Notes and Exercises from 'Introduction to Probability' by Blitzstein & Hwang

Ghassan Shahzad

May 14, 2024

1 Probability and Counting

1.1 Notes

- We express probability mathematically using sets.
- We use **sample spaces** specifically, where a sample space is the set of all possible **outcomes** in, say, an experiment.
- The sample space has subsets called **events**. We say an event has *occurred* if the outcome is *in* the event.
- The **mass** of an outcome is its probability.
- **Example**: Let's say we flip a coin twice. This lends itself to four possible outcomes, and we have a sample space $S = \{HH, TT, HT, TH\}$ containing those four outcomes. We're hoping for the event A that both flips are the same, which is $A \subseteq S = \{HH, TT\}$. Since coin flips are symmetric (i.e. equally likely by nature), the mass of each outcome is equal.
- Let A and B be events such that $A, B \subseteq S$. $A \cup B$ is the event that occurs if *either* A or B occur. $A \cap B$ is the event that occurs if *both* A and B occur. Don't let the phrasing confuse you; remember that an event is said to occur if the outcome is in the event, and nothing else.
- The **naive definition of probability** refers to the oldest definition of probability.
- Here, the number of possible outcomes $|S|$ and the total number of ways an event could occur $|A|$ were counted, and the latter was divided by the former to obtain the naive probability of event A: $P_{naive}(A)$. We can express this generally:
$$P_{naive}(A) = \frac{|A|}{|S|}$$
- We also find that the probability of a complement A^c occurring — basically the probability that A *doesn't* occur — is $1 - P(A)$. We can derive this as such:
$$P(A^c) = \frac{|A^c|}{|S|} = \frac{|S| - |A|}{|S|} = 1 = \frac{|A|}{|S|} = 1 - P(A)$$
- **Example:** Going back to our previous example. The probability of the event A — where both of our coinflips land on the same side — occurring is $\frac{2}{4}$. 2 is the number of outcomes within that event, and 4 is the total number of outcomes possible. The probability of our event not occurring, so A^c, is $1 - \frac{2}{4} = \frac{2}{4}$.

- The faults of the naive definition are: that it assumes each outcome has equal mass, which isn't always the case. It also assumes a finite, or countable, number of outcomes (that we can divide by).

- If you're looking for the number of possible subsets in a sample space S, or any set really, you get that number by 2^n, where n is the number of elements in the set.

- The number of possible outcomes can be counted in a number of ways. One technique is the **multiplication** rule. This rule states that in compound experiments, the total number of possible outcomes is the product of the total number of outcomes in each sub-experiment.

- **Example:** Referring back to our coin-flipping example. It is itself a compound experiment, with two experiments (coin-flips) performed sequentially. Each coin-flip has two possible outcomes. Therefore, to get the total possible outcomes, we do $2 \cdot 2 = 4$. If there were four coin-flips in our experiment, the total possible outcomes would be $2 \cdot 2 \cdot 2 \cdot 2 = 16$. It can be expressed as 2^n, where n is the number of coin-flip sub-experiments, and 2 is the total number of possible outcomes in each sub-experiment.

- Note that the order the sub-experiments are multiplied in doesn't matter since multiplication is — of course — commutative. That does not mean the order the sub-experiments are *performed* in doesn't matter, of course, since that depends on the experiment itself. In our coin-flipping experiment, however, the order is irrelevant.

- One use of the multiplication rule is in **sampling**. Sampling involves picking a group of k samples out of a larger group of n objects. The multiplication rule helps us determine how many *possible combinations* of samples we can pick out.

- There are two main types of sampling. Sampling **with replacement** means that, when we pick a sample, that does not preclude it from being picked again. To determine the total possible number of combinations in such a case, we use n^k; where n is the number of objects, and k is the number of samples we need.

- **Example:** Let's say we're picking marbles one-by-one from a jar with 20 marbles. We will pick the marble, examine it, put it back in, and then pick another marble. This is 'with replacement', since there is a chance that we may pick the same marble multiple times (we're putting it back in the jar, after all). We will repeat this process 10 times. The number of possible combinations, in this case, is $20^{10} = 1.024 \cdot 10^{13}$.

- Sampling **without replacement** means that an object, once chosen, will not be chosen again. Where n is the total number of objects, and

k the total number of samples, there are $n(n-1)...(n-k+1)$ possible outcomes, so long as $k \leq n$.

