This is a draft; please send me corrections and/or suggestions.

Conditional distributions

We have already seen the concept of conditional probability — $\mathrm{P}(A|B) = \frac{\mathrm{P}(A,B)}{\mathrm{P}(B)}$ — and have generalized this to the idea of a conditional PMF — $f_{X|Y=y}(x) = \frac{f_{XY}(x,y)}{f_Y(y)}$. Left implicit in these discussions has been the fact that a conditional PMF actually represents a probability distribution. This is not too difficult to show.

Theorem

Let X and Y be random variables with joint PMF f_{XY}. Then for any $y \in \mathrm{Supp}(Y)$, the conditional PMF $f_{X|Y=y}$ represents a probability distribution.

Proof: there are two features of a probability distribution which must be shown.

- $f_{X|Y=y}(x) \geq 0$. This will hold if and only if $\frac{f_{XY}(x,y)}{f_Y(y)} \geq 0$. But since $y \in \mathrm{Supp}(Y)$, we know $f_Y(y) > 0$. Since f_{XY} is a probability distribution, we know $f_{XY}(x,y) \geq 0$, and therefore $\frac{f_{XY}(x,y)}{f_Y(y)} \geq 0$.
- $\sum_{x \in \mathrm{Supp}(X)} f_{X|Y=y}(x) = 1$. We can see

$$\begin{aligned}
\sum_{x \in \mathrm{Supp}(X)} f_{X|Y=y}(x) &= \sum_{x \in \mathrm{Supp}(X)} \frac{f_{XY}(x,y)}{f_Y(y)} \\
&= \frac{1}{f_Y(y)} \sum_{x \in \mathrm{Supp}(X)} f_{XY}(x,y) \\
&= \frac{1}{f_Y(y)} \left(f_Y(y)\right) \qquad \text{(law of total probability)} \\
\sum_{x \in \mathrm{Supp}(X)} f_{X|Y=y}(x) &= 1.
\end{aligned}$$

The two defining features of a probability distribution are therefore satisfied, hence $f_{X|Y=y}$ represents a probability distribution.

□

This may not seem like such a big deal, but it allows us to speak generally about concepts like the expectation and variance of a conditional distribution just as we would with a "regular" distribution. In particular, the conditional expectation of X given $Y = y$ is

$$\mathbb{E}\left[X|Y=y\right] \equiv \sum_{x \in \mathrm{Supp}(X)} x f_{X|Y=y}(x).$$

Notice that this definition is identical to our standard definition of the expectation of a random variable, except instead of using the marginal distribution of the variable we are using the conditional distribution.

Conditional expectations have a nice property that we refer to as the *law of iterated expectation.*

Theorem

Let X and Y be random variables with joint PMF f_{XY}. Then

$$\mathbb{E}_Y\left[\mathbb{E}_X\left[g(X,y)|Y=y\right]\right] = \mathbb{E}_{XY}\left[g(X,Y)\right],$$

where $\mathbb{E}_S[\cdot]$ represents the expectation taken where the variables in S are random.

Proof: we can expand the left-hand side,

$$\begin{aligned}
\mathbb{E}_Y\left[\mathbb{E}_X\left[g(X,y)|Y=y\right]\right] &= \sum_{y\in\text{Supp}(Y)}\left(\sum_{x\in\text{Supp}(X)} g(x,y) f_{X|Y=y}(x)\right) f_Y(y) \\
&= \sum_{y\in\text{Supp}(Y)}\left(\sum_{x\in\text{Supp}(X)} g(x,y)\frac{f_{XY}(x,y)}{f_Y(y)}\right) f_Y(y) \\
&= \sum_{y\in\text{Supp}(Y)}\sum_{x\in\text{Supp}(X)} g(x,y) f_{XY}(x,y) \\
\mathbb{E}_Y\left[\mathbb{E}_X\left[g(X,y)|Y=y\right]\right] &= \mathbb{E}_{XY}\left[g(X,Y)\right].
\end{aligned}$$

□

The law of iterated expectations has a nice corollary regarding larger numbers of random variables.

