Introduction to Econometrics 2: Recitation 13

Seung-hun Lee

Columbia University

May 9th, 2020

Bootstraps and Multiple Hypothesis Tests

Better ways to do asymptotics

- There are cases when we are not certain about the characteristics of the distribution that generated the data.
- There are cases where we only know the asymptotic properties and not much about the finite sample properties.
- In this context, taking an asymptotic approach can lead to inaccurate estimations
- Therefore, we can look into conducting a bootstrap estimation, which is at least as better as classical asymptotics and sometimes better
  - Bootstrap estimation is also consistent and under some conditions, they have smaller approximation error.

Bootstraps and Multiple Hypothesis Tests

Better ways to do asymptotics

- We can conduct bootstraps by getting an IID sample $(X_1, ..., X_n)$ from the data and treating this as a population sample.
- We obtain a t-statistics from this sample.
- We make n draws with replacement and compute the new t-statistics.
- Re-draw the sample many times and keep track of each t-statistic.
- Then, generate an empirical distribution of these statistics and identify relevant cutoffs, (2.5% and 97.5%) for instance, that can be used to create upper and lower bounds for the bootstrap CI.
- It can be shown, using Edgeworth expansion, that bootstrap approximation error and the classical approximation error have the same size of error, $O\left(1/\sqrt{n}\right)$.
- If the test statistic is (asymptotically) pivotal, then bootstrapped estimation can do better since they have smaller approximation error of $O\left(1/n\right)$.

Bootstraps and Multiple Hypothesis Tests

Proper ways to run multiple tests

- When we are conducting multiple hypothesis tests naively, we can run into a mistake of making a false discovery - rejecting a true H_0
- This may lead to false discoveries
- FWE: Bonferroni and Holm
  - Bonferroni method uniformly applies a lower p-value cutoff
  - Holm method is a step-down method that considers from the hypothesis with the lowest p-value to the highest one
- FDR: Benjamini-Hochberg (step-up)
  - Plot the actual p-values of each hypotheses and the relevant cutoff p-values used to reject/not reject the null hypothesis.
  - We check whether the actual p-value lies above the cutoff, accept that H_0 as true and move on to the next highest hypothesis.
  - The process is stopped when the actual p-value starts to lie below the cutoff p-value. The remaining hypotheses are all considered to have a false H_0

Quantile Regression

Not everything is about conditional means

- We may be interested in not just the conditional expectation, but some other distributional properties
- To answer this, we run a quantile regression that is meant to capture different values of the parameter of interest depending on the location of the conditional distribution.
- Check function $\rho_\tau(u) = u(\tau - 1(u \leq 0))$ is very useful.
- For any τ, there is going to be a kink for the check function at $u = 0$.
- We make use of this property and obtain the quantile estimator by minimizing

$$E[\rho_\tau(Y_i - X_i\beta)|X_i] \text{ or } E[\rho_\tau(Y_i - X_i\beta)X_i]$$

Nonparametric Regression

Why do we even do this?

- We rarely have the exact knowledge about the characteristics of the distribution $P(y|x)$ that the observations are drawn from.
- When we carry out nonparametric estimation, we are not required to make any modeling assumptions related to the data.
- This helps address one of the key internal validity threats - using the wrong functional forms in the specification.
- We can also use nonparametric regression in tandem with parametric estimation - Run the model parametrically and run the same model in nonparametrics to confirm whether our parametric form is correct

Nonparametric Regression

Bandwidth is essential!

- Form of kernel density does not matter much. In any case, we estimate density of f with

  $$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right)$$

  where $K(\cdot)$ is the kernel density estimator of our choice.
- Bandwidth, on the other hand, is essential
  - When we undersmooth, we are including more variation than necessary and computation becomes difficult
  - When we oversmooth, we risk dropping even the essential traits about the distribution and mis-estimate the true density

Nonparametric Regression

Bandwidth is essential!

- The key to finding the optimal bandwidth is to balance between the bias and the variance.
  - Wider bandwidth increases bias but decreases variance.
  - Narrower bandwidth does the opposite.
- The answer can be found by minimizing the asymptotic mean integrated square error (AMISE)

$$\int E(\hat{f}(x) - f(x))^2 dx$$

  where $E(\hat{f}(x) - f(x))^2$ is the sum of variance and bias^2.
- In practice, we can choose the bandwidth based on Silverman's rule of thumb or from cross-validation.

Nonparametric Regression

What are the gains?

- One of the nonparametric methods we can use is local constant (Nadaraya and Watson Estimator), local linear, or local polynomial estimation.
- Key gains from nonparametrics compared to parametric estimation is the flexibility
  - The type of association between the covariates and the dependent variable is not bound by the functional form selected in parametric approach.
- We can model flexible relationship with just nonparametric estimation (albeit with some errors, which is why having a confidence interval is necessary)

Nonparametric Regression

The cost?

