# Chapter 4: Probabilities and Proportions

## Section 4.1 Introduction

In the real world, *variability is everywhere and in everything.*

Probability studies randomness, where random is not the same as haphazard. **Random** refers to a situation in which there are various possible outcomes, you don't know which one will occur, but there is a regular pattern in the results if you were to examine many repetitions.

**Example: Roll a fair die**
Before you roll the die do you know which one will occur? ________
But if I say what is the probability (chance) that the outcome will be a "4", you would say ________
Why?

- **Probability** is the PROPORTION of times the outcome would occur in many repeated trials of a random phenomenon.
- Probability is long term relative frequency.

## Section 4.2 Coin Tossing and Probability Models

If I toss a coin, what is the probability that it will turn up heads?

If I toss a coin 100 times, what is the probability that it will turn up 50 heads and 50 tails?

Read it!

## Section 4.3 Where Do Probabilities Come From?

- from models
- from data
- subjective probabilities

Read it!

## Section 4.4 Simple Probability Models

A probability model has two main parts:

1. a list of possible outcomes
2. probabilities assigned to each outcome (or a collection of outcomes)

**Definitions**:

- The **Sample Space**, $S$, of a random experiment is the set of all possible outcomes.
- An **Event** is an outcome or a set of outcomes of a random experiment, that is, a subset of the sample space.
- An event **occurs** if any outcome making up that event occurs.

**Example** Describe a sample space.

(a) Choose a student in class at random. Ask how much time spent studying in the past 24 hours.

$S =$

(b) In a test of a new package design, you drop a carton of a dozen eggs from a height of 1 foot and count the number of broken eggs.

$S =$

If we define the event $A =$ more than half break, then $A =$ ________________________

(c) A basketball player shoots two free throws. (Here we have some flexibility in defining the outcome.)

The possible outcomes for one free throw are

We can define the outcomes for two free throws as

$S =$

or

$S =$

**Events and Venn Diagram**

**Union of Two Events:**

**Intersection of Two Events:**

**Complement of An Event:**

**Definition**: Two events A, B are **Mutually Exclusive** (or **Disjoint**) if ...

We can "picture" disjoint events:

**Probability Distributions**

A list of numbers $p_1, p_2, p_3, \ldots$ is a **probability distribution** for sample space $S = \{s_1, s_2, s_3, \ldots\}$, if

1. the $p_i$'s lie between 0 and 1: $0 \leq p_i \leq 1$
2. the sum of all the $p_i$'s is 1: $p_1 + p_2 + p_3 + \ldots = 1$

Then $p_i$ is the probability that outcome $s_i$ occurs. Write $p_i = P(s_i)$.

We often list the probability distribution in a table:

| $s_1$ | $s_2$ | $s_3$ | $\cdots$ | $s_i$ | $\cdots$ |
|---|---|---|---|---|---|
| $p_1$ | $p_2$ | $p_3$ | $\cdots$ | $p_i$ | $\cdots$ |

**Probability of Events**

The probability of event $A$, $P(A)$ = sum of probabilities of all the outcomes in $A$.

For **equally likely outcomes**,

$$P(A) = \frac{\text{Number of outcomes in } A}{\text{Total number of outcomes in } S}$$

**Example:** Roll a fair die.

The sample space $S =$ ______________________

The probability distribution is

Event $A$= an even number = _______________

$P(A) =$ _______________

Event $B$= less than 3 = _______________

$P(B) =$ _______________

## Section 4.5 Probability Rules

Note: This is a combination of Sections 4.5, 4.6, 4.7 and some extra.

**Basic Probability Results**: $P(A) =$ ______________________

**1. The probability of any event A is:**

**2. The probability of the sample space is:**

**Example:** Probability of drawing each color of plain M&M's:

| Color | Brown | Red | Yellow | Green | Orange | Blue |
|---|---|---|---|---|---|---|
| Probability | 0.3 | 0.2 | 0.2 | 0.1 | 0.1 | ? |

What must be the probability of drawing a blue candy?

**3. The Complement Rule:**

**Example:** A sociologist studying social mobility in Denmark finds that the probability that the son of a lower-class father remains in the lower class is 0.46.
What is the probability that the son moves to one of the higher classes?

