Lecture 7

Quantitative Political Science

Prof. Bisbee

Vanderbilt University

Lecture Date: 2023/09/21

Slides Updated: 2023-09-26

Agenda

1. Sampling
2. The Central Limit Theorem
3. Estimation
4. An example

Our journey thus far

- Overarching goal: find the **best possible way** to make inferences about populations from samples
- Defined social phenomena as **experiments** that yield observations
    - Outcomes defined as **events**
    - Defined **sample space** as the set of all possible events
    - Used **set theory** to guide how we assign probabilities to events
- Then defined a **random variable** as a function mapping sample space to real numbers
    - Built on set theory to describe the **probability distribution** of an RV
    - And the probability distribution of a **function** of RVs

"Random" variables

- All observed social phenomena are **realizations of random variables**
- Put differently, political scientists study **random events**
- We mean "random" differently from the layperson
  - Layperson: "Random" $\Rightarrow$ something that cannot be anticipated
  - Us: an event that is **probabilistic** instead of deterministic
- In sum, **random variables** means we expect that the values we observe to be draws from an associated probability distribution

Expectations

- Powerful result: if the probability distribution is an accurate representation of the population frequency distribution, then the expected value of an RV is the population mean μ

Discrete case:

$$E(Y) \equiv \sum_{y} y p(y)$$

Continuous case:

$$E(Y) \equiv \int_{y} y f(y) dy$$

Putting this to work

- Fundamental challenge: **inference**
- What can we say about the **data we don't have**?
- Typically want to say something that **summarizes** the population
  - Central tendency is a very common target!
- These "things" we want to say are **parameters**
  - And we "estimate" them with **estimators**
- Definition: **Estimator**
  - A *rule* (often a formula) that tells us how to calculate an **estimate** from a sample
- We can come up with many of these, but how do we know if they're any good?

Estimates and Estimators

- What do we mean by "good"?
- Example: the observed sample mean
  - Draw random sample of n observations from a random variable Y

$$\bar{Y} \equiv \frac{y_1 + y_2 + \cdots + y_n}{n} = \frac{1}{n} \sum_i y_i$$

- Is this "good"?
  - Be precise! Is this a "good" estimate of the population parameter μ?
- It feels good...but why?

Simple Example

- Want to know the mean income of the American population
  - μ: Average income of **all** Americans
- But we can't ask everybody (takes too long, too expensive), so we run an **experiment** where we sample Americans **at random** and ask their income
- The response for the first person I ask is y_1
  - **NOTE:** y_1 is one of literally millions of responses I might have recorded
  - Thus it is **probabilistic**
  - Thus we can think of it as a realization of the random variable Y_1
- Denote response of second person I ask as y_2
  - AGAIN... y_2 is probabilistic and can be thought of as the realization of the random variable Y_2

Simple Example Cont'd

- Thus let observed sample of n observations be realizations of n random variables: $Y_1, Y_2, \ldots, Y_n$
- So what is $\bar{Y} = \frac{1}{n} \sum_i Y_i$?
  - One realization of many possible $\bar{Y}$ values
  - $\bar{Y}$ is a **function** of random variables, and therefore **itself** a random variable (recall definition of RV)
  - In other words, our observed sample produces $\bar{y}$, a realization of the random variable $\bar{Y}$
- Since $\bar{Y}$ is a random variable, it has a theoretical probability distribution
  - What does the probability distribution look like?

Magic time

- Thus far, we have being pretty **sketchy** about our distributions
  - Talked about *some* in concrete terms (Bernoulli, Binomial, Poisson, Uniform, Normal)
  - But mostly been **very** agnostic with our notation: $f(y)$; $F(y)$ could be (almost) anything
- But, thanks to a carefully defined experiment, we be **concrete** about the probability distribution of $\bar{Y}$
- Specifically, we are working with a **random sample**
  - Choose a set of n observations from a population of size N, producing $\binom{N}{n}$ possible samples
  - And each of these samples is **equiprobable**
- A random sample allows us to assume that the random variables $Y_1, Y_2, \ldots, Y_n$ are "i.i.d."
  - **I**ndependent and **I**dentically **D**istributed

IID

- **Independent**: $F(y_1, y_2, \ldots, y_n) = F_1(y_1) * F_2(y_2) * \cdots * F_n(y_n)$
- **Identically Distributed**: $F_1(y_1) = F_2(y_2) = \cdots = F_n(y_n) = F(y)$