- **Example:** There are k people in a room, and we are to find the probability that two or more people have the same birthday. Assume we ignore leap year birthdays, and that the naive definition of probability is justified.
  - There are 365 possible birthdays. Thus, there are thus 365^k possible combinations of birthdays among the group.
  - Within this set, the number of combinations where two or more people share a birthday is... hard to obtain.
  - Luckily, we can find its complement by sampling with replacement. This is because sampling with replacement gives us the number of possible combinations of birthdays where each birthday does not repeat — the opposite of what we're looking for.
  - Using the complement rule, we get the naive probability of two people sharing a birthday in said room as:

    $$1 - \frac{_{365}P_k}{365^k}$$

  - A surprising, and counterintuitive, fact is that if we were to assume 23 people in the room, and thus $k = 23$, we would get:

    $$1 - \frac{365(365-1)...(365-24)}{365^{23}} = 50.73\%$$

    Yes, a chance greater than 50% that two of them share a birthday!

- Sampling without replacement comes itself in two forms: one where *order* matters, and one where it doesn't. In the former case, the aforementioned method — which we call **permutation** — is used to determine the number of possible samples $_nP_r$. If order *doesn't* matter, then we use **combination** instead to get $_nC_r$.

- Similarly, we have the same two forms of sampling *with* replacement. We've already figured out how to determine the number of possible samples with replacement when order *does* matter: n^k. When order doesn't matter, we use the formula: $_{(n+k-1)}P_k$

- Make sure to label your experiments and outcomes. For instance, in our coin flip experiment, we ought to label the first coin flip A, and the second coin flip B. We also ought to label the heads outcome 'H' or 1, and tails 'T' or 0. This is good practice and helps you avoid mistakes.

- One thing to consider in counting problems is **overcounting**. In that case, you have to adjust the obtained value for overcounting by a factor c, by dividing it by c.

- **Example:** There are n people at a meeting. Each person has to shake hands with every other person in the meeting; so, $n-1$ people. To determine how many hands are getting shaked, we use the multiplication rule: n people multiplied by the $n-1$ other people that they're going to handshake for $n(n-1)$ total handshakes. But there is an overcounting problem here: for $n=2$, we get 2 total handshakes, and this is wrong. Two people at the meeting would only have to shake hands once. We're overcounting by a factor $c=2$, and we need to adjust for this. Thus, we get a final formula of:
$$\frac{n(n-1)}{2}$$

- A **story proof** is a proof that relies on the fact that there are usually multiple ways of counting things. Thus, we find two ways to count the same thing and intuit why they come out to the same result. These usually promote greater understanding and intuition of the problem in question than an algebraic proof.

- We have considered the naive definition of probability, and now we will create our own non-naive definition of probability:
  - Assuming a **probability space** P consists of *sample space* S, a **probability function** $P(A)$ takes an *event* A such that $A \subseteq S$ and returns a real number such that $0 \leq P(A) \leq 1$.
  - The probability of the *null set* $P(\emptyset)$ is 0; the probability of the sample space S, $P(S)$, is 1.
  - If the elements of *event* A $(A_1, A_2, ..., A_j)$ are *disjoint* (mutually exclusive), then:
$$P(\bigcup_{j=1}^{\infty} A_j) = \sum_{j=1}^{\infty} P(A_j)$$

- There are two ways to 'view' probability, which is seperate from a definition of probability. These have to do with how people think we can utilize probability.
  - The **frequentists** view probability as showing how frequently, in the long-run, an event will occur. For instance, the 50/50 probabilities on a coinflip mean that over a 1000 flips, 50% (500) will be heads, and the other 500 tails.
  - The **Bayesians** view probability as the degree of belief about an event. For instance, if I flip a coin they'll agree that there is a

50% chance it will be heads-up — their *prior possibility*. Once I've flipped the coin, and they've seen that it is indeed a heads, they'll say that the *posterior possibility* is 100%.

- There are three further theorems that help our definition:
  - $P(A^c) = 1 - P(A)$
  - If $A \subseteq B$, then $P(A) \leq B$
  - $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- The third theorem leads into the **inclusion-exclusion** principle. If you are counting the number of elements in the union of two sets A and B, you do it as such:

$$|A \cup B| = |A| + |B| - |A \cap B|$$

  If you added the number of elements in A and B, that would lead to you counting the elements in the *intersection* twice. Thus, you need to subtract it once from the sum.

