Lisa Yan and Jerry Cain
CS109

Lecture Notes #3
September 18, 2020

Probability

Based on a handout by Chris Piech and Lisa Yan

Pre-recorded lecture: Sections 1, 2, 3, and 5.1, 5.2 (skip Section 4).
In-lecture: Section 4 and Section 6.

It is that time in the quarter (it is still Week 1) when we get to talk about probability. We are again going to build this up from first principles. We will heavily use the rules of counting that we learned earlier this week.

1 Event Spaces and Sample Spaces

A **sample space**, S, is the set of all possible outcomes of an experiment. For example:

1. Coin flip: $S = \{\text{Heads}, \text{Tails}\}$
2. Flipping two coins: $S = \{(H, H), (H, T), (T, H), (T, T)\}$
3. Roll of 6-sided die: $S = \{1, 2, 3, 4, 5, 6\}$
4. Number of emails in a day: $S = \{x \mid x \in \mathbb{Z}, x \geq 0\}$ (non-negative integers)
5. Number of Netflix hours in a day: $S = \{x \mid x \in \mathbb{R}, 0 \leq x \leq 24\}$

An **event space**, E, is some subset of S that we ascribe meaning to. In set notation, $E \subseteq S$.

1. Coin flip is heads: $E = \{\text{Heads}\}$
2. ≥ 1 head on 2 coin flips: $E = \{(H, H), (H, T), (T, H)\}$
3. Roll of die is 3 or less: $E = \{1, 2, 3\}$
4. Number of emails in a day ≤ 20: $E = \{x \mid x \in \mathbb{Z}, 0 \leq x \leq 20\}$
5. "Wasted day" (≥ 5 Netflix hours): $E = \{x \mid x \in \mathbb{R}, 5 \leq x \leq 24\}$

We say that an event E occurs when the outcome of the experiment is one of the outcomes in E.

2 Probability

In the 20th century, people figured out one way to define what a probability is:

$$P(E) = \lim_{n \to \infty} \frac{n(E)}{n},$$

where n is the number of trials performed and $n(E)$ is the number of trials with an outcome in E. In English this reads: say you perform n trials of an experiment. The probability of a desired event E is defined as the ratio of the number of trials that result in an outcome in E to the number of trials performed (in the limit as your number of trials approaches infinity).

You can also give other meanings to the concept of a probability, however. One common meaning ascribed is that $P(E)$ is a measure of the chance of E occurring.

I often think of a probability in another way: I don't know everything about the world. As a result I have to come up with a way of expressing my belief that E will happen given my limited knowledge. This interpretation (often referred to as the *Bayesian* interpretation) acknowledges that there are two sources of probabilities: natural randomness and our own uncertainty. Later in the quarter, we will contrast the frequentist definition we gave you above with this other Bayesian definition of probability.

3 Axioms of Probability

Here are some basic truths about probabilities:

Axiom 1: $0 \leq P(E) \leq 1$
Axiom 2: $P(S) = 1$
Axiom 3: If E and F are mutually exclusive ($E \cap F = \emptyset$), then $P(E) + P(F) = P(E \cup F)$

You can convince yourself of the first axiom by thinking about the definition of probability above: when performing some number of trials of an actual experiment, it is not possible to get more occurrences of the event than there are trials (so probabilities are at most 1), and it is not possible to get less than 0 occurrences of the event (so probabilities are at least 0).

The second axiom makes intuitive sense as well: if your event space is the same as the sample space, then each trial must produce an outcome from the event space. Of course, this is just a restatement of the definition of the sample space; it is sort of like saying that the probability of you eating cake (event space) if you eat cake (sample space) is 1.

