \]

\[ = \sum_{x=0}^{n} \binom{n}{x} (pe^t)^x (1 - p)^{n-x} \]

We know that \((a + b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}\) for any \(x \) and \(y\).

Suppose we let \(a = pe^t\) and \(b = 1 - p\) Then

\[ M_X(t) = (1 - p + pe^t)^n \]

**Example 3.12.3** Suppose \(Y\) has an exponential pdf, where \(f_Y(y) = \lambda e^{-\lambda y}, y > 0\)

\[ M_Y(t) = E(e^{tY}) = \int_{0}^{\infty} e^{ty} \lambda e^{-\lambda y} \]

\[ = \int_{0}^{\infty} \lambda e^{-(\lambda-t)y} dy \]

Using substitution,
$$M_Y(t) = \int_{u=0}^{\infty} \lambda e^{-u} \frac{du}{\lambda - t}$$

$$= \frac{\lambda}{\lambda - t}$$

Note: Here $M_Y(t)$ if finite and nonzero only when $u = (\lambda - t)y > 0$, which implies that $t$ must be less than $\lambda$. For $t > \lambda$, $M_Y(t)$ fails to exist.

**Example 3.12.4.**
Memorize the result. When $Y$ is normally distributed with mean $\mu$ and variance $\sigma^2$ has $M_Y(t) = e^{\mu t + \sigma^2 t^2 / 2}$

**Practice: 3.12.3, 3.12.4, 3.12.5 and 3.12.6, 3.12.7**

**Using Moment Generating Functions to find moments**

**Theorem 3.12.1** Let $W$ be a random variable with probability density function $f_W(w)$. [ If $W$ is continuous, $f_W(w)$ must be sufficiently smooth
to allow the order of differentiation and integration to be interchanged. Let $M_Y(t)$ be the moment-generating function for $W$. Then provided the $r$th moment exists,

$$M_W^{(r)}(0) = E(W^r)$$

The proof is straightforward.

Example 3.12.5. For a geometric random variable $X$ with pdf

$$p_X(k) = (1 - p)^{k-1}p, \ k = 1, 2, \cdots$$

which from Example 3.12.1 has a moment-generating function

$$M_X(t) = \frac{pe^t}{1 - (1 - p)e^t}.$$  

Find the Expected value of $X$ by differentiation its moment-generating function.

Using the product rule, the first derivative of $M_X(t)$
is

\[
M_X^{(1)}(t) = p e^t (-1) [1 - (1 - p)e^t]^{-2} (-1)(1 - p)e^t \\
+ [1 - (1 - p)e^t]^{-1} pe^t \\
= \frac{p(1 - p)e^{2t}}{[1 - (1 - p)e^t]^2} + \frac{pe^t}{1 - (1 - p)e^t}
\]

Setting \( t = 0 \) shows that \( E(X) = \frac{1}{p} \):

\[
M_X^{(1)}(0) = E(X) = \frac{p(1 - p)}{p^2} + \frac{p}{p} = \frac{1}{p}
\]

**Example 3.12.6** Find the Expected value of an exponential random variable with pdf \( f_Y(y) = \lambda e^{-\lambda y}, y > 0 \). Use the fact that \( M_Y(t) = \frac{\lambda}{\lambda - 1} \)

\[
M_Y^{(1)}(t) = \lambda (-1) (\lambda - t)^{-2} (-1) \\
= \frac{\lambda}{(\lambda - t)^2}
\]

Set \( t = 0 \). then

\[
M_Y^{(1)}(0) = E(Y) = \frac{1}{\lambda}
\]
Using Moment Generating Functions to find Variances

Because \( \text{Var}(Y) = E(Y - \mu)^2 = E(Y^2) - [E(Y)]^2 \), when \( M_Y(t) \) is available,

\[ \text{Var}(Y) = M_Y^{(2)}(0) - [M_Y^{(1)}(0)]^2 \]

Example 3.12.9

A discrete random variable \( X \) is said to have a Poisson distribution if

\[ p_X(x) = P(X = x) = \frac{e^{-\lambda} \lambda^x}{x!}, \quad x = 0, 1, 2, \ldots \]

It can be shown (Question 3.12.7) that the moment-generating function for a Poisson random variable is given by

\[ M_X(t) = e^{-\lambda + \lambda e^t} \]

Use \( M_Y(t) \) to find \( E(X) \) and \( \text{Var}(X) \)

\[ M_Y^{(1)}(0) = E(X) = \lambda \]

\[ M_Y^{(2)}(0) = E(X^2) = \lambda + \lambda^2 \]
Var(\(X\)) = E(\(X^2\)) - [E(\(X\))]^2 = \lambda

