Lecs 13-14

10/4/17 11:40

Non Compliance & IVs

Recap: First departure from both unknown counterfactuals & complete control on treatment

Like a controlled experiment but units may choose to not comply with assigned treatment.

- Assignment to treatment $W_i \in \{0,1\}$
- Treatment exposure $Z_i(0), Z_i(1)$ ← potential outcomes for treatment exposure

Factual treatment exposure: $Z_i = Z_i(W_i) \in \{0,1\}$

- Outcome $Y_i(Z,W)$ ← potential outcome that depends both on treatment & our intervention

Factual outcome $Y_i = Y_i(Z_i(W_i), W_i) = Y_i(Z_i, W_i)$

↑ usual non-interference & consistency assumptions

W_i is known as an instrumental variable

Intent-to-treat analysis

$$ITTE_i = Y_i(Z_i(1), 1) - Y_i(Z_i(0), 0)$$

$$SAITTE = \frac{1}{n} \sum_i ITTE_i$$

$$\hat{\tau}_{ITT} = \frac{2}{n} \sum_i (-1)^{1+w_i} Y_i$$

Still unbiase, consistent, etc for SAITTE
just like a controlled experiment
nothing has changed

But it doesn't get at the isolated causal effect of treatment.

Exclusion restriction:

Assumption: Assignment only affects outcome via treatment:

$$Y_i(z,0) = Y_i(z,1) \quad \forall z \in \{0,1\}$$

For short: $Y_i(z) =$ [illegible]

Then we can conceive of the isolated effect of treatment:

$$TE_i = Y_i(1) - Y_i(0)$$

How can we estimate this type of effect?

As-treated analysis

$$\hat{\tau}_{AT} = \frac{\sum_{i:z_i=1} Y_i}{\sum_{i:z_i=1} 1} - \frac{\sum_{i:z_i=0} Y_i}{\sum_{i:z_i=0} 1}$$

Can be significantly biased!

E.g. those who would always refuse to go to job training are also the lazy job seekers who wouldn't benefit from treatment anyway → upward bias in estimate

So let's try & classify all modes of non compliance:

- Complier (C): $Z_i(0)=0, \; Z_i(1)=1$
- Always-takers (AT): $Z_i(0)=1=Z_i(1)$
- Never-takers (NT): $Z_i(0)=0=Z_i(1)$
- Defiers (D): $Z_i(0)=1, \; Z_i(1)=0$

This is an exhaustive list.

$$(Z_i(0), Z_i(1)) \in \{0,1\}^2$$

Note that for AT&NT, no matter what we do, we can never observe one of the outcome

→ SATE might not an achievable goal.

Alternative: $SLATE = \sum_{i:\text{Complier}} TE_i \Big/ \sum_{i:\text{Complier}} 1$

$\uparrow$
sample
local
average
treatment
effect

$$LATE = \mathbb{E}[SLATE]$$
$$= \mathbb{E}[TE \mid Z_i(0)=0, Z_i(1)=1]$$

Get a chance to observe both outcomes for a complier

But: We don't know who the compliers are ...

What can we measure:

- Compliers: $ITTE_i = Y_i(Z_i(1)) - Y_i(Z_i(0))$
  $= Y_i(1) - Y_i(0) = TE_i$
- AT: $ITTE_i = Y_i(1) - Y_i(1) = 0$
- NT: $ITTE_i = Y_i(0) - Y_i(0) = 0$
- Defiers: $ITTE_i = Y_i(0) - Y_i(1) = -TE_i$

Assumption (Monotonicity) No defiers

i.e. $Z_i(1) \geq Z_i(0)$

Then:

$$SAITTE = \frac{1}{n} \sum_i ITTE_i$$

$$= \frac{1}{n}\left(\sum_{\text{compliers}} TE_i + \sum_{AT} 0 + \sum_{NT} 0 \right)$$

$$= \frac{1}{n} \sum_{\text{compliers}} TE_i$$

$$= SLATE \cdot \underbrace{(\#\text{compliers}) \cdot \frac{1}{n}}_{p_c}$$

$$\Rightarrow SLATE = \underset{\substack{\uparrow \\ \text{can estimate} \\ \text{by } \hat{\tau}_{ITT}}}{SAITTE} \Big/ p_c$$

how can we estimate this?