**4. The General Addition Rule:**

From the *picture...*

**Example:** Household is "prosperous" if income $> \$100,000$. Household is "educated" if head of household completed college. Select an household at random. Event A = {household is prosperous}, Event B = {household is educated}. $P(A) = .134$, $P(B) = .254$, and $P(A \text{ and } B) = .080$. What is the probability that the household selected is either prosperous **OR** educated?

**Draw Venn Diagram:**

**Example: Prize Jar**

Each time a student gets 100% on a spelling test, name is placed in the Prize Jar. At the end of each quarter, one name is selected at random and they receive a prize. We have the following information just before a name will be selected:

| Student | Sarah | Michael | Jennifer | Elise | John |
|---|---|---|---|---|---|
| Number of 100% tests | 6 | 2 | 3 | 5 | 7 |

Also, 15 other students each received one 100% test.

(a) What is the probability that John will receive the prize?

(b) Laura is one of the students who received only one 100% test. Jennifer and Laura are good friends and agreed to share the prize if either wins. What is the probability that Jennifer or Laura will win?

**Special Case to Rule #4: Addition Rule for Mutually Exclusive Events**

If A, B are mutually exclusive (disjoint), then

Our next result shows us how to update probabilities about an event based on certain given information.

**Example:** Rolling a Fair Die

1. What is the probability of getting a "**1**"? ________

2. Suppose you are told the outcome is an **ODD** outcome, now what is the probability of getting a "**1**"?

In question 2, you have just computed a conditional probability.

**Conditional Probability:**

The **conditional probability of the event A, given the event B has occurred**, is given by:

$P(A \mid B) =$

This result gives us our 5th probability result called the Multiplication Rule.

**5. Multiplication Rule:** $P(A \text{ and } B) =$

**Think About It:**

What if you have two events A and B and you are told that:

$$P(A \mid B) = 0.60 \text{ and } P(A) = 0.60.$$

**What does this tell you?**

**Definition:**

Two events A, B are said to be **independent** if ______________________

**Notes:**

**1. Check the definitions.**

The definition for two events to be **disjoint** (mutually exclusive) was based on a ------- property.

The definition for two events to be **independent** is based on a ------------------ property.

You need to check if these definitions hold when asked to assess if two events are disjoint, or if two events are independent.

**2. Mutually Exclusive** vs **Independence**

Suppose the two events are person is a "**male**" and person is a "**female**". Also suppose that 50% of the population are male.

(i) Are the two events "male" and "female" **mutually exclusive**?

(ii) Are the two events "male" and "female" **independent**?

**Example: Customers of a Store**

A population consists of **200 customers** for a store. 120 are **regular** customers of which 50 pay with **cash**, and the rest pay with **credit**. Half of the 80 **non-regular** customers pay with cash, the rest pay with credit.

Display the information on **Payment Status (Cash or Credit)** and **Customer Status (Regular, Non-regular)** using the following table.

| | Cash | Credit | Total |
|---|---|---|---|
| Regular | | | |
| Non-Regular | | | |
| Total | | | |

1. What is the probability that a randomly selected customer from this population is a **Regular** customer?

2. What is the probability that a randomly selected customer from this population is a **Non-regular** customer who pays with **Cash**?

3. What is $P(\text{NR or Cash})$?

4. What is the probability that a **randomly selected Non-regular customer** pays with **Cash**?

5. Are the events "**Cash**" and "**Non-regular**" mutually exclusive? Explain.

6. Are the events "**Cash**" and "**Non-regular**" independent? Explain.

**Example: Blood Types**

Distribution of **blood types** is approximately: **37% type A, 13% type B, 44% type O, and 6% type AB**. Suppose that the blood types of married couples are **independent** and that both the husband and wife follow this distribution.

1. What is the probability that in a randomly chosen couple the **wife has type B blood and the husband has type A**?

2. What is the probability that **at least one** of a randomly chosen couple has **type O blood**?

Next we develop and apply two more probability rules: the **Partition Rule and Bayes' Rule**.

Consider the following sample space and event B shown via a Venn Diagram.

Suppose we wish to find the probability of an event B but the direct computation of this probability is not very easy. Perhaps we can break up our sample space into disjoint pieces, compute the probability of the event B on each piece and combine these probabilities appropriately to find the overall probability of the event B, $P(B)$.

In our example, these pieces, called here $A_1, A_2, A_3$, form a partition.

**Definition:**

The sets $A_1, A_2, ..., A_K$, form a **partition** if:

1. All of the sets are mutually exclusive (disjoint). (the intersection between any two of these sets is empty.)
2. The union of all of the sets equals the sample space $S$ (called exhaustive).

