STAT 305 Handouts

Kevin Ross

Sep 28, 2025

Table of contents

Preface

A selection of Handouts for STAT 305.

1 Randomness and Probability

Probability comes up in a wide variety of situations. Consider just a few examples.

1. The probability that you roll doubles in a turn of a board game.
2. The probability you win the next Powerball lottery if you purchase a single ticket, 4-8-15-16-42, plus the Powerball number, 23.
3. The probability that a "randomly selected" Cal Poly student is a California resident.
4. The probability that the high temperature in San Luis Obispo next Tuesday is above 90 degrees F.
5. The probability that the Philadelphia Eagles win the next Superbowl.
6. The probability that the Republican candidate wins the 2032 U.S. Presidential Election.
7. The probability that extraterrestrial life currently exists somewhere in the universe.
8. The probability that you ate an apple on April 17, 2009.

Example 1.1. How are the situations above similar, and how are they different? What is one feature that all of the situations have in common? Is the interpretation of "probability" the same in all situations? The goal here is to just think about these questions, and not to compute any probabilities (or to even think about how you would).

- A phenomenon is **random** if there are multiple potential outcomes, and there is **uncertainty** about which outcome will occur.
- Uncertainty is understood in broad terms, and in particular does not only concern future occurrences.
- Many phenomena involve physical randomness, like flipping a coin or drawing powerballs at random from a bin, or in statistical applications of random sampling or random assignment.
- But in many other situations, randomness just vaguely reflects uncertainty.
- Random does *not* mean haphazard. In a random phenomenon, while individual outcomes are uncertain, there is a *regular distribution of outcomes over a large number of (hypothetical) repetitions*.

- Also, random does *not* necessarily mean equally likely. In a random phenomenon, certain outcomes or events might be more or less likely than others.
- The **probability** of an event associated with a random phenomenon is a number in the interval $[0, 1]$ measuring the event's likelihood or degree of uncertainty. A probability can take any value in the continuous scale from 0% to 100%.
- There are two main interpretations of probability.
  - **Long run relative frequency.** The probability of an event can be interpreted as the proportion of times that the event would occur in a very large number of hypothetical repetitions of the random phenomenon.
  - **Subjective probability.** There are many situations where the outcome is uncertain, but it does not make sense to consider the situation as repeatable. In such situations, a subjective (a.k.a., personal) probability describes the degree of likelihood a given individual ascribes to a certain event. Think of subjective probabilities as measuring relative degrees of likelihood rather than long run relative frequencies.
- Fortunately, the mathematics of probability work the same way regardless of the interpretation. In either case, the same basic logical consistency requirements must be satisfied.
- A **simulation** involves an artificial recreation of the random phenomenon, usually using a computer. The probability of an event can be approximated by simulating the random phenomenon a large number of times and determining the proportion of simulated repetitions on which the event occurred out of the total number of repetitions in the simulation.

Example 1.2. One of the oldest documented problems in probability is the following: If three fair six-sided dice are rolled, what is more likely: a sum of 9 or a sum of 10?

1. Explain how you could conduct a *simulation* to investigate this question.

2. In 1 million repetitions of a simulation, a sum of 9 occurred in 115392 repetitions and a sum of 10 occurred in 125026 repetitions. Use the simulation results to approximate the probability that the sum is 9; repeat for a sum of 10.

3. It can be shown that the theoretical probability that the sum is 9 is $25/216 = 0.116$. Write a clearly worded sentence interpreting this probability as a long run relative

frequency.

4. It can be shown that the theoretical probability that the sum is 10 is $27/216 = 0.125$. How many times more likely is a sum of 10 than a sum of 9?

Example 1.3. The weather forecast calls for a 30% chance of rain in your city tomorrow. You ask Donny Don't to interpret the 30% as a long run relative frequency. Donny says: "it will rain in 30% of the city tomorrow". You ask him to elaborate; he says: "Well, there are many different locations in the city. In some of the locations it will rain, in some it won't. It will rain in 30% of the locations, and not in the other 70%. That is, rain will cover 30% of the area of the city, and the other 70% won't have rain." Do you agree? If not, how would you interpret the 30% as a long run relative frequency?

- When interpreting the long run, be careful to define the random phenomenon that is being repeated

Example 1.4. In the first 7 games of his NBA career, Paolo Banchero attempted 60 free throws and successfully made 44. Donny Don't says "the probability that Paolo Banchero successfully makes a free throw attempt is $44/60 = 0.733$." Do you agree? Explain.

- Distinguish between the short run and the long run
- Observed relative frequencies based on past data (sometimes called "empirical probabilities") are only short run approximations to theoretical probabilities which represent long run relative frequencies

Example 1.5. Your favorite local weatherperson forecasts a 30% chance of rain tomorrow and a 60% chance of rain the next day in your city.

1. Explain how these probabilities are subjective.

2. You ask Donny Don't to interpret these values as relative degrees of likelihood. Donny says: "Well, 30% is not that big, so it's not going to rain that hard tomorrow. Also, 60% is twice is big as 30%, so it's going to rain twice as hard two days from now as it will tomorrow". Do you agree? Explain.

3. Donny says: "Can't we just look at the data from all the days with weather conditions similar to the ones forecast for tomorrow, and see how often it rained on those days to find the probability of rain tomorrow? No subjectivity about that!" How would you respond?

- A **probabilistic forecast** combines observed data and statistical or mathematical models to make predictions.
- Rather than providing a single prediction such as "it will rain tomorrow", probabilistic forecasts provide a range of scenarios and their relative likelihoods.
- Such forecasts are subjective in nature, relying upon the data used and assumptions of the model.
- Changing the data or assumptions can result in different forecasts and probabilities.

Example 1.6. What is your subjective probability that Professor Ross (the author) has a TikTok account? Consider the following two bets, and suppose you must choose only one.

A) You win \$100 if Professor Ross has a TikTok account, and you win nothing otherwise.
B) A box contains 40 green and 60 gold marbles that are otherwise identical. The marbles are thoroughly mixed and one marble is selected at random. You win \$100 if the selected marble is green, and you win nothing otherwise.

1. Which of the above bets would you prefer? Or are you completely indifferent? What does this say about your subjective probability that Professor Ross has a Tik Tok account?
2. If you preferred bet B to bet A, consider bet C which has a similar setup to B but now there are 20 green and 80 gold marbles. Do you prefer bet A or bet C? What does this say about your subjective probability that Professor Ross has a Tik Tok account?
3. If you preferred bet A to bet B, consider bet D which has a similar setup to B but now there are 60 green and 40 gold marbles. Do you prefer bet A or bet D? What does this say about your subjective probability that Professor Ross has a Tik Tok account?
4. Continue to consider different numbers of green and gold marbles. Can you zero in on your subjective probability?

Example 1.7. As of Jun 13, FanGraphs listed the following probabilities for who will win the 2025 MLB World Series.

Team	Probability
Dodgers	21%
Yankees	17%
Tigers	10%
Mets	10%
Phillies	7%
Other	

According to FanGraphs (as of Jun 13):

1. Are the above percentages relative frequencies or subjective probabilities? Why?

2. What must be the probability that a team other than the above five teams wins the championship? That is, what value goes in the "Other" row in the table?

3. The Dodgers are how many times more likely than the Phillies to win?

4. What must be the probability that the Dodgers do *not* win the championship? How many times more likely are the Dodgers to not win than to win (this ratio is the "odds against" the Dodgers winning).

5. How could you construct a circular spinner (like from a kids game) to simulate the World Series champion according to these probabilities? According to this model, what would you expect the results of 10000 repetitions of a simulation of the champion to look like?

Example 1.8. Suppose your subjective probabilities for the 2025 World Series champion satisfy the following conditions.

- The Brewers and Yankees are equally likely to win
- The Phillies are 1.5 times more likely than the Yankees to win
- The Dodgers are 2 times more likely than the Phillies to win
- The winner is as likely to be among these four teams — Brewers, Yankees, Phillies, Dodgers — as not.

Construct a table of your subjective probabilities like the one in Example 1.7.

- The previous examples illustrate two interpretations of probability: long run relative frequencies and subjective probabilities.
- We will use these interpretations interchangeably.
- With subjective probabilities it is often helpful to consider what might happen in a simulation.
- It is also useful to consider long run relative frequencies in terms of relative degrees of likelihood.
- Fortunately, the mathematics of probability work the same way regardless of the interpretation.
- A probability takes a value in the sliding scale from 0 to 1 (or 0 to 100%).
- Don't just focus on computation; always remember to properly interpret probabilities.

