

# 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 as 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.