Recitation 8

Intro to Probability

Probability

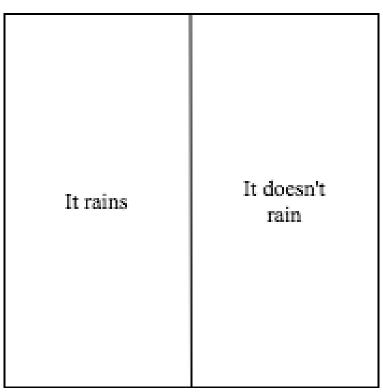
A **finite discrete probability space** is a pair $(S, \Pr)$ for some finite set S and some function $\Pr : S \to \mathbb{R}$, where, for all $x \in S$, $\Pr(x) > 0$, and $\sum_{x \in S} \Pr(x) = 1$. The set S is called the *sample space*, and the elements of S are called *outcomes*. The function Pr is called the *probability distribution function*.

Where do these defintions come from? It's all about predictions. The sample space S is supposed to represent a set of atomic outcomes—that is, the set of all things that can happen. All these outcomes are mutually exclusive.

An *event* is a subset of the sample space S, containing some atomic outcomes. For example, if I flip a coin 3 times, the outcomes are HHH, TTT, ..., so the sample space is $S = \{\text{HHH}, \text{TTT}, \text{TTH}, \dots\}$, with $2^3 = 8$ outcomes. An event can be: *We see exactly 1 H*, which we view as a set of outcomes $\{\text{HTT}, \text{THT}, \text{TTH}\}$. An event can also be the empty set: for example, the event *We see 4 Hs* is the empty set, since it is not a possible outcome (it does not correspond to anything in the sample space). Similarly, the event *We see an elephant* is the empty set for this probability space.

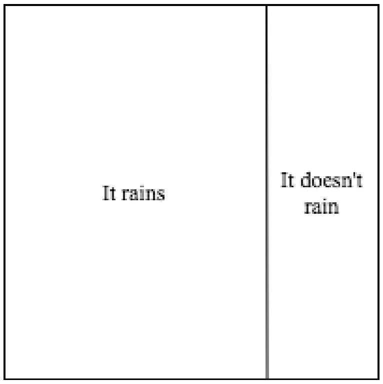
How many different events can the above sample space have? Since each event is a subset of the sample space, the number of events is equal to the number of possible subsets, which here is 2^8 (since the sample space has 8 elements).

To understand the probability distribution function, let's think about our sample space as a box of area 1. Then, each outcome occupies a certain amount of *area* within the sample space. That is, the area is *distributed* among the outcomes according to the probability distribution function! The larger the area, the more likely the outcome is.



In this example, we have that each outcome occupies the same amount of area in the box—each is equally likely. Because we have 2 outcomes, we have that the probability of each outcome is $\frac{1}{2}$.

Let's take a look at another way we could divide up the area of the box among the outcomes.



Here, the probability of it raining is $\frac{3}{4}$, so this outcome takes up $\frac{3}{4}$ of the area within the box. The remaining amount of area is $\frac{1}{4}$, which becomes the probability that it doesn't rain.

Task 1

a. If we have n outcomes in S, and Pr assigns each outcome an equal amount of area within our box, what is the probability of a particular outcome?

   Note: Such a division of the space is called a *uniform distribution*!

   > $1/n$

b. Consider $S = \{\text{It's sunny}, \text{It's raining}\}$, where the probability distribution is 0.8 and 0.2, respectively. Under what constraints on the world is this sample space reasonable?

   > If the world is only sunny or raining, then the sample space is fine. But, in the real world, we can have cloudy weather and snowy weather, too. A sample space should be *exhaustive* and also *mutually exclusive*, that is, exactly one of the outcomes has to happen each time we do the experiment.

c. Is the following probability distribution valid? Experiment: flipping a biased coin. $S = \{H, T\}$, $\Pr(H) = 1/3$, $\Pr(T) = 1/3$.

No, because the probability of all outcomes must add up to 1.

The Relationship Between Counting and Probability

Questions about probability are often closely tied to questions about counting. Often, we need to count two things—the set of outcomes in our event, and the set of total possible outcomes. In a uniform distribution, the probability of getting an event E is simply $\Pr(E) = |E|/|S|$.