Example 1.9. Consider a Cal Poly student who frequently has blurry, bloodshot eyes, generally exhibits slow reaction time, always seems to have the munchies, and disappears at 4:20 each day. Which of the following, A or B, has a higher probability? Assume the two probabilities are not equal.

- A: The student has a GPA above 3.0.
- B: The student has a GPA above 3.0 and smokes marijuana regularly.

- Warning! Your psychological judgment of probabilities is often inconsistent with the mathematical logic of probabilities.

Example 1.10. Ron and Leslie agree to the following bet. They'll ask Professor Ross if he has a TikTok account. If he does, Leslie will pay Ron \$200; if not, Ron will pay Leslie \$100. (Neither has any direct information about whether or not Professor Ross has a TikTok account.)

1. Given this setup, which of the following is being judged as more likely: that Professor Ross has a TikTok account, or that he does not? Why?

2. What are this bet's "odds"?

3. Ron and Leslie agree that this is a fair bet, and neither would accept worse odds. What is their subjective probability that Professor Ross has a TikTok account?

4. Suppose they were to hypothetically repeat this bet many times, say 3000 times. Given the probability from the previous part, how many times would you expect Leslie to win? To lose? What would you expect Leslie's net dollar winnings to be? In what sense is this bet "fair"? (Remember: Leslie's winnings are Ron's losses and vice versa.)

- The **odds** of an event is a ratio involving the probability that the event occurs and the probability that the event does not occur.
- Odds can be expressed as either "in favor" of or "against" the event occurring, depending on the order of the ratio.

2 Working with Probabilities

- It is often helpful to think of probabilities as percentages or proportions.
- Furthermore, when working with multiple percentages, it is also helpful to construct hypothetical **two-way tables** (a.k.a., contingency tables) of counts.
- For the purposes of constructing the table and computing related probabilities, any value can be used for the hypothetical total count.
- When dealing with percentages (or proportions or probabilities) be sure to ask "percent *of what*?" Thinking in fraction terms, be careful to identify the correct reference group which corresponds to the denominator.

Example 2.1. Do American adults (18+) think it's acceptable to curse out loud in public? Assume[1] that

- 62% of American adults age 18-29 think it is acceptable
- 45% of American adults age 30-49 think it is acceptable
- 24% of American adults age 50-64 think it is acceptable
- 11% of American adults age 65+ think it is acceptable

Also assume that among American adults (18+)

- 20% of American adults are age 18-29
- 33% of American adults are age 30-49
- 25% of American adults are age 50-64
- 22% of American adults are age 65+

1. Consider a hypothetical group of 10000 American adults and assume the percentages provided apply to this group. Fill in the counts in each of the cells of the following table.

	18-29	30-49	50-64	65+	Total
Acceptable					
Not acceptable					
Total					10000

[1]The values in this problem are based on a March 12, 2025 article by the Pew Research Center.

2. Randomly select an American adult *from this group of 10000.* Compute the probability that they think cursing out loud in public is acceptable.

3. Randomly select an *American adult.* Compute the probability that they think cursing out loud in public is acceptable. (Hint: did the 10000 matter?)

4. Compute the probability that an American adult who thinks cursing out loud in public is acceptable is age 18-29. (Answer with both an unreduced fraction and a demical/percent.)

5. Compute the probability that an American adult who is age 18-29 thinks cursing out loud in public is acceptable. (Answer with both an unreduced fraction and a demical/percent.)

6. Compute the probability that an American adult is age 18-29 and thinks cursing out loud in public is acceptable. (Answer with both an unreduced fraction and a demical/percent.)

7. Compare the unreduced fractions for the previous three parts. What is the same? What is different?

8. Suppose that we were only told that 35.67% of American adults overall think cursing out loud in public is acceptable, and that we not given the values 62%, 45%, 24%, 11%. Would we be able to complete the two-way table?

- **Warning!** In general, knowing probabilities of individual events alone is not enough to determine probabilities of combinations of them.

Example 2.2. Suppose that 47% of American adults[2] have a pet dog and 25% have a pet cat.

1. Donny Don't says that 72% (which is 47% + 25%) of American adults have a pet dog or a pet cat. Is that necessarily true? If not, is it even possible (in principle anyway) for this to be true? Under what circumstance (however unrealistic) would this be true? Construct a corresponding two-way table.

2. Given only the information provided, what is the smallest possible percentage of American who adults have a pet dog or a pet cat? Under what circumstance (however unrealistic) would this be true? Construct a corresponding two-way table.

3. Donny Don't says that 11.75% (which is 47% $\times$ 25%) of Americans have both a pet dog *and* a pet cat. Explain to Donny why that's not necessarily true. Without further information, what can you say about the percent of American adults who have both a pet dog and a pet cat?

[2]These values are based on the 2018 General Social Survey.

4. Suppose that 14% of American adults have both a pet dog *and* a pet cat. What is the percentage of American adults who have a pet dog *or* a pet cat? Construct a corresponding two-way table. Use your table to show Donny how to correct his error from part 1.

5. What percentage of American adults who have a pet dog also have a pet cat? Is it 25%?

6. What percentage of American adults who do not have a pet dog have a pet cat? Is this the same value as in the previous part?

7. What percentage of American adults who have a pet cat also have a pet dog? Is it 47%?

8. Describe in words the percentage that results from subtracting the answer to the previous part from 100%.

Example 2.3. A woman's chances of giving birth to a child with Down syndrome increase with age. The CDC estimates[3] that a woman in her mid-to-late 30s has a risk of conceiving a child with Down syndrome of about 1 in 250. A nuchal translucency (NT) scan, which involves a blood draw from the mother and an ultrasound, is often performed around the 13th week of pregnancy to test for the presence of Down syndrome (among other things). If the baby has Down syndrome, the probability that the test is positive is about 0.9. However, when the baby does not have Down syndrome, there is still a probability that the test returns a (false) positive of about[4] 0.05. Suppose that the NT test for a pregnant woman in her mid-to-late 30s comes back positive for Down syndrome. What is the probability that the baby actually has Down syndrome?

1. Before proceeding, make a guess for the probability in question.

   0-20% 20-40% 40-60% 60-80% 80-100%

2. Donny Don't says: 0.90 and 0.05 should add up to 1, so there must be a typo in the problem. Do you agree?

3. Construct a hypothetical two-way table of counts to represent the given information.

4. Use the table to find the probability in question: If NT test for a pregnant woman in her mid-to-late 30s is positive, what is the probability that the baby actually has Down syndrome?

5. The probability in the previous part might seem very low to you. Explain why the probability is so low.

[3]Source: http://www.cdc.gov/ncbddd/birthdefects/downsyndrome/data.html

[4]Estimates of these probabilities vary between different sources. The values in the exercise were based on https://www.ncbi.nlm.nih.gov/pubmed/17350315

6. Compare the probability of having Down Syndrome before and after the positive test. How much more likely is a baby who tests positive to have Down Syndrome than a baby for whom no information about the test is available?

- Remember to ask "percentage *of what*"? For example, the percentage of *babies who have Down syndrome* that test positive is a very different quantity than the percentage of *babies who test positive* that have Down syndrome.
- Probabilities are often conditional on information.
- Conditional probabilities (e.g., probability of Down Syndrome *given a positive test*) can be highly influenced by the original unconditional probabilities (e.g. probability of Down Syndrome), sometimes called the **base rates**. Don't neglect the base rates when evaluating probabilities.
- The example illustrates that when the base rate for a condition is very low and the test for the condition is less than perfect there will be a relatively high probability that a positive test is a *false positive.*

3 Interpreting Probabilities and "Expected" Values

- A probability takes a value in the sliding scale from 0 to 100%.
- Don't just focus on computation; always remember to properly interpret probabilities.

Example 3.1. In each of the following parts, which of the two probabilities, a or b, is larger, or are they equal? You should answer conceptually without attempting any calculations. Explain your reasoning.

1. Randomly select a man.

   a. The probability that a randomly selected man who is greater than six feet tall plays in the NBA.
   b. The probability that a randomly selected man who plays in the NBA is greater than six feet tall.

2. Randomly select a baby girl who was born in 1950.

   a. The probability that a randomly selected baby girl born in 1950 is alive today.
   b. The probability that a randomly selected baby girl born in 1950, who was alive at the end of 2020, is alive today.

