Intro to Probability Primer

BIOS 20172

Monday 4/24/23

Last week, a lot of equations were thrown at you and some probability concepts might still not entirely make sense yet, but that's okay – we'll walk through some of the fundamental probability concepts and review with some examples. If you find certain parts of this primer confusing or still have additional questions, you can email me (thuang5@uchicago.edu) or any of the other TAs; we'll be happy to help. This primer will cover most of what lecture discussed in the first two classes of Week 5. The slides accompanying this primer cover more of the lecture material from the end of Week 5/beginning of Week 6.

Why is probability theory useful?

Let's say we have a bunch of data. It's great that we can describe them with everything we've learned so far (summary statistics, data visualization, etc.). However, this data might only contain measurements taken from a *sample*, a subset of the *population*. With probability, we can:

- discuss *random* events – in a structured manner
- generalize information from our data like sample statistics to the population (since we might not know the true population *parameter*) – a.k.a. **inference**
- look at *uncertainty* (very important to report!), which can be formalized when we report these generalizations
- create models of processes we see in the world around us, such as the COVID pandemic

Quick notation + terminology guide!

Don't worry about any of the weird symbols or fancy notation. What's actually important is understanding what the terms mean.

Ω – **sample space**: all possible outcomes in our setting

- If our situation was flipping a fair coin twice (this is a **random** process!), $\Omega = \{HH, HT, TH, TT\}$. Here, I'm just listing out every possible outcome that could occur.

Event: any subset of our sample space Ω

- You're familiar with this – the textbook and slides will call events names like Event A or Event B. Let's continue with the coin flip example. If Event A is the event we get a tail on the first flip, it would look like this: $A = \{TH, TT\}$, since we're filtering our sample space down to just those that would be part of Event A – ones that get tails on the first flip.

P – **probability measure**: an operator assigning probability to every event

- You also know this already. You might write it as P or Pr, whichever you prefer! If we're talking about the *event* that we get a tail on the first flip (called Event A), I would just write $P(A) = 1/2$.

$\cap$ – "intersection of" two events, taking **both** ("AND")

$\cup$ – "union of" two events, taking **either** ("OR")

How do we calculate the probability of 2 events both occurring (jointly)?

We use the **Multiplicative Law**. This states that:

$$P(A \cap B) = P(A)P(B \mid A)$$
$$= P(B)P(A \mid B)$$

Independent Events

To calculate probability of independent events both occurring: use the multiplicative law – except in an easier way

When one event occurring does not inform you about how the other event might occur, we would consider this independent. If we take Event A as well as Event B, they would be independent if:

$$P(A \cap B) = P(A)P(B)$$

Where does this equation come from? The general **multiplicative law** tells us:

$$P(A \cap B) = P(A)P(B \mid A)$$
$$= P(B)P(A \mid B)$$

If A and B are independent events, then $P(B \mid A)$ – the probability of B **given** A should be the same as $P(B)$ – the probability of B in general. Same thing happens with A: $P(A \mid B) = P(A)$. From here, we can get the equation for **the probability of independent events jointly occurring:**

$$P(A \cap B) = P(A)P(B)$$

What if I want to find the probability of either A or B happening, not necessarily both together?

The **Additive Law:**
$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

The Additive Law is useful for when you're calculating Event A **OR** Event B happening. Add up the probabilities, but don't forget to *subtract the overlapping parts* so you don't double-count/end up with the a larger probability than what you should get. Examples of phrases you might see in a problem where you can use this rule would be "probability of majoring in Neuroscience OR Econ" or "probability of someone having at least # of jobs".

- The main idea here is not ALL of the possible options that fulfill your conditions need to occur for it to "count."
- Ex. I want to calculate the probability that a student here is a Neuroscience OR Econ major. We would first take $P(Neuro) + P(Econ)$. Then we subtract the overlapped part (double majors who have already been counted once in Neuro, then double-counted in Econ):

$$P(Neuro \cup Econ) = P(Neuro) + P(Econ) - P(Neuro \cap Econ)$$

Translation: The probability someone is a neuroscience OR econ major is equal to the sum of the probability someone is a neuroscience major and the probability someone is an econ major, minus the probability someone is majoring in both neuroscience and economics.