- This principle can be generalized:

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) - \sum_{i<j} P(A_i \cap A_j) + \sum_{i<j<k} P(A_i \cap A_j \cap A_k) +$$

$$\cdots + (-1)^{n-1} \sum_{i<\ldots<n} P\left(\bigcap_{i=1}^{n} A_i\right)$$

1.2 Exercises

1. How many ways are there to permute the letters in the word MISSISSIPPI? There are two ways to do this. The first involves accounting for overcounting, while the second uses the binomial coefficient:

$$\frac{11!}{4!4!2!} = \binom{11}{1}\binom{10}{4}\binom{6}{4}\binom{2}{2} = 34,650$$

2. (a) How many 7-digit phone numbers are possible, assuming that the first digit can't be a 0 or a 1? There are 10 digits to pick from — except for the first phone number digit, where 0 and 1 are not allowed — for our 7-digit phone number. We can express this using the multiplication rule:

$$8 \cdot 10^6 = 8,000,000$$

(b) Re-solve (a), except now assume also that the phone number is not allowed to start with 911. The number of outcomes where the phone number starts with 911 are: $1 \cdot 1 \cdot 1 \cdot 10^4 = 10,000$. Subtract this from the answer to (a), and we get $7,990,000$ possible phone numbers.

3. Fred is planning to go out to dinner each night of a certain week, Monday through Friday, with each dinner being at one of his ten favorite restaurants.

   (a) How many possibilities are there for Fred's schedule of dinners for that Monday through Friday, if Fred is not willing to eat at the same restaurant more than once? ${}_{10}P_5 = 30,240$ possible schedules.

   (b) How many possibilities are there for Fred's schedule of dinners for that Monday through Friday, if Fred is willing to eat at the same restaurant more than once, but is not willing to eat at the same place twice in a row (or more)? In terms of the multiplication rule:

$$10 \cdot 9^4 = 65,610 \text{ possible schedules.}$$

4. A round-robin tournament is being held with n tennis players; this means that every player will play against every other player exactly once.

   (a) How many possible outcomes are there for the tournament (the outcome lists out who won and who lost for each game)? Each player has to face $(n-1)$ opponents in matches, and there are n players, and adjusting for overcounting by a factor of 2, we get

$$\frac{n*(n-1)}{2} \text{ total matches}$$

   Each match has only two possible outcomes (assuming no draws). Thus, the number of total possible outcomes is:

$$2^{\frac{n*(n-1)}{2}}$$

   (b) How many games are played in total? We determined this previously:

$$\frac{n*(n-1)}{2} \text{ total matches}$$

5. A knock-out tournament is being held with 2^n tennis players. This means that for each round, the winners move on to the next round and the losers are eliminated, until only one person remains. For example, if initially there are $2^4 = 16$ players, then there are 8 games in the first round, then the 8 winners move on to round 2, then the 4 winners move on to round 3, then the 2 winners move on to round 4, the winner of which is declared the winner of the tournament. (There are various systems for determining who plays whom within a round, but these do not matter for this problem.)

(a) How many rounds are there? There are n rounds.

(b) Count how many games in total are played, by adding up the numbers of games played in each round.

$$2^{n-1} + \cdots + 2^{n-n}$$

6. There are 20 people at a chess club on a certain day. They each find opponents and start playing. How many possibilities are there for how they are matched up, assuming that in each game it does matter who has the white pieces (in a chess game, one player has the white pieces and the other player has the black pieces)? $\frac{20!}{10!}$ possible matches.

7. Two chess players, A and B, are going to play 7 games. Each game has three possible outcomes: a win for A (which is a loss for B), a draw (tie), and a loss for A (which is a win for B). A win is worth 1 point, a draw is worth 0.5 points, and a loss is worth 0 points.

   (a) How many possible outcomes for the individual games are there, such that overall player A ends up with 3 wins, 2 draws, and 2 losses? Encoding these results as a string gives us: WWWDDLL. We need to figure out the possible ways to permute this, and those are the possible outcomes. Thus:

$$\binom{7}{3}\binom{4}{2}\binom{2}{2} = 210$$

   (b) How many possible outcomes for the individual games are there, such that A ends up with 4 points and B ends up with 3 points?