Gov 50: 15. Probability

Matthew Blackwell

Harvard University

Fall 2018

1/ Today's agenda

Logistics

- DataCamp Assignment 5: due Thursday.
- HW 4: out today, due next Thursday.
- Final project:
  - Survey about groups due Thursday, 11/1.
  - Paragraph describing data, proposed analyses due 11/21.
  - See “Final Project” link on Canvas for pointers on data sets.
  - Feel free to email us or come to office hours to talk about ideas.

Where are we? Where are going?

- Up to now: how to estimate things.
  - Causal effects.
  - Measurements of concepts.
  - Predictions about unknown quantities.
- Problem: how do we know our estimates are “real” or just due to random chance?
  - Could have randomly selected a different treatment/control group.
  - Could have randomly selected a different sample.
- We need a way to talk about random variability/chance: **probability**.

2/ Probability

Sample spaces & events

- Probability formalizes chance variation or uncertainty in outcomes.
  - It might rain or be sunny today, we don't know which.
  - To formalize, we need to define the set of possible outcomes.
- **Sample space**: Ω the set of possible outcomes.
- **Event**: any subset of outcomes in the sample space

Example: gambling

- A standard deck of playing cards has 52 cards:
  - 13 rank cards: (2,3,4,5,6,7,8,9,10,J,Q,K,A)
  - in each of 4 suits: ($\clubsuit$, $\spadesuit$, $\heartsuit$, $\diamondsuit$)
- Hypothetical trial: pick a card, any card.
  - Uncertainty: we don't know which card we're going to get.
- One possible outcome: picking a $4\clubsuit$
- Sample space:

$$
\begin{array}{ccccccccccccc}
2\clubsuit & 3\clubsuit & 4\clubsuit & 5\clubsuit & 6\clubsuit & 7\clubsuit & 8\clubsuit & 9\clubsuit & 10\clubsuit & J\clubsuit & Q\clubsuit & K\clubsuit & A\clubsuit \\
2\spadesuit & 3\spadesuit & 4\spadesuit & 5\spadesuit & 6\spadesuit & 7\spadesuit & 8\spadesuit & 9\spadesuit & 10\spadesuit & J\spadesuit & Q\spadesuit & K\spadesuit & A\spadesuit \\
2\heartsuit & 3\heartsuit & 4\heartsuit & 5\heartsuit & 6\heartsuit & 7\heartsuit & 8\heartsuit & 9\heartsuit & 10\heartsuit & J\heartsuit & Q\heartsuit & K\heartsuit & A\heartsuit \\
2\diamondsuit & 3\diamondsuit & 4\diamondsuit & 5\diamondsuit & 6\diamondsuit & 7\diamondsuit & 8\diamondsuit & 9\diamondsuit & 10\diamondsuit & J\diamondsuit & Q\diamondsuit & K\diamondsuit & A\diamondsuit
\end{array}
$$

- An event: picking a Queen, $\{Q\clubsuit, Q\spadesuit, Q\heartsuit, Q\diamondsuit\}$

What is probability?

- The probability $\mathbb{P}(A)$ represents how likely event A occurs.
- If all outcomes equally likely, then:

$$\mathbb{P}(A) = \frac{\text{number of elements in } A}{\text{number of elements in } \Omega}$$

- Example: randomly draw 1 card:
  - probability of drawing $4\clubsuit$: $\frac{1}{52}$
  - probability of drawing any $\clubsuit$: $\frac{13}{52}$
- Same math, but different interpretations:
  - **Frequentist**: probabilities reflect relative frequency in a large number of trials.
  - **Bayesian**: probabilities are subjective beliefs about outcomes.
- Not our fight $\rightsquigarrow$ stick to frequentism in this class.

Probability axioms

- Probability quantifies how likely or unlikely events are.
- We'll define the probability $\mathbb{P}(A)$ with three axioms:

1. (Nonnegativity) $\mathbb{P}(A) \geq 0$ for every event A
2. (Normalization) $\mathbb{P}(\Omega) = 1$
3. (Addition Rule) If two events A and B are mutually exclusive

$$\mathbb{P}(A \text{ or } B) = \mathbb{P}(A) + \mathbb{P}(B).$$

Gambling 102

- What's the $\mathbb{P}(\text{Q card})$ if a single card is randomly selected from a standard deck?
  - "randomly selected" $\rightsquigarrow$ all cards have prob. $1/52$
- "4 card" event = $\{Q\clubsuit \text{ or } Q\spadesuit \text{ or } Q\heartsuit \text{ or } Q\diamondsuit\}$
- Union of mutually exclusive events $\rightsquigarrow$ use addition rule
  - $\rightsquigarrow \mathbb{P}(\text{Q card}) = \mathbb{P}(Q\clubsuit) + \mathbb{P}(Q\spadesuit) + \mathbb{P}(Q\heartsuit) + \mathbb{P}(Q\diamondsuit) = \frac{4}{52}$

