Probability

**Law of large numbers:** if we observe more and more repetitions of any chance process, the proportion of times that a specific outcome occurs approaches a single number.

**Probability:** probability of any 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.

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

(a) **Explain what that means:** If you asked a large sample of U.S. adults whether they usually eat breakfast, about 61% of them would answer “Yes.”

(b) **Why doesn’t this probability say that if 100 U.S. adults are chosen at random, exactly 61 of them usually eat breakfast?** In a random sample of 100 adults, we would expect that around 61 of them will usually eat breakfast. However, the exact # will vary from sample to sample.

**Simulation:** imitation of chance behavior, based on a model that accurately reflects the situation.

**Steps:**

1) What is the question of interest about some chance process?
2) 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) Perform many repetitions.
4) Use the results of your simulation to answer the question of interest.

**Probability Rules**

**Sample Space:** \( (S) \) is a set of all possible outcomes.

**Probability Model:** a description of some chance process that consists of two parts: a sample space \( S \) and a probability for each outcome.

Sample space for tossing a coin: \{H, T\} for rolling two dice \{36 outcomes\}

**Event:** any collection of outcomes from some chance process. That is, an event is a subset of the sample space. Events are usually designated by capital letters, like A, B, C, and so on.

**Basic Rules of Probability:**

- The probability of any event is a number between 0 and 1.
- All possible outcomes together must have probabilities whose sum is 1.
- If all outcomes in the sample space are equally likely, the probability that event A occurs can be found by: \( P(A) = \frac{\# \text{ of outcomes corresponding to event } A}{\text{total } \# \text{ of outcomes in } S} \)
- The probability that an event does not occur is 1 minus the probability that the event does occur. (Complement).
- If two events have no outcomes in common, the probability that one or the other occurs is the sum of their individual probabilities.
**Mutually Exclusive (disjoint):** Two events are mutually exclusive (disjoint) if they have no outcomes in common and so can never happen together. \( P(A \text{ or } B) = P(A) + P(B) \)

**Example:** Distance-learning courses are rapidly gaining popularity among college students. Randomly select an undergraduate student taking a distance-learning course for credit and record the student’s age.

<table>
<thead>
<tr>
<th>Age group (yr)</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>18 to 23</td>
<td>0.57</td>
</tr>
<tr>
<td>24 to 29</td>
<td>0.17</td>
</tr>
<tr>
<td>30 to 39</td>
<td>0.14</td>
</tr>
<tr>
<td>40 and over</td>
<td>0.13</td>
</tr>
</tbody>
</table>

(a) Show that this is a legitimate probability model.
(b) Find the probability that the chosen student is 30 and over.
(c) Find the probability that the chosen student is not in the traditional college age group (18 to 23 years).

**Example:** Students in college statistics class wanted to find out how common it is for young adults to have their ears pierced. They recorded data on two variables—gender and whether the student had a pierced ear—for all 178 people in the class. The two-way table below displays the data:

```
<table>
<thead>
<tr>
<th>Gender</th>
<th>Yes</th>
<th>No</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>19</td>
<td>71</td>
<td>90</td>
</tr>
<tr>
<td>Female</td>
<td>84</td>
<td>4</td>
<td>88</td>
</tr>
<tr>
<td>Total</td>
<td>103</td>
<td>75</td>
<td>178</td>
</tr>
</tbody>
</table>
```

Suppose we choose a student from the class at random. Find the probability that the student:

(a) Has pierced ears
(b) Is a male with pierced ears
(c) Is a male or has pierced ears

**Example:** In an apartment complex, 40% of the residents read *USA Today*. Only 25% read the *New York Times*. Five percent of the residents read both papers. Suppose we select a resident of the apartment complex at random and record which of the two papers the person reads.

