CHAPTER 5 PROBABILITY

Section 5.1 - Randomness, Probability & Simulation (pp. 287-299)

1. The Idea of Probability

- Random samples and randomized experiments
- Avoid bias by allowing chance to decide what individuals get selected.
- Chance behavior is *unpredictable in the* ***short run*** *but has a regular and predictable pattern in the* ***long run***.
- This is the basis for **probability**.

Application - Random Babies

Suppose a stork randomly delivers four babies to four different houses. We are going to ***simulate*** this situation and count the number of correct deliveries.

(www.rossmanchance.com/applets/randomBabies/Babies.html)

Plot the proportion of proportion of trials where there were 0 matches. What are we observing when we run the simulation a large number of times?

Law of Large Numbers - The fact that the proportion of trials that had no babies delivered to the right house converges to 0.375 is guaranteed by the **law of large numbers**. This result says that if we observe more and more repetitions of a chance process, the proportion of times that a specific outcome occurs approaches a single value. The single value is called **probability**.

What implications does this have on sampling design and experimental design?

Definition: The **probability** of any outcome of a chance process is a number between 0 and 1 that describes the proportion of times the outcome would occur in a very long series of repetitions.

Example. How much should a company charge for an extended warranty for a specific type of cell phone? Suppose that 5% of these cell phones under warranty will be returned, and the cost to replace the phones is $150. If the company knew which phones would go bad, it could charge $150 for these phones and $0 for the rest. However, since the company cannot know which phones will be returned but knows that about 1 in every 20 will be returned. How much should they charge for the extended warranty?

Other examples:

CHECK YOUR UNDERSTANDING

1. According to the "Book of Odds," the probability that a randomly selected U.S. adult usually eats breakfast is 0.61.

(a) Explain what probability 0.61 means in this setting.

(b) Why doesn't this probability say that if 100 U.S. adults are chosen at random, exactly 61 of them usually eat breakfast?

2. Probability is a measure of how likely an outcome is to occur. Match one of the probabilities that follow with each statement. Be prepared to defend your answer.

0 0.01 0.3 0.6 0.99 1

(a) This outcome is impossible. It can never occur.

(b) This outcome is certain. It will occur on every trial.

(c) This outcome is very unlikely, but it will occur once in a while in a long sequence of trials.

(d) This outcome will occur more often than not.

---

2. Myths about Randomness - The idea of probability seems straightforward. It answers the question "What would happen if we did this many times?" In fact, both the behavior of random phenomena and the idea of probability are a bit subtle. We meet chance behavior constantly, and psychologists tell us we deal with it poorly.

Application - Suppose that a basketball announcer suggests that a certain player is *streaky*. That is, the announcer believes that if the player makes a shot, then he is more likely to make his next shot. As evidence, he points to a recent game where the player took 30 shots and had a streak of 7 in a row. Is this evidence of streakiness or could it have occurred by chance? Assuming the player makes 50% of his shots and the results of a shot do not depend on previous shots, how likely is it for the player to have a streak of 7 or more made shots in a row?

Solution (4 Step Process):

1. **State** - How likely is it for the player to have a streak of 7 or more made shots in a row?

2. **Plan** - Use the random number generator on the calculator to generate 30 random 0s and 1s. 0 = misses shot, 1 = makes shot. Record the outcome of each trial. Record the longest streaks on a dot plot.

3. **Do** - Have each student do this 2 times.

0 2 4 5 6 7 8 9 10 11 12

Longest Streak

4. **Conclude**

---

3. Simulations - The application that we just conducted is called a **simulation**. In fact, the vaunted Hyena Problem on day 1 was a simulation. To perform a simulation, we are going to use the venerable 4 Step Process.

Performing a Simulation

1. **State** - What is the question of interest about some chance process?

2. **Plan** - Describe how to use a *chance device* to imitate one repetition of the process. Explain clearly how to identify the outcomes of the chance process and what variable to measure.

3. **Do** - Perform *many* repetitions of the simulation. (At least 30.)

