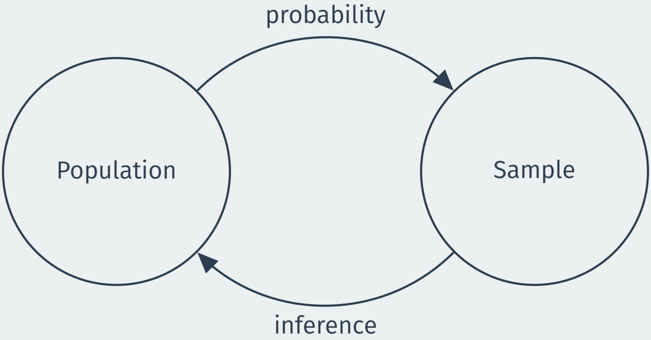
Gov 51: Probability

Matthew Blackwell

Harvard University

Learning about populations



- **Probability**: formalize the uncertainty about how our data came to be.
- **Inference**: learning about the population from a sample of data.

Why probability?

- Probability quantifies chance variation or uncertainty in outcomes.
    - It might rain or be sunny today, we don't know which.
- We estimated a treatment effect of 7.2, but what if we reran history?
    - Weather changes $\rightsquigarrow$ slightly different estimated effect.
- Statistical inference is a **thought experiments** about uncertainty.
    - Imagine a world where the treatment effect were 0 in the population.
    - What types of estimated effects would we see in this world by chance?
- Probability to the rescue!

Sample spaces & events

- To formalize chance, 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♣: $\frac{1}{52}$
  - probability of drawing any ♣: $\frac{13}{52}$
- Same math, but different interpretations:
  - **Frequentist**: $\mathbb{P}()$ reflects relative frequency in a large number of trials.
  - **Bayesian**: $\mathbb{P}()$ 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 is $\mathbb{P}(\text{Q card})$ if a single card is randomly selected from a 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: $\mathbb{P}(\text{not } A) = 1 - \mathbb{P}(A)$
  - 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 ♣?
  - $\frac{4}{52} + \frac{13}{52} - \frac{1}{52} = \frac{16}{52}$

Conjunction fallacy

Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations.

- What is more probable?
    1. Linda is a bank teller?
    2. Linda is a bank teller and is active in the feminist movement?
- Famous example of the **conjuction fallacy** called the Linda problem.
    - Majority of respondents chose 2, but this is impossible!
- **Law of total probability** for any events A and B:

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

    - $\mathbb{P}(\text{teller and feminist}) = \mathbb{P}(\text{teller}) - \mathbb{P}(\text{teller and not feminist})$

Law of Total Probability

	Democrats	Republicans	Independents	Total
Men	28	44	2	74
Women	17	9	0	26
Total	45	53	2	100

- What's the probability of randomly selecting a woman senator?

$$\begin{aligned}\mathbb{P}(\text{woman}) &= \mathbb{P}(\text{woman \& a Democrat}) + \mathbb{P}(\text{woman \& not a Democrat}) \\ &= \frac{17}{100} + \frac{9}{100} = \frac{26}{100}\end{aligned}$$