Corollary

Let X, Y, and Z be random variables with joint PMF f_{XYZ}. Then

$$\mathbb{E}_Y\left[\mathbb{E}_X\left[g(X,y,z)|Y=y\right]|Z=z\right] = \mathbb{E}_{XY}\left[g(X,Y,z)|Z=z\right].$$

We will apply these features in the following question.

Differing levels of effort

One interesting feature of class grades is that they often display a *bimodal* distribution (i.e., one with two peaks). Figure 1 is an example. While not universal, this regularity is seen often enough that there are some theories regarding its structure. One simple explanation is that there are two types of people, "capable" and "incapable;" conditional on their type, grade distributions are identical, but those who are capable have some fixed quantity added to their grade while those who are incapable have some fixed quantity subtracted from their grade[1]; consider Figure 2 as an example. That is, if G a random variable denoting the final grade, C is 1 if a student is capable and -1 if they are incapable, and X is some randomness, we might have

$$G = kC + X.$$

We will consider a slightly different and more charitable version here.

Let Y represent whether a student is a studier (S) or a non-studier (N), and let X be the student's grade on an exam, $\text{Supp}(X) = \{\text{A}, \text{B}, \text{C}, \text{D}\}$. The joint probability of X and Y is given by

[1] See, e.g., Dehnadi and Bornat, 2006, *The camel has two humps.*

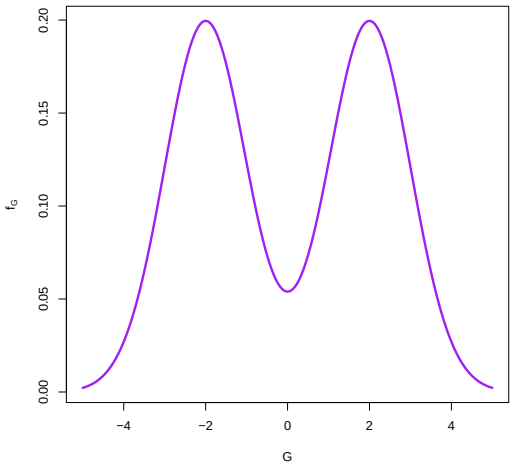

Figure 1: an example bimodal distribution.

	A	B	C	D
S	$\frac{2}{9}$	$\frac{1}{9}$	0	0
N	0	$\frac{1}{9}$	$\frac{1}{3}$	$\frac{2}{9}$

(a) *What is the conditional distribution of X when $Y = S$? When $Y = N$?*

Solution: we know that $\mathrm{P}(X = x|Y = \mathrm{S}) = \frac{\mathrm{P}(X=x, Y=\mathrm{S})}{\mathrm{P}(Y=\mathrm{S})}$. So se begin by computing

$$\mathrm{P}(Y = \mathrm{S}) = \mathrm{P}(X = \mathrm{A}, Y = \mathrm{S}) + \mathrm{P}(X = \mathrm{B}, Y = \mathrm{S}) = \frac{1}{3} \quad \Longrightarrow \quad \mathrm{P}(Y = \mathrm{N}) = \frac{2}{3}.$$

Then the conditional distributions are given by

	A	B	C	D
S	$\frac{2}{3}$	$\frac{1}{3}$	0	0

	A	B	C	D
N	0	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{1}{3}$

(b) *Let G represent the GPA associated with X (i.e., $X = A$ implies $G = 4.0$). What is the conditional expectation of G when $Y = S$? When $Y = N$?*

Solution: we can see

$$\begin{aligned}
\mathbb{E}\left[G|Y = \mathrm{S}\right] &= \sum_{g \in \mathrm{Supp}(G)} g\mathrm{P}(G = g|Y = \mathrm{S}) \\
&= 4\left(\frac{2}{3}\right) + 3\left(\frac{1}{3}\right) + 2\left(0\right) + 1\left(0\right) \\
\mathbb{E}\left[G|Y = \mathrm{S}\right] &= \frac{11}{3}.
\end{aligned}$$