4. **Conclude** - Use the results of the simulation to answer the question of interest.

Examples of 4-Step Process

Refer to the table on p. 296

Examples on pp. 296-297

Application - At a department picnic, 18 students in the mathematics/statistics department at a university decide to play a softball game. Twelve of the 18 students are math majors and 6 are stats majors. To divide into two teams of 9, one of the professors put all the players' names into a hat and drew out 9 players to form one team, with the remaining 9 players forming the other team. The players were surprised when one team was made up entirely of math majors. Is it possible that the names were not adequately mixed in the hat, or could this have happened by chance? Design and carry out a simulation to help answer this question.

Section 5.2 - Probability Rules (pp. 305-314)

1. Definitions

- **Sample Space** -
- **Probability Model** -
- **Event** -
- **Mutually Exclusive** -

Example:

2. Basic Rules of Probability

- The probability of any event is

- All possible outcomes together must have

- If all possible outcomes in a sample space are equally likely, the probability that event A occurs can be found using the formula

- The probability that an even does not occur is

- If two events have no outcomes in common, the probability that one or the other occurs is

Basic Probability Rules

- For any event A, $0 \leq P(A) \leq 1$.
- If S is the sample space in a probability model, $P(S) = 1$.
- In the case of equally likely outcomes,

$$P(A) = \frac{number\ of\ outcomes\ corresponding\ to\ event\ A}{total\ number\ of\ outcomes\ in\ the\ sample\ space}$$

- Complement rule: $P(A^c) = 1 - P(A)$.
- Addition rule for mutually exclusive events: If A and B are mutually exclusive,

$$P(A\ or\ B) = P(A) + P(B)$$

Example - Randomly select a student who took the 2013 AP Statistics exam and record the student's score. Here is the probability model:

Score	1	2	3	4	5
Probability	0.233	0.183	0.235	0.224	0.125

(a) Show that this is a legitimate probability model.

(b) Find the probability that the chosen student scored 3 or better.

CHECK YOUR UNDERSTANDING

Choose an American adult at random. Define two events:

A = the person has a cholesterol level of 240 milligrams per deciliter of blood (mg/dl) or above (high cholesterol)
B = the person has a cholesterol level of 200 to 239 mg/dl (borderline high cholesterol)

According to the American Heart Association, $P(A) = 0.16$ and $P(B) = 0.29$.

1. Explain why events A and B are mutually exclusive.
2. Say in plain language what the event "A or B" is. What is $P(A \text{ or } B)$?
3. If C is the event that the person chosen has normal cholesterol (below 200 mg/dl), what's $P(C)$?

3. Two-Way Tables and Probability

When we are trying to find probabilities involving two events, a two-way table can display the sample space in a way that makes probability calculations easy.

Example - What is the relationship between educational achievement and home ownership? A random sample of 500 people who participated in the 2000 census was chosen. Each member of the sample was identified as a high school graduate (or not) and as a homeowner (or not). The two-way table displays the data.

	High School Grad	Not a HS Grad	Total
Homeowner	221	119	340
Not a homeowner	89	71	160
Total	310	190	500

Suppose we choose a member of the sample at random. Find the probability that the member

(a) is a high school graduate.

(b) is a high school graduate and owns a home.

(c) is a high school graduate or owns a home.

General Addition Rule for Two Events

If A and B are any two events resulting from some chance process, then

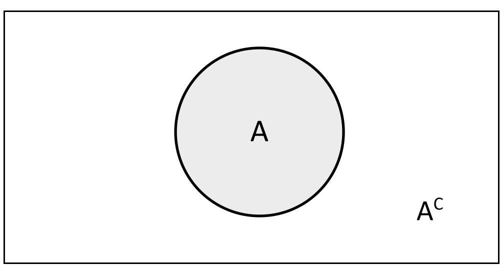
$$P(A\ or\ B) = P(A) + P(B) - P(A\ and\ B)$$