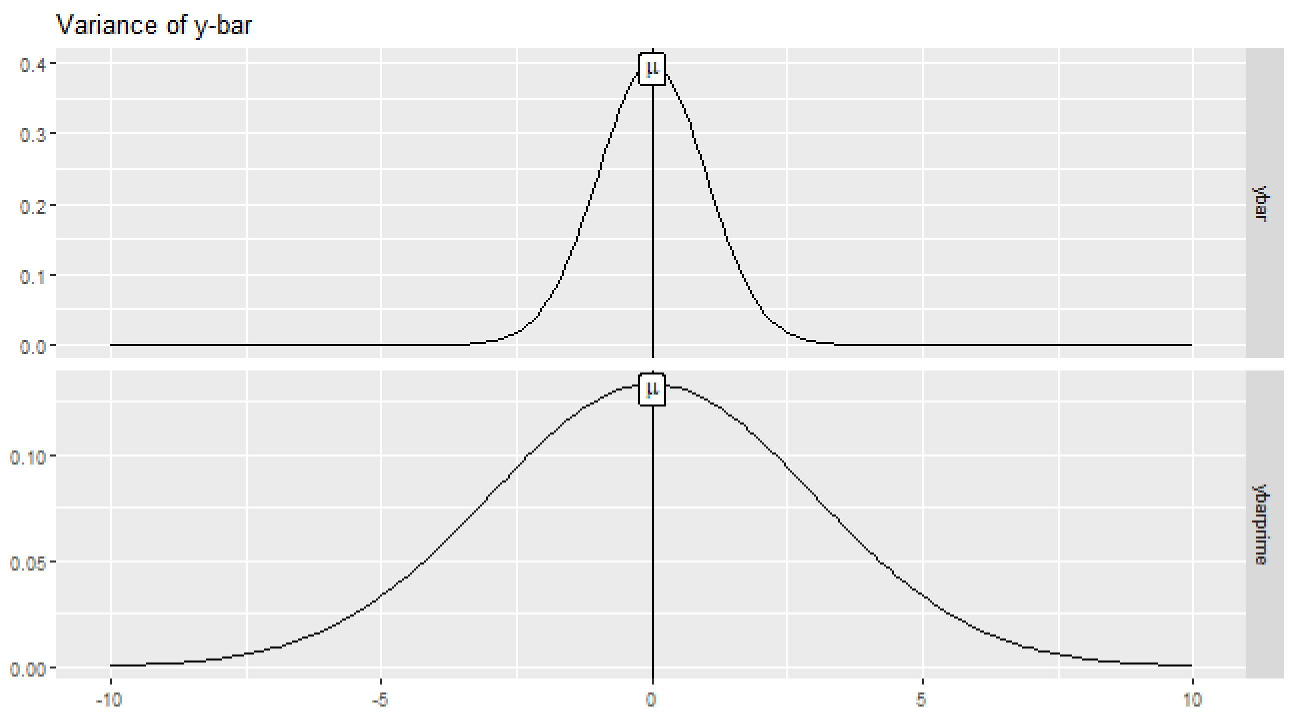
How good is $\bar{Y}$?

- We will refer to $\bar{Y} = \frac{1}{n}\sum_i Y_i$ as a **sample statistic**
  - A function of the random variables $Y_1, Y_2, \ldots, Y_n$ and "known constants" (in this case, $\frac{1}{n}$)
- We can prove that $E(\bar{Y}) = \mu$ thanks to the **identicality assumption**
  - So we can say $\bar{Y}$ is "good" on average because it will be μ on average
- But how far off might a given sample's $\bar{Y}$ be?
  - This is just the standard deviation of $\bar{Y}$, or $\sigma_{\bar{Y}}$
  - $\sigma_{\bar{Y}} = \sqrt{VAR(\bar{Y})} = \frac{\sigma}{\sqrt{n}}$ (we can prove with **identicality and independence**)

How good is $\bar{Y}$?