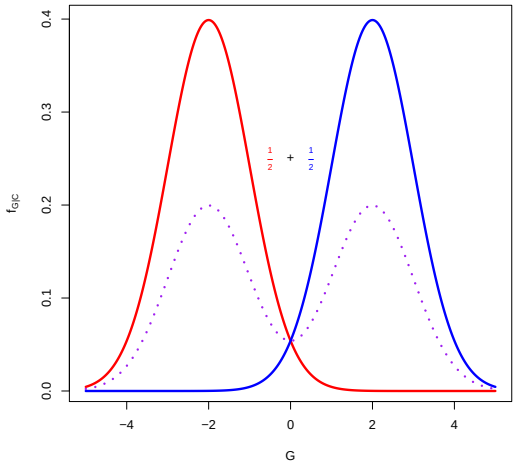

Figure 2: the bimodal distribution expressed as a weighted sum of two unimodal distributions.

$$\begin{aligned}\mathbb{E}\left[G|Y=\mathrm{N}\right] &= \sum_{g\in\mathrm{Supp}(G)} g\mathrm{P}(G=g|Y=\mathrm{S}) \\ &= 4\left(0\right)+3\left(\frac{1}{6}\right)+2\left(\frac{1}{2}\right)+1\left(\frac{1}{3}\right) \\ \mathbb{E}\left[G|Y=\mathrm{N}\right] &= \frac{11}{6}.\end{aligned}$$

(c) *Use (b) and the law of iterated expectations to find* $\mathbb{E}\left[G\right]$.

Solution: the law of iterated expectations says that

$$\mathbb{E}\left[G\right] = \mathbb{E}_Y\left[\mathbb{E}_G\left[G|Y=y\right]\right].$$

Expanding, we have

$$\begin{aligned}\mathbb{E}\left[G\right] &= \sum_{y\in\mathrm{Supp}(Y)} \mathbb{E}_G\left[G|Y=y\right]\mathrm{P}(Y=y) \\ &= \left(\frac{11}{3}\right)\left(\frac{1}{3}\right)+\left(\frac{11}{6}\right)\left(\frac{2}{3}\right) \\ \mathbb{E}\left[G\right] &= \frac{22}{9}.\end{aligned}$$

(d) *What is the conditional distribution of* Y *when* $X=A$*? Compute the same for all other values in the support of* X.

Solution: we can see

$$\begin{aligned}\mathrm{P}(Y=\mathrm{S}|X=\mathrm{A}) &= \frac{\mathrm{P}(Y=\mathrm{S},X=\mathrm{A})}{\mathrm{P}(X=\mathrm{A})} \\ &= \frac{\frac{2}{9}}{\frac{2}{9}} \\ \mathrm{P}(Y=\mathrm{S}|X=\mathrm{A}) &= 1 \\ \implies \mathrm{P}(Y=\mathrm{N}|X=\mathrm{A}) &= 0,\end{aligned}$$

$$\begin{aligned}
\mathrm{P}(Y = \mathrm{S}|X = \mathrm{B}) &= \frac{\mathrm{P}(Y = \mathrm{S}, X = \mathrm{B})}{\mathrm{P}(X = \mathrm{B})} \\
&= \frac{\frac{1}{9}}{\frac{1}{9} + \frac{1}{9}} \\
\mathrm{P}(Y = \mathrm{S}|X = \mathrm{B}) &= \frac{1}{2} \\
\implies \mathrm{P}(Y = \mathrm{N}|X = \mathrm{B}) &= \frac{1}{2},
\end{aligned}$$

$$\begin{aligned}
\mathrm{P}(Y = \mathrm{S}|X = \mathrm{C}) &= \frac{\mathrm{P}(Y = \mathrm{S}, X = \mathrm{C})}{\mathrm{P}(X = \mathrm{C})} \\
&= \frac{0}{\frac{1}{3}} \\
\mathrm{P}(Y = \mathrm{S}|X = \mathrm{C}) &= 0 \\
\implies \mathrm{P}(Y = \mathrm{N}|X = \mathrm{C}) &= 1,
\end{aligned}$$

$$\begin{aligned}
\mathrm{P}(Y = \mathrm{S}|X = \mathrm{D}) &= \frac{\mathrm{P}(Y = \mathrm{S}, X = \mathrm{D})}{\mathrm{P}(X = \mathrm{D})} \\
&= \frac{0}{\frac{2}{9}} \\
\mathrm{P}(Y = \mathrm{S}|X = \mathrm{D}) &= 0 \\
\implies \mathrm{P}(Y = \mathrm{N}|X = \mathrm{D}) &= 1.
\end{aligned}$$