Useful probability facts

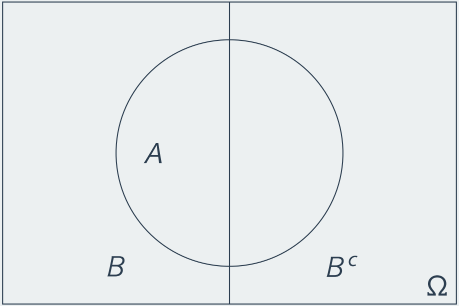
- Probability of the complement: $\mathbb{P}(A^c) = \mathbb{P}(\text{not } A) = 1 - \mathbb{P}(A)$
  - "The probability of something equals 1 minus the chance of the opposite happening."
  - Probability of **not** drawing a Queen is $1 - \frac{4}{52} = \frac{48}{52}$
- **General addition rule** for any events, A and B:

$$\mathbb{P}(A \text{ or } B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \text{ and } B)$$

  - Probability of drawing Queen or ♣?
  - $\frac{4}{52} + \frac{13}{52} - \frac{1}{52} = \frac{16}{52}$
- **Law of total probability** for any events A and B:

$$\mathbb{P}(A) = \mathbb{P}(A \text{ and } B) + \mathbb{P}(A \text{ and } B^c)$$

Law of Total Probability



- How to calculate probability of A?
- **Law of total probability**:

$$\mathbb{P}(A) = \mathbb{P}(A \text{ and } B) + \mathbb{P}(A \text{ and } B^c)$$

3/ Conditional Probability

Conditional probability

- **Conditional probability**: if we know that B has occurred, what is the probability of A?
    - Conditioning our analysis on B having occurred.
- Examples:
    - What is probability of two states going to war *if* they are both democracies?
    - What is the probability of a judge ruling in a pro-choice direction conditional on having daughters?
    - What is the probability that there will be a coup in a country conditional on having a presidential system?
- Conditional probability is a huge part of what we do in the empirical social sciences.

Conditional Probability definition

- Definition: If $\mathbb{P}(B) > 0$ then the **conditional probability** of A given B is

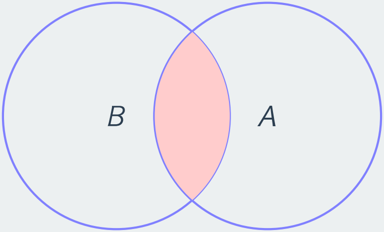
$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \text{ and } B)}{\mathbb{P}(B)}.$$

- How often A and B occur divided by how often B occurs.
- **WARNING!** $\mathbb{P}(A \mid B)$ does **not**, in general, equal $\mathbb{P}(B \mid A)$.
    - $\mathbb{P}(\text{smart} \mid \text{in gov 50})$ is high
    - $\mathbb{P}(\text{in gov 50} \mid \text{smart})$ is low.
    - There are many many smart people who are not in this class!
- If all outcomes equally likely:

$$\mathbb{P}(A \mid B) = \frac{\text{number of outcomes in both } A \text{ and } B}{\text{number of outcomes in just } B}$$

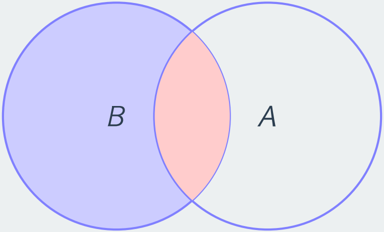
Conditional probability

$P(A \text{ and } B)$



Conditional probability

$$P(A \mid B)$$



US Senate example

	Democrats	Republicans	Independents	Total
Men	39	42	2	83
Women	12	5	0	17
Total	51	47	2	100