Task 2

a. Suppose S is the set of all binary strings of length n, and Pr is a uniform distribution. What is the probability of getting a string with k ones (where $0 \leq k \leq n$)?

> There are $\binom{n}{k}$ strings with k ones, and 2^n total strings, so the answer is $\binom{n}{k}/2^n$.

b. For what value(s) of k is the probability of getting a string with k ones the largest? Try it out for some values and see if you find a pattern.

> When is $\binom{n}{k}$ largest? When k is closest to $\frac{n}{2}$

c. Explain how we can use n-ary (not just binary) strings to help us with other problems that involve, say, flipping a fair coin m times, or rolling a die m times.

> We can form bijections to count things! For example, let 0 denote tail, head denote 1. Then, there is a bijection between the possible binary strings and the possible outcomes of m coin flips. If we were using dice instead, though, we can use a senary (6 valued) string.

Checkpoint 1 — call over a TA!

Conditional Probability and Independence

We sometimes want to know what the probability of something is, given that we know an outcome from certain subset of outcomes has happened (that is, no outcome outside that subset can happen).

Warmup

Prof. Lewis has 2 coins: one fair, and one double heads. He picks one uniformly at random (that is, each coin has probability of 1/2 of being picked), but he doesn't tell us which one he picks. Prof. Lewis then flips the coin he picks, and he then shouts out the result.

Suppose Prof. Lewis shouts out "Heads." What is the probability he flipped the fair coin, given that we know the coin flip resulted in heads (H)?

Hint: How many ways can the coin flip result in heads, and how many of them start with the fair coin?

> 1/3

We can generalize our work here. Given two events $A, B \subseteq S$ (the sample space) where $\Pr(B) > 0$, we define the *conditional probability of A given B* as

$$\Pr(A \mid B) = \frac{\Pr(A \cap B)}{\Pr(B)}.$$

We know that the only possible outcomes are those in B (since B is supposed to have happened). We therefore make $\Pr(B)$ our denominator to *renormalize* our universe.

Looking at this formula, we also get a general definition for $Pr(A \cap B)$:

$$\Pr(A \cap B) = \Pr(A)\Pr(B|A) = \Pr(B)\Pr(A|B)$$

This is a derivation of **Bayes' Theorem**, and it will be your friend in the probability unit.

There are certain special situations where $\Pr(B|A) = \Pr(B)$; that is, where the fact that A happened does not affect the probability of B happening. We have a name for this situation.

Definition: Two events $A, B \subseteq S$ are *(pairwise) independent* if

$$\Pr(A|B) = \Pr(A)$$

If $\Pr(B) = 0$, B and any other event are independent.

Equivalently, A and B are independent if

$$\Pr(A \cap B) = \Pr(A)\Pr(B).$$

Note that these two definitions are equivalent—try to convince yourself by substituting the definition of conditional probability!

Task 3

a. If two events A and B have an empty intersection, what is the probability of A AND B?

> 0

b. For each of the following pairs of events A and B, identify whether they are independent, and justify why or why not.

   i. Suppose we roll a fair die, and suppose A = rolling an even number, and B = rolling a number greater than three.

   > Not independent: $\Pr(A) = \frac{1}{2}$, $\Pr(B) = \frac{1}{2}$. $A \cap B = \{4, 6\}$, so $\Pr(A \cap B) = \frac{1}{3}$, but $\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$

   ii. *Optional:* Suppose we flip a fair coin three times, and suppose A = the last coin is a tails, and B = there is a run of exactly two tails (that is, two, but not three, tails are flipped in a row).

   > Independent: $\Pr(A) = \frac{1}{2}$, and $B = \{TTH, HTT\}$ so $\Pr(B) = \frac{2}{8} = \frac{1}{4}$. $A \cap B = \{HTT\}$ so $\Pr(A \cap B) = \frac{1}{8}$ which is $\frac{1}{2} \times \frac{1}{4}$

c. *Optional:* If two events A and B are *mutually exclusive*, are they *independent*? You can assume $\Pr(A)$ and $\Pr(B)$ are nonzero.