- A probability is a measure of the likelihood or degree of uncertainty or plausibility of an event.
- A "conditional" probability revises this measure to reflect any additional information about the outcome of the underlying random phenomenon.
- In a sense, all probabilities are conditional upon some information, even if that information is vague ("well, it has to be one of these possibilities"). Be careful to clearly identify what information is reflected in probabilities
- When interpreting probabilities, consider the conditions under which the probabilities were computed, in the proper direction

Example 3.2. In each of the following parts, which of the two probabilities, a or b, is larger, or are they equal? You should answer conceptually without attempting any calculations. Explain your reasoning.

1. Flip a coin *which is known to be fair* 10 times.

a. The probability that the results are, in order, HHHHHHHHHH.
b. The probability that the results are, in order, HHTHTTTHHT.

2. Flip a coin which is known to be fair 10 times.

   a. The probability that all 10 flips land on H.
   b. The probability that exactly 5 flips land on H.

3. In the Powerball lottery there are roughly 300 million possible winning number combinations, all equally likely.

   a. The probability you win the next Powerball lottery if you purchase a single ticket, 4-8-15-16-42, plus the Powerball number, 23
   b. The probability you win the next Powerball lottery if you purchase a single ticket, 1-2-3-4-5, plus the Powerball number, 6.

4. Continuing with the Powerball

   a. The probability that the numbers in the winning number are not in sequence (e.g., 4-8-15-16-42-23)
   b. The probability that the numbers in the winning number are in sequence (e.g., 1-2-3-4-5-6)

5. Continuing with the Powerball

   a. The probability that you win the next Powerball lottery if you purchase a single ticket.
   b. The probability that someone wins the next Powerball lottery. (FYI: especially when the jackpot is large, there are hundreds of millions of tickets sold.)

- When interpreting probabilities, be careful not to confuse "the particular" with "the general".

  - **"The particular:"** A very specific event, surprising or not, often has low probability.
  - **"The general:"** While a very specific event often has low probability, if there are many like events their combined probability can be high.

- Even if an event has extremely small probability, given enough repetitions of the random phenomenon, the probability that the event occurs on *at least one* of the repetitions is often high.
- In general, even though the probability that something very specific happens to you today is often extremely small, the probability that something similar happens to someone some time is often quite high.

- When assessing a numerical probability, always ask “probability of what”? Does the probability represent “the particular” or “the general”? Is it the probability that the event happens in a single occurrence of the random phenomenon, or the probability that the event happens at least once in many occurrences?
- Also distinguish between assumption and observation. For example, if you assume that a coin is fair and the flips are independent, then all possible H/T sequences are equally likely. However, if you observe the coin landing on heads on 20 flips in a row, then that might cast doubt on your assumption that the coin is fair.

Example 3.3. Shuffle a standard deck of 52 playing cards (13 face values in each of 4 suits) and deal two cards without replacement.

1. What is the probability that the first card dealt is a heart?
2. What is the probability that the second card dealt is a heart?
3. What is the probability that the second card dealt is a heart if the first card dealt is a heart?
4. What is the probability that the second card dealt is a heart if the first card dealt is not a heart?
5. Revisit part 2. What is the probability that the second card dealt is a heart? Create a two-way table to answer this question.

- Be careful to distinguish between conditional and unconditional probabilities.
- A conditional probability reflects additional information about the outcome of the random phenomenon.
- In the absence of such information, we must continue to account for all the possibilities.
- When computing probabilities, be sure to only reflect information that is known. Especially when considering a phenomenon that happens in stages, don't assume that when considering what happens second that you know what happened first.

Example 3.4. Within both the colleges of Agriculture and Architecture at Cal Poly, about 49% of admitted students are female, about 84% of admitted students went to high school in CA, and the median GPA of admitted students is about 4.1.

An orientation group of 100 newly admitted Cal Poly students includes 75 students in Agriculture and 25 students in Architecture. A student is randomly selected from this group. The selected student is Maddie, who is female, went to high school in CA, and had a high school GPA of 4.1.

Donny Don't says, "The information about Maddie applies equally well to Agriculture or Architecture and doesn't help us decide which college she's in, so it's just 50/50. Given the information about Maddie, the conditional probability that she is in Agriculture is 0.5." Do you agree? If not, what is the conditional probability that Maddie is in the college of Agriculture given the information about her?

Example 3.5. This is a very simplified example illustrating the basic idea of how insurance works. Every year an insurance company sells many thousands of car insurance policies to drivers within a particular risk class. Each policyholder pays a "premium" of \$1000 at the start of the year, and the insurance company agrees to pay for the cost of all damages that occur during the year. Suppose that each policy incurs damage of either \$0, \$5000, \$20000, or \$50000 with the following probabilities.

Amount of damage (\$)	Profit (\$)	Probability
0	1000	0.910
5000	-4000	0.070
20000	-19000	0.019

Amount of damage ($)	Profit ($)	Probability
50000	-49000	0.001

The insurance company's profit on a policy at the end of the year is the difference between the premium of $1000 and any damage paid out. For example, a policy that incurs no damage results in a profit of $1000; a policy that incurs $5000 in damage results in a profit of -$4000 (that is, a loss of $4000) for the insurance company.

1. Interpret the probabilities 0.91, 0.07, 0.019, and 0.001 as long run relative frequencies in this context.

2. Compute the probability that a policy results in a positive profit for the insurance company.

3. Imagine 100,000 hypothetical policies. How many of these policies would you expect to result in a profit of $1000? -$4000? -$19000? -$49000?

4. What do you expect the total profit for these 100,000 policies to be?

5. What do you expect the average profit per policy for these 100,000 policies to be?

6. Compute the probability that a policy has a profit equal to the value from part 5.

7. Compute the probability that a policy has a profit greater than the value from part 5.

8. Is the value from part 5 the most likely value of profit for a single policy?

9. Is the value from part 5 the profit you would expect for a single policy?

10. Explain in what sense the value from part 5 is "expected".

- The long run average value of a random quantity is called its "expected value".
- Be careful: the term "expected value" is somewhat of a misnomer.
- The expected value is *not* necessarily the value we expect on a single repetition of the random phenomenon, nor the most likely value (or even a possible value).
- Rather, the expected is the value we expect to see on average in the long run over many repetitions.
- A probability can be interpreted as a long run relative frequency; an expected value can be interpreted as a long run average value.

Example 3.6. Continuing Example 3.5. We considered what we would expect for 100000 hypothetical policies, but what about an unspecified large number of policies?

1. Imagine that we have recorded the profit for each of a large number of policies (not necessarily 100000). Explain in words the process by which you would compute the average profit per policy. (In other, more general, words: how do you compute an average of a list of numbers?)

2. Given that the profit of any policy is either 1000, -4000, -19000, or -49000, how could we simplify the calculation of the sum in the previous part? Write a general expression for the average profit per policy in this scenario.

3. What do you think the expression in the previous part converges to in the long run?

4. Explain how the value in the previous part is a "probability-weighted average value".

5. Compute the expected value of damage (not profit) as a probability-weighted average value.

6. Interpret the value from the previous part as a long run average value in this context.

7. How is the expected value of profit related to the expected value of damage? Does this make sense? Why?

- An expected value can be computed as a “probability-weighted average value”
- But this is just a more compact way of computing an average in the usual way: add up all the values and divide by the number of values.

4 Outcomes, Events, and Random Variables

Probability models can be applied to any situation in which there are multiple potential outcomes and there is uncertainty about which outcome will occur.

4.1 Outcomes

- Due to the wide variety of types of random phenomena, an **outcome** can be virtually anything
- In particular, an outcome does *not* have to be a number.
- The **sample space**, denoted Ω (the uppercase Greek letter "omega"), is the set of all possible outcomes of a random phenomenon. An **outcome**, denoted ω (the lowercase Greek letter "omega"), is an element of the sample space: $\omega \in \Omega$.
- Mathematically, the sample space Ω is a *set* containing all possible outcomes, while an individual outcome ω is a *point* or *element* in Ω.
- A random phenomenon is modeled by a *single* sample space, with respect to which all objects (events, random variables) are defined. Whenever possible, a sample space outcome should be defined to provide the maximum amount of information about the outcome of random phenomenon.
- In practice we rarely enumerate the sample space as we'll for some of the examples in this class. Nonetheless, there is always some underlying sample space corresponding to all possible outcomes of the random phenomenon.

Example 4.1. Roll a four-sided die twice, and record the result of each roll in sequence as an ordered pair. For example, the outcome $(3, 1)$ represents a 3 on the first roll and a 1 on the second; this is not the same outcome as $(1, 3)$.

1. How many possible outcomes are there? Identify the sample space.

2. We might be interested in the sum of the two dice. Explain why it is still advantageous to define the sample space as in the previous part, rather than as $\Omega = \{2, \dots, 8\}$.

