Lisa Yan
CS 109

Probability

Based on a handout by Chris Piech

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.

4 Provable Identities of Probability

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

Identity 1:

$$P(E^c) = 1 - P(E) \qquad (= P(S) - P(E))$$

Identity 2:

$$\text{If } E \subseteq F, \text{ then } P(E) \leq P(F)$$

Identity 3:

$$P(E \cup F) = P(E) + P(F) - P(EF)$$

General Inclusion-Exclusion Identity:

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

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 all single events, then pairs of events, then 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. The less-than signs ensure 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):

$$\begin{aligned} P(E_1 \cup E_2 \cup E_3) = {}& P(E_1) + P(E_2) + P(E_3) \\ & - P(E_1 E_2) - P(E_1 E_3) - P(E_2 E_3) \\ & + P(E_1 E_2 E_3) \end{aligned}$$
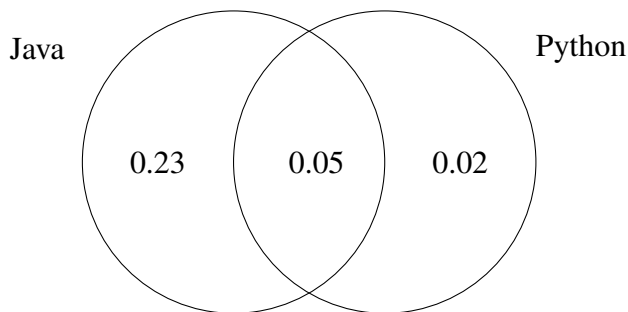
4.1 Example 1

Problem: On a university campus, 28% of all students program in Java, 7% program in Python, and 5% program in both Java and Python. You meet a random student on campus. What is the probability that they do not program in Java or Python?

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

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

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

Java Python

0.23   0.05   0.02

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 2

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 3

Problem: There are 4 oranges and 3 apples in a bag. 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 $P(E) = 72/210 = 12/35$.

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 divide by the number of overcounted cases. However, probability being a ratio means that this constant number of overcounted cases gets 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 space, and then (2) defining the sample space and event space to consistently be ordered or unordered.

5.3   *Example 4*

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).