- Ex. I want to calculate the probability that someone has *at least* 2 jobs. I might add up the probabilities that someone has exactly 2 jobs, that someone has exactly 3, that someone has exactly 4, etc.

Mutually Exclusive Events

Two events are **mutually exclusive** if they don't overlap anywhere. What does this mean? It means if Event A and Event B are mutually exclusive, Event A occurring means that Event B will not occur.

- **Recall the definition for independent events.** If each event's probability > 0, **mutually exclusive events are NOT independent because one event occurring will inform us about the probability of the second event occurring.**

One cool thing about the additive law is that it works no matter if your events are mutually exclusive or not. Once again, we have our equation:

$$P(A) + P(B) - P(A \cap B)$$

However, the last part of this formula ($P(A \cap B)$) is always going to be 0 for mutually exclusive events, since the probability that both events occur will be 0 – there are no intersecting values ($A \cap B = \emptyset \implies P(A \cap B) = 0$).

That simplifies our equation to:

$$P(A) + P(B)$$

If you picture a typical Venn Diagram with 2 overlapping sets, think of 2 mutually exclusive events as 2 disjoint sets – the circles don't overlap. That is why you can just add up $P(A) + P(B)$ without subtracting the overlap, $P(A \cap B)$.

Knowing if events are mutually exclusive can help us simplify our calculations in other ways as well.

- Let's say we want to find the probability that someone has at least 2 jobs, from our example in the last section. Instead of adding up every single probability for 2 jobs, 3 jobs, 4 jobs, etc., we can do something easier. The probability that everyone has 0 jobs or 1 job or 2 jobs and so on is 1 – since this includes everyone. There might be some people out there who have a LOT of jobs, and adding up the probabilities for exactly each number of jobs up to the maximum number of jobs will take too much time. So we subtract from 1 (this is everyone!) the probability that *someone has less than 2 jobs*, which is the same as the probability that someone has **exactly** *0 jobs* or **exactly** *1 job*:

$$P(jobs \geq 2) = 1 - P(jobs = 0) - P(jobs = 1)$$

- *A very important note:* in these kinds of problems, don't forget about the probability that the number of jobs or any other x = 0. It's easy to forget to add the probability that someone has 0 jobs, but we need it in order for us to calculate the probability by subtracting from 1, as 1 is inclusive of everyone no matter how many jobs they have $(0, 1, 2, \dots)$.

- Wording is *very* important. Pay attention to the difference between **at least** and **exactly** (or **only**).

Conditional Probability

The **conditional probability** of an event is exactly what it sounds like: the probability of an event *conditional* (or *given*) on another event.

You'll see "probability of event A given B" as something like:

$$P(A \mid B)$$

This is the same as:

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

if the probability of event B occurring is greater than 0 – a.k.a. "Out of all of the times event B occurs, how many times does event A also occur?"

Example: Returning to our example with 2 coin flips, we can try calculating the probability we get tail on our second flip – Event A – *given* we get a tail on our first flip – Event B.

Possible outcomes (sample space) of flipping a fair coin twice: $\{HH, HT, TH, TT\}$

- Event A is the event we get a T second: $\{HT, TT\}$
- Event B is the event we get a T first: $\{TH, TT\}$
- $P(A) = 1/2$
- $P(B) = 1/2$
- $P(A \mid B) =$ out of all the outcomes in Event B (TH or TT), the probability we get Event A (HT or TT)

It's pretty clear from looking at all the outcomes that $P(A \mid B)$ is just $1/2$. We're basically limiting ourselves to just TH or TT instead of all 4 original possibilities, and out of those, the probability we get HT or TT is just $1/2$.

