Lecture 9

Quantitative Political Science

Prof. Bisbee

Vanderbilt University

Lecture Date: 2023/09/28

Slides Updated: 2023-09-27

Agenda

1. Recap of where we are
2. σ^2 and consistency
3. Slutzky's Theorem
4. Fun example time!

Where we started

- Wanted to identify a "good" estimator for the population mean μ...
- ...based on a **random sample** of data from that population
- We proposed $\bar{Y} \equiv \frac{1}{n} \sum_i Y_i$ which seemed intuitive
- We redefined the sampling process for a size n as a series of random variables $Y_1, Y_2, \ldots, Y_n$
- Crucially, these are IID, meaning that they all have the same:
    1. CDF $F()$
    2. Mean μ
    3. Variance σ^2
- This allowed us to establish that $\bar{Y}$ is an **unbiased estimator** of μ: $E(\bar{Y}) = \mu$
- And that $VAR(\bar{Y}) = \frac{\sigma^2}{n}$, and thus $\sigma_{\bar{Y}} = \frac{\sigma}{\sqrt{n}}$

Where we went

- We wanted to know how close, on average, $\bar{Y}$ is to μ
- CLT tells us that the **sampling distribution** of $\bar{Y}$ is distributed Normal as $n \to \infty$
- We also know that the standardized version of $U_n \equiv Z \equiv \frac{\bar{Y}-\mu}{\sigma_{\bar{Y}}}$ converges to the *standard* Normal distribution
- This allowed us to find values of α and $z_{\alpha/2}$ which satisfy

$$ P(\bar{Y} - z_{\alpha / 2}\sigma_{\bar{Y}} \leq \mu \leq \bar{Y} + z_{\alpha/2}\sigma_{\bar{Y}}) = 1 - \alpha $$

- And since we know that $\sigma_{\bar{Y}} \equiv \frac{\sigma}{\sqrt{n}}$, we should be good to go! Right?

Where we are now

- Not quite! We need to confront the fact that we don't know σ!
- This is something of a Catch-22
  - We want to describe an interval estimate that contains the true population parameter μ
  - We have an estimator $\bar{Y}$, a standard normal distribution which gives us $z_{\alpha/2}$, and the sample size n
  - But we need $\frac{\sigma}{\sqrt{n}}$!
- We propose using $S_U^2 \equiv \frac{\sum_i (Y_i - \bar{Y})^2}{n-1}$, our **unbiased** estimator for σ^2

Consistency

- But wait! Before we can plug in S_U, we need to prove it is both unbiased and **consistent**
- We already know how to prove unbiasedness
- Consistency: as the sample size used to construct the estimator gets large, the probability of it being measured with error gets small
- Denote $\hat{\theta}_n$ as the estimate for a given sample size n
  - In the extreme: $\lim_{n \to \infty} P(|\hat{\theta} - \theta| > \epsilon) = 0$ where ϵ is any positive number
  - Can also express as " $\hat{\theta}_n$ converges in probability to θ ", or $\hat{\theta}_n \xrightarrow{p} \theta$
- In practice, we can evaluate this property by checking whether $VAR(\hat{\theta})$ approaches zero as n gets large (see pg. 450 for proof)
  - $\lim_{n \to \infty} VAR(\hat{\theta}) = 0$

Consistency

- Apply to $\bar{Y}$ for intuition

$$VAR(\bar{Y}) = \frac{\sigma^2}{n}$$

$$\lim_{n \to \infty} \frac{\sigma^2}{n} = 0$$

- Note that this **by itself** is insufficient to claim $\bar{Y} \xrightarrow{p} \mu$...we need to also prove unbiasedness (which we did last class)
- In other words, an estimator might be **consistent** but **biased**
- Or an estimator might be **unbiased** but not **consistent**
- Need to check both!