Necessary assumption: compliers exist $p_c > 0$

$\Rightarrow LATE = AITTE / \text{prob of compliance}$

Put another way, assignment has *some* effect on treatment exposure

$$\mathbb{E}[Z(1) - Z(0)] \neq 0$$

Can't observe who's a complier
but w/o defiers:

- can observe AT in intended control gp

$$\hat{p}_{AT} = \frac{2}{n} \sum_{i: W_i = 0} \mathbb{I}[Z_i = 1]$$

- Can observe NT in intended treatment gp

$$\hat{p}_{NT} = \frac{2}{n} \sum_{i:W_i=1} \mathbb{I}[Z_i=0]$$

w/o defiers:

$$\begin{aligned}
\hat{p}_C &= 1 - \hat{p}_{AT} - \hat{p}_{NT} \\
&= (1-\hat{p}_{NT}) - \hat{p}_{AT} \\
&= \frac{2}{n}\sum_{i:W_i=1} \mathbb{I}[Z_i=1] - \frac{2}{n}\sum_{i:W_i=0} \mathbb{I}[Z_i=1] \\
&= \frac{2}{n}\sum_i (-1)^{1+W_i} Z_i
\end{aligned}$$

vs $p_C = \frac{1}{n}\sum_i \mathbb{I}[\text{Complier}] = \frac{1}{n}\sum_i (Z_i(1) - Z_i(0))$

intent to treat estimate of the effect of assignment W_i on exposure Z_i

$\Rightarrow$ $\hat{p}_C$ is unbiased & consistent for p_C

Get: The Wald estimator for SLATE:

$$\hat{\tau}_{LATE} = \frac{\hat{\tau}_{ITT}}{\hat{p}_C} = \frac{\sum_i (-1)^{1+W_i} Y_i}{\sum_i (-1)^{1+W_i} Z_i} = \frac{\widehat{Cov}_n(W,Y)}{\widehat{Cov}_n(W,Z)}$$

Consistent for SLATE (under similar Lyapunov conditions for fixed sample path)

$\hat{\tau}_{LATE} - SLATE = o_p(1)$

for LATE (under random sampling)

Corollary of Slutsky's thm

But $\hat{\tau}_{LATE}$ is biased!

How to reduce bias? Reduce variance of $\hat{p}_C$.

How to reduce variance of $\hat{p}_c$?

If's a controlled experiment on Z as outcome — just do balancing.

$$\hat{Z}_i = \frac{Z_i(0)+Z_i(1)}{2} = \begin{cases} 1 & \text{AT} \\ 1/2 & \text{Comp} \\ 0 & \text{NT} \end{cases}$$

compliance status

$$C_i = \begin{cases} 1 & \text{AT} \\ 0 & \text{comp} \\ -1 & \text{NT} \end{cases} = 2\hat{Z}_i - 1$$

$$p_c - \hat{p}_c = \frac{2}{n}\sum_i (-1)^{1+W_i}\hat{Z}_i = \frac{1}{n}\sum_i (-1)^{1+W_i} C_i$$

→ so we need to balance

$$\hat{f}_c(\cancel{x}) = \mathbb{E}[C_i \mid X=x]$$

but we also want to balance

$$\hat{f}_y(x) = \mathbb{E}\left[\frac{Y_i(1)+Y_i(0)}{2} \,\middle|\, X=x\right]$$

If we have two different beliefs $\|\cdot\|', \|\cdot\|''$ on $\hat{f}_y$ & $\hat{f}_c$ (resp)

then by creating the new

$$\|f\|^2 = \inf_{f=g+h} \|g\|'^2 + \|h\|''^2$$

Example: Sexton & Hebel '84 JAMA

Study pregnant women who usually smoke at least 10 cigs/day

Randomly assign (CR):