(a) Make a two-way table that displays the sample space of this chance process.
(b) Find the probability that the person reads at least one of the two papers.
(c) Find the probability that the person does not read either paper.

```
<table>
<thead>
<tr>
<th>Reads USA Today</th>
<th>Yes</th>
<th>No</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reads NY Times</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>5</td>
<td>20</td>
<td>25</td>
</tr>
<tr>
<td>No</td>
<td>35</td>
<td>40</td>
<td>75</td>
</tr>
<tr>
<td>Total</td>
<td>40</td>
<td>60</td>
<td>100</td>
</tr>
</tbody>
</table>
Blood Types. All human blood can typed as one of O, A, B, or AB, but the distribution of the types varies a bit with race. Here is the distribution of the blood type of a randomly chosen African American:

<table>
<thead>
<tr>
<th>Blood type:</th>
<th>O</th>
<th>A</th>
<th>B</th>
<th>AB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability:</td>
<td>0.49</td>
<td>0.27</td>
<td>0.20</td>
<td>?</td>
</tr>
</tbody>
</table>

(a) What is the probability of type AB blood? Why?
(b) What is the probability that the person chosen does not have type A blood?
(c) Maria has type B blood. She can safely receive blood transfusions from people with blood types O and B. What is the probability that a randomly chosen African American can donate blood to Maria?

Who Eats Breakfast? The two-way table below describes the members of the U.S. senate in a recent year.

<table>
<thead>
<tr>
<th></th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>Democrats</td>
<td>47</td>
<td>13</td>
</tr>
<tr>
<td>Republicans</td>
<td>36</td>
<td>4</td>
</tr>
</tbody>
</table>

(a) Who are the individuals? What variables are being measured?
(b) If we select a U.S. Senator at random, what is the probability that we choose

- A democrat?
- A female?
- A female democrat?
- A female or a Democrat?

Vegetarians. Twenty of the samples of 275 students say they are vegetarians. Of the vegetarians, 9 eat both fish and eggs, 3 eat eggs but not fish, and 8 eat neither. Choose one of the vegetarians at random. What is the probability that the chosen student eats neither fish nor eggs?

Myspace versus Facebook. A recent survey suggests that 85% of college students have posted a profile on Facebook, 54% use MySpace regularly, and 42% do both. Suppose we select a college student at random.

(a) Assuming there are 20 million college students, make a two-way table for this chance process.
(b) Find the probability that the chosen student has posted a profile on Facebook or not a MySpace user.
(c) Find the probability that the chosen student has posted a profile on at least one of these two sites?
Conditional probability and Independence

**Conditional Probability:** The probability that one event happens given that another event is already known to have happened. $P(B|A)$.

**Independent Events:** 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. Events $A$ and $B$ are independent if: $P(A|B) = P(A)$ and $P(B|A) + P(B)$.

Is there a relationship between gender and handedness? To find out, we used CensusAtSchool’s Random Data Selector to choose an SRS of 50 Australian high school students who completed a survey. The two-way table displays the gender and dominant hand of each student.

<table>
<thead>
<tr>
<th>Dominant Hand</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>Right</td>
<td>Left</td>
<td>Total</td>
</tr>
<tr>
<td>Male</td>
<td>20</td>
<td>3</td>
<td>23</td>
</tr>
<tr>
<td>Female</td>
<td>23</td>
<td>4</td>
<td>27</td>
</tr>
<tr>
<td>Total</td>
<td>43</td>
<td>7</td>
<td>50</td>
</tr>
</tbody>
</table>

Are the students “male” and “left-handed” independent? Justify your answer.

$P(LH|male) = 3/23 = 0.13 \quad P(LH) = 7/50 = 0.14$ Since these two probabilities are not equal the two events “left-handed” and “male” are not independent.

**General Multiplication Rule:** $P(A \cap B) = P(A) \cdot P(B|A)$

**Tree Diagrams:**

The Pew Internet and American Life Project finds that 93% of teenagers (ages 12 to 17) use the internet and that 55% of online teens have posted a profile on a social-networking site. What percent of teens are online and have posted a profile?

We know that $P(online) = 0.93$ and $P(profile|online) = 0.55$. 
