

3.1 Brief Overview of Probability

PLSC30500, Fall 2021

co-taught by Molly Offer-Westort & Andy Eggers

(This lecture with references to Aronow & Miller (2019) and Wasserman (2004))

What do we do with data

Now that we've gotten started on working with data...what do we want to get from that data?

- Describe what's going on in the data
 - We can do a pretty good job of this with the data visualization tools we have, along with summary statistics for numerical descriptions

This is already a really useful start, but beyond just describing the data we see in front of us, we may have other goals. We may want to:

- Make generalizations to a larger population
- Make informed guesses about how things would have turned out differently, if something different had happened

These latter two are called *inference*, and we'll need some additional tools and assumptions to make headway on them

Probability

Why do we use probability theory?

- Probability theory allows us to talk about *random* events in structured way.
 - Often, we see only part of the picture we'd like to talk about. We only see some of the data, or we can only see one version of events.
 - Probability theory gives us a way to describe the process that results in the data that we observe.
 - It also allows us to *formalize our uncertainty* when making inference.

What does it mean to describe real world events as "random"?

- Probability theory is an abstract construct, but it is useful for empirical research to create a model of the world in which events are probabilistic.
 - If we are conducting measurement on some population but can only observe a sample, we assume that there is some randomness to who we observe and who we don't observe.
- If we are describing events with counterfactuals, such as turning out to vote or testing positive for COVID, it can be useful to describe those events as probabilistic.
- These are *models* of how the world works, and they help us make sense of the fundamentally squishy nature of social science research. With limited information about the world, we operate with uncertainty. Assigning probabilities conditional on the information we *do* have helps us formalize that uncertainty. Even if we don't necessarily believe that human behavior is "random."

Flipping a coin twice

Suppose we are flipping a coin twice, and the coin is fair. This is a random process, and we will describe the probability space associated with this process.

Useful terms in probability

- Ω : Sample space. Describes all possible outcomes in our setting.
 - ω : Generic notation for the realized outcomes in the sample space.[1]
 - Here, $\Omega = \{HH, HT, TH, TT\}$.
- Event: a subset of Ω .
 - We will often use terms like A or B to define events.
 - In our example, the event that we get a head on the first flip is $A = \{HT, HH\}$.
- \mathcal{S} : Event space. Describes all subsets of events, including the null set. [\[Full event space\]](#)
 - We use this in addition to the sample space, so we can describe all types of events that we can define the probability for.
- P : Probability measure. A function that assigns probability to all of the events in the event space.
 - Here, since our coin is fair, for the event that we get a head on the first flip, $P(A) = 1/2$.

[1]In probability, we often use uppercase terms to denote the random process, and lowercase terms to denote specific outcomes of that random process.

An aside on mathematical notation

Mathematical notation is a tool that gives us a common language to express concepts with precision. There are different conventions in different communities, and no approach is "right" or "wrong"--it's just a question of whether your notation is appropriately communicating to your audience what you want it to.

To use mathematical notation in R Markdown, write LaTeX typesetting commands inside of dollar signs.

For example,

```
$Y = \beta_0 + \beta_1 X$
```

is rendered as

$$Y = \beta_0 + \beta_1 X.$$

We have just used probability notation as a way to fully describe any random generative process.

We can simulate our double coin flip process in R, using the `sample()` function. There are four possible outcomes, and all are equally likely.

```
Omega <- c('HH', 'HT', 'TH', 'TT')
probs <- c(0.25, 0.25, 0.25, 0.25)

sample(x = Omega,
       size = 1,
       prob = probs)
```

```
## [1] "HH"
```

We can run this simulation many times, and our results should *approximately* follow the probabilities we assigned.

```
n <- 1000
result_n <- sample(x = Omega,
                  size = n,
                  prob = probs,
                  replace = TRUE)

table(result_n)
```

```
## result_n
##  HH  HT  TH  TT
## 247 241 258 254
```


Independent events

- Two events are *independent* if

$$P(AB) = P(A)P(B)$$

Notational aside: The event AB is that both A and B happen. There are other ways to write this, including $A \cap B$.