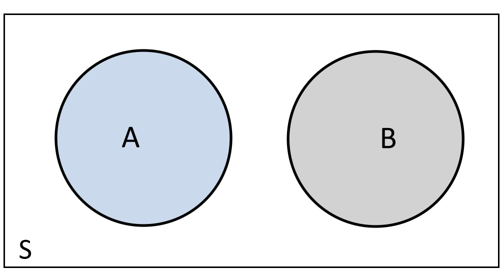
CHECK YOUR UNDERSTANDING

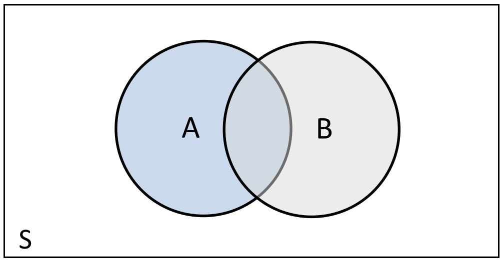
A standard deck of playing cards (with jokers removed) consists of 52 cards in four suits—clubs, diamonds, hearts, and spades. Each suit has 13 cards, with denominations ace, 2, 3, 4, 5, 6, 7, 8, 9, 10, jack, queen, and king. The jack, queen, and king are referred to as "face cards." Imagine that we shuffle the deck thoroughly and deal one card. Let's define events A: getting a face card and B: getting a heart.

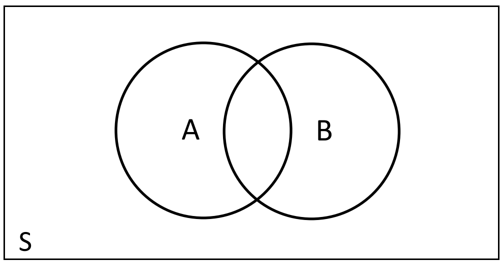
1. Make a two-way table that displays the sample space.
2. Find $P(A \text{ and } B)$.
3. Explain why $P(A \text{ or } B) \neq P(A) + P(B)$. Then use the general addition rule to find $P(A \text{ or } B)$.

4. Venn Diagrams and Probability









Examples – pp. 312-13

Application: According to the National Center for Health Statistics, in December 2012, 60% of US households had a traditional landline telephone, 89% of households had cell phones, and 51% had both. Suppose we randomly selected a US household in December 2012.

a. Make a two-way table that displays the sample space of this chance process.
b. Construct a Venn diagram to represent the outcomes of this chance process.
c. Find the probability that the household has at least two types of phones.
d. Find the probability that the household has a cell phone only.

Section 5.3 - Conditional Probability & Independence (pp. 318-333)

1. Conditional Probability - Let's return to the setting of the homeowners example in Section 5.2.

	High School Grad	Not a HS Grad	Total
Homeowner	221	119	340
Not a homeowner	89	71	160
Total	310	190	500

If we know that a person owns a home, what is the probability that the person is a high school graduate?

If we know that a person is a high school graduate, what is the probability that the person owns a home?

These questions involve **conditional probabilities**. The name comes from the fact that we are trying to find the probability that one event will happen under the *condition* that some other event is already known to have occurred. We often use the phrase *"given that"* to signal the condition.

Definition: The probability that one event happens given that another event is already known to have happened is called a **conditional probability**. Suppose we know that event A has happened. Then the probability that A happens *given that* event B has happened is denoted by P(A | B).

Using this notation, we can restate the answers to our two previous questions:

- P(HS grad | Homeowner) =
- P(Homeowner | HS grad)=

---

2. Calculating Conditional Probabilities

Conditional Probability Formula

To find the conditional probability P(B | A), use the formula

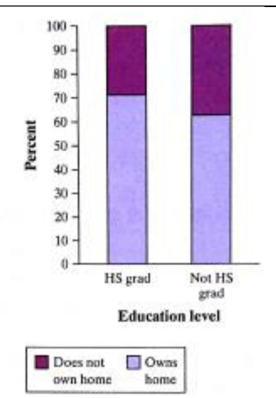
$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

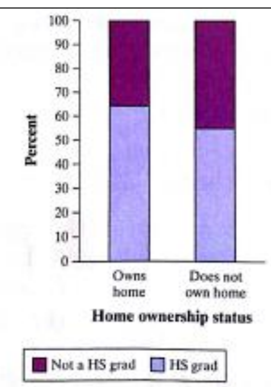
Example: Given the table below, what is the probability that a randomly selected household with a landline also has a cell phone?

