Econ 210 - Review

Sidharth Sah

September 8, 2023

Estimators

- Estimators are functions of samples of data that are used to make “guesses” of parameters of probability distributions
- Estimators are *random* objects - they are functions of random samples, so they themselves are random
- For any parameter, there are infinite possible estimators - we rely on the properties of these estimators to evaluate their usefulness

Unbiasedness

- An estimator, $\hat{\theta}_n$ is unbiased for parameter θ if $E[\hat{\theta}_n] = \theta$
- If you want to prove biasedness/unbiasedness of an estimator, when in doubt, take the expectation!
- Useful expectation tricks
  - For constant a and rv X

$$E[aX] = aE[X] \ \& \ E[a + X] = a + E[X]$$

  - $E[\sum_{i=1}^{m} X_m] = \sum_{i=1}^{m} E[X_m]$
  - For identically dist $X_i \sim X$, $E[X_i] = E[X]$
  - For rv X and set A, $E[\mathbb{1}\{X \in A\}] = P(X \in A)$

(Finite-Sample) Variance

- The variance of an estimator tells us how disperse its distribution is. If we have an unbiased estimator, lower variance therefor means we're more likely to have estimates "close" to the actual value - want smaller variance where possible
- Useful variance tricks
  - For a,b scalars: $Var(a + bX) = b^2 Var(X)$
  - $Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y)$
  - For $X_1, ..., X_m$ independent,

$$Var(\sum_{i=1}^{m} X_i) = \sum_{i=1}^{m} Var(X_i)$$

Consistency

- An estimator, $\hat{\theta}_n$ is consistent for parameter θ if $\hat{\theta}_n \xrightarrow{p} \theta$ - aka the estimator gets arbitrarily close to the parameter in probability, as sample size gets large
- Proving consistency is generally a two step process:
  - Apply WLLN: this says that "sample means" converge in probability to the expectation of the thing within as $n \to \infty$. Find the sample mean(s) and apply it (checking the assumptions if you've got time)
  - Apply CMT (if necessary): This says that continuous functions of things that converge in probability also converge in probability. Is the expression for the estimator a continuous function of one or more sample means? You'll need the CMT!

Limiting Distribution

- Usually take the form

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} N(0, \sigma_\theta^2)$$

  Allows finite-sample inference because it tells us how likely it is that our estimate is "far away" from the true value
- Deriving the limiting distribution is generally two steps:
  - Apply CLT: this says that expressions like $\sqrt{n}(\frac{1}{n}\sum_{i=1}^{n} X_i - E[X])$ converge in distribution to $N(0, Var(X))$ as $n \to \infty$. Find the sample mean and apply it (checking the assumptions if you've got time)
  - Apply Slutsky (if necessary): This says that "continuous" functions of things that converge in dist and things that converge in prob also converge in dist. Is the expression for the estimator a continuous function of one or more sample means? You'll need Slutsky!

Linear Regression

- We're interested in the relationship between some random variable Y and some random variable/vector X, and want to learn about them through the parameter β in the equation:

$$Y = X'\beta + U$$

- We can define the β and U in the equation above in multiple ways - which definition we choose informs how we interpret the estimate $\hat{\beta}_n$
- In order for equation (1) to be estimable, we need $E[XU] = 0$ (among other assumptions). Thus, we can pick any definition that allows us to assume that (generally prefer causal interp to descriptive interps and lin cond exp interp to BLP interp)

Linear Regression

- Three interpretations:
  - Linear Conditional Expectation: Defines β as:

    $$E[Y|X] = X'\beta$$

    and $U = Y - E[Y|X]$. This requires assuming that $E[Y|X]$ is linear in X. Descriptive interpretation. Implies that $E[XU] = E[U|X] = 0$
  - Best Linear Predictor: Defines β as solving:

    $$\min_b E[(Y - X'b)^2]$$

    and $U = Y - X'\beta$. No real assumptions required. Descriptive interpretation. Implies that $E[XU] = 0$

Linear Regression

- Three interpretations:
  - Causal Model: Defines β as:

    $$Y = g(X, U) = X'\beta + U$$

    where $g(X, U)$ is a causal model of Y. Defines U as all of the non-X determinants of Y. Assuming $E[XU] = 0$ requires making an argument about the context - why is X uncorrelated with the other determinants of Y?