σ^2

- Remember what we're doing here!
    - We know that $U_n \equiv \frac{\bar{Y}-\mu}{\sqrt{\sigma^2/n}} \sim \mathcal{N}(\mu, \sigma^2)$
    - But can we be sure that $\hat{\theta} \equiv \frac{\bar{Y}-\mu}{\sqrt{S_U^2/n}} \sim \mathcal{N}(\mu, \sigma^2)$?
- Note that, in the original setting, σ^2 is a **parameter** whereas in our sample setting S_U^2 is a **random variable**

$$F\left(\frac{\bar{Y}-\mu}{S_U/\sqrt{n}}\right) \xrightarrow{p} \Phi$$

σ^2

- So let's examine whether S_U^2 is a **consistent** estimator for σ^2

$$
\begin{aligned}
S_U^2 &= \frac{\sum_i (Y_i - \bar{Y})^2}{n-1} \\
&= \frac{1}{n-1}\left(\sum_i Y_i^2 + \sum_i \bar{Y}^2 - \sum_i 2Y_i\bar{Y}\right) \\
&= \frac{1}{n-1}\left((\sum_i Y_i^2) + n\bar{Y}^2 - 2n\bar{Y}^2\right) \\
&= \frac{1}{n-1}\left((\sum_i Y_i^2) - n\bar{Y}^2\right) \\
&= \frac{n}{n-1}\left(\frac{1}{n}\sum_i Y_i^2 - \bar{Y}^2\right)
\end{aligned}
$$

σ^2

- So let's examine whether S_U^2 is a **consistent** estimator for σ^2

$$S_U^2 = \frac{n}{n-1}\left(\frac{1}{n}\sum_i Y_i^2 - \bar{Y}^2\right)$$

$$\begin{aligned}\lim_{n\to\infty}\frac{1}{n}\sum_i Y_i^2 - \bar{Y}^2 &= \lim_{n\to\infty}\frac{1}{n}\sum_i Y_i^2 - \lim_{n\to\infty}\frac{1}{n}\sum_i \bar{Y}^2 \\ &= \mu_{Y^2} - \mu_Y^2 \\ &= E[Y^2] - \mu^2 \\ &= \sigma^2 \\ \text{So: } \; S_U^2 &= \frac{n}{n-1}(\sigma^2)\end{aligned}$$

- But $\lim_{n\to\infty}\frac{n}{n-1} = 1$, meaning $S_U^2 \xrightarrow{p} \sigma^2$

Consistency

- Why did we need "consistency"?
- We know from the CLT that the standardized version of $\bar{Y}$ converges in probability to the standard Normal

$$F\left(\frac{\bar{Y}-\mu}{\sigma/\sqrt{n}}\right) \xrightarrow{p} \Phi$$

- We need to prove that the logic of the CLT works when we replace σ with S_U

$$F\left(\frac{\bar{Y}-\mu}{S_U/\sqrt{n}}\right) \xrightarrow{p} \Phi$$

Slutzky's Theorem

- If:

  1. $F(U_n) \xrightarrow{p} \Phi$

  2. $F(W_n) \xrightarrow{p} 1$

- Then:
  - $F\left(\frac{U_n}{W_n}\right) \xrightarrow{p} \Phi$

- In words: the ratio of a function that converges to the Standard Normal over a function that converges to 1 itself converges to the Standard Normal

- **OUR GOAL**: Prove $F\left(\frac{\bar{Y}-\mu}{S_U/\sqrt{n}}\right) \xrightarrow{p} \Phi$

Proof

- Start by re-writing our standardized sampling distribution as follows (dropping the $F(\cdot)$ for legibility):

$$\begin{aligned}
\frac{\bar{Y}-\mu}{S_U/\sqrt{n}} &= \sqrt{n}\left(\frac{\bar{Y}-\mu}{S_U}\right) \\
&= \sqrt{n}\left(\frac{\bar{Y}-\mu}{S_U}\right)\frac{\sigma}{\sigma} \\
&= \sqrt{n}\left(\frac{\bar{Y}-\mu}{\sigma}\right)\frac{\sigma}{S_U} \\
&= \frac{\sqrt{n}\left(\frac{\bar{Y}-\mu}{\sigma}\right)}{\frac{S_U}{\sigma}}
\end{aligned}$$

- From CLT: $\sqrt{n}\left(\frac{\bar{Y}-\mu}{\sigma}\right) \xrightarrow{p} \Phi$

- So need to prove that $\frac{S_U}{\sigma} \xrightarrow{p} 1$

Proof

$$
\begin{aligned}
\frac{S_U}{\sigma} &= \sqrt{\frac{S_U^2}{\sigma^2}} \\
&= \sqrt{\frac{S_U^2 \xrightarrow{p} \sigma^2}{\sigma^2 \xrightarrow{p} \sigma^2}} \\
&= \sqrt{1} \\
&= 1
\end{aligned}
$$