(e) *Suppose that the student takes two exams; call the first grade X_1 and the second grade X_2. The student either studies for both exams or for neither. What is the conditional distribution of X_2 given $X_1 = A$? Compute the same for all other values in the support of X.*

(hint: X_1 provides information about the probability that the student is studying; given that we know this probability, X_1 conveys no further information. That is, $\mathrm{P}(X_2|X_1, Y) = \mathrm{P}(X_2|Y)$. You will need to apply the law of total probability to make this hint useful.)

Solution: we can see

$$\begin{aligned}
\mathrm{P}(X_2|X_1 = \mathrm{A}) &= \mathrm{P}(X_2, Y = \mathrm{S}|X_1 = \mathrm{A}) + \mathrm{P}(X_2, Y = \mathrm{N}|X_1 = \mathrm{A}) \\
&= \mathrm{P}(X_2|X_1 = \mathrm{A}, Y = \mathrm{S})\mathrm{P}(Y = \mathrm{S}|X_1 = \mathrm{A}) + \mathrm{P}(X_2|X_1 = \mathrm{A}, Y = \mathrm{N})\mathrm{P}(Y = \mathrm{N})|X_1 = \mathrm{A}) \\
&= \mathrm{P}(X_2|Y = \mathrm{S})\mathrm{P}(Y = \mathrm{S}|X_1 = \mathrm{A}) + \mathrm{P}(X_2|Y = \mathrm{N})\mathrm{P}(Y = \mathrm{N}|X_1 = \mathrm{A}).
\end{aligned}$$

The right-hand multiples were computed in part (d) above; the left-hand multiples were computed in part (a) above. Putting these together, we get

	A	B	C	D
$X_2\|X_1 = \mathrm{A}$	$\frac{2}{3}$	$\frac{1}{3}$	0	0
$X_2\|X_1 = \mathrm{B}$	$\frac{1}{3}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{6}$
$X_2\|X_1 = \mathrm{C}$	0	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{1}{3}$
$X_2\|X_1 = \mathrm{D}$	0	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{1}{3}$

(f) *What is the conditional expectation of G_2 when $X_1 = A$? Compute the same for all other values in the support of X.*

Solution: with the data from part (d), it is easy to compute

$$\begin{aligned}\mathbb{E}\left[G_2|X_1=\mathrm{A}\right] &= \sum_{g\in\mathrm{Supp}(G_2)} g\mathrm{P}(G_2=g|X_1=\mathrm{A}) \\ &= 4\left(\frac{2}{3}\right)+3\left(\frac{1}{3}\right)+2\left(0\right)+1\left(0\right) \\ \mathbb{E}\left[G_2|X_1=\mathrm{A}\right] &= \frac{11}{3},\end{aligned}$$

$$\begin{aligned}\mathbb{E}\left[G_2|X_1=\mathrm{B}\right] &= \sum_{g\in\mathrm{Supp}(G_2)} g\mathrm{P}(G_2=g|X_1=\mathrm{B}) \\ &= 4\left(\frac{1}{3}\right)+3\left(\frac{1}{4}\right)+2\left(\frac{1}{4}\right)+1\left(\frac{1}{6}\right) \\ \mathbb{E}\left[G_2|X_1=\mathrm{B}\right] &= \frac{11}{4},\end{aligned}$$

$$\begin{aligned}\mathbb{E}\left[G_2|X_1=\mathrm{C}\right] &= \sum_{g\in\mathrm{Supp}(G_2)} g\mathrm{P}(G_2=g|X_1=\mathrm{C}) \\ &= 4\left(0\right)+3\left(\frac{1}{6}\right)+2\left(\frac{1}{2}\right)+1\left(\frac{1}{3}\right) \\ \mathbb{E}\left[G_2|X_1=\mathrm{C}\right] &= \frac{11}{6},\end{aligned}$$

$$\begin{aligned}\mathbb{E}\left[G_2|X_1=\mathrm{D}\right] &= \sum_{g\in\mathrm{Supp}(G_2)} g\mathrm{P}(G_2=g|X_1=\mathrm{D}) \\ &= 4\left(0\right)+3\left(\frac{1}{6}\right)+2\left(\frac{1}{2}\right)+1\left(\frac{1}{3}\right) \\ \mathbb{E}\left[G_2|X_1=\mathrm{D}\right] &= \frac{11}{6},\end{aligned}$$

