STAT 305 Handouts

Kevin Ross

Sep 19, 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. 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.

1. If you are trying to decide which college Maddie is in, is the information that she is female, went to high school in CA, and had a high school GPA of 4.1 helpful? Why?

2. 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? Hint: what was the last sentence before this example!

Example 3.4. 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 3.5. 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 3.6. 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
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 3.7. Continuing Example 3.6. 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.

References