- Thus!

$$
\begin{aligned}
\frac{\bar{Y}-\mu}{S_U/\sqrt{n}} &= \frac{\sqrt{n}\left(\frac{\bar{Y}-\mu}{\sigma}\right) \xrightarrow{p} \Phi}{\frac{S_U}{\sigma} \xrightarrow{p} 1} \\
\frac{\bar{Y}-\mu}{S_U/\sqrt{n}} &\xrightarrow{p} \Phi
\end{aligned}
$$

Large-Sample CI

- So we can use S_U in the standard sampling distribution!
  - (When n is large...if n isn't large, then these asymptotic properties don't hold)
- Therefore: $P\left(\bar{Y} - z_{\alpha/2}\frac{S_U}{\sqrt{n}} \leq \mu \leq \bar{Y} + z_{\alpha/2}\frac{S_U}{\sqrt{n}}\right) \approx 1 - \alpha$
- Quiz: why did we spend that time with **consistency**?

Example Time!

- American Community Study (ACS) sampled 350,000 NY households with a sample mean of 76,247 household income and an unbiased sample standard deviation (i.e., the unbiased estimate of the population standard deviation) of $S_U = 61,427$. What is the 90% CI associated with this estimate?

- We want to write our 90% CI as $\bar{Y} \pm z_{\alpha/2}\sigma_{\bar{Y}}$

  - What is $\bar{Y}$?

  - What is $z_{\alpha/2}$?

  - What is $\sigma_{\bar{Y}}$?

  - What can we replace σ with in $\frac{\sigma}{\sqrt{n}}$?

Example Time!

- CNN poll of 1,038 randomly sampled adults revealing that Biden's approval rating is at 41%, meaning of those asked if they approve of Biden's performance as president, 41% said yes. What is the 95% CI associated with this estimate?
- Trickier! Still want to write $\hat{\theta} \pm z_{\alpha/2}\sigma_{\hat{\theta}}$
- Our parameter of interest θ is no longer μ but p
- Our estimator $\hat{\theta}$ is no longer $\bar{Y}$ but $\hat{p} = \frac{Y}{n} = 0.41$
- So we have $\hat{p}$ and we can get $z_{\alpha/2}$ the standard way (i.e., using `qnorm(.025) = 1.96`)
- What about $\sigma_{\hat{p}}$?
- Recall that $VAR(\hat{p}) = VAR\left(\frac{Y}{n}\right) = \frac{1}{n^2}VAR(Y) = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n}$
- So $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$

Example Time!

- Poll of 1,006 adults between Sep. 15 and 20, 2023 asking about a hypothetical vote choice if the election were held tomorrow, found that 50% of respondents indicated they would support Trump, and 46% indicated they would support Biden. This marks a reduction in Trump support from a previous tracking poll fielded a week earlier of 1,203 adults who indicated 52% support for Trump and 46% support for Biden.

- How confident are we that the change in Trump's support over this period is not due to sampling error?

- Parameter we seek is $p_1 - p_2$ where p_1 is Trump's **true** support in the first poll and p_2 is his **true** support in the second poll. Consider the polls as binomial experiments in which Y_1 is the number of "successes" (here, the # favoring Trump) in the first poll and Y_2 is the number of "successes" in the second poll.

- Intuitive estimator: $\hat{p}_1 - \hat{p}_2$. Is this unbiased?

- Calculate estimator's standard errors: $\sqrt{VAR(\hat{p}_1 - \hat{p}_2)} = \sqrt{VAR(\hat{p}_1) + VAR(\hat{p}_2)} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

Example Time!

- Continuing from the previous example, what is the 95% confidence interval for this estimator?
- Does this interval include zero? How can we interpret that?
- What about the 90% confidence interval? Does it still include zero?
- At what level of confidence would we conclude Trump's support changed between the two surveys?
- **Think**: want to find α (call it α^*) s.t. the *lower bound of the CI is greater than zero*

$$\hat{p}_1 - \hat{p}_2 - z_{\alpha^*/2}\left(\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right) > 0$$

$$-z_{\alpha^*/2}\left(\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right) > -(\hat{p}_1 - \hat{p}_2)$$

$$z_{\alpha^*/2} < \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$