- **Multiplication principle for counting** Suppose that stage 1 of a process can be completed in any one of n_1 ways. Further, suppose that for each way of completing the stage 1, stage 2 can be completed in any one of n_2 ways. Then the two-stage process can be completed in any one of $n_1 \times n_2$ ways.
- This rule extends naturally to a ℓ-stage process, which can then be completed in any one of $n_1 \times n_2 \times n_3 \times \cdots \times n_\ell$ ways.
- It is not important whether there is a "first" or "second" stage. What is important is that there are distinct stages, each with its own number of "choices".

Example 4.2. The "matching problem" is one well known probability problem. The general setup involves n distincts objects labeled $1, \dots, n$ which are placed in n distinct boxes labeled $1, \dots, n$, with exactly one object placed in each box.

1. Consider the matching problem with $n = 3$. Label the objects 1, 2, 3, and the spots 1, 2, 3, with spot 1 the correct spot for object 1, etc. Specify an appropriate definition of an outcome, determine the number of outcomes, and specify the sample space.

2. For a general n, how many possible outcomes are there?

Example 4.3. Regina and Cady plan to meet for lunch They will definitely arrive between noon and 1, but their exact arrival times are uncertain. Rather than dealing with clock time, it is helpful to represent noon as time 0 and measure time as minutes after noon, including fractions of a minute, so that arrival times take values in the continuous interval [0, 60].

Specify an appropriate definition of an outcome and draw a picture representing the sample space.

4.2 Events

- An event is something that could happen or might be true.
- An *event* is a collection of outcomes that satisfy some criteria.
- Mathematically, an **event** A is a *subset* of the sample space: $A \subseteq \Omega$.
- Events are typically denoted with capital letters near the start of the alphabet, with or without subscripts (e.g. A, B, C, A_1, A_2). Events can be composed from others using basic set operations like unions ($A \cup B$), intersections ($A \cap B$), and complements (A^c).
  - Read A^c as "not A".
  - Read $A \cap B$ as "A and B"
  - Read $A \cup B$ as "A or B". Note that unions ($\cup$, "or") are always inclusive. $A \cup B$ occurs if A occurs but B does not, B occurs but A does not, or both A and B occur.
- A collection of events $A_1, A_2, ...$ are **disjoint** (a.k.a. mutually exclusive) if $A_i \cap A_j = \emptyset$ for all $i \neq j$. That is, multiple events are disjoint if none of the events have any outcomes in common.
- If the sample space outcomes are represented by rows in a spreadsheet, then an event is a subset of rows that satisfies some criteria

Example 4.4. Matching problem. For $n = 3$, objects labeled 1, 2, 3, are placed at random in spots labeled 1, 2, 3, with spot 1 the correct spot for object 1, etc. Using the sample space from Example 4.2, identify the following events.

1. B, the event that no objects are put in the correct spot.

2. In words, what does B^c represent?

3. A, the event that all objects are put in the correct spot.

4. C, the event that exactly 2 objects are put in the correct spot.

5. A_3, the event that object 3 is put (correctly) in spot 3.

6. For a general n let A the event that all objects are put in the correct spot, let B the event that no objects are put in the correct spot, and let A_i be the event that object i is put (correctly) in spot i, $i = 1, \dots, n$. What is the relationship between A and $A_1, \dots, A_n$? What is the relationship between B and $A_1, \dots, A_n$?

Example 4.5. Using the sample space from Example 4.3, identify the following events using pictures.

1. Identify A, the event that Regina arrives after Cady.

2. Identify B, the event that either Regina or Cady arrives before 12:30.

3. Identify C, the event that they arrive within 15 minutes of each other.

4. Identify D, the event that Regina arrives before 12:24.

4.3 Random variables

- Roughly, a *random variable* assigns a number measuring some quantity of interest to each outcome of a random phenomenon.
- Mathematically, a **random variable (RV)** X is a *function* that takes an outcome in the sample space as input and returns a real number as output
- The random variable itself is typically denoted with a capital letter (X); possible values of that random variable are denoted with lower case letters (x).
  - Think of the capital letter X as a label standing in for a formula like "the number of heads in 4 flips of a coin" and
  - x as a dummy variable standing in for a particular value like 3.
- **Discrete random variables** take at most countably many possible values (e.g., $0, 1, 2, ...$). They are often counting variables (e.g., the number of Heads in 10 coin flips).
- **Continuous random variables** can take any real value in some interval (e.g., $[0, 1]$, $[0, \infty)$, $(-\infty, \infty)$.). That is, continuous random variables can take uncountably many different values. Continuous random variables are often measurement variables (e.g., height, weight, income).
- A function of a random variable is also a random variable: if X is a RV then so is $g(X)$
- Sums and products, etc., of random variables *defined on the same sample space* are random variables. IfX and Y are RVs defined on the same sample space then so are $X + Y$, $X - Y$, XY
- If the sample space outcomes are represented by rows in a spreadsheet, then random variables correspond to columns.
- Expressions like $X = x$ or $\{X = x\}$ represent *events*: for which outcomes is the value of the random variable X equal to the value x

Example 4.6. Roll a four-sided die twice, and record the result of each roll in sequence. Recall the sample space from Example 4.1. Let X be the sum of the two dice, and let Y be the larger of the two rolls (or the common value if both rolls are the same).

Outcome	X	Y
(1, 1)		
(1, 2)		
(1, 3)		
(1, 4)		
(2, 1)		
(2, 2)		
(2, 3)		
(2, 4)		
(3, 1)		
(3, 2)		
(3, 3)		
(3, 4)		
(4, 1)		
(4, 2)		
(4, 3)		
(4, 4)		

1. Construct a table identifying the values of X and Y for each outcome in the sample space.

2. Identify the possible values of X.

3. Identify the possible values of Y.

4. Identify the possible values of the pair (X, Y).

5. Identify $\{Y = 1\}$.

6. Identify $\{Y = 2\}$.

7. Identify $\{Y = 3\}$.

8. Identify $\{Y = 4\}$.

9. Identify $\{X \leq 4\}$.

10. Identify $\{X = 4, Y = 3\}$

Example 4.7. Matching problem. For $n = 3$ objects labeled 1, 2, 3, are placed at random in spots labeled 1, 2, 3, with spot 1 the correct spot for object 1, etc. Recall the sample space from Example 4.2. Let the random variable X count the number of objects that are put in the correct spot. Let I_1 be equal to 1 if object 1 is placed (correctly) in spot 1, and define I_2, I_3 similarly.

Outcome	X	I_1	I_2	I_3
123				
132				
213				
231				
312				
321				

1. Construct a table identifying the value of X, I_1, I_2, I_3 for each outcome in the sample space.
2. Identify the possible values of X.

3. What is the relationship between I_3 and event A_3 from Example 4.4?

4. How can you express X in terms of I_1, I_2, I_3?

- The **indicator (a.k.a., Bernoulli) random variable** corresponding to event A is equal to 1 if A occurs and 0 otherwise
- Indicator random variables can be used for incremental counting and are often useful in problems involving "find the expected number of"

Example 4.8. Regina and Cady will definitely arrive between noon and 1, but their exact arrival times are uncertain. Recall the sample space from Example 4.3. Let R be the random variable representing Regina's arrival time (minutes after noon), and Y for Cady.

1. What does the random variable $T = \min(R, Y)$ represent? What are the possible values of T?

2. What does the random variable $W = |R - Y|$ represent? What are the possible values of W?

3. Let N be the number of people (out of 2) who arrive before 12:30. How can you represent N in terms of R and Y. (Hint: use indicators.)

4. Identify each of the random variables above as discrete or continuous.

5. Interpret each of the following in words and draw a picture representing it.

   a. $\{R > Y\}$.

   b. $\{T < 30\}$.

   c. $\{W < 15\}$.

d. $\{R < 24\}$.