The third form is the most useful for us in terms of analytically computing probability. More generally, for any sequence of mutually exclusive events $E_1, E_2, \ldots$, it can be shown that:

$$P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i)$$

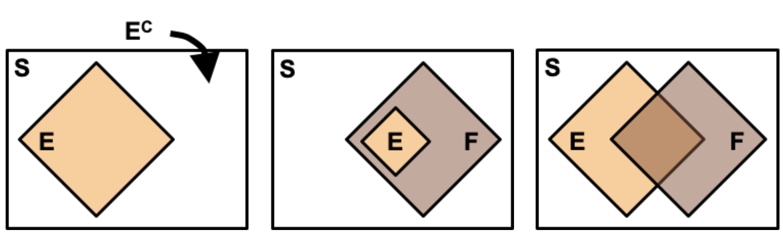
4 Provable Identities of Probability

We often refer to these as corollaries that are directly provable from the three axioms given above.

In this class, for an event E we define E^C to be the **complement** of event E, containing all outcomes that are not in E. It follows that E and E^C are mutually exclusive.

For each identity, we include a graphic in Figure 1 and sketch out the key points of the proof below.

Corollary 1: $P(E^c) = 1 - P(E) \qquad (= P(S) - P(E))$
Corollary 2: If $E \subseteq F$, then $P(E) \leq P(F)$
Corollary 3: $P(E \cup F) = P(E) + P(F) - P(EF)$



(a) Corollary 1 (b) Corollary 2 (c) Corollary 3

Figure 1: In the above pictures, E and F are events in sample space S.

Proof of Corollary 1, for any event E:

E, E^C are mutually exclusive	Definition of E^C
$P(E \cup E^C) = P(E) + P(E^C)$	Axiom 3
$S = E \cup E^C$	Everything must either be in E or E^C, by definition
$1 = P(S) = P(E) + P(E^C)$	Axiom 2
$P(E^C) = 1 - P(E)$	Rearrange

Proof of Corollary 2, where $E \subseteq F$:

$F = E \cup E^C F$	$E \subseteq F$
$E, E^C F$ are mutually exclusive	Definition of E^C
$P(F) = P(E) + P(E^C F)$	Axiom 3
$P(E^C F) \geq 0$	Axiom 1
$P(E) \leq P(F)$	Rearrange

Proof of Corollary, 3: We refer you to the Ross textbook (Section 2.4, Proposition 4.3).

General Inclusion-Exclusion Identity

$$P\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{r=1}^{n} (-1)^{r+1} \sum_{i_1 < \cdots < i_r} P(E_{i_1} E_{i_2} \ldots E_{i_r})$$

This last rule is somewhat complicated, but the notation makes it look far worse than it is. What we are trying to find is the probability that any of a number of events happens. The outer sum loops over the possible sizes of event subsets (that is, first we look at $r = 1$—all single events—then $r = 2$—pairs of events—then $r = 3$—subsets of events of size 3—etc.). The "−1" term tells you whether you add or subtract terms with that set size. The inner sum sums over all subsets of that size. While it is written as a single summation, this innermost sum is actually an r-way summation, where the inequality $<$ ensures that you don't count a subset of events twice, by requiring that the indices $i_1, \ldots, i_r$ are in ascending order.

Here's how that looks for three events (E_1, E_2, E_3):

$$\begin{aligned} P(E_1 \cup E_2 \cup E_3) = {} & P(E_1) + P(E_2) + P(E_3) \\ & - P(E_1 E_2) - P(E_1 E_3) - P(E_2 E_3) \\ & + P(E_1 E_2 E_3) \end{aligned}$$

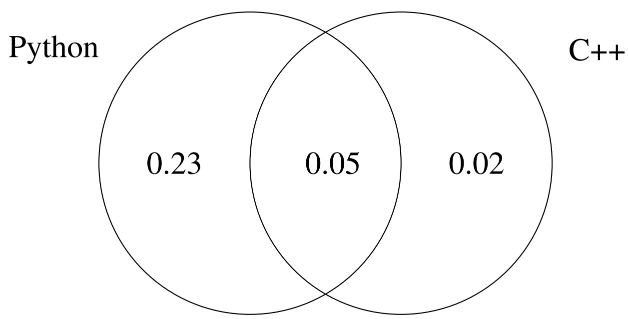
4.1 Example: Student programmers

Problem: On a university campus, 28% of all students program in Python, 7% program in C++, and 5% program in both Python and C++. You meet a random student on campus. What is the probability that they do not program in Python nor C++?