Identification

Up until this point, we have been concerned with what data can tell us about the values of a partiuclar parameter. For example, we look at a random sample and determine the sample mean as a way of estimating the population mean; we can then estimate the variance of the underlying random variable to help build confidence intervals and test hypotheses. This portion of econometrics is referred to as *inference*, and addresses questions about what values a particular parameter might take.

However, when making statistical inferences we are implicitly assuming that there is something to be said about the parameter! There is no guarantee that the data will tell us what we want to know. Attempts to address this conundrum are referred to as *identification*, and this area of econometrics provides us the tools and methods necessary to determine the parameters for which the data will provide inferences.

As a simple example, suppose that X represents the height of a randomly-drawn Econ 41 student. We believe that the mean of this random variable is $\mathbb{E}\left[X\right]=h$. Using what we've learned so far, we can estimate h by computing the sample mean of a series of X_i observations. In this case, h is *identified*.

Now suppose that we believe the mean of this random variable is $\mathbb{E}\left[X\right]=h+k$. Using what we've learned so far, we can still estimate $h+k$ by computing the sample mean of a series of observations, but at this point we are stuck — we cannot disentangle h from k! That is, if the sample mean is 68 inches, we don't know

if $h = 0$ and $k = 68$, or if $h = 34$ and $k = 34$, or if $h = 136$ and $k = -68$, etc. While we can estimate the left-hand side of the equation, we have no direct way of knowing what values of h and k combine to reach the sample mean; in this case, neither h nor k are identified.

Clearly, this example is contrived: making the assumption that $\mathbb{E}[X] = h + k$ is ridiculous on its face. However, analogues of this issue is extremely common in applied economics; since economics is, at its heart, about providing advice as to how certain features of the economy affect other features, we need to be sure that parameters that we estimate are absent issues such as these. Identification gives us a way of codifying the exact assumptions we are making when we claim that a parameter takes a particular value.

Before we continue, it is useful to introduce some standard notation that we've so far avoided in class.

Definition

Given two random variables X and Y, if X is independent of Y we write $X \perp\!\!\!\perp Y$.

We will also reference the concept of *conditional independence*; this definition is **not** necessary for Econ 41 in general[2] but unfortunately we will need it here.

Definition

Given three random variables X, Y, and Z, we say that X and Y are *conditionally independent* given Z — written $X \perp\!\!\!\perp Y|Z$ — if

$$\begin{aligned} & \mathrm{P}(X = x, Y = y|Z) = \mathrm{P}(X = x|Z)\mathrm{P}(Y = y|Z) \\ \iff \quad & f_{XY|Z=z}(x, y) = f_{X|Z=z}(x) f_{Y|Z=z}(y). \end{aligned}$$

Equivalently,

$$\begin{aligned} & \mathrm{P}(X = x|Y = y, Z) = \mathrm{P}(X = x|Z) \\ \iff \quad & f_{X|Y=z,Z=z}(x) = f_{X|Z=z}(x). \end{aligned}$$

That conditional independence is different from general independence is somewhat nuanced and beyond the context of this course. The key takeaway is that conditional independence says that, fixing a random variable Z, X contains no information about Y and vice-versa.

Example

There are two types of people in the world: people who eat at Burger King (BK) and people who eat at Whole Foods (WFM); everyone who doesn't eat at BK eats at WFM, and vice versa. Let Y be a random variable concerning the frequency of heart attacks: $Y = 1$ if a person has a heart attack in the next year, and $Y = 0$ otherwise.