5 Probability Models

- A **probability measure**, typically denoted P, assigns probabilities to *events* to quantify their relative likelihoods according to the assumptions of the model of the random phenomenon.
- The probability of event A, computed according to probability measure $\mathrm{P}(A)$, is denoted $\mathrm{P}(A)$.
- A valid probability measure P must satisfy the following three logical consistency "axioms".
  - For any event A, $0 \le \mathrm{P}(A) \le 1$.
  - If Ω represents the sample space then $\mathrm{P}(\Omega) = 1$.
  - (*Countable additivity.*) If $A_1, A_2, A_3, \dots$ are disjoint then

$$\mathrm{P}(A_1 \cup A_2 \cup A_2 \cup \cdots) = \mathrm{P}(A_1) + \mathrm{P}(A_2) + \mathrm{P}(A_3) + \cdots$$

- Additional properties of a probability measure follow from the axioms
  - *Complement rule.* For any event A, $\mathrm{P}(A^c) = 1 - \mathrm{P}(A)$.
  - *Subset rule.* If $A \subseteq B$ then $\mathrm{P}(A) \le \mathrm{P}(B)$.
  - *Addition rule for two events.* If A and B are any two events

$$\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \cap B)$$

  - *Law of total probability.* If $C_1, C_2, C_3 \dots$ are disjoint events with $C_1 \cup C_2 \cup C_3 \cup \cdots = \Omega$, then

$$\mathrm{P}(A) = \mathrm{P}(A \cap C_1) + \mathrm{P}(A \cap C_2) + \mathrm{P}(A \cap C_3) + \cdots$$

- A **probability model** (or **probability space**) is the collection of all outcomes, events, and random variables associated with a random phenomenon along with the probabilities of all events of interest under the assumptions of the model.
- The axioms of a probability measure are minimal logical consistent requirements that ensure that probabilities of different events fit together in a valid, coherent way.
- A single probability measure corresponds to a particular set of assumptions about the random phenomenon.
- There can be many probability measures defined on a single sample space, each one corresponding to a different probability model for the random phenomenon.
- Probabilities of events can change if the probability measure changes.

Example 5.1. Consider a *single* roll of a four-sided die. (Careful: don't confuse these examples with other examples that involve two rolls.) The sample space is $\{1, 2, 3, 4\}$. Table 5.1 lists all possible *events*.

1. Add a description in words for each of the events

2. Suppose the die is fair, and let P denote the correponding probability measure. Compute $\mathrm{P}(A)$ for each event in Table 5.1.

Table 5.1: All possible events associated with a single roll of a four-sided die.

A	Description	$\mathrm{P}(A)$	$\mathrm{Q}(A)$	$\tilde{\mathrm{Q}}(A)$?
$\emptyset$				
$\{1\}$				
$\{2\}$				
$\{3\}$				
$\{4\}$				
$\{1, 2\}$				
$\{1, 3\}$				
$\{1, 4\}$				
$\{2, 3\}$			0.5	
$\{2, 4\}$				
$\{3, 4\}$			0.7	
$\{1, 2, 3\}$			0.6	
$\{1, 2, 4\}$				
$\{1, 3, 4\}$				
$\{2, 3, 4\}$				
$\{1, 2, 3, 4\}$				

Example 5.2. Consider a *single* roll of a four-sided die, but suppose the die is weighted so that the outcomes are no longer equally likely. Let Q denote the probability measure corresponding to a particular weighting. A few probabilities are provided in Table 5.1;

compute $\mathrm{Q}(A)$ for all other events in Table 5.1. In what particular way is the die weighted? That is, what is the probability of each the four possible outcomes?

Example 5.3. Consider a single roll of a different weighted four-sided die. Suppose that

- Rolling a 1 is twice as likely as rolling a 4
- Rolling a 2 is three times as likely as rolling a 4
- Rolling a 3 is 1.5 times as likely as rolling a 4

Let $\tilde{\mathrm{Q}}$ denote the probability measure corresponding to this particular weighting. Compute $\tilde{\mathrm{Q}}(A)$ for all events in Table 5.1.

In what particular way is the die weighted? That is, what is the probability of each the four possible outcomes?

Example 5.4. The general meeting problem involves multiple people, but we'll first consider the arrival time of just a single person, who we'll call Han. Suppose that Han arrives "uniformly at random" at a time in $[0, 60]$.

1. Compute the probability that Han arrives before 12:15.

2. Compute the probability that Han arrives after 12:45.

3. Let P denote the corresponding probability measure. Suggest a general formula for $\mathrm{P}([a, b])$, the probability that Han arrives between a and b minutes after noon for $0 \le a < b \le 60$.

4. Compute the probability that Han's arrival time, truncated to the nearest minute, is 0 minutes after noon; that is, find the probability that Han arrives between 12:00 and 12:01.

5. Continue to compute the probability that Han's arrival time truncated to the nearest minute is $1, 2, 3, \ldots, 59$, and sketch a plot with arrival time (truncated minutes after noon) on the horizontal axis and probability on the vertical axis. Is the plot what you would expect for arriving "uniformly at random"?

Example 5.5. Assume that the probability that Han arrives between a and b minutes after noon is $(b/60)^2 - (a/60)^2$, for $0 \le a < b \le 60$. Let Q denote the corresponding probability measure; notice that $\mathrm{Q}([0, 60]) = (60/60)^2 - (0/60)^2 = 1$. Compute the following probabilities and compare your answers to the corresponding parts from Example 5.4.

1. Compute the probability that Han arrives before 12:15.

2. Compute the probability that Han arrives after 12:45.

3. Compute the probability Han's arrival time, truncated to the nearest minute, is 0 minutes after noon; that is, find the probability that Han arrives between 12:00 and 12:01.

4. Compute the probability Han's arrival time, truncated to the nearest minute, is 59 minutes after noon; that is, find the probability that Han arrives between 12:59 and 1:00.

5. Continue to compute the probability that Han's arrival time truncated to the nearest minute is $1, 2, 3, \dots, 59$, and sketch a plot with arrival time (truncated minutes after noon) on the horizontal axis and probability on the vertical axis. What assumptions about Han's arrival time does this probability measure reflect?

Example 5.6. Continuing with the uniform probability measure of Example 5.4.

1. Compute the probability that Han arrives between 12:00 and 12:01, within 1 minute after noon.

2. Compute the probability that Han arrives between 12:00:00 and 12:00:01, within 1 second after noon.

3. Compute the probability that Han arrives between 12:00:00.000 and 12:01:00.001, within 1 millisecond after noon.

4. Compute the probability that Han arrives at the exact time 12:00:00.00000... (with infinite precision).

5. What is the probability that Han arrives "at noon"? Discuss.

Example 5.7. Continuing with the non-uniform probability measure of Example 5.5.

1. Compute the probability that Han arrives between 12:00 and 12:01, within 1 minute after noon.

2. Compute the probability that Han arrives between 12:00:00 and 12:00:01, within 1 second after noon.

3. Compute the probability that Han arrives between 12:00:00.000 and 12:01:00.001, within 1 millisecond after noon.

4. Compute the probability that Han arrives at the exact time 12:00:00.00000... (with infinite precision).

5. Compute the probability that Han arrives between 12:59 and 1:00, within 1 minute before 1:00.

6. Compute the probability that Han arrives between 12:59:59 and 1:00:00, within 1 second before 1:00.

7. Compute the probability that Han arrives between 12:59:59.999 and 1:00:00.000, within 1 millisecond before 1:00.

8. Compute the probability that Han arrives at the exact time 1:00:00.00000… (with infinite precision).

9. Which is more likely: that Han arrives "at noon" or "at 1:00"? Discuss.

- For a continuous sample space, the probability of any particular outcome is 0.
- Particular outcomes represent "infinite precision" which is not practical in real applications
- For continuous sample spaces it makes more sense to consider "close to" probabilities

rather than "equals to" probabilities.
- "Close to" events correspond to *intervals* of reasonable practical precision and these intervals can have non-zero probability.
- Certain outcomes can be more likely than others in the "close to" sense.

Example 5.8. Back to the Regina, Cady meeting problem. Assume that Regina and Cady each arrive at a time uniformly at random between noon and 1:00, independently of each other, so that they arrive "uniformly at random" in the sample space of Example 4.3. Let P denote the corresponding probability measure.

Let R be the random variable representing Regina's arrival time (minutes after noon), and Y for Cady, and let $T = \min(R, Y)$ and $W = |R - Y|$.

Compute and interpret the following.

1. $\mathrm{P}(R > Y)$

2. $\mathrm{P}(T < 30)$

3. $\mathrm{P}(W < 15)$

4. $\mathrm{P}(R < 24)$

5. $\mathrm{P}(W < 1)$

6. $\mathrm{P}(W = 0)$

7. What is the probability that Regina and Cady arrive "at the same time"? Discuss.

6 Distributions of Random Variables (A Brief Introduction)

- The **joint (probability) distribution** of a collection of random variables identifies the possible values that the random variables can take and their relative likelihoods.
- We will see many ways of describing a distribution, depending on how many random variables are involved and their types (discrete or continuous).
- In the context of multiple random variables, the distribution of any one of the random variables is called a **marginal distribution**.

Example 6.1. Roll a fair four-sided die twice. Let X be the sum of the two dice, and let Y be the larger of the two rolls (or the common value if both rolls are the same).

