## Contents

2 **Theory of Probability**  
2.1 Interpretation ............................................. 2  
2.2 Probability and Inference ................................. 3  
2.3 A Review of Set Notation .................................. 4  
2.4 Probabilistic Model (the discrete case) .................. 10  
2.5 Sample-point method ..................................... 15  
2.6 The Equally-likely Outcomes Model and Counting Rules 19  
2.7 Conditional Probability and Independence .............. 36  
2.8 Laws of Probability ....................................... 41  
2.9 Law of Total Probability and Bayes’ Rule ............... 52  
2.10 Introduction to Random Variables ....................... 57
2 Theory of Probability

2.1 Interpretation

- **Probability** of an event is the proportion of times that the event will occur in the long run. It is a measure of how likely the event will occur. A probability is a value between 0 and 1.

- **Example**: If the probability of rain is 0.5, this means that there is a 50% chance that we will get rain. If the probability of rain is 1, this means it will rain for sure. If the probability is 0, then it will not rain (without doubt).

- The **relative frequency** is the observed number of successful events for a finite sample of trials. It is a meaningful measure of our belief in the occurrence of an event.

- **Example**: Suppose we toss a fair coin 50 times and have 27
heads and 23 tails. We define a head as a success. The relative frequency of heads is: \( \frac{27}{50} = 54\% \). The probability of a head is 50\%. \textbf{Q: what’s the cause of the difference?}

2.2 Probability and Inference

- Suppose that tossing a coin for 10 times yields 10 heads. Is the coin fair?
  That is, we want to test the hypothesis: the coin is fair.

- If the hypothesis is true, how likely this event (10 heads for 10 tossing) will happen?

- We need a theory of probability that permits us to calculate such probability of observing a specified outcome, assuming that our hypothesized model is correct. This provides the foundation for modern statistical inference.
2.3 A Review of Set Notation

A set is a collection of elements, denoted by \( A = \{a_1, a_2, a_3, \cdots \} \).

We use notation \( a_1 \in A \) to denote that \( a_1 \) is one element of \( A \).

We use capital letter, \( A, B, C, \cdots \) to denote sets of points.

Some useful definitions:

- **Universal** set: set of all possible elements under consideration, denoted by \( S \).
- **Subset**: \( A \) is a subset of \( B \), denotes as \( A \subset B \), if every element of \( A \) belongs to \( B \).
- **Empty** or **Null** set: a set that has no elements, denoted by \( \emptyset \).
- **Union** of sets: the set of all points in \( A \) or \( B \) or both, denoted as \( A \cup B \).
- **Intersection** of sets: set of all points in both \( A \) and \( B \), denoted
as \( A \cap B \).

- **Complement** of set: If \( A \subset S \), the complement of \( A \), denoted by \( \bar{A} \), contains elements that are in \( S \) but not in \( A \).

- **Disjoint or Mutually exclusive** sets: \( A \) and \( B \) are disjoint if \( A \cap B = \emptyset \). \( A \) and \( \bar{A} \) are mutually exclusive.

**Example 2.3.1** Suppose a family has 2 children of different ages. We are interested in the gender of these children. Let a pair \( FM \) denote the older child is female and the younger is male.

1. Universal set: \( S = \{ \} \)

2. Let \( A \) the subset containing two males, \( B \) the subset containing at least one male, and \( C \) denote the subset containing no males. List the elements of
   (a) \( A = \{ \} \)
(b) \( B = \{ \} \)

(c) \( C = \{ \} \)

(d) \( A \cap C = \{ \} \)

(e) \( A \cup C = \{ \} \)

(f) \( B \cap C = \{ \} \)

(g) \( B \cup C = \{ \} \)

(h) \( A \cap B = \{ \} \)

(i) \( A \cup B = \{ \} \)

(j) \( B \cup \overline{A} = \{ \} \)

**Distributive laws:**

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

\[
A \cap (B \cup C) = (A \cap B) \cup (A \cap C)
\]
DeMorgan’s laws:

\[(A \cup B) = \overline{A} \cap \overline{B}, \text{ and } (A \cap B) = \overline{A} \cup \overline{B}\]

**Example 2.3.2** For the Example 2.3.1 verify the DeMorgan’s laws.
Example 2.3.3 Roll a balanced die, define $A = \text{outcome is even}$, $B = \text{outcome is a prime number}$, $C = \text{outcome} \leq 3$. Verify the distributive laws:

\[ A \cup (B \cap C) = (A \cup B) \cap (A \cup C) \]
\[ A \cap (B \cup C) = (A \cap B) \cup (A \cap C) \]
Example 2.3.4 Simplify (a) $A \cap (B \cap \bar{A})$; (b) $B \cup (A \cap \bar{B})$. 
2.4 Probabilistic Model (the discrete case)

Some fundamental concepts:

1. An **experiment** is the process by which an observation is made, e.g., coin/dice tossing, measuring the height of a person.