Solution: Let E = the event that a randomly selected student programs in Python and F = the event that a randomly selected student programs in C++. We would like to compute $P((E \cup F)^C)$:

$$\begin{aligned} P((E \cup F)^C) &= 1 - P(E \cup F) && \text{Identity 1} \\ &= 1 - [P(E) + P(F) - P(EF)] && \text{Identity 3} \\ &= 1 - (0.28 + 0.07 - 0.05) = 0.7 \end{aligned}$$

We can confirm this by drawing a Venn diagram as below:



5 Equally Likely Outcomes

Some sample spaces have outcomes that are all equally likely. We like those sample spaces; they make it simple to compute probabilities. Examples of sample spaces with equally likely outcomes:

1. Coin flip: $S = \{\text{Heads}, \text{Tails}\}$
2. Flipping two coins: $S = \{(H, H), (H, T), (T, H), (T, T)\}$
3. Roll of 6-sided die: $S = \{1, 2, 3, 4, 5, 6\}$

Probability with equally likely outcomes: For a sample space S in which all outcomes are equally likely,

$$P(\text{Each outcome}) = \frac{1}{|S|}$$

and for any event $E \subseteq S$,

$$P(E) = \frac{\text{number of outcomes in } E}{\text{number of outcomes in } S} = \frac{|E|}{|S|}$$

5.1 Example: Roll two dice

Problem: You roll two six-sided dice. What is the probability that the sum of the two rolls is 7?

Solution: Define the sample space as a space of pairs, where the two elements are the outcomes of the first and second dice rolls, respectively. The event is the subset of this sample space where the sum of the paired elements is 7.

$$\begin{aligned} S = \{&(1,1),(1,2),(1,3),(1,4),(1,5),(1,6),\\ &(2,1),(2,2),(2,3),(2,4),(2,5),(2,6),\\ &(3,1),(3,2),(3,3),(3,4),(3,5),(3,6),\\ &(4,1),(4,2),(4,3),(4,4),(4,5),(4,6),\\ &(5,1),(5,2),(5,3),(5,4),(5,5),(5,6),\\ &(6,1),(6,2),(6,3),(6,4),(6,5),(6,6)\}\\ E = \{&(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\} \end{aligned}$$

Since all outcomes are equally likely, the probability of this event is:

$$P(E) = \frac{|E|}{|S|} = \frac{6}{36} = \frac{1}{6}$$

5.2 *Example: Drawing items out of a bag*

Problem: There are 4 oranges and 3 apples in a bag (or 4 cats and 3 sharks, depending on how large your bag is). You draw out 3. What is the probability that you draw 1 orange and 2 apples?

Solution 1: If we treat the oranges and apples as indistinct, we do not have a space with equally likely outcomes. We therefore treat all objects as distinct.

Suppose we treat each outcome in the sample space as an *ordered* list of three distinct items. The size of the sample space, S, is simply the total number of ways to order 3 of 7 distinct items: $|S| = 7 \cdot 6 \cdot 5 = 210$. We can then decompose the event, E, into three mutually exclusive events, where we pick the orange first, second, or third, respectively: $|E| = 4 \cdot 3 \cdot 2 + 3 \cdot 4 \cdot 2 + 3 \cdot 2 \cdot 4 = 72$. The probability of our event is therefore $P(E) = 72/210 = 12/35$.

Note that experiments with indistinguishable objects often have sample spaces that are not equally likely. However, if we make the objects distinguishable, then we can create a sample space that with equally likely outcomes, because we do not need to consider overcounting/non-unique outcomes. In probability, since we are just looking to take a ratio of number of outcomes in a sample space with equally likely outcomes of our own design, we do not need to explicitly recreate the indistinguishable case.