**General multiplication rule:** \[ P(\text{online and have profile}) = P(\text{online}) \cdot P(\text{profile|online}) = (0.93)(0.55) = 0.5115. \] Or use tree diagram.

Video-sharing sites, led by YouTube, are popular destinations on the Internet. Let’s look at adult Internet users, aged 18 and over. About 27% of adult Internet users are 18 to 29 years old, another 45% are 30 to 49 years old, and the remaining 28% are 50 and over. The Pew Internet and American Life Project finds that 70% of Internet users aged 18 to 29 have visited a video-sharing site, along with 51% of those aged 30 to 49 and 26% of those 50 and older. What percent of Internet users visit YouTube and similar sites?

**Tree diagram:**

- \( P(18) = 0.27 \)
- \( P(30) = 0.45 \)
- \( P(50+) = 0.28 \)

**Conditional probabilities:**

- \( P(\text{video yes|18}) = 0.70 \)
- \( P(\text{video yes|30}) = 0.51 \)
- \( P(\text{video yes|50+}) = 0.26 \)

\[ P(\text{video yes}) = \text{Mult Rule} \]

\[ (.27)(.7) + (.45)(.51) + (.28)(.26) = 0.4913 \]

**Multiplication Rule for independent events:** If \( A \) and \( B \) are independent events, then the probability that \( A \) and \( B \) both occur is \( P(\text{A} \cap \text{B}) = P(A) \cdot P(B) \)

On January 28, 1986, Space Shuttle *Challenger* exploded on takeoff. All seven crew members were killed. Following the disaster, scientists and statisticians helped analyze what went wrong. They determined that the failure of O-ring joints in the shuttle’s booster rockets was to blame. Under the cold conditions that day, experts estimated that the probability that an individual O-ring joint would function properly was 0.977. But there were six of these O-ring joints and all six had to function properly for the shuttle to launch safely. Assuming that O-ring joints succeed or fail independently, find the probability that the shuttle would launch safely under similar conditions.

\[ P(\text{O-ring 1 OK and O-ring 2OK and ... O-ring 6 OK}) = P(1) \cdot P(2) \cdot \cdots \cdot P(6) = (0.977)(0.977)\cdots(0.977) = 0.87 \]

There is an 87% chance that the shuttle would launch safely under similar conditions (and a 13% chance that it would not).

**Note:** As a result of the statistical analysis following the *Challenger* disaster, NASA made important safety changes to the design of the shuttle’s booster rockets.

Finding at least one: \( 1 - P(\text{none}) \)

Many people who come to clinics to be tested for HIV, the virus that causes AIDS, do not come back to learn the test results. Clinics now use “rapid HIV tests” that give a result while the patient waits. In a
clinic in Malawi, for example, use of rapid tests increased the percent of patients who learned their test results from 69% to 99.7%.

The trade-off for fast results is that rapid tests are less accurate than slower laboratory tests. Applied to people who have no HIV antibodies, one rapid test has probability about 0.004 of producing a false positive (that is, of falsely indicating that antibodies are present).

If a clinic tests 200 people who are free of HIV antibodies, what is the chance that at least one false positive will occur?

\[ P(\text{at least one positive}) = 1 - P(\text{no positives}) \]

\[ P(\text{no positives}) = 1 - 0.004 = 0.996. \]

\[ P(\text{no positives in 200 people}) = (0.996)^{200} = 0.4486 \]

\[ P(\text{at least one positive}) = 1 - 0.4486 = 0.5514. \]

The probability is greater than \( \frac{1}{2} \) that at least 1 of the 200 people will test positive for HIV, even though no one has the virus.

Conditional Probability Formula:

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

From Internet user problem, what is the percent of adult Internet users who visit video-sharing sites are aged 18 to 29?

\[ P(18 \text{ to } 29|\text{video yes}) = \frac{P(18 \text{ to } 29 \text{ and video yes})}{P(\text{video yes})} = \frac{0.1890}{0.4913} = 0.3847 \]

Athletes and Drug Testing: Over 10,000 athletes competed in the 2008 Olympic Games in Beijing. The International Olympic Committee (IOC) wanted to ensure that the competition was as fair as possible. So the committee administered more than 5,000 drug tests to the athletes. All medal winners were tested, as well as other randomly selected competitors. Suppose that 2% of athletes had actually taken (banned) drugs. No drug test is perfect. Sometimes the test says that an athlete took drugs, but the athlete actually did not (false positive). Other times, the drug test says an athlete is “clean” but the athlete actually took drugs (false negative). Suppose that the testing procedure used at the Olympics has a false positive rate of 1% and a false negative rate of 0.5%. What is the probability that an athlete who tests positive actually took drugs?