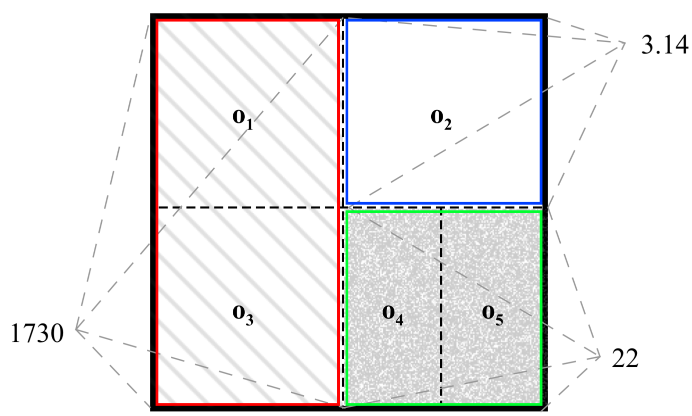
> No. By definition, $Pr(A \cap B) = 0$, so if the probability of each individual event occurring is nonzero, then $\Pr(A \cap B) \neq \Pr(A)\Pr(B)$.

Random Variables

From a probability space (S, p), we can create a new probability space (S', p'), where S' is a partition of S, that is, $S' = \{P_1, P_2, ...P_n\}$, and where $p' = \Pr(P_i)$ for all i.

Visually, we can take a box partitioned into m outcomes, and then partition these m outcomes into n groups. The amount of area each group takes up is just the sum of the outcomes within the group, so this box with m groups is just like having a probability space with m outcomes.

This leads us to one of the biggest, most famous misnomers in all of mathematics: random variables. A **random variable** is a function from a set of outcomes S to $\mathbb{R}$ or $\mathbb{Z}$. We can think of this random variable as partitioning S, where each group is a set of outcomes that are all assigned the same value. We can also think of the random variable as assigning each group a different value. This diagram shows what we mean:



Here, $o_1, \ldots, o_5$ are outcomes in S. Our random variable assigns two of them to 1730, one of them to 3.14, and two of them to 22. Thus, the random variable partitions the outcomes into 3 groups. Let's walk through some other more concrete examples of random variables.

Task 4

Let random variable X be on the coin flip sample space $\{H, T\}$, where $X(H) = 1$, and $X(T) = 0$. (This random variable is also known as the indicator random variable of event H.)

a. If $\Pr(H) = 1/2$, and $\Pr(T) = 1/2$, then what is $\Pr(X = 1)$? How about $\Pr(X = 0)$?

> 1/2, 1/2

b. We could also have the random variable Y, where the domain of Y is S, all sequences of coin flips of length n, and where $Y(s)$ = the number of heads in s.

If S is uniformly distributed, then what is $\Pr(Y = k)$, where $0 \leq k \leq n$?

Hint: You computed this value in last week's recitation.

> $\binom{n}{k}/2^n$

c. Let C_i be a random variable for the ith coin flip, s_i, in our sequence, s, of coin flips of length n, where $C_i(H) = 1$ and $C_i(T) = 0$. Let $C(s) = C_1(s_1) + C_2(s_2) + ...C_n(s_n)$. Explain why $C(s) = Y(s)$.

> Think about s as a 0/1 string. The sum of the digits in s (which is what $C(s)$ is) is the number of 1s (that is, heads) we got in s (which is what $Y(s)$ is)

d. *Optional:* Explain why $\Pr(C = k) = \binom{n}{k} \Pr(C_i = 1)^k \Pr(C_i = 0)^{n-k}$.

> There are two ways to see it. We could show it reduces to $\binom{n}{k}/2^n$, and because $\Pr(Y = k) = \Pr(C = k)$, we are done. Or we could be a bit more general and explain that there are $\binom{n}{k}$ ways to choose k 1's, and so we sum up the probabilities of these different ways

e. *Optional:* Given an expression for $\Pr(C \geq k)$.

> $\sum_{j=k}^{n} \binom{n}{j} \times \Pr(C_i = 1)^j \times \Pr(C_i = 0)^{n-j}$

f. *Optional:* Compute $\Pr(C = k | C_1 = 1)$.

> $\binom{n-1}{k-1} \Pr(C_i = 1)^{k-1} \Pr(C_i = 0)^{n-1-(k-1)}$

Checkpoint 2 — Call over a TA!