Solution 2: Another approach is to treat each outcome in the sample space as an *unordered* group. The size of the sample space, S, is the total number of ways to choose any 3 of 7 distinct items: $|S| = \binom{7}{3}$. The event space is then the way to pick 1 distinct orange (out of 4) and 2 distinct apples (out of 3), which we combine with the product rule: $|E| = \binom{4}{1}\binom{3}{2}$. The probability of our event is therefore $P(E) = \frac{\binom{4}{1}\binom{3}{2}}{\binom{7}{3}} = 12/35$.

The reason we can choose either an ordered or unordered approach is because probability is a *ratio*. As we saw last time, any unordered counting task can be generated by first creating an ordered list, splitting the list at marked intervals, then dividing out by the overcounted cases due to ordering. If our sample space is ordered, then our event (being a subset of the sample space) is also ordered, and therefore we should account for the overcounted cases. However, probability being a ratio means that these overcounted cases get cancelled out.

The key to solving many of this section's problems involves (1) deciding whether to count distinct objects to create an equally likely outcome sample space, and then (2) defining the sample space and event to consistently be ordered or unordered.

6 Some more examples

Once we have this definition of probability in sample spaces with equally likely outcomes, we can analyze the probabilities of many different situations. Here are just a few to get you started.

6.1 Example: Poker Straight

Part A: Consider drawing 5-card poker hands, where you are equally likely to get any 5-card poker hand. What is the probability of you drawing a "poker straight", which is defined as 5 consecutive rank cards of any suit?

Solution: Our sample space S is the ways to draw 5 unordered cards of a 52-card standard deck: $|S| = \binom{52}{5}$. Define E as the event where we draw a straight. We can determine the 5 cards that go into our straight by first noting that the ranks of any straight is determined by lowest in the straight, because the ranks of the other 4 cards must be consecutive. There are 10 choices for lowest rank in straight (anything from A to 10). Since A doubles as the lowest or highest rank, we have choices $(A, 2, 3, 4, 5), \ldots, (10, J, Q, K, A)$. Next, there are 4 choices of suit for each of the 5 cards in our hand; therefore $|E| = 10 \cdot \binom{4}{1}^5$, and $P(E) = \frac{10 \cdot 4^5}{\binom{52}{5}} \approx 0.00394$.

Part B: If we consider official poker rules, we must differentiate between a "straight" and a "straight flush," where the latter is a 5 consecutive rank cards of the *same* suit. What is the probability of you drawing a poker straight, but not a straight flush?

Solution: Define F as the event where we draw a straight flush. Note that $E = F \cup F^C$, where E is defined as in Part A, and F^C is the event whose probability we would like to compute. The number of ways to choose a straight flush is $|F| = 10 \cdot \binom{4}{1}$, where like before we have 10 choices of lowest rank, but we are now restricted to a single suit for all 5 cards. Therefore $P(F^C) = \frac{10 \cdot \binom{4}{1}^5 - 10 \cdot \binom{4}{1}}{\binom{52}{5}} \approx 0.00392$. Either way—that's a pretty good hand!

6.2 Example: Chip defects

Problem: A factory that manufactures n computer chips knows that exactly one of the chips is defective, but it is not sure each one. However, since it would be infeasible to test all n chips (assume n is very large) for defects, the factory instead randomly selects k chips for testing. What is the probability that exactly one defective chip is in the k test chips?

Solution: Suppose we consider that the factory draws chips in unordered groups. The sample space S is therefore all the ways to pick k test chips from n total chips, where $|S| = \binom{n}{k}$. Define E as the event where our unordered group of k chips contains the single defective chip (and $k-1$ regular chips). There is only 1 way to choose the single defective chip, and there are $\binom{n-1}{k-1}$ ways to choose $k-1$ chips from the $n-1$ functional chips. By the product rule, $|E| = \binom{1}{1} \cdot \binom{n-1}{k-1}$. With some simplification:

$$P(E) = \frac{|E|}{|S|} = \frac{\binom{n-1}{k-1}}{\binom{n}{k}} = \frac{\frac{(n-1)!}{(k-1)!(n-k)!}}{\frac{n!}{k!(n-k)!}} = \frac{k}{n}$$

An alternative solution is to redefine the experiment as follows: There is exactly 1 way to choose k of n chips if all n are *indistinct*, so we just choose k chips and mark them as our test group. Then, we throw a dart at our chips to make one defective (where we have an equal probability of hitting any of our n chips). The probability of hitting one of the k test chips out of the n possible chips is $\frac{k}{n}$.

