R Lab 3

topics:

- IPTW estimator
- positivity assumption
- missing data

IMPT REMINDERS

- Josh: remember to record this
- R HW 3 is posted soon (already?)
- Use TEMPLATE for R HW 3 so you won't miss any Qs
- Solutions to this lab will be posted soon too; R code will be very helpful

- Groups for final project - size is flexible - find your group and let LBB know! Discussion groups often keep working together if you haven't thought about it yet

IPTW intuition

IPTW intuition

Create weights based on covariate strata and exposure status, use to re-weight outcome differences.

In a sense, it 'encodes' covariate information and strata size into the weights. [red]

Compare to g-comp, which 'encodes' covariate information into the outcome regression model (aka conditional mean model), then averages over strata. [blue]

$$\mathbb{P}_0(O) = \mathbb{P}_0(W, A, Y) = \mathbb{P}_0(W)\mathbb{P}_0(A|W)\mathbb{P}_0(Y|A, W)$$

IPTW estimand

$$\hat{\Psi}(\mathbb{P}_n)_{IPW} = \frac{1}{n}\sum_{i=1}^{n} \frac{\mathbb{I}(A_i = 1)}{\hat{\mathbb{P}}(A = 1|W_i)} Y_i - \frac{1}{n}\sum_{i=1}^{n} \frac{\mathbb{I}(A_i = 0)}{\hat{\mathbb{P}}(A = 0|W_i)} Y_i$$

Average among treated, weighted

Average among untreated, weighted

Up-weights treatment observations that are unusual for that covariate strata

Up-weights control observations that are unusual for that covariate strata

See full proof of equivalence to g-comp estimand in Lecture 7 slides.

IPTW issues… 1

$$\hat{\Psi}(\mathbb{P}_n)_{IPW} = \frac{1}{n}\sum_{i=1}^{n}\frac{\mathbb{I}(A_i = 1)}{\hat{\mathbb{P}}(A = 1|W_i)}Y_i - \frac{1}{n}\sum_{i=1}^{n}\frac{\mathbb{I}(A_i = 0)}{\hat{\mathbb{P}}(A = 0|W_i)}Y_i$$

If $\hat{\mathbb{P}}(A = a \mid W_i)$ gets very small (or zero)… weights can get Very Large. “Near”/“Practical” positivity violation.

Example: If *W* is age, and if very few young children are getting a treatment, the outcomes for the few who DO get the treatment will have a LOT of impact on the resulting ATE estimate. High variance based on that small subset of the population.

And if we have *no* children getting treatment, only control, we can’t do estimation non-parametrically.

Very common in applied settings. Tough decisions need to be made about whether/how to trim weights.

IPTW issues… 2

$$\hat{\Psi}(\mathbb{P}_n)_{IPW} = \frac{1}{n}\sum_{i=1}^{n} \frac{\mathbb{I}(A_i = 1)}{\hat{\mathbb{P}}(A = 1|W_i)} Y_i - \frac{1}{n}\sum_{i=1}^{n} \frac{\mathbb{I}(A_i = 0)}{\hat{\mathbb{P}}(A = 0|W_i)} Y_i$$

If $\hat{\mathbb{P}}(A = a \mid W_i)$ not estimated consistently, biased estimator. As dimension of *W* grows (and if confounding by *W* is strong), consistent estimation more difficult.

“Stabilized” Horvitz-Thompson IPTW can sometimes be better in small samples but suffers some of the same issues.

Practical notes

- Don't want to use covariates that have a strong effect on A but not on W (instruments)
- Sensitivity analyses on impact of different levels of truncation for large weights are a good thing to pre-specify

R Lab 3 tasks

- Create a logistic regression model for $\mathbb{P}(A = a \mid W_1, W_2)$
- Use it to predict $\hat{\mathbb{P}}(A = 1 \mid W_i)$ and $\hat{\mathbb{P}}(A = 0 \mid W_i)$ for each observation
- Calculate the weights and the ATE based on the IPTW estimator.
- Assess how sensitive the results are to weight truncation.
- Lots of clear directions in the instructions.

Connection to missing data

- Intuition:
  - Up-weighting rare exposure/covariate combinations adjusts for imbalance
  - Up-weighting rare missing data/covariate combinations adjusts for differential missingness among strata of exposures and covariates
- Note that we only have *Y* values for non-missing outcomes, of course

$$\hat{\Psi}_{IPTW}(\mathbb{P}_n) = \frac{1}{n}\sum_{i=1}^{n} \frac{\mathbb{I}(A_i = 1, \Delta_i = 1)}{\hat{\mathbb{P}}(A = 1|W_i)\hat{\mathbb{P}}(\Delta = 1|A_i, W_i)} Y_i - \frac{1}{n}\sum_{i=1}^{n} \frac{\mathbb{I}(A_i = 0, \Delta_i = 1)}{\hat{\mathbb{P}}(A = 0|W_i)\hat{\mathbb{P}}(\Delta = 1|A_i, W_i)} Y_i$$

Connection to missing data

$$\hat{\Psi}_{IPTW}(\mathbb{P}_n) = \frac{1}{n}\sum_{i=1}^{n} \frac{\mathbb{I}(A_i = 1, \Delta_i = 1)}{\hat{\mathbb{P}}(A = 1|W_i)\hat{\mathbb{P}}(\Delta = 1|A_i, W_i)} Y_i - \frac{1}{n}\sum_{i=1}^{n} \frac{\mathbb{I}(A_i = 0, \Delta_i = 1)}{\hat{\mathbb{P}}(A = 0|W_i)\hat{\mathbb{P}}(\Delta = 1|A_i, W_i)} Y_i$$

For the lab and HW 3:

- Create a model for $\mathbb{P}(A \mid W)$ and $\mathbb{P}(\Delta \mid A, W)$
- Use to predict probabilities needed for denominator, multiply to get weights
- Implement IPTW based on those weights
- Lots of clear instructions in the lab assignment.

R HW 3

Most of R HW 3 is very similar to the lab, except for the last question, where you have to:

- Compare the H-T IPTW and regular IPTW on bias/variance/MSE
- Develop your own estimator that's better than them for a particular DGP.
- Note: Only trying to estimate $\mathbb{E}_0[Y_1]$, not the typical $\mathbb{E}_0[Y_1 - Y_0]$ (ATE)
- A couple of hints on the next slide…

R HW 3

Hints:

- Variance of a Bernoulli random variable (like a single coin flip) with probability of success *p* is: $p(1-p)$

- $Var(c \cdot X) = c^2 Var(X)$ for any constant *c*. For example, *c* might be 1000. :)

- Write down the values the IPTW estimator will take when *A* = 0 and 1, *W* = 0 or 1, and look at the probabilities of occurrence for each combo

- Centering the values around 0 or 0/1 will reduce variance…

Coming attractions

So far, g-comp and IPTW have problems:

1. How to correctly specify models once our covariate dimension gets large?

2. How to deal with extreme weights?

3. How to get statistical inference? Smallest possible standard errors?

4. … probably other things I'm forgetting…

Using Super Learner will help with (1), TMLE will help with (2) and (3)!

IMPT REMINDERS

- Josh: stop the recording
- Use TEMPLATE for R HW 3 so that you won't forget any Qs!
- I didn't cover EVERYTHING in R Lab 3; please look over the answer key. R code will be very helpful, plus more detail and context