Task 5

The Brown Review is a test-prep company that publishes books helping high school students prepare for the upcoming Accelerated Placement tests. Recently, they published a study with 20000 students nationwide showing that using their books to prepare for the AP Statistics exam resulted were 5% more likely to pass than those who studied using their rival, Karron's. However, The Brown Review did not publish all of their data, and you uncover the following data table:

	The Brown Review	Karron's	Total
Took a statistics class	$\frac{8550}{9000} = 95\%$	$\frac{4450}{4500} = 99\%$	$\frac{13000}{13500} = 96\%$
Did not take a statistics class	$\frac{750}{1000} = 75\%$	$\frac{4350}{5500} = 79\%$	$\frac{5100}{6500} = 78\%$
Total	$\frac{9300}{10000} = 93\%$	$\frac{8800}{10000} = 88\%$	$\frac{18100}{20000} = 90\%$

a. Make an argument as to why The Brown Review appears to be a better choice for test prep.

> *Answer:* The Brown Review has a higher total percentage of success.

b. Make an argument as to why Karron's is the better choice.

> *Answer:* Given that you are a student who took a statistics class, the chance you succeed with Karron's is higher. This is true eve if you didn't take a statistics class.
>
> *Prodding Question:* Let's pretend all students taking the test took a statistics class. Which test-prep book is better? What happens if we asssume all students weren't taking a statistics class?

c. Given that it is very likely that a student who takes a class before the exam succeeds more often than students who did not take a class, how could the Brown Review manipulate sample sizes such that their final percentage looks higher?

> *Answer:* By letting more of their sample draw from the students that took a statistics class, they are able to get more students that were going to succeed anyways into their sample. This would inflate the number of successes.
>
> *Prodding questions:*
>
> - Look at the denominators of each cell. Does something feel off about The Brown Review's?
> - Let's say I have two bags of 10 balls each. The first bag has 8 red balls and the second one has 5 red balls. Which bag would I reach out of if I want to increase the chances of getting a red ball? How does this situation apply if we replace red balls with passing the test and the bags with whether or not a student took the statistics class?

This contradiction of conclusions is known as Simpson's Paradox, which you can read more about here. This occurs when comparisons of two variables in separate groups yield one conclusion, but comparisons of the variables overall yield a different result due to a more significant trend working in the background, like how students taking or not taking the class determines outcomes.

d. It is very common for news services to read the conclusion of a scientific study and report the final results directly. It is also extremely common for readers to look at a news headline and not read the article. How do these practices encourage misinformation? Why do details of a study matter?

> *Pass condition:* Students should be able to explain why summarizing a study or article in a sentence or two results in a sizeable loss of nuance and information, leading to inaccurate or broad conclusions.
>
> *Prodding questions:*
>
> - Would saying that simply stating the conclusion - that the Brown Review is the superior option - give a clear view of the results of the study?
> - Do you think all studies are equally rigorous in their treatment of their data set and analysis?
> - How could researchers potentially manipulate the results of the same study to reach differing conclusions?

e. With the growing popularity of data analysis and machine learning, our society increasingly relies on statistical tools to explain the world we live in. Why is it extremely important to never take any conclusions we garner from these methods at face value? What are the dangers in becoming over-reliant on these methods to solve complex problems such as policing or policymaking?

Pass condition: Students should be able to explain the dangers of accepting "facts" or statistics at face value, without questioning the data that backs the claims. They should also recognize that real life is not so black and white, where we can apply a single statistic to all situations (e.g. this can lead to stereotyping, etc.)

Prodding questions:

- How much should we consider a person/company/entity's motive when they make claims? Can we trust everything they say?
- How can a study's sampling method, experimental procedure, etc. affect its results? How can confounding variables contribute to inaccurate conclusions?
- How can policing or policymaking properly be hindered by a reliance simply on statistics?

Task 6

Carmen goes to PVDonuts and gets four boxes of donuts for everyone. One box has two chocolate donuts, two boxes have one chocolate and one glazed, and one box has two glazed donuts.

a. Carmen picks a box at random, and gives a random donut to Tyler. If that donut is chocolate, what is the probability the other donut in the box is glazed?