2. The outcomes of an experiment are called **events**.

3. An event that can be decomposed is called a **compound event**.

4. An event that cannot be decomposed is called a **simple event**.

5. Each simple event corresponds to one **sample point**.

6. **Sample space**: the set consisting of all possible outcomes (sample points) of an experiment, denoted by $S$.

7. **Discrete sample space**: contains either finite or countable number of distinct sample points.

8. An **event** in a discrete sample space $S$ is a subset of $S$. 
Example 2.4.1 Consider the “Gender of Children” Example 2.3.1.

1. Experiment: Pick two children at random from a pool of families who have two children.

2. Recall events A: two males; B: at least one male; C: no males
   - Event A is a _________ event.
   - Event B is a _________ event.
   - Event C is a _________ event.

3. Sample space is ________________.
Example 2.4.2 Suppose 3 fair coins are tossed.

1. List the simple events and the sample space.
2. How many sample points are in the event $A$: at most two heads?
3. How many sample points are in the event $B$: at least two heads?
4. Find $A \cup B$ and $A \cap B$. 

Basic axioms of probability

- We will develop methods to determine the probability of the occurrence of an event.

- Let $P(A)$ denote the probability that the event $A$ occurs when the experiment is performed. Here $A$ is a subset of $S$.

- The probability $P(A)$, a number, must satisfy a set of axioms:
  - Axiom 1: $P(A) \geq 0$
  - Axiom 2: $P(S) = 1$.
  - Axiom 3: If $A_1, A_2, \cdots$ form a sequence of disjoint events in $S$, then
    $$P(A_1 \cup A_2 \cup \cdots) = \sum_{i=1}^{\infty} P(A_i).$$
Some remarks about the axioms:

- Axioms 1 and 2 imply that probabilities are between 0 and 1.

- An outcome with zero probability is called an impossible event, whereas an outcome with probability 1 is called a certain event.

- Axiom 3 implies that if $A_i$ are disjoint,

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) = \sum_{i=1}^{n} P(A_i).$$

- Suppose that $A$ is an event in $S$, consisting of single events $E_1, E_2, \cdots, E_k$. Then $P(A) = P(E_1) + \cdots + P(E_k)$. 
2.5 Sample-point method

**Example 2.5.1** In the previous Example 2.4.2 of tossing three fair coins, assign reasonable probabilities to the sample points in $S$, and calculate

1. $P(A)$

2. $P(B)$

3. $P(A \cup B)$

4. $P(A \cap B)$
The above procedure for finding probabilities of an event is called the **sample-point method**:

1. Define the experiment.
2. List all simple events. Define the sample space.
3. Assign reasonable probabilities to the sample points in $S$.
4. Define the event of interest $A$ as a specific collection of sample points in $S$.
5. Calculate $P(A)$ by summing the probabilities of the sample points in $A$. 
Q: In the previous Example 2.4.2 of tossing three fair coins, what if the coins are unfair? e.g. $P(\text{head})=2/3$ and $P(\text{tail})=1/3$. 
Example 2.5.2 A manufacturer has five seemingly identical computer terminals, of which two (unknown ones) are defective. A particular order is filled by randomly selecting two of the five.

1. List the sample space for this experiment.

2. Let $A$ denote the event that the order is filled with two nondefective terminals. List the sample points in $A$.

3. Suppose each pair is chosen with equal chance. Find the probability of even $A$. 
2.6 The Equally-likely Outcomes Model and Counting Rules

**Equally-Likely Outcomes Model**

An experiment has $N$ outcomes in its sample space. Suppose it is reasonable to assume that the $N$ outcomes are equally likely to occur. Therefore,

- the probability for each single event is $\frac{1}{N}$;
- suppose event $A$ in $S$ consists of exactly $n$ sample points, then

$$P(A) = \frac{1}{N} + \cdots + \frac{1}{N} = \frac{n}{N}.$$
Example 2.6.1 An urn contains 3 red marbles and 2 blue marbles. A marble is drawn at random (i.e. assume that all outcomes are equally likely). What is the probability that the marble is red?
Example 2.6.2 An urn contains 3 red marbles and 2 blue marbles. Two marbles are drawn at random. What is the probability that one marble is red and the other is blue?
Counting rules