- Curse of dimensionality.
- As the dimension of covariates become larger, larger sample size is required to do any sort of asymptotic inference.
- Put it differently, the estimators converge to its asymptotic values in a much slower pace than in parametric setup.
- One of the middle grounds that is used in an attempt to have both traits of nonparametric and parametric regression is the semiparametric regression. In class, we learned partially linear models where we run

$$Y_i = X_i\beta + g(W_i) + \epsilon_i$$

Treatment Effect

Fundamental Problem in Treatment Effects

- We are interested in
$$Y_i(1) - Y_i(0)$$
- Unfortunately, there are always a problem of a missing data - you cannot be treated and untreated at the same time.
- Specifically, if we are to identify what an average treatment effect is,

$$\begin{aligned} E[Y_i(1) - Y_i(0)] = \Pr(D_i = 1)E[Y_i(1)|D_i = 1] + \Pr(D_i = 0)\color{red}{E[Y_i(1)|D_i = 0]} \\ - \Pr(D_i = 0)E[Y_i(0)|D_i = 0] - \Pr(D_i = 1)\color{red}{E[Y_i(0)|D_i = 1]} \end{aligned}$$

- We can get all the other conditional expectations from the data except the ones in red, preventing us from going further
- The type of assumptions we set determine what the conditional expectation in red can be expressed as.

Treatment Effect

Fundamental Problem in Treatment Effects

- The treated population and the untreated population may be different in some unobservable dimensions.
- In such case, conditioning the outcome on the treatment status alone would not give us an accurate assessment of the treatment effect.
- In such case, we are left to rely on instrumental variables to back out the treatment effect- MTE and LATE
- Depending on the assignment mechanism used, we may be able to use this to our advantage.
- If the assignment was based on some sort of a cutoff, then there is bound to be a jump in probability of being treated.
- This discontinuity allows us to utilize regression discontinuity, one of the front-runner methods in applied economics.

Treatment Effect

Assumptions

- Random assignment: $(Y_i(1), Y_i(0)) \perp\!\!\!\perp D_i$
- Conditional independence assumption: $(Y_i(1), Y_i(0)) \perp\!\!\!\perp D_i | X_i$
- Selection on unobservables: we need to find an instrument Z_i that satisfies validity ($(Y_i(1), Y_i(0), u_i) \perp\!\!\!\perp Z_i | X_i$), relevancy, and monotonicity (No defiers are observed)
  - MTE: Z_i should have a continuous support. It would also be ideal to have the propensity score take the value from $[0, 1]$
  - LATE: A binary variable for a Z_i would be sufficient
- Regression discontinuity: Jump in propensity, continuous distribution for X_i, W_i. Also, the continuity of potential outcomes at $W_i = c$

$$E[Y_i(d)|X_i, W_i = c^+] = E[Y_i(d)|X_i, W_i = c^-] = E[Y_i(d)|X_i, W_i = c]$$

Treatment Effect

Identification Tactics

- Random assignment: In this case, we can back out ATE of all observation using $E[Y_i(d)|D_i = d] = E[Y_i(d)|D_i = d']$
- Conditional independence assumption: We can back out is the ATE for those whose $X_i = x$.
- Selection on unobservables: We identify ATE only on compliers
  - MTE: $E[Y_i(1) - Y_i(0)|u_i = p, X_i] = \frac{\partial E[Y_i|p(x,z)=p,X_i]}{\partial p}$
  - LATE: $E[Y_i(1) - Y_i(0)|p(x,z) < u_i < p(x,z'), X_i] = \frac{E[Y_i|z',X_i]-E[Y_i|z,X_i]}{E[D_i|z',X_i]-E[D_i|z,X_i]}$
- Regression discontinuity: ATE for the observations within the bandwidth around the cutoff of the forcing variables

$$E[Y_i(1) - Y_i(0)|X_i, W_i = c] = \frac{E[Y_i|X_i, c^+] - E[Y_i|X_i, c^-]}{\Pr(D_i = 1|X_i, c^+) - \Pr(D_i = 1|X_i, c^-)}$$

Treatment Effect

Implementation

- Random assignment: OLS $Y_i = \beta + \gamma D_i + \epsilon_i$
- Conditional independence assumption: Matching, IPW, DID
  - DID when there is a clear before and after and separation of treatment and control
- Selection on unobservables: Regress D on Z and X to get $\hat{p}$
  - MTE: Regress Y onto X and $\hat{p}$ flexibly (hopefully have squared or cubed terms) or with local linear regression
  - LATE: Regress Y on X and the predicted D in a 2SLS estimation
- Regression discontinuity: No overlaps above and below the cutoff
  - Nonparametrics: Local polynomial with IK bandwidth
  - Parametrics: Run regressions separately or in one go with

$$Y_i = X_i\beta + X_i \cdot 1(W_i \geq c)\gamma + X_i \cdot (c - W_i)\delta + X_i \cdot (c - W_i) \cdot 1(W_i \geq c)\mu + \epsilon_i$$

Final words

Last minute tips

- Have resources at the ready: Lecture notes, Recitation notes, external sources you found through searching...
- Nail down on which program you would like to use
- Even if seemingly new concepts pop up, they can be mapped to what we have learned in class.
- Moreover, you have enough time to write up your answers. Do not panic and work through the question step-by-step
- You have worked very hard throughout this year. You have what it takes to succeed in the course this semester.
- So just don't lose your mind for one week
- And after it is all over, take at least a week or two off from your work
  - ... Normally, I would've recommended getting out of NYC and travelling around, but we are in a pandemic....so do something else..

... and see you (hopefully face-to-face) on September!