1. Construct a table and plot displaying the marginal distribution of Y.

2. Describe the distribution of Y in terms of long run relative frequency.

3. Describe the distribution of Y in terms of relative degree of likelihood.

4. Construct a table and plot displaying the joint distribution of X and Y.

5. Construct a table and plot displaying the marginal distribution of X.

- The expected value $\mathrm{E}(X)$ of a discrete random variable X is defined by the probability-weighted average according to the underlying probability measure.
- The expected value of a random variable can be interpreted as the long-run average value of the random variable

Exercise 6.1. Consider the matching problem with $n = 4$: objects labeled 1, 2, 3, 4, are placed at random in spots labeled 1, 2, 3, 4, with spot 1 the correct spot for object 1, etc. Let the random variable X count the number of objects that are put back in the correct spot. Let P denote the probability measure corresponding to the assumption that the objects are equally likely to be placed in any spot, so that the 24 possible placements are equally.

1. Find the distribution of X by creating an appropriate table and plot.

2. Compute $\mathrm{E}(X)$.

3. Is the value from part 2 the most likely value of X? Explain.

4. Is the value from part 2 the value that we would "expect" to see for X in a single repetition of the phenomenon? Explain.

5. Explain in what sense the value from part 2 is “expected”.

7 Conditioning

- Conditioning concerns how probabilities of events or distributions of random variables are influenced by information about the occurrence of events or the values of random variables.
- A probability is a measure of the likelihood or degree of uncertainty of an event. A *conditional probability* revises this measure to reflect any "new" information about the outcome of the underlying random phenomenon.

Example 7.1. The probability that a randomly selected American adult believes in human-driven climate change[1] is 0.54.

The probability[2] that a randomly selected American adult is a Democrat is 0.28.

1. Donny Don't says that the probability that a randomly selected American adult both (1) is a Democrat, *and* (2) believes in human-driven climate change is equal to 0.28×0.54. Do you agree?

2. Suppose that the probability that a randomly selected American adult both is a Democrat and believes in human-driven climate change is 0.19. Construct an appropriate two-way table of probabilities.

3. Compute the probability that a randomly selected American adult *who is a Democrat* believes in human-driven climate change.

[1] Probabilities are estimated based on this 2024 survey.
[2] Estimate based on Gallup poll

4. Compute the probability that a randomly selected American adult *who believes in human-driven climate change* is a Democrat.

5. How can the probability in the two previous parts be written in terms of the probabilities provided (0.54, 0.28, 0.19)?

- The **conditional probability of event A given event B**, denoted $\mathrm{P}(A|B)$, is defined as (provided $\mathrm{P}(B) > 0$)
$$\mathrm{P}(A|B) = \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)}$$
- The conditional probability $\mathrm{P}(A|B)$ represents how the likelihood or degree of uncertainty of event A should be updated to reflect information that event B has occurred.
- In general, knowing whether or not event B occurs influences the probability of event A. That is,
$$\text{In general, } \mathrm{P}(A|B) \neq \mathrm{P}(A)$$
- Be careful: order is essential in conditioning. That is,
$$\text{In general, } \mathrm{P}(A|B) \neq \mathrm{P}(B|A)$$
- Within the context of two events, we have joint, conditional, and marginal probabilities.
  - Joint: unconditional probability involving both events, $\mathrm{P}(A \cap B)$.
  - Conditional: conditional probability of one event given the other, $\mathrm{P}(A|B)$, $\mathrm{P}(B|A)$.
  - Marginal: unconditional probability of a single event $\mathrm{P}(A)$, $\mathrm{P}(B)$.
- The relationship $\mathrm{P}(A|B) = \mathrm{P}(A \cap B)/\mathrm{P}(B)$ can be stated generically as
$$\text{conditional} = \frac{\text{joint}}{\text{marginal}}$$
- In many problems conditional probabilities are provided or can be determined directly.

Example 7.2. Suppose that

- 67% of Democrats believe in human-driven climate
- 46% of Independents believe in human-driven climate
- 34% of Republicans believe in human-driven climate

Also suppose that

- 28% of American adults are Democrats
- 42% of American adults are Independents
- 30% of American adults are Republicans

1. Define the event A to represent "believes in human-driven climate change" and D, I, R to correspond to affiliation in each of the parties. If the probability measure P corresponds to selecting an American adult uniformly at random, write all the percentages above as probabilities using proper notation.

2. Construct an appropriate two-way table of probabilities.

3. Now suppose that the randomly selected American believes in human-driven climate change. How does this information change the probability that the selected American belongs to each political party? Answer by computing appropriate probabilities (and using proper notation).

4. How many times more likely is it for an American *adult* to believe in human-driven climate change and be Independent than to:

    a. believe in human-driven climate change and be Democrat

b. believe in human-driven climate change and be Republican

5. How many times more likely is it for an American adult *who believes in human-driven climate change* to be Independent than to be:

   a. Democrat

   b. Republican

6. What do you notice about the answers to the two previous parts?

- The process of conditioning can be thought of as **"slicing and renormalizing".**
  - Extract the "slice" corresponding to the event being conditioned on (and discard the rest). For example, a slice might correspond to a particular row or column of a two-way table.
  - "Renormalize" the values in the slice so that corresponding probabilities add up to 1.
- Slicing determines the *shape*; renormalizing determines the *scale*.
- Slicing determines relative probabilities; renormalizing just makes sure they add up to 1.

Example 7.3. Roll a fair four-sided die twice. Let X be the sum of the two dice, and let Y be the larger of the two rolls (or the common value if both rolls are the same). The table below represents the joint distribution of X and Y.

$x \setminus y$	1	2	3	4
2	1/16	0	0	0
3	0	2/16	0	0
4	0	1/16	2/16	0
5	0	0	2/16	2/16
6	0	0	1/16	2/16
7	0	0	0	2/16
8	0	0	0	1/16

1. Compute and interpret in context $\mathrm{P}(X = 5|Y = 4)$.

2. Construct a table to represent the conditional distribution of X given $Y = 4$ by "slicing and renormalizing".

3. Interpret the the conditional distribution of X given $Y = 4$ as a long run relative frequency distribution.

4. Compute $\mathrm{E}(X|Y = 4)$.

5. Interpret the value from the previous part as a long run average value in context.

6. Construct a table to represent the conditional distribution of X given $Y = 3$, and

compute $\mathrm{E}(X|Y=3)$

7. Construct a table to represent the conditional distribution of X given $Y = 2$, and compute $\mathrm{E}(X|Y=2)$

8. Construct a table to represent the conditional distribution of X given $Y = 1$, and compute $\mathrm{E}(X|Y=1)$

8 Probability Rules

8.1 Multiplication rule

- Rearranging the definition of conditional probability we get the **Multiplication rule:** the probability that two events both occur is

$$\begin{aligned} \mathrm{P}(A \cap B) &= \mathrm{P}(A|B)\mathrm{P}(B) \\ &= \mathrm{P}(B|A)\mathrm{P}(A) \end{aligned}$$

- The multiplication rule says that you should think "multiply" when you see "and".
- However, be careful about *what* you are multiplying: to find a joint probability you need an unconditional and an appropriate conditional probability.
- You can condition either on A or on B, provided you have the appropriate marginal probability; often, conditioning one way is easier than the other.
- Be careful: the multiplication rule does *not* say that $\mathrm{P}(A \cap B)$ is the same as $\mathrm{P}(A)\mathrm{P}(B)$.

- The multiplication rule is useful in situations where conditional probabilities are easier to obtain directly than joint probabilities.

Example 8.1. A standard deck of playing cards has 52 cards, 13 cards (2 through 10, jack, king, queen, ace) in each of 4 suits (hearts, diamonds, clubs, spades). Shuffle a deck and deals cards one at a time without replacement.

1. Find the probability that the first card dealt is a heart.

2. If the first card dealt is a heart, determine the conditional probability that the second card is a heart.

3. Find the probability that the first two cards dealt are hearts.

4. Find the probability that the first two cards dealt are hearts and the third card dealt is a diamond.

- The multiplication rule extends naturally to more than two events (though the notation gets messy). For three events, we have

$$\mathrm{P}(A_1 \cap A_2 \cap A_3) = \mathrm{P}(A_1)\mathrm{P}(A_2|A_1)\mathrm{P}(A_3|A_1 \cap A_2)$$

- And in general,

$$\mathrm{P}(A_1 \cap A_2 \cap A_3 \cap A_4 \cap \cdots) = \mathrm{P}(A_1)\mathrm{P}(A_2|A_1)\mathrm{P}(A_3|A_1 \cap A_2)\mathrm{P}(A_4|A_1 \cap A_2 \cap A_4)\cdots$$

- The multiplication rule is useful for computing probabilities of events that can be broken down into component "stages" where conditional probabilities at each stage are readily available. At each stage, condition on the information about all previous stages.