Most of the time it is not possible to enumerate the entire sample space. However using the counting rules can help in counting the number of outcomes in a sample space/event without listing of the sample space. **Rule 1: the \( m \times n \) Rule**

**Theorem 1** Suppose conducting one experiment requires two operations. If you can do one operation in \( m \) ways, and a second operation in \( n \) ways, then you can do both operations in \( m \times n \) ways. This method extends to three or more operations.

e.g., carry out operation 1 in 3 ways and operation 2 in 2 ways. Then there are total \( m \times n = 3 \times 2 = 6 \) ways for carrying out both operations.
Example 2.6.3 (Counting words).

- If letters may be repeated, how many different three-letter "words" can be formed using the letters a, b, c, d, and e?

- If letters may NOT be repeated, how many different three-letter "words" can be formed using the letters a, b, c, d, and e?

- Let $A$ denote the event of three-unique-letter "words" formed with letters a, b, c, d, and e. Suppose every combination of letters is equally likely. Calculate $P(A)$. 
Example 2.6.4 (coin-tossing). Recall in the experiment of tossing three coins (Example 2.4.2), we found the total number of sample points was eight. Use the $m \times n$ rule to confirm this result.

Example 2.6.5 (Counting selections). From a committee of 7 people select a President, a Vice President and a Secretary-Treasurer. No one can serve more than one role. How many ways can these officers be filled?
Rule 2: Permutations

Example 2.6.6 (counting arrangements). How many ways can we arrange/order 3 objects, labeled a, b, c, in a row?

Example 2.6.7 (counting arrangements). How many ways can we arrange/order n objects in a row?
• **Permutation**: an ordered arrangement of distinct objects.

• The number of permutations (arrangement/ways of ordering) of \( n \) objects taken \( n \) at a time is \( n! \), where \( n! = n(n - 1) \cdots 1 \), \( 0! = 1 \).

**Theorem 2**  The number of permutations of \( n \) distinct objects taken \( r \) at a time is

\[
P_r^n = n(n - 1) \cdots (n - r + 1) = \frac{n!}{(n - r)!}, \quad r \leq n.
\]
Example 2.6.8 The names of 3 employees are randomly drawn, without replacement, from the names of 30 employees. Those whose names are drawn first, second and third will receive $100, $50 and $25, respectively. How many sample points are associated with this experiment?
Rule 3: Counting Combinations

Example 2.6.9 A club consists of 7 members. Two of these 7 will be selected at random to form a committee (of size 2). How many ways can this be done? Note that there is no ordering within the committee. A member is either in or not in.

Theorem 3 The number of combinations of $n$ distinct objects taken $r$ at a time is

$$C^n_r = \binom{n}{r} = \frac{n!}{r!(n-r)!}$$
The notation \( \binom{n}{r} \) is called “n choose r”. It is the number of different subsets of size \( r \) that can be formed from a set of \( n \) distinct objects.

Some useful formula to compute \( P^n_r \) and \( C^n_r \):

1. \( C^n_0 = P^n_0 = 1 \)
2. \( P^n_r = n(n-1) \cdots (n-r+1) \). Take product going \( r \) steps “backwards” from \( n \).
3. \( C^n_r = \frac{n(n-1) \cdots (n-r+1)}{r(r-1) \cdots 1} \).
4. \( C^n_r = C^n_{n-r} \).
5. \( C^n_r + C^n_{r+1} = C^{n+1}_{r+1} \).

**Exercise**: compute the following

1. \( P^8_3 \)
2. \( C^8_3 \)
3. \( \binom{100}{97} \)
Example 2.6.10 Suppose a company has 9 equally qualified workers. There is a job to be done that requires 5 workers. How many ways can the company assign the 5 workers to this job?

Example 2.6.11 A jury panel has 25 members. A jury consists of 12 members. How many different juries can be formed?
Rule 4: Counting Partitions

Theorem 4  Consider partitioning $n$ distinct objects into $k$ distinct groups containing $n_1, n_2, \ldots, n_k$ objects, respectively, where $\sum_{i=1}^{k} n_i = n$. The number of ways this can be done is

\[
\binom{n}{n_1, n_2, \ldots, n_k} = \frac{n!}{n_1!n_2! \cdots n_k!}
\]

Proof.  Use the extension of $m \times n$ rule.
Example 2.6.12 Job Assignments.