After careful observation, some public health researchers conclude that

$$\mathrm{P}(Y = 1|X = \mathrm{WFM}) = 2\%, \qquad \mathrm{P}(Y = 1|X = \mathrm{BK}) = 10\%.$$

Appropriately, the government is appalled at the frequency of heart attacks among the part of the population which eats every meal at BK. It enlists the help of some economists to figure out how best to address the

[2] That is, don't worry too much about it.

growing crisis. After much discussion, economists provide the expected answer: the government should initiate a so-called "Whopper tax" on all purchases at BK. Basic economics tells us that, barring some silly cases like Giffen goods and assuming that everyone has to eat, placing a tax on BK will increase prices at BK, decreasing demand at BK and increasing demand at WFM. Assume that the tax is large enough that everyone who is now eating at BK would give it up and start eating only at WFM.

Will this reduce the annual risk of heart attack in the group of people who currently eat at BK to the annual risk of those who currently eat at WFM?

Let's consider the two possible answers to this question.

- **Yes.** All the data in the model suggests that eating at WFM implies that your annual risk of heart attack is 2%. Provided that your TA is being honest, there's no evidence that we should expect anything else.
- **No.** On the surface, it may look like this is the case, but there may be something peculiar about people who eat at BK which will not change if we simply change where they eat. After all, even WFM sells chicken wings; you can't change a burger-eater to a kale-eater overnight.

Henceforth, we'll assume that **no** is the reasonable response; in the end, economics is about applying models to the real world, and a researcher who brazenly claimed that people who eat at BK are otherwise the same as people who eat at WFM would be roundly rebuked.

To quantify this, let's expand our outcome space to consider the feature $Z \in \{\mathrm{X}, \mathrm{N}\}$ representing whether someone exercises (X) or does not exercise (N). As it turns out, exercise is perfectly correlated with eating at WFM: everyone who eats at WFM exercises, and everyone who eats at BK does not.[3] In this light, we should reconsider the effects of the Whopper tax.

Will taxing BK change people's exercise habits?

Barring the idea that maybe BK-eaters don't exercise because of constant indigestion, it's probable that introducing this tax will not significantly affect an individual's proclivity toward exercise. That is, through the tax we can affect the consumption behavior in the market affected by the tax, but that's all; we cannot change behavior outside of this area. While this is certainly an assumption, it doesn't seem out-and-out unreasonable.

Why does this matter? Before we considered exercise, the introduction of a tax on BK purchases modified agents so that

$$(\mathrm{WFM}) \xrightarrow{\Delta\mathrm{tax}} (\mathrm{WFM}), \qquad (\mathrm{BK}) \xrightarrow{\Delta\mathrm{tax}} (\mathrm{WFM}).$$

That is, the tax changed BK-eaters to WFM-eaters. Once we consider exercise, the introduction of a tax on BK purchases modifies agents so that

$$(\mathrm{WFM}, \mathrm{X}) \xrightarrow{\Delta\mathrm{tax}} (\mathrm{WFM}, \mathrm{X}), \qquad (\mathrm{BK}, \mathrm{N}) \xrightarrow{\Delta\mathrm{tax}} (\mathrm{WFM}, \mathrm{N}).$$

We are now transforming BK-eating non-exercisers into WFM-eating non-exercisers.

Problem is we have no information about people who eat at WFM but never leave the couch. It could be that the *only* thing causing BK-eaters heart trouble was eating at BK, so their new annual risk of heart attack is 0%; on the other hand, it could be that they are all allergic to amaranth flour (in a way that causes heart attacks) and will die immediately, so their new annual risk of heart attack is 100%. Since we've

[3] Ridiculous? In a past life, I went to WFM every day to buy a fresh-made sandwich for lunch; when I got home after work, I would ride my bicycle for exercise. In 2006, BK released a series of video games which you could buy cheaply if you purchased a value meal. Naturally I immediately started eating at BK; when I got home after work, I would play the video games. I continued to do so until I had beaten the games, at which point I switched back to my WFM/exercise routine. Anecdotes are not statistics, but you get the point.

collected no data on this kind of person, we'll have to introduce some restrictions if we want to make any useful predictions about this population's new heart attack risk.

Roughly speaking, identification is the art of doing just this. We will now consider a set of individual assumptions we might make which will allow us to say something more about the outcome.

(a) $X \perp\!\!\!\perp Y|Z$. That is, the probability of a heart attack is independent of where you eat, conditional on how much you exercise.