Example 8.2. The birthday problem concerns the probability that at least two people in a group of n people have the same birthday[1]. Ignore multiple births and February 29 and assume that the other 365 days are all equally likely[2].

1. If $n = 30$, what do you think the probability that at least two people share a birthday is: 0-20%, 20-40%, 40-60%, 60-80%, 80-100%? How large do you think n needs to be in order for the probability that at least two people share a birthday to be larger than 0.5? Just make guesses before proceeding to calculations.

[1] You should really click on this birthday problem link.

[2] Which isn't quite true. However, a non-uniform distribution of birthdays only increases the probability that at least two people have the same birthday. To see that, think of an extreme case like if everyone were born in September.

2. Now consider $n = 3$ people, labeled 1, 2, and 3. What is the probability that persons 1 and 2 have different birthdays?

3. What is the probability that persons 1, 2, and 3 all have different birthdays *given* that persons 1 and 2 have different birthdays?

4. What is the probability that persons 1, 2, and 3 all have different birthdays?

5. When $n = 3$. What is the probability that at least two people share a birthday?

6. For $n = 30$, find the probability that none of the people have the same birthday.

7. For $n = 30$, find the probability that at least two people have the same birthday.

8. Write a clearly worded sentence interpreting the probability in the previous part as a long run relative frequency.

9. When $n = 100$ the probability is about 0.9999997.
   If you are in a group of 100 people and no one shares your birthday, should you be surprised? Discuss.

8.2 Law of total probability

Example 8.3. Suppose that

- 67% of Democrats believe in human-driven climate
- 46% of Independents believe in human-driven climate
- 34% of Republicans believe in human-driven climate

Also suppose that

- 28% of American adults are Democrats
- 42% of American adults are Independents
- 30% of American adults are Republicans

Randomly select an American adult. Compute the probability that the selected person believes in human-driven climate change.

- **Law of total probability.** If $C_1, \dots, C_k$ are disjoint with $C_1 \cup \cdots \cup C_k = \Omega$, then

$$\begin{aligned} \mathrm{P}(A) &= \sum_{i=1}^{k} \mathrm{P}(A \cap C_i) \\ &= \sum_{i=1}^{k} \mathrm{P}(A|C_i)\mathrm{P}(C_i) \end{aligned}$$

- The events $C_1, \dots, C_k$, which represent the "cases", form a *partition* of the sample space; each outcome $\omega \in \Omega$ lies in exactly one of the C_i.
- The law of total probability says that we can interpret the unconditional probability $\mathrm{P}(A)$ as a probability-weighted average of the case-by-case conditional probabilities $\mathrm{P}(A|C_i)$ where the weights $\mathrm{P}(C_i)$ represent the probability of encountering each case.

- Conditioning and using the law of probability is an effective strategy in solving many problems, even when the problem doesn't seem to involve conditioning.
- For example, when a problem involves iterations or steps it is often useful to *condition on the result of the first step*.

Example 8.4. You and your friend are playing the "lookaway challenge".

The game consists of possibly multiple rounds. In the first round, you point in one of four directions: up, down, left or right. At the exact same time, your friend also looks in one of those four directions. If your friend looks in the same direction you're pointing, you win! Otherwise, you switch roles and the game continues to the next round — now your friend points in a direction and you try to look away. As long as no one wins, you keep switching off who points and who looks. The game ends, and the current "pointer" wins, whenever the "looker" looks in the same direction as the pointer.

Suppose that each player is equally likely to point/look in each of the four directions, independently from round to round. What is the probability that you win the game?

1. Why might you expect the probability to not be equal to 0.5?

2. If you start as the pointer, what is the probability that you win in the first round?

3. If p denotes the probability that the player who starts as the pointer wins the game, what is the probability that the player who starts as the looker wins the game? (Note: p is the probability that the person who starts as pointer wins the whole game, not just the first round.)

4. Let A be the event that the person who starts as the pointer wins the game, and B be the event that the person who starts as the pointer wins in the first round. What is $\mathrm{P}(A|B)$?

5. Find a simple expression for $\mathrm{P}(A|B^c)$ in terms of p. The key is to consider this question: if the player who starts as the pointer does not win in the first round, how does the game behave from that point forward?

6. Condition on the result of the first round and set up an equation to solve for p.

7. Interpret the probability from the previous part.

8.3 Bayes Rule

Example 8.5. Suppose that

- 67% of Democrats believe in human-driven climate
- 46% of Independents believe in human-driven climate
- 34% of Republicans believe in human-driven climate

Also suppose that

- 28% of American adults are Democrats
- 42% of American adults are Independents
- 30% of American adults are Republicans

Randomly select an American adult. Compute the conditional probability that the selected adult is a Democrat given that they believe in human-driven climate change.

- **Bayes' rule for events** specifies how a prior probability $P(H)$ of event H is updated in response to the evidence E to obtain the posterior probability $P(H|E)$.

$$\mathrm{P}(H|E) = \frac{\mathrm{P}(E|H)\mathrm{P}(H)}{\mathrm{P}(E)}$$

- Event H represents a particular hypothesis (or model or case)
- Event E represents observed evidence (or data or information)
- $\mathrm{P}(H)$ is the unconditional or **prior probability** of H (prior to observing evidence E)
- $\mathrm{P}(H|E)$ is the conditional or **posterior probability** of H after observing evidence E.
- $\mathrm{P}(E|H)$ is the **likelihood** of evidence E given hypothesis (or model or case) H

Example 8.6. Continuing the previous example. Randomly select an American adult.

1. Consider the conditional probability that the selected adult is a Democrat given that they believe in human-driven climate change. Identify the prior probability, hypothesis, evidence, likelihood, and posterior probability.

2. Compute the conditional probability that the selected adult is a Republican given that they believe in human-driven climate change.

3. How many times more likely is it for an American adult to be a Democrat than be be a Republican?

4. How many times more likely is it for an American adult to believe in human-driven climate change given that they are a Democrat than given that they are a Republican?

5. How many times more likely is it for an American adult who believes in human-driven climate change to be a Democrat than to be a Republican?

6. How are the values in the three previous parts related?

7. Compute the conditional probability that the selected adult is an Independent given that they believe in human-driven climate change.

8. How many times more likely is it for an American adult to be an Independent than be be a Republican?

9. How many times more likely is it for an American adult to believe in human-driven climate change given that they are an Independent than given that they are a Republican?

10. How many times more likely is it for an American adult who believes in human-driven climate change to be an Independent than to be a Republican?

11. How are the values in the three previous parts related?

- Bayes rule is often used when there are multiple hypotheses or cases. Suppose $H_1, \dots, H_k$ is a series of distinct hypotheses which together account for all possibilities, and E is any event (evidence).
- Combining Bayes' rule with the law of total probability,

$$
\begin{aligned}
P(H_j|E) &= \frac{P(E|H_j)P(H_j)}{P(E)} \\
&= \frac{P(E|H_j)P(H_j)}{\sum_{i=1}^{k} P(E|H_i)P(H_i)}
\end{aligned}
$$

$$
P(H_j|E) \propto P(E|H_j)P(H_j)
$$

- The symbol $\propto$ is read "is proportional to". The relative *ratios* of the posterior probabilities of different hypotheses are determined by the product of the prior probabilities and the likelihoods, $P(E|H_j)P(H_j)$. The marginal probability of the evidence, $P(E)$, in the denominator simply normalizes the numerators to ensure that the updated conditional probabilities given the evidence sum to 1 over all the distinct hypotheses.
- **In short, Bayes' rule says**

$$
\textbf{posterior} \propto \textbf{likelihood} \times \textbf{prior}
$$

Example 8.7. Suppose that you are presented with six boxes, labeled 0, 1, 2, ..., 5, each containing five marbles. Box 0 contains 0 green and 5 gold marbles, box 1 contains 1 green and 4 gold, and so on with box i containing i green and $5-i$ gold. One of the boxes is chosen uniformly at random (perhaps by rolling a fair six-sided die), and then you will randomly select marbles from that box, without replacement. Imagine the boxes appear identical and you can't see inside; all you observe is the color of the marbles you select. Based on the colors of the marbles selected, you will update the probabilities of which box had been chosen.

1. Suppose that a single marble is selected and it is green. Which box do you think is the most likely to have been chosen? Make a guess for the posterior probabilities for each box. Then construct a Bayes table to compute the posterior probabilities. How do they compare to the prior probabilities?