Suppose a company has 9 equally qualified workers. There are 3 jobs to be done, the first of which requires 2 workers, the second 4, and the third 3. How many ways can the company assign the 9 workers to these 3 jobs?
Example 2.6.13 DNA sequences.

DNA consists of very long sequences of four nucleotides denoted by $T$, $C$, $A$ and $G$. Consider the collection of 10 nucleotides:

$$TTTTCCAAAG$$

How many different DNA sequences of length 10 can be formed from this collection? E.g. one possible sequence is $TCTCATAAGTA$. 
Equally-Likely Outcomes Model Examples

Example 2.6.14 Consider the experiment of selecting 2 job candidates out of five. Imagine that the applicants vary in competence, 1 being the best, 2 second best, and so on for 3, 4, 5.

1. Find the total number of sample points in $S$.

2. Let $A$ denote the event that exactly one of the two best applicants appears in the selection. Find the number of sample points in $A$ and $P(A)$.
Example 2.6.15  Jury composition probability.

A jury panel has 20 members, 14 men and 6 women. A jury of 12 is selected AT RANDOM. What is the probability that there are exactly 2 women on the jury?
2.7 Conditional Probability and Independence

The probability of an event will sometimes depend on whether we know that other events have occurred.

For example, the *unconditional probability* of getting a 1 in the toss of a die is \[ \text{unconditional probability} \]

If we know that an odd number has fallen, then the probability of occurrence of a 1 is \[ \text{new probability} \] which is \[ \text{comparison} \] than the unconditional probability.

**Definition.** The *conditional probability* of an event \( A \) given that an event \( B \) has occurred is equal to

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

provided \( P(B) > 0 \).
Example 2.7.1 Deal or No deal  The game host puts $10K, $100 and $1 in three briefcases. You, the player, can select any case you like. Then you are offered the second chance to select another case and the case will be opened. Suppose you see the second case does not have $10k, then what is the chance that the first chosen case has $10K dollars?
Additional Example: A black box contains 10 white, 5 yellow and 10 black marbles.

1. A marble is chosen at random, what’s the probability that the chosen one is yellow?

2. Suppose we know that the marble chosen is not one of the black marbles. Then what’s the probability that the chosen one is yellow?

Note that in general \( P(A|B) \neq P(A)/P(B) \). For instance, suppose a roll a fair six-sided die. Let \( A \): observe a prime number and \( B \): observe an even number. Verify that \( P(A|B) \neq P(A)/P(B) \).
Independent Events

Definition. Two events and are said to be independent if and only if any one of the following 3 conditions hold:

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

Otherwise, the events are said to be dependent.

Example 2.7.2 Consider the “Gender of Children” Example 2.3.1. Let \( A \) be the event that the younger child is female, and \( B \) be the event that the older child is male. Are \( A \) and \( B \) dependent? Verify.
Example 2.7.3 The probability that a NCSU student has a visa credit card is 0.48, a MasterCard is 0.64, and both a visa and a MasterCard is 0.35.

1. Calculate the conditional probability that a student has a Visa given he/she has a MasterCard.

2. The information that a student has a MasterCard ________ the probability that the student has a Visa.
   (1) increases (2) decreases

3. The events having a Visa and having a MasterCard are _____.
   (1) independent (2) dependent
2.8 Laws of Probability

Sometimes probabilities of compound events can be obtained by using multiplicative and additive rules.

**Multiplicative Law of Probability:**

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

Notice that if \( A \) and \( B \) are independent, then

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

The multiplicative law can be extended to cover 3 or more events:

\[ P(A \cap B \cap C) = P(A \cap B)P(C|A \cap B) = P(A)P(B|A)P(C|A \cap B). \]
Example 2.8.1 An urn contains 10 marbles, 4 are red (R) and 6 are black (B).

1. If 2 are randomly chosen from the urn, what is the probability that both are black?

2. If 3 are randomly chosen from the urn, what is the probability that all three are black?
Q: is it always true that $P(A|B) \geq P(A)$?

Example 2.8.2 In Neverland, men constitute 60% of the labor force. The rates of unemployment is 4.5% among females and 6% among males. A person is randomly selected from Neverland’s labor force. Suppose the randomly selected person is unemployed, what is the probability that a female is selected?
Sometimes we can **ASSUME** that two events are independent in order to simplify probability computations.