Returning to our example of flipping two fair coins. Let's say:

- Event A : we get a head on the first coin flip; $A = \{HT, HH\}$.
- Event B : we get a head on the second coin flip; $B = \{TH, HH\}$.
- We can see the event AB as the overlap in their respective sets,
 $AB = \{HH\}$

The coin flips are unrelated, so the events should be independent. We can check this mathematically.

- First, we know that all of the outcomes $\Omega = \{HH, HT, TH, TT\}$ are equally likely.

$$P(A) = P(\{HT\}) + P(\{HH\}) = 0.25 + 0.25 = 0.5$$

$$P(B) = P(\{TH\}) + P(\{HH\}) = 0.25 + 0.25 = 0.5$$

Then, we can calculate the product of the probabilities, and the probability of the joint event.

$$P(A)P(B) = 0.5 \times 0.5 = 0.25$$

$$P(AB) = P(\{HH\}) = 0.25$$

We see that they are the same, so we have independence.

$$P(A)P(B) = P(AB)$$

We can also check if observed proportions in our simulations show the same thing.

```
Omega <- c('HH', 'HT', 'TH', 'TT'); probs <- c(0.25, 0.25, 0.25, 0.25)
result_n <- sample(x = Omega,
                   size = n,
                   prob = probs,
                   replace = TRUE)
```

```
(observed_props <- prop.table(table(result_n)))
```

```
## result_n
##      HH      HT      TH      TT
## 0.273 0.264 0.243 0.220
```

```
(PA <- mean(result_n == 'HT' | result_n == 'HH'))
```

```
## [1] 0.537
```

```
(PB <- mean(result_n == 'TH' | result_n == 'HH'))
```

```
## [1] 0.516
```

```
(PAB <- mean(result_n == 'HH'))
```

```
## [1] 0.273
```

```
PA*PB
```

```
## [1] 0.277092
```

The proportions look pretty close.

What about for a case where A and B are not independent?

- Let's say A is still the event that we get a head on the first coin flip; $A = \{HT, HH\}$.
- But event B is now the event that we get a head on both coin flips; $B = \{HH\}$.
- AB is the intersection of these two sets, which is just $AB = \{HH\}$.

```
(observed_props <- prop.table(table(result_n)))
```

```
## result_n  
##      HH      HT      TH      TT  
## 0.273 0.264 0.243 0.220
```

```
(PA <- mean(result_n == 'HT' | result_n == 'HH'))
```

```
## [1] 0.537
```

```
(PB <- mean(result_n == 'HH'))
```

```
## [1] 0.273
```

```
(PAB <- mean(result_n == 'HH'))
```

```
## [1] 0.273
```

```
PA*PB
```

```
## [1] 0.146601
```

The proportions are then quite different.

Conditional probability

If $P(B) > 0$, the *conditional probability* of an event A occurring, given event B has occurred is:

$$P(A|B) = \frac{P(AB)}{P(B)}$$

This can also be read as, out of all of the times event B occurs, how many times does event A also occur?

For our coin flip example, we'll stick with:

- A is the event that we get a head on the first coin flip; $A = \{HT, HH\}$.
- B is the event that we get a head on the both coin flips; $B = \{HH\}$.
- AB is $\{HH\}$

$$\begin{aligned} P(A|B) &= \frac{P(AB)}{P(B)} \\ &= \frac{P(\{HH\})}{P(\{HH\})} \\ &= 1 \end{aligned}$$

What is $P(B|A)$?

Bayes Rule

A useful theorem to return to is *Bayes Rule*

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Why is Bayes rule so useful? It basically tells us how we update probability based on observed data.

Bayes Rule

Example: Suppose everyone in the University of Chicago community is given a new rapid test for COVID.

- We are concerned with *false negatives*, when we get a negative test for a person who actually *is* infected.
- And *false positives*, when we get a positive test for a person who *is not* infected.

Bayes Rule

Example: Suppose everyone in the University of Chicago community is given a new rapid test for COVID.