	Cell Phone	No Cell Phone	Total
Landline	0.60	0.18	0.78
No Landline	0.20	0.02	0.22
Total	0.80	0.20	1.00

Is there a connection between *conditional probability* and the *conditional distribution* from Chapter 1?

The answer is **yes**. The two segmented bar graphs below display the conditional distributions for the Homeowners example.





CHECK YOUR UNDERSTANDING

Students at the University of New Harmony received 10,000 course grades last semester. The two-way table below breaks down these grades by which school of the university taught the course. The schools are Liberal Arts, Engineering and Physical Sciences, and Health and Human Services.

	Grade Level		
School	**A**	**B**	**Below B**
Liberal Arts	2,142	1,890	2,268
Engineering and Physical Sciences	368	432	800
Health and Human Services	882	630	588

(This table is based closely on grade distributions at an actual university, simplified a bit for clarity.)[10]

College grades tend to be lower in engineering and the physical sciences (EPS) than in liberal arts and social sciences (which includes Health and Human Services). Consider the two events E: the grade comes from an EPS course, and L: the grade is lower than a B.

1. Find $P(L)$. Interpret this probability in context.
2. Find $P(E \mid L)$ and $P(L \mid E)$. Which of these conditional probabilities tells you whether this college's EPS students tend to earn lower grades than students in liberal arts and social sciences? Explain.

3. The General Multiplication Rule

General Multiplication Rule

The probability that events A and B both occur can be found using the **general multiplication rule**

$$P(A \cap B) = P(A) \cdot P(B|A)$$

Where P(B|A) is the conditional probability that event B occurs given that event A has already occurred.

79. **Free downloads?** Illegal music downloading has become a big problem: 29% of Internet users download music files, and 67% of downloaders say they don't care if the music is copyrighted.[15] What percent of Internet users download music and don't care if it's copyrighted? Write the information given in terms of probabilities, and use the general multiplication rule.

pg 319

4. Tree Diagrams and the General Multiplication Rule

Shannon hits the snooze bar on her alarm clock on 60% of school days. If she does not hit the snooze bar, there is a 0.90 probability that she makes it to class on time. However, if she hits the snooze bar, there is only 0.70 probability that she makes it to class on time. In a randomly chosen day, what is the probability that Shannon is late to class?

CHECK YOUR UNDERSTANDING

A computer company makes desktop and laptop computers at factories in three states—California, Texas, and New York. The California factory produces 40% of the company's computers, the Texas factory makes 25%, and the remaining 35% are manufactured in New York. Of the computers made in California, 75% are laptops. Of those made in Texas and New York, 70% and 50%, respectively, are laptops. All computers are first shipped to a distribution center in Missouri before being sent out to stores. Suppose we select a computer at random from the distribution center.[12]

1. Construct a tree diagram to represent this situation.
2. Find the probability that the computer is a laptop. Show your work.

5. Conditional Probability and Independence

Suppose you toss a fair coin twice. Define events A: first toss is a head, and B: second toss is a head. P(A) = 0.5 and P(B) = 0.5. What is P(A|B)? It is the conditional probability that the second toss is a head given that the first toss was a head. The coin has no memory, so P(A|B) = 0.5. In this case P(A|B) = P(A).

Let's contrast the coin-toss scenario with our earlier homeowner example. The events of interest were A: is a high school graduate and B: owns a home. We already learned that P(B) = 340/500 = 0.68 and P(B|A) = 221/310 = 0.712. That is, we know that a randomly selected member of the sample has a 0.68 probability of owning a home. However, if we know that the randomly selected member is a high school graduate, the probability of owning a home increases to 0.712.