**Aside:** Suppose we had the following test results for three classes.

| Class | Average Test Score | Number of students in the Class |
|---|---|---|
| 1 | 60 | 20 |
| 2 | 70 | 30 |
| 3 | 80 | 50 |

How would you find the overall average test score for all 100 students?

The idea of partitioning is useful in two-stage experiments.

**Example: Two-Stage Experiment**

Suppose we have 3 baskets as shown below.

Consider the following experiment:

**Stage 1:** Roll a fair die.

If outcome is ______ then pick Basket 1,

If outcome is ______ then pick Basket 2,

If outcome is ______ then pick Basket 3.

**Stage 2:** From the selected basket, pick 1 ball at random.

**Question: What is the probability the ball will be RED?**

**Partition Rule:**

$$P(B) = P(B \mid A_1)P(A_1) + P(B \mid A_2)P(A_2) + ... + P(B \mid A_k)P(A_k)$$

where $A_1, A_2, ..., A_K$, form a **partition.**

**Example: Automobile Insurance**

An insurance company believes that people can be divided into two classes: those who are accident prone and those who are not. Their statistics show that an accident-prone person will have an accident at some time within a fixed 1-year period with probability 0.4, whereas this probability decreases to 0.2 for a non-accident prone person. Assume that 30 percent of the population is accident prone.

a. What is the probability that a new policyholder will have an accident within a year of purchasing a policy?

b. Given that a policy holder has had an accident within a year of purchasing a policy, what is the probability the policy holder is accident prone?

**Bayes' Rule**

$P(A \mid B) =$

**General Bayes' Rule:**

$P(A_j \mid B) =$

**Example: Diagnostic Testing**

Consider the following information about a diagnostic test for some disease (HIV).

- When the disease is **present**, the **test is positive 98%** of the time.
- When the disease is **absent**, the **test is negative 93%** of the time.
- It is estimated that about **1% of the population have the disease**.

**Questions:**
What is the **probability that a person actually has the disease if the test is positive**?

Before we solve the problem, let's set up some notation:

D = ND =

'+' = '-' =

**1. Convert these statements into probability statements of events.**

When the disease is **present**, the **test is positive 98%** of the time.

When the disease is **absent**, the **test is negative 93%** of the time.

It is estimated that about **1% of the population have the disease**.

**2. Find the complements of each probability.**

**3. Convert the question of interest into a probability statement of events.**

"What is the **probability that a person actually has the disease if the test is positive**?"

Or equivalently,

"Given a person tests positively, what is the probability he/she will have the disease?"

**Three ways to solve this problem:**

- Solution by Bayes' Rule
- Solution by Tree Diagram
- Solution by Two-Way Table

**Solution by Bayes' Rule:**

**Solution by Tree Diagram:**

**Solution by Two-Way Table:**

**Summary of Probability Rules**

**Complement Rule:** $P(A^c) = 1 - P(A)$.

**Addition Rule:** $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$.

If A, B are **disjoint (mutually exclusive)**, then $P(A \text{ or } B) = P(A) + P(B)$.

**Multiplication Rule:** $P(A \text{ and } B) = P(A)P(B|A) = P(B)P(A|B)$.

If A, B are **independent** then $P(A \text{ and } B) = P(A)P(B)$ or equivalently, $P(B \mid A) = P(B)$.

**Conditional Probability:** $P(A \mid B) = \frac{P(A \text{ and } B)}{P(B)}$

**Partition Rule:** If $A_1, A_2, ..., A_k$ form a partition,

$$P(B) = P(B \mid A_1)P(A_1) + P(B \mid A_2)P(A_2) + ... + P(B \mid A_k)P(A_k).$$

**Bayes' Rule:** If $A$ and $B$ are any events whose probabilities are not 0 or 1,

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)} = \frac{P(B \mid A)P(A)}{P(B \mid A)P(A) + P(B \mid A^c)P(A^c)}.$$

**General Bayes' Rule:** If $A_1, A_2, ..., A_k$ form a partition, then

$$P(A_j \mid B) = \frac{P(A_j \text{ and } B)}{P(B)} = \frac{P(B \mid A_j)P(A_j)}{P(B \mid A_1)P(A_1) + P(B \mid A_2)P(A_2) + ... + P(B \mid A_k)P(A_k)}.$$