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 1**

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 \( \Pr(X = 0) \)?

\[ \frac{1}{2}, \frac{1}{2} \]

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 \( 0 \leq k \leq n? \)

\[ \text{Hint: You computed this value in last week’s recitation.} \]

\[ \binom{n}{k}/2^n \]

c. Let \( C_i \) be a random variable for the \( i \)th 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. 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 1 — Call over a TA!
Expected Value

Intuitively, the expected value is the weighted average of values, kind of a mass center of the probability distribution.

More formally, the expected value of a random variable is denoted $\mathbb{E}[X]$ and is defined as

$$\mathbb{E}[X] = \sum_{s \in S} X(s) \Pr(s) = \sum_{r \in X(S)} r \Pr(X = r).$$

We define the conditional expected values as follows: Given that event $E$ has occurred, the expectation of random variable $X$ is

$$\mathbb{E}[X \mid E] = \sum_{r \in X(S)} r \Pr(X = r \mid E). \quad (1)$$

Moreover, the linearity of expectation can be very useful in calculating expected value: Given that $Z, X, Y$ are three random variables defined on a sample space $S$ and $a$ and $b$ are two real numbers such that $Z = aX + bY$, we know that $\mathbb{E}[Z] = a\mathbb{E}[X] + b\mathbb{E}[Y]$ must be true.

Let’s practice this through a task:

Task 2

Tim and Joe are playing a game. They flip a fair coin 3 times. When the coin is heads, Joe pays $1 to Tim; and when the coin is tails, Tim pays $1 to Joe.

a. Let $G_i$ be a random variable representing what Tim gains on the $i$-th round. For instance, $G_3 = -1$ if the coin is tails. What is the expected value of $G_i$?

Since $\Pr(G_i = 1) = \frac{1}{2}$ and $\Pr(G_i = -1) = \frac{1}{2}$, $\mathbb{E}[G_i] = \frac{1}{2} - \frac{1}{2} = 0$.

b. Let $G$ be a random variable that represents Tim’s total gain in this game. What is the expected value of $G$?

$$\mathbb{E}[G] = \mathbb{E}[G_1] + \mathbb{E}[G_2] + \mathbb{E}[G_3]$$

where $G_i$ is the expected gain from flip $i$. We already know that $\mathbb{E}[G_i] = 0$, so $\mathbb{E}[G] = 0 + 0 + 0 = 0$ as well.

c. What is the expected value of $G$ if the coin is biased and the probability of heads is $p$? in other words, generalize your solution from part b in terms of $p$. 

3
\[
\begin{align*}
\text{Now, } E[G_i] &= p \cdot 1 + (1-p) \cdot (-1) = p - 1 + p = 2p - 1. \quad \text{Thus, } E[G] = E[G_1] + E[G_2] + E[G_3] = (2p - 1) + (2p - 1) + (2p - 1) = 6p - 3.
\end{align*}
\]

d. Tim and Joe are still using the biased coin from part c. Let \( H_1 \) be the event that the first coin is heads. What is \( E[G | H_1] \)?

\[
E[G | H_1] = 1 + (2p - 1) + (2p - 1) = 4p - 1.
\]
\( G_1 \) is 1 from the first round guaranteed, and the other terms are from part c.

e. Use your answers to calculate \( E[G] \) and \( E[G | H_1] \) when \( p = 0.7 \).

\[
\begin{align*}
E[G] &= 6p - 3 = 6 \cdot 0.7 - 3 = 1.2 \\
E[G | H_1] &= 4p - 1 = 4 \cdot 0.7 - 1 = 1.8
\end{align*}
\]

f. Optional: Assume that \( p = 0.7 \) and let’s say we want to change the game to make it “fair.” If the flip is tails, then Tim pays a dollar to Joe—how much should Joe pay Tim on Heads so that for any number of flips we know \( E[G] = 0 \)?

Each \( E[G_i] \) is \( 0.3 \cdot (-1) + 0.7 \cdot x \), where \( x \) is the number of dollars that Joe should pay Tim on Heads. Thus, we have \( E[G] = (0.3 \cdot (-1) + 0.7 \cdot x) \cdot 3 = 0 \). Solving for \( x \), we have \( x = \frac{3}{7} \). So, Joe should pay Tim $0.43 on heads so that the game is fair.