Special Case of Simple Linear Regression

- Potential outcomes Y_1 and Y_0 tells us what value of outcome Y someone *would* have if they had treatment values $X = 1$ versus $X = 0$
- If we have SLR with $X \in \{0, 1\}$ and potential outcomes $(Y_0, Y_1) \perp\!\!\!\perp X$, then we can interpret the β as the ATE:

$$\beta = ATE = E[Y_1 - Y_0]$$

  the average difference in potential outcomes across the whole population!

Special Case of Multivariate Linear Regression

- If we have causal model

$$Y = \beta_0 + \beta_1 D + U$$

with $E[DU] \neq 0$, but we have a *control* C such that

$$E[U|D, C] = E[U|C] = \gamma_0 + \gamma_2 C$$

then we can consistently estimate the new regression

$$Y = (\beta_0 + \gamma_0) + \beta_1 D + \gamma_2 C + V$$

to recover the causal parameter β_1 and the non-causal parameter γ_2

Assorted Regression Sub-topics

- R^2 is a measure of fit - how well does the estimated $\hat{\beta}_n$ fit the data, on a scale from 0 to 1? This is purely descriptive. R^2 will mechanically increase with additional regressors. For this reason, some prefer the adjusted R^2 which may increase *or* decrease with additional regressors
- Homoskedasticity/heteroskedasticity: When we do finite-sample inference with OLS, need to assume that U is homoskedastic or else allow for heteroskedasticity. Homoskedasticity means that $Var(U|X) = Var(U)$ - the error term has the same dispersion everywhere - usually a weird assumption!

Assorted Regression Sub-topics

- Standard error: This is a measure of the precision with which we estimated β (or subcomponents of β). It decreases with n and increases with the variance of the limiting distribution. When it's smaller, we'll have smaller confidence intervals - aka more precise estimates
- Omitted Variable Bias: If we have a causal model with multiple independent variables, but we omit 1 or more, this gives us a formula telling us how the OLS estimator will diverge from the causal parameter we are after:

$$\hat{\beta}_1 \xrightarrow{p} \beta_1 + \beta_2 \frac{Cov(X_1, X_2)}{Var(X_1)}$$

Assorted Regression Sub-topics

- Perfect Multicollinearity: In order to estimate multivariate linear regression, we need there to not be perfect multicollinearity in X. Perfect multicollinearity occurs when some element of X, X_i, can be written as a linear function of one or more of the other subcomponents of X

Instrumental Variables

- If we want to estimate a causal model

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

  but we have $E[XU] \neq 0$, one option is to use an instrument Z. Z is a variable that only interacts with Y indirectly through X. That is, it satisfies:
  - Instrument Exogeneity: $E[ZU] = 0$ - so Z doesn't covary with the other determinants of Y
  - Instrument Relevance: $Cov(Z, X) \neq 0$ - so Z does covary with X

LATE

- Under special circumstances, we can interpret the β_1 that we estimate using IV as a LATE. Those circumstances are $X, Z \in \{0, 1\}$ and, for potential treatments X_1 and X_0 and potential outcomes Y_1 and Y_0:
  - $(Y_1, Y_0, X_1, X_0) \perp\!\!\!\perp Z$
  - $X_1 \neq X_0$ sometimes
  - $X_1 \geq X_0$ always (monotonicity)
- Then, we'll have that

$$\beta_1 = E[Y_1 - Y_0 | X_1 > X_0]$$

this is the average treatment effect of X on Y among the sub-population for whom Z actually has an effect on X (the *compliers*)