$W_i=1$: encourage & assist to stop smoking through a new intervention

$W_i=0$: usual care

Real data:

	$W_i=+1$	$W_i=0$
# subjects	1000	1000
% smoke	18.7%	44.9%
Avg birth weight (oz)	116.14	111.84

Effect of new intervention on smoking cessation:
26.2 reduction in prob (% pts) of smoking while pregnant

Effect of new intervent on birthweight:
4.3 oz increase in birth weight

→ probably a good intervention to implementation

But we also want to know if to go further & design a more intensive intervention. Should we?
How to design? What to address?

Key question: effect of smoking itself on birth weight.

How to analyze? IVs!

Wald estimate of smoking's effect on birth weight?

$$\frac{4.3}{.262} = 16.412 \text{ oz}$$

Conclusion: should definitely address smoking by pregnant women and try further smoking cessation — not just this intervention

Q: How to understand this estimate?
Did we satisfy all our assumptions?
Does it generalize?

IVs in natural experiments:

What is the effect of having a baby on a cis woman's income?

population: cis women w/ ≥ 2 kids
treatment: have/don't have 3rd kid
instrument: are the first two kids of same or different gender?

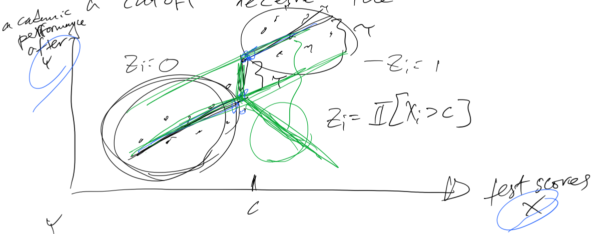
Regression Discontinuity Designs

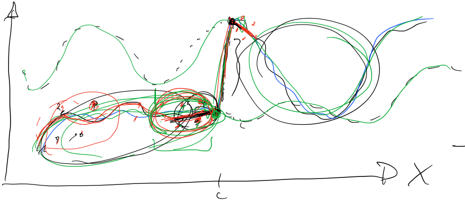
The original experiment where RDDs were devised:

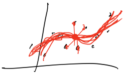
Thistlewaite & Campbell '60

What's the effect of a merit award on academic outcomes.

- students w/ test scores above a cutoff receive the award.







$\propto e^{-\|x-x'\|^2/\sigma}$

$$\tau = CATE(c) = \mathbb{E}[TE \mid X=c]$$
$$= \mathbb{E}[Y(1) \mid X=c] - \mathbb{E}[Y(0) \mid X=c]$$

Asn: $\mathbb{E}[Y(z) \mid X=x]$ is continuous at $x=c$

$$CATE(c) = \tau = \lim_{x \downarrow c} \mathbb{E}[Y(1) \mid X=x] - \lim_{x \uparrow c} \mathbb{E}[Y(0) \mid X=x]$$
$$= \lim_{x \downarrow c} \mathbb{E}[Y \mid X=x] - \lim_{x \uparrow c} \mathbb{E}[Y \mid X=x]$$

$\mathbb{E}[Y \mid X=x]$ is not continuous

only $\mathbb{E}[Y(z) \mid X=x]$ is continuous

$$\hat{\tau} = \lim_{x \downarrow c} \hat{\mathbb{E}}[Y \mid X=x] - \lim_{x \uparrow c} \hat{\mathbb{E}}[Y \mid X=x]$$

regression estimate

So: regress factual outcome Y on X near c
& look for the discontinuity

One option: linear regression

$$Y \sim \alpha + \beta X + \tau \mathbb{1}[X>c] + \varepsilon$$

$\hat{\tau}$ is our estimate for this coefficient

other options: any local nonparametric regression

usually: kernel/local poly regression

Opportunity for ML to learn useful representations to learn value near c — especially for more personalized estimates (i.e. conditioned on more covariates)