**Homework:** 3.12.9, 3.12.11, 3.12.16

**Using Moment-Generating Functions to Identify Pdfs**

Finding moments is not the only application of moment-generating functions. They are also used to identify the pdf of sums of random variables- that is, finding \(f_W(w)\), where \(W = W_1 + W_2 + \cdots + W_n\)

**Theorem 3.12.2** Suppose that \(W_1\) and \(W_2\) are random variables for which \(M_{W_1}(t) = M_{W_2}(t)\) for some interval of \(t\)'s containing 0. Then \(f_{W_1}(w) = f_{W_2}(w)\)

**Theorem 3.12.3**

a. Let \(W\) be a random variable with moment generating function \(M_W(t)\). Let \(V = aW + b\), then

\[M_V(t) = e^{bt} M_W(at)\]
b. Let \( W_1, W_2, \cdots, W_n \) be independent random variables with moment generating function 
\( M_{W_1}(t), M_{W_2}(t), \cdots, M_{W_n}(t) \) respectively. Let 
\( W = W_1 + W_2 + \cdots + W_n \). Then 
\[
M_W(t) = M_{W_1}(t) \cdot M_{W_2}(t) \cdots M_{W_n}(t)
\]

**Example 3.12.10** Suppose that \( X_1 \) and \( X_2 \) are two independent Poisson random variables with parameters \( \lambda_1 \) and \( \lambda_1 \) respectively. That is
\[
p_{X_1}(x) = P(X_1 = x) = \frac{e^{-\lambda_1} \lambda_1^x}{x!}, x = 0, 1, 2, \cdots
\]
and
\[
p_{X_2}(x) = P(X_2 = x) = \frac{e^{-\lambda_2} \lambda_2^x}{x!}, x = 0, 1, 2, \cdots
\]
Let \( X = X_1 + X_2 \). What is the pdf \( X \)?

\[
M_{X_1}(t) = e^{-\lambda_1 + \lambda_1 e^t}
\]
and
\[
M_{X_2}(t) = e^{-\lambda_2 + \lambda_2 e^t}
\]
Moreover when \( X = X_1 + X_2 \).
\[ M_X(t) = M_{X_1}(t) \cdot M_{X_2}(t) \]
\[ = e^{-\lambda_1 + \lambda_1 e^t} \cdot e^{-\lambda_2 + \lambda_2 e^t} \]
\[ = e^{-(\lambda_1 + \lambda_2) + (\lambda_1 + \lambda_2)e^t} \quad (3.12.1) \]

By inspection, 3.12.1 is the moment generating function for Poisson with \( \lambda = \lambda_1 + \lambda_2 \)

**Example 3.12.**

The normal random variable, \( Y \), with mean \( \mu \) and variance \( \sigma^2 \) has pdf

\[ f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left[ -\frac{1}{2} \left( \frac{y - \mu}{\sigma} \right)^2 \right], \quad -\infty < y < \infty \]

and mgf

\[ M_Y(t) = e^{\mu t + \sigma^2 t^2/2} \]

By definition, the standard normal distribution denoted as \( Z \), have \( \mu = 0 \) and \( \sigma^2 = 1 \). Then

\[ f_Z(z) = \frac{1}{\sqrt{2\pi}} \exp \left[ -\frac{1}{2} (z)^2 \right], \quad -\infty < y < \infty \]

\[ M_Z(t) = e^{t^2/2} \]
Show that the ratio \( \frac{Y - \mu}{\sigma} \) is a standard normal variable, \( Z \).

Write \( \frac{Y - \mu}{\sigma} \) as \( \frac{1}{\sigma} Y - \frac{\mu}{\sigma} \).

From the theorem 3.12.3 which stated

\[
M_{\frac{Y - \mu}{\sigma}}(t) = e^{bt} M_Y(at)
\]

\[
= e^{-\mu t/\sigma} M_Y \left( \frac{t}{\sigma} \right)
\]

\[
= e^{-\mu t/\sigma} \cdot e^{-\mu t/\sigma + \sigma^2 (t/\sigma)^2/2}
\]

\[
= e^{t^2/2}
\]

But \( M_Z(t) = e^{t^2/2} \) then it follows that the pdf for \( \frac{Y - \mu}{\sigma} \) is the same as the pdf for \( f_Z(z) \). We call \( \frac{Y - \mu}{\sigma} \) a \( Z \) transformation.