The big takeaway of this example is that probability **does not take into account chronological order**. It is true that in real life, one defective chip was already defective, and then our task was to try to pick it as one of our k test chips. However, the event we care about is simply us having a defective chip in k test chips—it does not matter if the defect occurs before or after we have chosen our set. We will see next week that our power to redefine an experiment (specifically, rearrange steps of our experiment) will be very powerful in computing probability of unknown causes, when only the effects are observable.

6.3 Example: Serendipitous meetings

Problem: The population of Stanford is $n = 17,000$ people, and suppose you are friends with $r = 100$ other Stanford people. Suppose you walk into CS109 and see $k = 200$ random people, where every possible group of k people from Stanford is equally likely to be in that room. What is the probability that you see at least one friend in the room?

Solution: Define our sample space S to be all equally likely outcomes of seeing groups of k Stanford people in the room, then $|S| = \binom{n}{k}$. We would like to compute the probability of event E, the event that we see at least one friend in the room.

There are two common strategies to solving this problem. The first approach uses Corollary 1, where we can compute $P(E) = 1 - P(E^C)$. Note that E^C (the complement of E) is the event that we see no one we know in the room, and therefore

$$P(E) = 1 - P(E^C) = 1 - \frac{\binom{n-r}{k}}{\binom{n}{k}} \approx 0.6948$$

The second, longer approach is to calculate the probability of you seeing exactly 1 friend, seeing exactly 2 friends, . . . , seeing exactly r friends, and sum up the probabilities of those mutually exclusive events:

$$P(E) = \sum_{i=1}^{r} \frac{\binom{r}{i}\binom{n-r}{k-i}}{\binom{n}{k}} \approx 0.6948$$

The first approach is more straightforward and often numerically easier to compute. Also, making friends is exciting!

6.4 Example: Flipping Cards

Problem: In a 52-card deck, cards are flipped one at a time. After the first ace (of any suit) appears, consider the next card. Is the next card more likely to be the Ace of Spades than the 2 of Clubs? (This problem is based on Example 5j in Chapter 2.5 of Ross's textbook, 10th Edition.)

Solution: No; the probabilities are **equal**. The difficulty of this problem stems from defining an experiment that gives equally likely outcomes while preserving the specifications of the original problem. An incorrect approach is to define the experiment as just drawing a pair of cards (first ace, next card) because then we discard all information about the cards flipped prior to the pair. Instead, consider the experiment to be shuffling the full 52-card deck, where $|S| = 52!$. We can then reconstruct all outcomes of the pairs of cards that we care about (if we so wish—but we just care about getting an equally likely outcome sample space).

Define E_{AS} as the event where the next card is the Ace of Spades. To construct a 52-card order where this event holds, we first take out the Ace of Spades, then shuffle the remaining 51 cards (51! ways), then insert the Ace of Spades immediately after the first ace (1 way). By the product rule, $|E_{AS}| = 51! \cdot 1$. Then define E_{2C} as the event where the next card is the 2 of Clubs. To construct a 52-card order where this event holds, we perform exactly the same steps, but with the 2 of Clubs instead. Then $|E_{2C}| = 51! \cdot 1$. Therefore $P(E_{AS}) = 51!/52! = P(E_{2C})$.

For many readers, it may seem apparent that the first ace drawn could very well be the Ace of Spades, and so it is less likely that the next card is the Ace of Spades. Yet by a similar train of thought, the 2 of Clubs could very well have been drawn prior to the first ace drawn, and so we must consider all of those cases as well. This example serves to highlight the difficulty of probability: Mathematics often trumps intuition (no pun intended).