2. Now suppose a second marble is selected from the same box, without replacement, and its color is gold. Which box do you think is the most likely to have been chosen given these two marbles? Make a guess for the posterior probabilities for each box. Then construct a Bayes table to compute the posterior probabilities, *using the posterior probabilities from the previous part after the selection of the green marble as the new prior probabilities before seeing the gold marble.*

3. Now construct a Bayes table corresponding to the original prior probabilities (1/6 each) and the combined evidence that the first ball selected was green and the second was gold. How do the posterior probabilities compare to the previous part?

- Bayesian analysis is often an iterative process.
- Posterior probabilities are updated after observing some information or data. These probabilities can then be used as prior probabilities before observing new data.
- Posterior probabilities can be sequentially updated as new data becomes available, with the posterior probabilities after the previous stage serving as the prior probabilities for the next stage.
- The final posterior probabilities only depend upon the cumulative data. It doesn't matter if we sequentially update the posterior after each new piece of data or only once after all the data is available; the final posterior probabilities will be the same either way. Also, the final posterior probabilities are not impacted by the order in which the data are observed.

9 Independence

Example 9.1. Consider the following hypothetical data.

	Democrat (D)	Not Democrat (D^c)	Total
Loves puppies (L)	180	270	450
Does not love puppies (L^c)	20	30	50
Total	200	300	500

Suppose a person is randomly selected from this group. Consider the events

$$L = \{\text{person loves puppies}\}$$
$$D = \{\text{person is a Democrat}\}$$

1. Compute and interpret $\mathrm{P}(L)$.

2. Compute and interpret $\mathrm{P}(L|D)$.

3. Compute and interpret $\mathrm{P}(L|D^c)$.

4. What do you notice about $\mathrm{P}(L)$, $\mathrm{P}(L|D)$, and $\mathrm{P}(L|D^c)$?

5. Compute and interpret $\mathrm{P}(D)$.

6. Compute and interpret $\mathrm{P}(D|L)$.

7. Compute and interpret $\mathrm{P}(D|L^c)$.

8. What do you notice about $\mathrm{P}(D)$, $\mathrm{P}(D|L)$, and $\mathrm{P}(D|L^c)$?

9. Compute and interpret $\mathrm{P}(D \cap L)$.

10. What is the relationship between $\mathrm{P}(D \cap L)$ and $\mathrm{P}(D)$ and $\mathrm{P}(L)$?

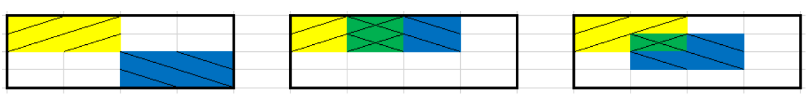
11. When randomly selecting a person from this particular group, would you say that events D and L are independent? Why?

- Events A and B are **independent** if the knowing whether or not one occurs does not change the probability of the other.
- For events A and B (with $0 < \mathrm{P}(A) < 1$ and $0 < \mathrm{P}(B) < 1$) the following are equivalent. That is, if one is true then they all are true; if one is false, then they all are false.

$$\begin{aligned}
A \text{ and } B &\text{ are independent}\\
\mathrm{P}(A \cap B) &= \mathrm{P}(A)\mathrm{P}(B)\\
\mathrm{P}(A^c \cap B) &= \mathrm{P}(A^c)\mathrm{P}(B)\\
\mathrm{P}(A \cap B^c) &= \mathrm{P}(A)\mathrm{P}(B^c)\\
\mathrm{P}(A^c \cap B^c) &= \mathrm{P}(A^c)\mathrm{P}(B^c)\\
\mathrm{P}(A|B) &= \mathrm{P}(A)\\
\mathrm{P}(A|B) &= \mathrm{P}(A|B^c)\\
\mathrm{P}(B|A) &= \mathrm{P}(B)\\
\mathrm{P}(B|A) &= \mathrm{P}(B|A^c)
\end{aligned}$$

Example 9.2. Each of the three Venn diagrams below represents a sample space with 16 equally likely outcomes. Let A be the yellow / event, B the blue \ event, and their intersection $A \cap B$ the green $\times$ event. Suppose that areas represent probabilities, so that for example $\mathrm{P}(A) = 4/16$.

In which of the scenarios are events A and B independent?



- Do not confuse "disjoint" with "independent".
- Disjoint means two events do not "overlap". Independence means two events *"overlap in just the right way"*.
- You can pretty much forget "disjoint" exists; you will naturally apply the addition rule for disjoint events correctly without even thinking about it.

- Independence is much more important and useful, but also requires more care.

Example 9.3. Roll two fair six-sided dice, one green and one gold. There are 36 total possible outcomes (roll on green, roll on gold), all equally likely. Consider the event $E = \{\text{the green die lands on 1}\}$. Answer the following questions by computing and comparing appropriate probabilities.

1. Consider $A = \{\text{the gold die lands on 6}\}$. Are A and E independent?

2. Consider $B = \{\text{the sum of the dice is 2}\}$. Are B and E independent?

3. Consider $C = \{\text{the sum of the dice is 7}\}$. Are C and E independent?

- Independence concerns whether or not the occurrence of one event affects the *probability* of the other.
- Given two events it is not always obvious whether or not they are independent.

- Independence depends on the underlying probability measure. Events that are independent under one probability measure might not be independent under another.
- Independence is often assumed. Whether or not independence is a valid assumption depends on the underlying random phenomenon.

Example 9.4. You have just been elected president (congratulations!) and you need to choose one of four people to sing the national anthem at your inauguration: Alicia, Ariana, Beyonce, or Billie. You write their names on some cards — *each name on possibly a different number of cards* — shuffle the cards, and draw one. Let A be the event that either Alicia or Ariana is selected, and B be the event that either Alicia or Beyonce is selected.

The following questions ask you to specify probability models satisfying different conditions. You can specify the model by identifying how many cards each person's name is written on.

For each model, find the probabilities of A, B, and $A \cap B$, and verify whether or not events A and B are independent according to the model.

1. Specify a probability model according to which the events A and B are independent.

2. Specify a different probability model according to which the events A and B are independent.

3. Specify a probability model according to which the events A and B are not independent.

Example 9.5. Flip a fair coin twice. Let

- A be the event that the first flip lands on heads
- B be the event that the second flip lands on heads,
- C be the event that both flips land on the same side.

1. Are the two events A and B independent?

2. Are the two events A and C independent?

3. Are the two events B and C independent?

4. Are the three events A, B, and C independent?

- Events $A_1, A_2, A_3, \ldots$ are **independent** if:
  - any pair of events $A_i, A_j, (i \neq j)$ satisfies $\mathrm{P}(A_i \cap A_j) = \mathrm{P}(A_i)\mathrm{P}(A_j)$,
  - and any triple of events A_i, A_j, A_k (distinct i, j, k) satisfies $\mathrm{P}(A_i \cap A_j \cap A_k) = \mathrm{P}(A_i)\mathrm{P}(A_j)\mathrm{P}(A_k)$,
  - and any quadruple of events satisfies $\mathrm{P}(A_i \cap A_j \cap A_k \cap A_m) = \mathrm{P}(A_i)\mathrm{P}(A_j)\mathrm{P}(A_k)\mathrm{P}(A_m)$,
  - and so on.
- Intuitively, a collection of events is independent if knowing whether or not any combination of the events in the collection occur does not change the probability of any other event in the collection.

Example 9.6. A certain system consists of four identical components. Suppose that the probability that any particular component fails is 0.1, and failures of the components occur independently of each other. Find the probability that the system fails if:

1. The components are connected in *parallel*: the system fails only if *all* of the components fail.

2. The components are connected in *series*: the system fails whenever *at least one* of the components fails.

3. Donny Don't says the answer to the previous part is $0.1+0.1+0.1+0.1=0.4$. Explain the error in Donny's reasoning.

- When events are independent, the multiplication rule simplifies greatly.

  $$\mathrm{P}(A_1 \cap A_2 \cap A_3 \cap \cdots \cap A_n) = \mathrm{P}(A_1)\mathrm{P}(A_2)\mathrm{P}(A_3)\cdots\mathrm{P}(A_n) \quad \text{if } A_1, A_2, A_3, \ldots, A_n \text{ are independent}$$

- When a problem involves independence, you will want to take advantage of it. Work with "and" events whenever possible in order to use the multiplication rule.
- For example, for problems involving "at least one" (an "or" event) take the complement to obtain "none" (an "and" event).

References