**Task 3**

Tim and Joe are now playing a similar, but different, game. This time they flip a coin 2 times. Let \( X \) be the random variable that is equal to the number of heads and \( Y \) the random variable that is equal to the number of tails. At the end of the game, Joe pays Tim \( X^2 \) dollars. Once again, let \( G \) be the random variable for Tim’s gain.

a. Calculate \( E[X^2] \) when the probability of heads is \( p = 0.7 \).

\[
E[X] = 0.7 + 0.7 = 1.4. \quad \text{So, } E[X^2] = 1.4^2 = 1.96
\]

b. Calculate \( E[G] \) when \( p = 0.7 \).

\[
E[G] = E[X^2] = (0.7)^2(4) + \binom{2}{1}(0.3)(0.7)(1) = 2.38
\]

No!
In general for any random variable with nonzero variance we have $E[X^2] > (E[X])^2$. But you will learn about variance in more advanced courses!

d. Find an example that shows that $E[XY] = E[X]E[Y]$ does not hold where $X$ and $Y$ are not independent variables.

Hint: Try $X$ and $Y$ as defined above.

We know that $X$ and $Y$, as defined in this problem, are not independent variables. We know that $E[X]E[Y] = 1 \cdot 1 = 1$. We can also calculate $E[XY]$ to be $(\frac{1}{4})(2)(0) + (\frac{1}{2})(1)(1) = (\frac{1}{4})(0)(2) = \frac{1}{2}$.
Thus, this is an example of where $X$ and $Y$ are dependent variables where $E[XY] \neq E[X]E[Y]$.

e. Optional: Prove that if $X$ and $Y$ are two independent random variables, then $E[XY] = E[X]E[Y]$.

Assume $X$ and $Y$ are independent. So for any $x$ and $y$ in their range we have $\Pr(X = x \wedge Y = y) = \Pr(X = x) \Pr(Y = y)$. We now use the definition of expected value.

\[
E[XY] = \sum_{x \in X(S)} \sum_{y \in Y(S)} xy \Pr(X = x \wedge Y = y) \quad (2)
\]

\[
= \sum_{x \in X(S)} \sum_{y \in Y(S)} xy \Pr(X = x) \Pr(Y = y) \quad (3)
\]

\[
= \left( \sum_{x \in X(S)} x \Pr(X = x) \right) \left( \sum_{y \in Y(S)} y \Pr(Y = y) \right) \quad (4)
\]

\[
= E[X]E[Y] \quad (5)
\]
Bayes Rule

The Bayes Rule can be summarized as

\[ P(A|B) = \frac{P(B|A)P(A)}{P(B)} \]

where \( A \) and \( B \) are events and \( P(B) \neq 0 \).

Task 4

Assume Brown’s CS department has an evaluation system for CS courses based on student evaluations. In any class, the students can fill the evaluation form and give a score of 0, 1, or 2 to the course. Let \( X \) be the random variable of this score. The students of CS0220 either like the course with probability 3/4 (Event \( L \)) or they do not like the course with probability 1/4 (Event \( \text{NOT } L \)).

Assume that the conditional probability distribution of \( X \) given \( L \) is

\[
\begin{align*}
\Pr(X = 0 \mid L) &= 1/8 \\
\Pr(X = 1 \mid L) &= 1/4 \\
\Pr(X = 2 \mid L) &= 5/8
\end{align*}
\]

and given that they do not like the course (NOT \( L \)) it is

\[
\begin{align*}
\Pr(X = 0 \mid \text{NOT } L) &= 9/10 \\
\Pr(X = 1 \mid \text{NOT } L) &= 1/10 \\
\Pr(X = 2 \mid \text{NOT } L) &= 0.
\end{align*}
\]

a. If a student has given score of 0 to CS0220, what is the probability that they do not like the course?

\[
\Pr(\text{NOT } L \mid X = 0) = \frac{\Pr(X=0|\text{NOT } L)\Pr(\text{NOT } L)}{\Pr(X=0)} = \frac{9/10 \cdot 1/4}{1/8} = \frac{12}{17}.
\]

The formula for conditional probability \( \Pr(\text{NOT } L \mid X = 0) = \frac{\Pr(\text{NOT } L \cap X = 0)}{\Pr(X=0)} \) gives the same answer.