> Let G be the event that the other donut is glazed, and C be the event that Carmen chooses a chocolate donut. We want to find $\Pr(G|C) = \Pr(G \cap C)/\Pr(C)$.
>
> $G \cap C$ is just the event that Carmen picks a box with a chocolate and glazed donut $(\frac{2}{4})$ *and* Carmen picks the chocolate donut first $(\frac{1}{2})$, for a total probability of $\frac{1}{4}$.
>
> C is the probability that Carmen picked a chocolate donut, which is $\frac{1}{2}$. This is because there are four chocolate donuts and four glazed donuts in total to pick from in the equally likely boxes, so there's a $\frac{1}{2}$ that a chocolate donut is picked by Carmen.
>
> So, the total probability is
>
> $$\Pr(G|C) = \frac{\left(\frac{1}{4}\right)}{\left(\frac{1}{2}\right)} = \frac{1}{2}$$
>
> .

b. Let's consider another scenario, where Tyler picks a donut box at random. Before he opens it, Carmen tells him one of the donuts in the box is chocolate.

What is the probability the other donut is glazed?

Hint: Your answer should be different from part a.

> Let G be the event that the other donut is glazed, and C be the event that Tyler chooses a box with at least one chocolate donut.
>
> $G \cap C$ is the event where Tyler chooses one of the chocolate/glazed boxes, for a $\frac{1}{2}$ chance. C is the event where Tyler chooses any box with chocolate inside, with probability $\frac{3}{4}$.
>
> So, $\Pr(G|C) = \frac{\frac{1}{2}}{\frac{3}{4}} = 2/3$.

***Optional:* Task 7**

Suppose that, in a certain family, the probability of each child being born with a cat allergy is 1/2. You can assume one of the parents have it and it gets inherited with 1/2 probability. Assume this family has two children, Tyler and Allie.

a. What is the probability that both Tyler and Allie have a cat allergy?

b. Consider a scenario where you go to the family's house and meet one of the children. They tell you they have a cat allergy, but not their name. What is the probability the other child is allergic?

c. The child you meet says their name is Tyler. Does the probability of the other child being allergic change?

d. Given that at least one of the children is allergic and was born on a Tuesday, what is the probability the family has 2 allergic children? You may assume that the probability a child is born on a given day of the week is 1/7.

a. $\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$.

b. Note that you are given that one of the children (without knowing which one) has a cat allergy. Hence, from here there are three equally likely ways for this outcome to occur: both have it (the other child has the allergy with probability $\frac{1}{2}$), Allie has it but not Tyler (the other child does not have the allergy with probability $\frac{1}{2}$), and Tyler has it but not Allie (same as previous case). So, it's $\frac{1}{3}$ (the probability that first case discussed occurs). Note that this problem doesn't ask you to factor in the probability of meeting the child with the allergy.

c. Yes, it changes. Now the only options are Tyler has it but not Allie, and both having it. So it now becomes $\frac{1}{2}$.

d. If the **first** person is allergic and born on a Tuesday, there are 14 options for allergy/day combinations for the second person.

   If the **second** person is allergic and born on a Tuesday, there are again 14 options for the first person.

   However, there are only 27 unique outcomes, since we counted the outcome where both are allergic and born on a Tuesday in both cases. So, our total number of outcomes is 27.

   Out of these, there are 7 options in the first case where the other person is allergic, and 7 options in the second case. These also overcount the case where both are allergic, so the number of times where both are allergic is 13.

The final probability is 13/27, not quite 1/2!

Alternate Solution:

Let A be the event where both are allergic, and B be the event where at least one child is allergic and was born on a Tuesday.

Using Bayes' rule:

- $\Pr(B|A)$ can be reduced to the probability at least one person was born on a Tuesday. This is $1 - \frac{6}{7}^2 = 36/49$ by the complement rule.
- $\Pr(A) = \frac{1}{4}$ as from before.
- $\Pr(B)$ is the probability *no* child is allergic and born on a Tuesday. By the complement rule, this is $1 - (\frac{13}{14})^2 = 27/196$.

$\Pr(A|B) = \Pr(B|A) Pr(A) / \Pr(B) = 13/27$.

Checkoff — Call over a TA!