Now let's put it in calculation terms!

- One thing we have to look at now is $P(A \cap B)$, or probability of Event A *and* Event B both occurring – both flips are tails.
- Out of $\{HH, HT, TH, TT\}$, the probability we see both tails would be $1/4$; $P(A \cap B) = 1/4$.

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$
$$= \frac{1/4}{1/2}$$
$$= 1/2$$

This may be obvious – we know intuitively that the probability of the first coin flip being tails won't do anything to the probability of the second coin flip being tails. However, we can also use this calculation for conditional probability to see if the two events are **independent.**

$$P(A \mid B) = 1/2$$
$$P(A) = 1/2$$
$$P(A) = P(A \mid B)$$

- We see that Event A and Event B are independent because Event A is equal to the probability of Event A **given** Event B. Even after Event B occurs, the probability of Event A doesn't change in this example, making it independent!
- Remember our definition for independent events? **When one event occurring does not inform you about how the other event might occur**, the two events are independent.

Another example: Let's find the probability that we get tails on *both flips* (Event C) given that we get tails on the *first flip* (Event D) .

Possible outcomes (sample space) of flipping a fair coin twice: $\{HH, HT, TH, TT\}$

- Event C is the event we get T on both flips: $\{TT\}$
- Event D is the event we get a T first: $\{TH, TT\}$
- They overlap in one outcome: TT.
- $P(C) = 1/4$
- $P(D) = 1/2$

- $P(C \cap D) = 1/4$ (where C and D would both occur, a.k.a. TT)
- $P(C \mid D)$ = out of all the outcomes in Event D (TH or TT), the probability we get Event C (TT)

Just by looking at the outcomes, we can see that $P(C \mid D)$ will be 1/2. In calculation form:

$$P(C \mid D) = \frac{P(C \cap D)}{P(D)}$$
$$= \frac{1/4}{1/2}$$
$$= 1/2$$

Now let's see if Event C and Event D are independent.

$$P(C \mid D) = 1/2$$
$$P(C) = 1/4$$
$$P(C) \neq P(C \mid D)$$

They are **not** independent! When Event D occurs/when we get tails on the first flip, that will change things + inform us about the probability of Event C – the probability that we get tails on both flips.

Given Event A and Event B:

Notation	Meaning
$P(A \cup B)$	probability of Event A **or** Event B occurring (union)
$P(A \cap B)$	probability of Event A **and** Event B occurring (intersection)
$P(A \mid B)$	probability of Event A occurring **given** Event B

	Formula	Interpretation
Additive Law	$P(A \cup B) = P(A) + P(B) - P(A \cap B)$	probability of Event A **or** Event B occurring (union)
if events are mutually exclusive	$P(A \cup B) = P(A) + P(B)$	
Multiplicative Law	$P(A \cap B) = P(A)P(B \mid A) = P(B)P(A \mid B)$	probability of Event A **and** Event B occurring (intersection)
if events are independent	$P(A \cap B) = P(A)P(B)$	
Conditional Probability*	$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$	probability of Event A occurring **given** Event B

*Note that the formula for conditional probability is just the multiplicative law rewritten.

Revisiting conditional probability: Bayes' Theorem

Bayes' Theorem may seem more complicated than it really is at first. It also doesn't look very intuitive as a formula compared to everything else you've learned so far. However, you can find Bayes' Theorem by taking the multiplicative law and rewriting it to find conditional probability. This makes sense since Bayes' is also a calculation for some type of conditional probability. Below we begin with the multiplicative law and end up with Bayes' Theorem for $P(A \mid B)$:

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

The review slides go more into depth on the later lecture material from end of Week 5+beginning of Week 6–more about Bayes' Theorem, the Law of Total Probability, an example implementing both, and some extra practice problems (with solutions).

References

Material drawn from the course textbook, slides from Prof. Haddadian, and notation/examples from Prof. Offer-Westort.