Note: Drawing out a diagram like this may be helpful for this problem:
b. Use the definition of conditional expected value (Equation 1) and find $E[X \mid \text{NOT } L]$

$$E[X \mid \text{NOT } L] = \sum r \ Pr(X = r \mid \text{NOT } L) = 2 \cdot 0 + 1 \cdot \frac{1}{10} + 0 \cdot \frac{9}{10} = \frac{1}{10}.$$ 

\[ E[X] \]

\[= 2 \cdot \left( \Pr(X = 2 \mid L) \cdot \Pr(L) + \Pr(X = 2 \mid \text{NOT } L) \cdot \Pr(\text{NOT } L) \right) + \Pr(X = 1 \mid L) \cdot \Pr(L) + (\Pr(X = 1 \mid \text{NOT } L) \cdot \Pr(\text{NOT } L)) \]
\[= 2 \cdot \left( \frac{3}{4} \cdot \frac{5}{8} + 1 \cdot \frac{3}{4} \cdot \frac{1}{4} + 1 \cdot \frac{1}{4} \cdot \frac{1}{10} \right) = \frac{23}{20} = 1.15. \]

Alternative solution: $E[X] = \Pr(L) E[X \mid L] + \Pr(\text{NOT } L) E[X \mid \text{NOT } L] = \frac{3}{4} \cdot \left( \frac{5}{8} \cdot 2 + \frac{1}{4} \cdot 1 \cdot \frac{1}{8} \cdot 0 \right) + \frac{1}{4} \cdot \left( 0 \cdot \frac{2}{10} + \frac{1}{10} \cdot 1 \cdot \frac{9}{10} \cdot 0 \right) = \frac{23}{20} = 1.15$ 

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:

<table>
<thead>
<tr>
<th></th>
<th>The Brown Review</th>
<th>Karron’s</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Took a statistics class</td>
<td>8550 / 9000 = 95%</td>
<td>4450 / 4500 = 99%</td>
<td>13000 / 13500 = 96%</td>
</tr>
<tr>
<td>Did not take a statistics class</td>
<td>750 / 1000 = 75%</td>
<td>4350 / 5500 = 79%</td>
<td>5100 / 6500 = 78%</td>
</tr>
<tr>
<td>Total</td>
<td>9300 / 10000 = 93%</td>
<td>8800 / 10000 = 88%</td>
<td>18100 / 20000 = 90%</td>
</tr>
</tbody>
</table>

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 even 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 assume 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

Joe 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. Joe picks a box at random, and gives a random donut to Tim. 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 Joe 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 Joe picks a box with a chocolate and glazed donut (\( \frac{2}{4} \)) and Joe picks the chocolate donut first (\( \frac{1}{2} \)), for a total probability of \( \frac{1}{4} \).

\( C \) is the probability that Joe 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 Joe.

So, the total probability is

\[
\Pr(G|C) = \frac{\left(\frac{1}{2}\right)}{\left(\frac{1}{2}\right)} = \frac{1}{2}
\]

b. Let’s consider another scenario, where Tim picks a donut box at random. Before he opens it, Joe 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 Tim chooses a box with at least one chocolate donut.

\( G \cap C \) is the event where Tim chooses one of the chocolate/glazed boxes, for a \( \frac{1}{2} \) chance. \( C \) is the event where Tim chooses any box with chocolate inside, with probability \( \frac{3}{4} \).

So, \( \Pr(G|C) = \frac{1}{4} = 2/3 \).
Optional: Task 7

Suppose that, in a certain family, the probability of each child being born with a cat allergy is \( \frac{1}{2} \). You can assume one of the parents have it and it gets inherited with \( \frac{1}{2} \) probability. Assume this family has two children, Jiahua and Ken.

a. What is the probability that both Jiahua and Ken 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 Jiahua. 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 \( \frac{1}{7} \).

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 \( \frac{1}{2} \)), Ken has it but not Jiahua (the other child does not have the allergy with probability \( \frac{1}{2} \)), and Jiahua has it but not Ken (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 Jiahua has it but not Ken, 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 $1 - \frac{6^2}{7} = 36/49$ 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 $1 - \left(\frac{13}{14}\right)^2 = 27/196$.

$$
\Pr(A|B) = \frac{\Pr(B|A)\Pr(A)}{\Pr(B)} = \frac{13}{27}.
$$

Checkoff — Call over a TA!