A student gets a positive test back. What is the probability that they have COVID?

- Doctors know that the probability of having COVID is 3% in this population : $P(A)$

```
PA <- 0.03
```

- The overall rate of positive tests is 5% : $P(B)$

```
PB <- 0.05
```

- If you have COVID, your test will turn up positive 95% of the time. : $P(B|A)$

```
PB_if_A <- 0.95
```

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

```
( PB_if_A * PA )/PB
```

```
## [1] 0.57
```

Prosecutor's fallacy

$$P(A|B) \stackrel{?}{=} P(B|A)$$

As a general rule, we cannot assume the conditional probability of A given B is the same as the probability of B given A

- In a court case, the probability that a person is guilty given that we see a DNA match is NOT the same as the probability of a DNA match given that they are guilty.
- This fallacy often occurs when we observe evidence that we are very unlikely to see among innocent people, and very likely to observe among guilty people
- But the overall number of innocent people in the population is very large, so the number of false positives is much higher than the number of true positives

Why thinking about conditional probability is so important in social science research

"20% of people hospitalized with COVID-19 are vaccinated"

Does this tell us that vaccines aren't very effective?

Some things we should think about:

- What percentage of the population in areas served by hospitals are vaccinated?
- Are people who are vaccinated at higher risk for breakthrough cases? Or more likely to be hospitalized? Are they older, or have pre-existing conditions?

Why thinking about conditional probability is so important in social science research

What data do you need to answer the question: "Do white officers shoot minority citizens at a higher rate than non-white officers?"

Why thinking about conditional probability is so important in social science research

Officer characteristics and racial disparities in fatal officer-involved shootings

David J. Johnson^{a,b,1}, Trevor Tress^b, Nicole Burkel^b, Carley Taylor^b, and Joseph Cesario^b

^aDepartment of Psychology, University of Maryland at College Park, College Park, MD 20742; and ^bDepartment of Psychology, Michigan State University, East Lansing, MI 48824

- In a 2020 paper, Johnson and co-authors evaluated the relationship between race of victims shot by police, and the characteristics of the police shooters
- The authors claim, "White officers are not more likely to shoot minority civilians than non-White officers"--and the paper has been used in congressional testimony to support the claim that diversity in police forces would not be beneficial in reducing bias in officer-involved shootings
- Their database only has data on fatal shootings--not on cases where people were NOT shot.
- Their results report "whether a person fatally shot was more likely to be Black (or Hispanic) than White" conditional on the race of the officer involved

"White officers are not more likely to shoot minority civilians than non-White officers"

$$\frac{P(\text{shot} | \text{White officer, minority civilian})}{P(\text{shot} | \text{Minority officer, minority civilian})}$$

What data are we missing to address this question?

- Dean Knox & Jonathan Mummolo wrote a letter critiquing the original article, based on the authors' failure to appropriately apply Bayes Rule;
- The article was eventually retracted
- (It's a bit more complicated than that, but know your conditional probability when thinking about difficult subjects!)

Some reference texts

1. Aronow, P., & Miller, B. (2019). Foundations of Agnostic Statistics. Cambridge: Cambridge University Press. Chapter 1. Probability Theory.
2. Wasserman, L. (2004). All of Statistics: A Concise Course in Statistical Inference. New York: Springer. Chapter 1. Probability Theory and Chapter 2. Random Variables.
3. Hernán MA, Robins JM (2020). Causal Inference: What If. Boca Raton: Chapman & Hall/CRC.

Flipping two coins event space:

$$\begin{aligned} S = \{ & \emptyset, \\ & \{HH\}, \{HT\}, \{TH\}, \{TT\}, \\ & \{HH, HT\}, \{HH, TH\}, \{HH, TT\}, \{HT, TH\}, \{HT, TT\}, \{TH, TT\} \\ & \{HH, HT, TH\}, \{HH, HT, TT\}, \{HH, TH, TT\}, \{HT, TH, TT\}, \\ & \{HH, HT, TH, TT\} \} \end{aligned}$$

[\[Back to terms\]](#)