**Practice 3.12.23,**

Appendix 3.A.1 : R Application

Downloading R
http://www.r-project.org/

For binomial Distribution
http://www.stat.umn.edu/geyer/old/5101/rlook.html#dbinom

For Exponential Distribution and other continuous distributions
http://msenux.redwoods.edu/math/R/ContinuousDistributions.php

http://stat.ethz.ch/R-manual/R-patched
#3.13 R application

# Binomial Distribution

# dbinom - evaluate P(X=k) where X is Binomial

```r
dbinom(10, size=30, prob=0.60) # dbinom(x,n,p)
```

```r
x=seq(0,30,by=1)
y=dbinom(x, size=30, prob=0.60)
plot(x,y,main='Binomial pdf n=30, p=0.60')
```

```r
# plot(dbinom(x, size=30, prob=0.60)) # dbinom(x,n,p)
```

```r
require(graphics)
```

```
# Compute P(5 < X < 18) for X Binomial(30,0.6)
sum(dbinom(6:17, 30, 0.6))
```

```
# Compute P(-1 < X < 16) for X Binomial(30,0.6); cdf F_X(15)
sum(dbinom(0:15, 30, 0.6))
```
pbinom(15, size=30, prob=0.6) # F_X(15)
pbinom(15, 30, 0.6)

cdf.bi=pbinom(x, 30, 0.6) #F_X(x)
cdf.bi

plot(x,cdf.bi)

#################################################################
#
#qbinom(q, n, p) = x value such that P(X<=x)=q
# q == pbinom(qbinom(q, n, p))
#q=is quantile # see ?

qbinom(0.1, 30, 0.6)
qbinom(0.17, 30, 0.6)
qbinom(0.2, 30, 0.6)
# and so forth, or all at once with
qbinom(seq(0.1, 0.9, 0.1), 10, 1/3)
# Exponential Distribution

\#dexp(x, rate = 1, log = FALSE); f_X(x)
\#pexp(q, rate = 1, lower.tail = TRUE, log.p = FALSE); F_X(x)
\#qexp(p, rate = 1, lower.tail = TRUE, log.p = FALSE)

x=seq(0,4,length=200)
y=dexp(x,rate=1)  # lambda=1
plot(x,y,type="l",lwd=2,col="red",ylab="p")

x=seq(0,4,length=200)
y=dexp(x,rate=0.5)  # lambda=1
plot(x,y,type="l",lwd=2,col="red",ylab="p")

#cdf ; F_X(1)
pexp(1, rate = 1, lower.tail = TRUE, log.p = FALSE)

# inverse probability
qexp(p=0.632, rate = 1, lower.tail = TRUE, log.p = FALSE)

# Normal Distribution
# REF: http://msenux.redwoods.edu/math/R/normal.php
#
# standard normal Distribution
x=seq(-4,4,length=200)
y=1/sqrt(2*pi)*exp(-x^2/2)
plot(x,y,type="l",lwd=2,col="red")

# Alternative approach
x=seq(-4,4,length=200)
y=dnorm(x,mean=0,sd=1)
plot(x,y,type="l",lwd=2,col="red")
\begin{verbatim}
x=seq(-8,8,length=500)
y1=dnorm(x,mean=0,sd=1)
plot(x,y1,type="l",lwd=2,col="red")
y2=dnorm(x,mean=0,sd=2)
lines(x,y2,type="l",lwd=2,col="blue")
legend("topright",c("sigma=1","sigma=2"),lty=c(1,1),col=c("red","blue"))

#########################################
#Poisson

#dpois(x, lambda, log = FALSE)
#ppois(q, lambda, lower.tail = TRUE, log.p = FALSE)
#qpois(p, lambda, lower.tail = TRUE, log.p = FALSE)

#Poisson Distribution

#The Poisson distribution is the probability
# of independent event occurrences in an interval.
#If \( \lambda \) is the mean occurrence per interval, the
# probability of having \( x \) occurrences within
\end{verbatim}
\[ f(x) = (\lambda)^x e^{-\lambda} / x! \quad \text{where} \quad x = 0, 1, 2, 3, \ldots \]

# Problem

# If there are twelve cars crossing a bridge per minute on average, find the probability of having seventeen or more cars crossing the bridge in a particular minute.

# Solution

# The probability of having sixteen or less cars crossing the bridge in a particular minute is given by the function `ppois`.

\[
\text{ppois}(16, \lambda=12) \quad \# \text{lower tail}
\]

#[1] 0.89871

# Hence the probability of having seventeen or more cars crossing the bridge in a minute is in the upper tail.

\[
\text{ppois}(16, \lambda=12, \text{lower}=\text{FALSE}) \quad \# \text{upper tail}
\]

#[1] 0.10129
#Answer

If there are twelve cars crossing a bridge per minute on average, the probability of having seventeen or more cars crossing the bridge in a particular minute is 10.1%.