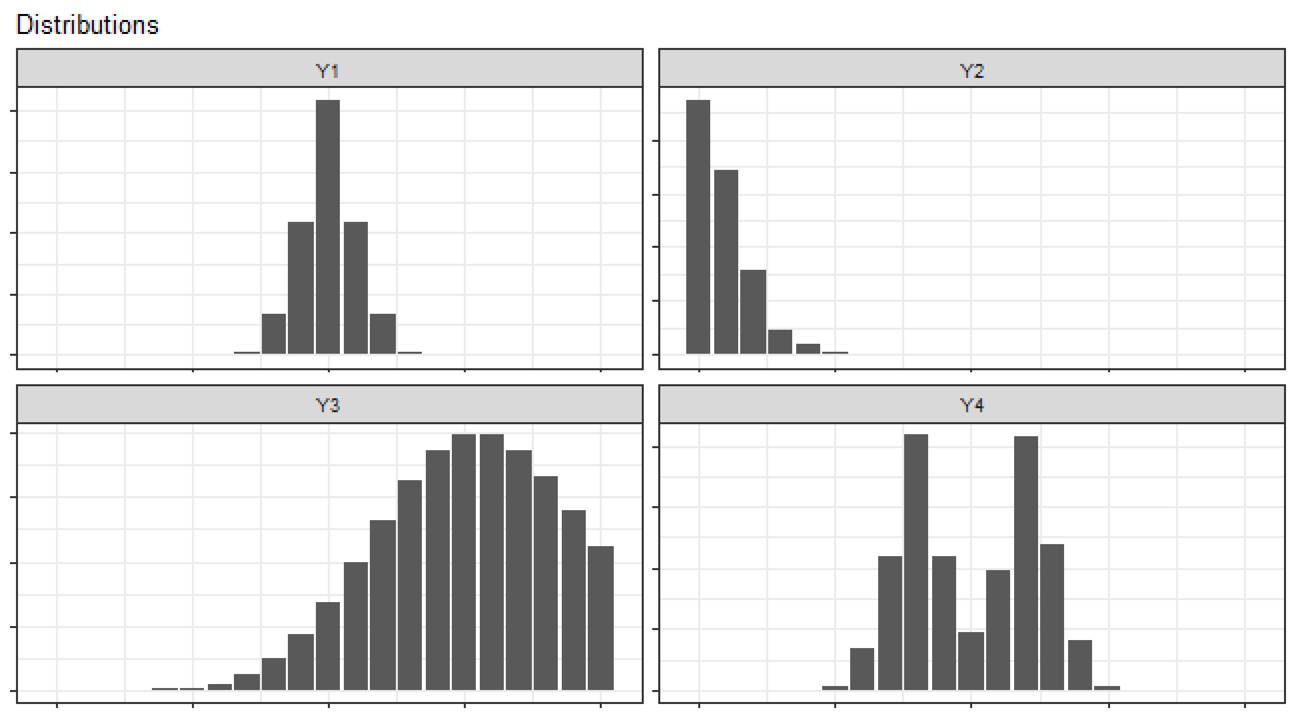
- Remember that $\bar{Y}$ is a random variable and thus it has a probability distribution?
- In general, any **sample statistic** is a random variable and has a probability distribution
- We call these probability distributions **sampling distributions**
    - Literally just the probability distribution for a sample statistic
    - They are **theoretical models** for the possible values of the sample statistic we would expect to see through repeated random sampling
- We might have one sampling distribution for $\bar{Y}$ and another for $\bar{Y}'$
    - The "better" sampling distribution is the one whose estimates are closer to the true parameter of interest μ
    - I.e., the one whose variance is **smaller**

How good is $\bar{Y}$?



Shapes

- But the previous was just an example, wasn't it? There's no way the **sampling distribution** would look like that
    - After all, other random variables can have *any* shape



Central Limit Theorem

- We can rely on a **powerful** result of the math thus far:
- If $Y_1, Y_2, \ldots, Y_n$ are i.i.d., then $\bar{Y}$'s sampling distribution is approximately normal
- See some examples in the handout!
- Formally, let $Y_1, Y_2, \ldots, Y_n$ be i.i.d. random variables with:
  - $E(Y_i) = \mu$
  - $VAR(Y_i) = \sigma^2$
- Define $U_n \equiv \frac{\bar{Y} - \mu}{\sigma / \sqrt{n}}$ as the standardized $\bar{Y}$, and denote $F_{U_n}(u)$ as the CDF of this standardized random variable
- We know that $\lim_{n \to \infty} F_{U_n}(u) \int_{-\infty}^{u} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt \;\; \forall \; u$
  - You don't need to know this proof, but see Section 7.4 in WMS if interested

Central Limit Theorem

- In plain language: We know the (asymptotic) sampling distribution of $\bar{Y}$ **without requiring any assumptions about the probability distribution of** Y
- We can now actually **calculate** probabilities!
- This is inference!
- This is magic!
- THIS IS SPARTA!