Definition. Two events A and B are **independent** if the occurrence of one event has no effect on the chance that the other event will happen. In other words, events A and B are independent if $P(A|B) = P(A) \text{ and } P(B|A) = P(B)$

Example - Is there a relationship between gender and having allergies? To find out, we used the random the CensusAtSchool web site to randomly select 40 U.S. high school students who completed a survey. The two-way table shows the gender of each student and whether the student has allergies.

	Female	Male	Total
Allergies	10	8	18
No Allergies	13	9	22
Total	23	17	40

Are the events "female" and "allergies" independent?

CHECK YOUR UNDERSTANDING

For each chance process below, determine whether the events are independent. Justify your answer.

1. Shuffle a standard deck of cards, and turn over the top card. Put it back in the deck, shuffle again, and turn over the top card. Define events *A*: first card is a heart, and *B*: second card is a heart.
2. Shuffle a standard deck of cards, and turn over the top two cards, one at a time. Define events *A*: first card is a heart, and *B*: second card is a heart.
3. The 28 students in Mr. Tabor's AP Statistics class completed a brief survey. One of the questions asked whether each student was right- or left-handed. The two-way table summarizes the class data. Choose a student from the class at random. The events of interest are "female" and "right-handed."

	Gender	
Handedness	**Female**	**Male**
Left	3	1
Right	18	6

5. Independence: A Special Multiplication Rule - What happens to the general multiplication rule when events A and B are independent?

Multiplication Rule for Independent Events

If A and B are independent events, then the probability that A and B both occur is

$$P(A \cap B) = P(A) \cdot P(B)$$

Example: In baseball, a perfect game is when a pitcher does not allow any hitters to reach base in all nine innings. Historically, pitchers throw a perfect inning - an inning where no hitters reach base - about 40% of the time. So, to throw a perfect game, a pitcher needs to have nine perfect innings in a row.

What is the probability that a pitcher throws nine perfect innings in a row, assuming the pitcher's performance in an inning is independent of his performance in the other innings.

Example: The First Trimester Screening is a noninvasive test given during the first trimester of pregnancy to determine if there are specific chromosomal abnormalities in the fetus. According to the *New England Journal of Medicine* in November 2005, approximately 5% of normal pregnancies will receive a false positive result.

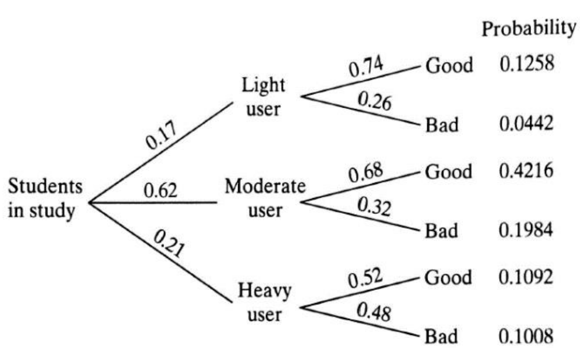
Among 100 women with normal pregnancies, what is the probability that there will be *at least* 1 false positive?

CHECK YOUR UNDERSTANDING

1. During World War II, the British found that the probability that a bomber is lost through enemy action on a mission over occupied Europe was 0.05. Assuming that missions are independent, find the probability that a bomber returned safely from 20 missions.

2. Government data show that 8% of adults are full-time college students and that 30% of adults are age 55 or older. Since (0.08)(0.30) = 0.024, can we conclude that about 2.4% of adults are college students 55 or older? Why or why not?

Example: Given the diagram below, what percent of youth with good grades are heavy users of media?



Example: Many employers require prospective employees to take a drug test. A positive result indicates that the prospective employee uses illegal drugs. However, not all people who test positive actually use drugs. Suppose that 4% of the prospective employees use drugs, the false positive rate is 5% and the false negative rate is 10%.

What percent of people who test positive actually use drugs?