Tree Diagram vs. Two-way table

<table>
<thead>
<tr>
<th>Took Drugs</th>
<th>Positive test</th>
<th>Yes</th>
<th>No</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>199</td>
<td>98</td>
<td>297</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>1</td>
<td>9,702</td>
<td>9,703</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>200</td>
<td>9,800</td>
<td>10,000</td>
<td></td>
</tr>
</tbody>
</table>

\[ P(\text{took drugs}|\text{Test +}) = \frac{P(\text{took drugs} \cap \text{test +})}{P(\text{test +})} = \frac{(0.02)(0.995)}{(0.02)(0.995) + (0.98)(0.01)} = \frac{0.0199}{0.0297} = 0.67 \]
67% of athletes who test positive actually took drugs.

**Probability work:**

**Get Rich** A survey of 4,826 randomly selected young adults (aged 19 to 25) asked, “What do you think are the chances you will have much more than a middle-class income at age 30?” The two-way table shows the responses. Choose a survey respondent at random.

<table>
<thead>
<tr>
<th>Opinion</th>
<th>Female</th>
<th>Male</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Almost no chance</td>
<td>96</td>
<td>98</td>
<td>194</td>
</tr>
<tr>
<td>Some chance but probably not</td>
<td>426</td>
<td>286</td>
<td>712</td>
</tr>
<tr>
<td>A 50-50 chance</td>
<td>696</td>
<td>720</td>
<td>1416</td>
</tr>
<tr>
<td>A good chance</td>
<td>663</td>
<td>758</td>
<td>1421</td>
</tr>
<tr>
<td>Almost certain</td>
<td>486</td>
<td>597</td>
<td>1083</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>2367</td>
<td>2459</td>
<td>4826</td>
</tr>
</tbody>
</table>

(a) Given that the person selected is male, what is the probability that he answered “almost certain”?

(b) If the person selected said “some chance but probably not,” what is the probability that the person is female?

**Tall people and basketball players** Select an adult at random. Define events T: person over 6 feet tall, and B: person is a professional basketball player. Rank the following probabilities from smallest to largest. Justify your answer.

\[
P(T) \quad P(B) \quad P(T \mid B) \quad P(B \mid T)
\]

**Box of chocolates** According to Forrest Gump, “Life is like a box of chocolates. You never know what you’re gonna get.” Suppose a candy maker offers a special “Gump box” with 20 chocolate candies that look the same. In fact, 14 of the candies have soft centers and 6 have hard centers. Choose 2 of the candies from a Gump box at random.

(a) Draw a tree diagram that shows the sample space of this chance process.

(b) Find the probability that one of the chocolates has a soft center and the other one does not.

**Going Pro** Only 5% of male high school basketball, baseball, and football players go on to play at the college level. Of these, 1.7% enter major league professional sports. About 40% of the athletes who compete in college and then reach the pros have a career of more than 3 years. What is the probability that a high school athlete who plays basketball, baseball or football competes in college and then goes on to have a pro career of more than 3 years? Show your work.

**Medical Risks** Morris’ kidneys are failing, and he is awaiting a kidney transplant. His doctor gives him this information for patients in his condition: 90% survive the transplant and 10% die. The transplant succeeds in 60% of those who survive, and the other 40% must return to kidney dialysis. The proportions who survive five years are 70% for those with a new kidney and 50% for those who return to kidney dialysis.
(a) Make a tree diagram to represent this setting.

(b) Find the probability that Morris will survive for five years. Show your work.

**Late Shows** Some TV shows begin after their scheduled times when earlier programs run late. According to a network's records, about 3% of its shows start late. To find the probability that three consecutive shows on this network start on time, can we multiply \(0.97\)(0.97)(0.97)? Why or why not?