Estimation

- With the CLT in hand, let's return to **estimators**
  - Remember the definition?
  - A rule (often a formula) that tells us how to calculate an estimate of a population parameter
- Two types:
  1. **Point Estimates**: a single value (i.e., a "point") is given as the estimate of the parameter of interest
  2. **Interval Estimates**: two values are used to construct an range (i.e., an "interval") in which the parameter of interest exists

Estimation

- What is $\bar{Y} = \frac{1}{n} \sum_i Y_i$?
  - A point estimate
- But there are many other candidates
- Consider $\bar{Y}_B = \frac{1}{n} \sum_i (Y_i + 1)$
- Is this "good"? No...why not?
- It is "biased"

Aside on notation

- We have been interested in μ which is a population parameter
- But there are other quantities of the population that we might be interested in
  - Central tendency parameters: median, mode
  - Dispersion parameters: range, variance
  - (Preview) Relationship parameters: coefficients
- Generically, denote a population parameter with θ and the proposed estimator for this parameter as $\hat{\theta}$

Bias

- Math of expectations can help us formalize bias
- Define "bias" as an estimator that is equal to the parameter it claims to estimate **in expectation**
  - $\hat{\theta}$ is unbiased for θ if $E(\hat{\theta}) = \theta$
  - (i.e., $E(\bar{Y}) = \mu$)
  - If $E(\hat{\theta}) \neq \theta$, then $\hat{\theta}$ is **biased**
  - Denote bias as $B(\hat{\theta}) = E(\hat{\theta}) - \theta$
- We can prove an estimator is unbiased using expectations!

Bias

- Does $E(\hat{Y}_B) = \mu$?

$$
\begin{aligned}
E(\hat{Y}_B) &= E\left[\frac{1}{n}\sum_{i}^{n}(Y_i + 1)\right] \\
&= \frac{1}{n}\left[\sum_{i}^{n} E(Y_i + 1)\right] \\
&= \frac{1}{n}\left[\left(\sum_{i}^{n} E(Y_i)\right) + \left(\sum_{i}^{n} E(1)\right)\right] \\
&= \frac{1}{n}\left[\sum_{i}^{n} \mu + \sum_{i}^{n} 1\right] \\
&= \frac{1}{n}(n\mu + n) \\
&= \mu + 1 \\
&\neq \mu
\end{aligned}
$$

Variance

- Recall from earlier when we talked about how close a random sample's $\bar{Y}$ would be to μ
- This is equivalent to saying we want to minimize the variance of the sampling distribution
- Recall the definition of variance of a random variable

$$VAR(\hat{\theta}) = E[(\hat{\theta} - E(\hat{\theta}))^2]$$

- We want to make this as small as possible

"Good" is bias and variance

- **Bias-variance tradeoff**: we often will find ourselves caught between wanting to reduce the bias of an estimator and reducing its variance
- Evaluate this tradeoff using the **mean squared error (MSE)**

$$MSE(\hat{\theta}) = E[(\hat{\theta} - \theta)^2]$$

- MSE can be rewritten in terms of both bias and variance

$$MSE(\hat{\theta}) = VAR(\hat{\theta}) + B(\hat{\theta})^2$$

- (Ideas on why this includes the square of the estimator's bias?)

Example

- Consider a new parameter of interest, defined as $\theta = \mu_1 - \mu_2$
    - We are interested in **the difference in means of two different populations**
    - Note that we are using Y_1 and Y_2 to represent random variables from **different populations** here (don't get confused from the notation above)
- Is $\hat{\theta} = \bar{Y}_1 - \bar{Y}_2$ unbiased?
    - Easy proof!

Example

- What is $\hat{\theta}$'s variance?

$$\begin{aligned} VAR(\bar{Y}_1 - \bar{Y}_2) &= VAR(\bar{Y}_1) + VAR(\bar{Y}_2) + 2COV(\bar{Y}_1, \bar{Y}_2) \\ &= VAR(\bar{Y}_1) + VAR(\bar{Y}_2) + 2 * 0 \quad \text{(bc } \bar{Y}_1, \bar{Y}_2 \text{ are indep)} \\ &= \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} \end{aligned}$$

- We denote the standard error of the estimator $\hat{\theta}$ with $\sigma_{\hat{\theta}}$, which is just the square root of the variance
    - Thus $\sigma_{\bar{Y}_1 - \bar{Y}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

Final CLT Variant

- Turns out (beyond the scope of this class) that a variant of the CLT tells us

$$\bar{Y}_1 - \bar{Y}_2 \sim \mathcal{N}\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right) \text{ as } n_1, n_2 \to \infty$$

- (Notation reminder!
    - $Y \sim \mathcal{N}(\mu, \sigma^2)$ is how we write "is distributed normal with mean μ and standard deviation σ^2)