If we make this assumption, we can see

$$\begin{aligned}\mathrm{P}(Y=1|X=\mathrm{WFM},Z=\mathrm{N}) &= \mathrm{P}(Y=1|Z=\mathrm{N})\\ &= \mathrm{P}(Y=1|X=\mathrm{BK},Z=\mathrm{N})\\ &= 10\%.\end{aligned}$$

That is, by taxing BK we will have switched where people eat but we will not have affected their risk of heart attack at all! This makes intuitive sense *given the assumption*: if the probability of a heart attack is independent of where you eat, it is completely determined by how much you exercise. Since the outcome is precise, we say that the value is *point identified.*

(b) $Z \perp\!\!\!\perp Y|X$. That is, the probability of a heart attack is independent of whether or not you exercise, conditional on where you eat.

If we make this assumption, we can see

$$\begin{aligned}\mathrm{P}(Y=1|X=\mathrm{WFM},Z=\mathrm{N}) &= \mathrm{P}(Y=1|X=\mathrm{WFM})\\ &= \mathrm{P}(Y=1|X=\mathrm{WFM},Z=\mathrm{X})\\ &= 2\%.\end{aligned}$$

That is, by taxing BK we will have switched where people eat, and the fact that they continue to not exercise doesn't matter at all. Again, this makes intuitive sense *given the assumption*: if the probability of a heart attack is independent of how much you exercise, it is completely determined by where you eat.

(c) *Suppose that Z represents more exercise than Z'. Then*

$$P(Y=1|X,Z) \leq P(Y=1|X,Z').$$

An assumption of this sort is referred to as a *monotonicity assumption*: the probability of an event is monotonic in one of the conditioning variables. Although in this particular setup, "Z represents more exercise than Z'" immediately implies that $Z=\mathrm{X}$ and $Z'=\mathrm{N}$, it is not hard to imagine a more general setup where people could exercise 30 minutes per week, 1 hour per week, etc.

Under this assumption, we can see

$$\mathrm{P}(Y=1|X=\mathrm{WFM},Z=\mathrm{N}) \geq \mathrm{P}(Y=1|X=\mathrm{WFM},Z=\mathrm{X}) = 2\%.$$

So the probability that a given one of the people who switched from BK to WFM has a heart attack within the next year is no lower than 2%. This result is intuitive, but far less precise than the point estimates above.

(d) *Suppose that X represents a healthier diet than X'. Then*

$$P(Y=1|X,Z) \leq P(Y=1|X',Z).$$

Under this assumption, we can see

$$\mathrm{P}(Y=1|X=\mathrm{WFM},Z=\mathrm{N}) \leq \mathrm{P}(Y=1|X=\mathrm{BK},Z=\mathrm{N}) = 10\%.$$

So the probability that a given one of the people who switched from BK to WFM has a heart attack within the next year is no higher than 10%. Again, this result is intuitive, but far less precise than the point estimates above.

(e) *(c) and (d).* We can now apply both the bounds to find

$$2\% \leq \mathrm{P}(Y = 1|X = \mathrm{WFM}, Z = \mathrm{N}) \leq 10\%.$$

Key in this set of derivations is the amount of work it takes to get a "reasonable" answer. When you naïvely consider this question, the natural answer to "What is $\mathrm{P}(Y = 1|X = \mathrm{WFM}, Z = \mathrm{N})$?" is that it's somewhere between 2% (the risk for "healthy" people) and 10% (the risk for "unhealthy" people). That is, by making the unhealthy people a little healthier, we shouldn't increase their risk of heart attack; and in the same way, we shouldn't see that they are at a lower risk than those who are still healthier (because they exercise). That is, while the answer is incredibly natural, it takes two substantial assumptions to formally reach the result.

Although our identification problems and approaches will grow more complex as the quarter comes to a close, in the end this is the nature of all identification issues: we would like to say something about an unknown parameter, and then look for ways that we can place assumptions which imply that we have the information we need. A fair bit of economics is devoted to dissecting these assumptions, to ensure that once the inference stage is reached the results are valid; as in (a) and (b) above, two different assumptions can yield wildly different results! Making assumptions intelligently is key, but our approach will only be to see whether or not an assumption identifies a parameter of interest.