- Choose one senator at random from this population
- What is the probability of choosing a female?
  - $\mathbb{P}(\text{Female}) = \frac{17}{100} = 0.17$
- What is the probability of choosing a female Republican?
  - $\mathbb{P}(\text{Female and Republican}) = \frac{5}{100} = 0.05$
- What is the probability that a randomly selected Republican is Female:
  - $\mathbb{P}(\text{Female} \mid \text{Republican}) = \frac{\mathbb{P}(\text{Female and Republican})}{\mathbb{P}(\text{ Republican})} = \frac{5/100}{47/100} = \frac{5}{47} = 0.106$

Conditional probability rules

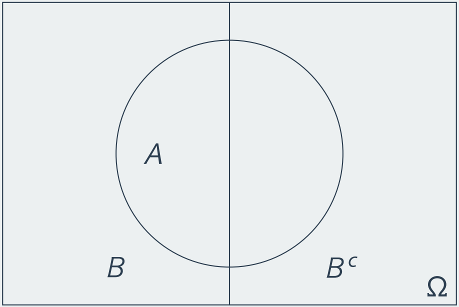
- Multiplication rule:

$$\mathbb{P}(A \text{ and } B) = \mathbb{P}(A \mid B)\mathbb{P}(B) = \mathbb{P}(B \mid A)\mathbb{P}(A)$$

- Law of total probability:

$$\mathbb{P}(A) = \mathbb{P}(A \mid B)\mathbb{P}(B) + \mathbb{P}(A \mid \text{not } B)\mathbb{P}(\text{not } B)$$

Law of Total Probability



- How to calculate probability of A?
- **Law of total probability**:

$$\mathbb{P}(A) = \mathbb{P}(A \text{ and } B) + \mathbb{P}(A \text{ and } B^c)$$

- **Law of total probability** (with the multiplication rule):

$$\mathbb{P}(A) = \mathbb{P}(A \mid B)\mathbb{P}(B) + \mathbb{P}(A \mid B^c)\mathbb{P}(B^c)$$

Multiplication rule, example

- Let's say we draw two cards at random from a deck and don't put them back.
- What's the probability that we draw two Aces?

$$\mathbb{P}(\text{Ace}_1 \text{ and } \text{Ace}_2) = \mathbb{P}(\text{Ace}_1)\mathbb{P}(\text{Ace}_2 \mid \text{Ace}_1)$$

- What are these probabilities?
  - $\mathbb{P}(\text{Ace}_1) = \frac{4}{52}$
  - 4 Aces to pick out of 52 cards
  - $\mathbb{P}(\text{Ace}_2 \mid \text{Ace}_1) = \frac{3}{51}$
  - 3 Aces left in the 51 remaining cards
- Thus,

$$\mathbb{P}(\text{Ace}_1 \text{ and } \text{Ace}_2) = \frac{4}{52} \times \frac{3}{51} = 0.0045$$

4/ Independence

Independence

- Two events are **independent** if one occurring has no bearing on the probability of the other occurring.
  - Formally, $\mathbb{P}(A \text{ and } B) = \mathbb{P}(A)\mathbb{P}(B)$.
- Intuitively, A and B are independent if knowing that B occurred has no impact on the probability of A occurring:

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \text{ and } B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(A)\mathbb{P}(B)}{\mathbb{P}(B)} = \mathbb{P}(A)$$

Sampling and independence

- Sampling > 1 with replacement: **independent draws**
  - Randomly draw 1 card, note the card, then put it back in deck.
  - Shuffle, randomly draw 2nd card, note the card.
  - First draw doesn't affect second $\rightsquigarrow$ independence
- Sampling > 1 without replacement: **dependent draws**
  - Randomly pick 1st card, note it, leave it out.
  - Randomly pick 2nd card from remaining 51 cards.
  - Getting an Ace in first card changes the probability of drawing an Ace for the second card.

Independence and the lottery

Every week you buy a ticket in a lottery that offers one chance in a million of winning. What is the chance that you never win, even if you keep this up for 10 years?

- Each week lottery results are independent.
- Probability of not winning in each week: $\frac{999{,}999}{1{,}000{,}000}$
- Probability of not winning in week 1 and in week 2:

$$\frac{999,999}{1,000,000} \times \frac{999,999}{1,000,000}$$

- Probability of not winning for 520 weeks (10 years)?

$$\left(\frac{999,999}{1,000,000}\right)^{520}$$

```
(999999/1000000)^(520)
```

```
## [1] 0.999
```

- Still very small.

Wrap up

- Starting to think about how to quantitatively summarize chance variation.
- Next time: random variables.