**Example 2.8.3** *Flip a fair coin 3 times, find the probability of observing 3 heads (HHH). We assume the outcome of flips are independent.*
Additive Law of Probability:

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

**Proof.** Hint: by Axiom 3.

Recall Axiom 3: if \( A \) and \( B \) are disjoint, then
\[ P(A \cup B) = P(A) + P(B). \]
Example 2.8.4 Draw 5 cards randomly from a standard deck of 52 cards. Calculate the probabilities of the following events

1. 5 cards contain two aces or two kings or both;
2. there is at least one ace in the 5 cards.
Complementary Law of Probability:

A special case of additive law is obtained by taking $B = \overline{A}$, then

$$P(A) + P(\overline{A}) = 1.$$ 

Generalized Additive Law of Probability:

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$$
Example 2.8.5 The probability that a NCSU student has a visa credit card is 0.48, a MasterCard is 0.64, and both a visa and a MasterCard is 0.35. Find the probability that a student has a Visa or MasterCard (or both).
Example 2.8.6 Common Birthdays: calculate the probability that at least two of 40 students have the same birthday.
Example 2.8.7  

[de M’er’e’s problem](http://www.daviddarling.info/encyclopedia/D/de_Meres_problem.html)  

This marked the birth of probability theory.

In 17th century, once De M’er’e (a French nobleman and inveterate gambler) asked Blaise Pascal (a French mathematician), which is more likely:

A: rolling at least one six in four throws of a single die, or,

B: rolling at least one double six in 24 throws of a pair of dice?

De M’er’e told Pascal he had actually observed that event B occurred less often than event A, but he was at a loss to explain why...
Example 2.8.8

\(A_i = \text{“Component } i \text{ works”}, \ P(A_i) = 0.9, \ i = 1, \cdots, 4.\) Suppose components work or fail independently.

Calculate the probability that the system works (system reliability), i.e. either both 1 and 2 work, or both 3 and 4 work.
2.9 Law of Total Probability and Bayes’ Rule

Theorem 5 (The Law of Total Probability) Assume that

\[ S = B_1 \cup B_2 \cdots \cup B_k, \text{ where } P(B_i) > 0 \text{ for } i = 1, \cdots, k \text{ and } B_i \cap B_j = \emptyset \text{ for } i \neq j. \]

Then for any event \( A \),

\[ P(A) = \sum_{i=1}^{k} P(B_i)P(A|B_i). \]
Theorem 6 (The Bayes’ Rule) Assume that
\[ S = B_1 \cup B_2 \cdots \cup B_k, \] where \( P(B_i) > 0 \) for \( i = 1, \cdots, k \) and \( B_i \cap B_j = \emptyset \) for \( i \neq j \). Then
\[
P(B_j | A) = \frac{P(B_j)P(A | B_j)}{\sum_{i=1}^{k} P(B_i)P(A | B_i)}.
\]

Proof. By the definition of conditional probability and Theorem 5.
Example 2.9.1 False positive paradox.

Suppose a rare disease infects 1 out of 1000 people in a population. There is a good, but not perfect test for this disease: if a person has the disease, the test comes back positive 99% of the time. On the other hand, the test also produces 2% false positives. What is your chance of having the disease if your got a positive test result?

Conclusion: Despite of the high accuracy of the test, _____ of those who test positive actually have the disease!
Example 2.9.2 *Bayesian spam filter*

From experience I know that 70% of my email messages are spam and that the word “enhanced” occurs in 10% of my spam messages, but in only 1% of my legitimate messages. What is the probability that the next email message I receive is spam, given that it contains the word “enhanced”?
Review of Probability Formulae

- Complementary Law
- Multiplicative Law
- Additive Law
- Law of Total Probabilities
- Bayes’ Rule
### 2.10 Introduction to Random Variables

**Some concepts of random variables**

- Given the EXPERIMENT, OUTCOME, SAMPLE SPACE, EVENT setup, a **Random Variable** is a rule that associates a numerical value with each outcome in the sample $S$.

- More formally, a Random Variable ($rv$) is a real-valued function whose domain is the sample space $S$ and whose range is a subset of real numbers.

- We generally denote random variables by capital letters from the end of the alphabet, e.g. $X, Y, W, Z$. 
Types of Random Variables

- Discrete random variable: takes a finite (or countably infinite) set of values, most often a set of integers representing counts. Any examples?

- Continuous random variable: takes values consisting of an entire interval on the numerical line. These variables often represent measurements like heights, weights, speeds etc. Any examples?