Modern Sampling Methods

Class 5: Multi-Wave Experiments

January 10, 2022

Outline

- Choice of Propensity Score
- Two-Stage Experiment
- Adaptive Choice of Propensity Score
  (based on Hahn, Hirano, & Karlan 2011)

Choice of Propensity Score

Consider a **1-stage experiment or observational study** with:

1. Unconfoundness: $D_i \perp (Y_{0i}, Y_{1i}) | X_i$.
2. Overlap: $0 < P(D_i = 1 | X_i) < 1 \quad \forall X_i$.

Assume X_i is <u>discrete</u> or <u>discretized</u>.

In experiments, 1 & 2 can be guaranteed by design.

Propensity score: $p(x) = P(D_i = 1 | X_i = x)$.

Semiparametric Efficiency Bound

Theorem

(Hahn, 1998) Suppose that $\hat{\beta}$ satisfies

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N(0, V),$$

and is regular*. Then its variance satisfies*

$$V \geq E\left[\frac{\sigma_1^2(X_i)}{p(X_i)} + \frac{\sigma_0^2(X_i)}{1 - p(X_i)} + (\beta(X_i) - \beta)^2\right],$$

where

$$\begin{aligned} \beta(x) &= E[Y_{1i} - Y_{0i} | X_i = x], \\ \sigma_0^2(x) &= Var[Y_{0i} | X_i = x], \\ \sigma_1^2(x) &= Var[Y_{1i} | X_i = x]. \end{aligned}$$

An efficient estimator:

Hahn (1998):

$$\hat{\beta} = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{r}_1(X_i) - \hat{r}_0(X_i)\right),$$

where $\hat{r}_0(x)$ and $\hat{r}_1(x)$ are nonparametric estimators of

$$r_1(x) = E[Y_i|D_i = 1, X_i = x],$$

$$r_0(x) = E[Y_i|D_i = 0, X_i = x].$$

Another efficient estimator:

Hirano, Imbens, and Ridder (HIR, 2003):

$$\hat{\beta} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{D_i Y_i}{\hat{p}(X_i)} - \frac{(1-D_i)Y_i}{1-\hat{p}(X_i)}\right),$$

where $\hat{p}(x)$ is a nonparametric estimator of the propensity score.

When X_i is discrete: Hahn and HIR estimators are identical.

Now suppose we can <u>choose</u> $p(x)$ based on knowledge of $\sigma_0(x)$ and $\sigma_1(x)$.

$$\min_{p(\cdot)} E\left[\frac{\sigma_1^2(X_i)}{p(X_i)} + \frac{\sigma_0^2(X_i)}{1-p(X_i)} + (\beta(X_i)-\beta)^2\right]$$

First order conditions for a minimum imply:

$$p^*(x) = \frac{\sigma_1(x)}{\sigma_0(x)}\left(1+\frac{\sigma_1(x)}{\sigma_0(x)}\right)^{-1}.$$

Constrained version: minimize variance subject to:

$$E[p(X_i)] = p.$$

Interior solution satisfies:

$$\lambda = -\frac{\sigma_1^2(x)}{p(x)^2} + \frac{\sigma_0^2(x)}{(1 - p(x))^2},$$

where λ is the Lagrange multiplier.

(Can be solved by numerical methods).

Intuition

- Efficiency bound involves conditional variances.
- Suppose X_i binary, and

$$\sigma_0^2(0) = \sigma_0^2(1) = \sigma_1^2(0) = \sigma_1^2(1) = 1.$$

(*Homoskedasticity*)

- Then optimal propensity score is $p(x) = p$.

Intuition

- Now suppose same setup, except that $\sigma_1^2(0) = 10$.
- This means $\text{Var}[Y_i|X_i = 0, D_i = 1]$ high. (*Heteroskedasticity*)
- $\Rightarrow$ Hard to estimate $E[Y_{1i}|X_i = 0]$.
- $\Rightarrow$ Want more observations with $X_i = 0$, $D_i = 1$.
- $\Rightarrow$ $p(0)$ should be relatively large

- So if we knew $\sigma_0(x)$ and $\sigma_1(x)$, we could pick $p(x)$ to minimize the variance bound.
- Not feasible in one-stage experiments.
- But in a two-stage experiment, we could try to estimate $\sigma_0(x)$ and $\sigma_1(x)$ from first-round data.
- We cannot change π_1, but we can choose $\pi_2(x)$ to make overall propensity score optimal.

Two-Stage Experiment

Stage 1:

- Draw n_1 subjects from population.
- Assign to treatment 1 with probability π_1 (**fixed**).
- Observe outcome Y (and D and X).

Stage 2:

- Draw n_2 subjects from population.
- Assign treatment 1 with probability $\hat{\pi}_2(X)$.
  ("hat": can use Stage 1 data to determine the rule.)
- Observe outcome Y (and D and X).

Finally, use all data to estimate effect of treatment 1 vs 0.

Two-Stage Experiment

Budget Constraint: overall treatment probability fixed at p.

Let

$$n = n_1 + n_2,$$
$$\kappa = \frac{n_1}{n}.$$

We require

$$p = \kappa\pi_1 + (1-\kappa)E_X[\hat{\pi}_2(X_i)].$$

($E_X[\cdot]$ = expectation WRT distribution of X_i.)

Adaptive Procedure

1. Using data from Stage 1, estimate conditional variances:

$$\widehat{\sigma}_0^2(x), \quad \widehat{\sigma}_1^2(x).$$

2. Choose $\hat{\pi}_2(x)$ to minimize:

$$E\left[\frac{\widehat{\sigma}_1^2(X_i)}{p(X_i)} + \frac{\widehat{\sigma}_0^2(X_i)}{1 - p(X_i)} + (\beta(X_i) - \beta)^2\right]$$

where

$$p(x) = \kappa\pi_1 + (1 - \kappa)\hat{\pi}_2(x).$$

possibly subject to:

$$E[p(X_i)] = p.$$

Adaptive Procedure

3. Use solution $\hat{\pi}_2(x)$ to determine assignment probabilities in second stage.

4. After collecting all data, pool the two stages and estimate with Hahn/HIR:

$$\hat{\beta} = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{D_i Y_i}{\hat{p}(X_i)} - \frac{(1 - D_i) Y_i}{1 - \hat{p}(X_i)} \right).$$

Note:

$\hat{\pi}_2(x)$: the "true" assignment probability in stage 2.
$\hat{p}(x)$: nonparametric estimate using pooled data.

Asymptotic Theory for Adaptive Procedure

Suppose that:

- $n_1 \to \infty$ and $n_2 \to \infty$, with $n_1/n \to \kappa$.
- $\hat{\sigma}_0^2(x)$ and $\hat{\sigma}_1^2(x)$ are sample analogs based on first-stage data.
- Let
$$\pi_2^*(x) \equiv \text{plim } \hat{\pi}_2(x).$$

Then
$$\sqrt{n}\left(\hat{\beta} - \beta\right) \xrightarrow{d} N(0, V^*),$$

where
$$V^* = E\left[\frac{\sigma_1^2(X_i)}{\pi^*(X_i)} + \frac{\sigma_0^2(X_i)}{1-\pi^*(X_i)} + (\beta(X_i) - \beta)^2\right].$$

Example 1: Karlan and List (2007)

"Does Price Matter in Charitable Giving? Evidence from a Large-scale Natural Field Experiment," AER

- Political non-profit organization
- Mailed solicitations for donations to 50,000 prior donors
- Treatment:
  $T = 1$: donation will be matched by someone else
  $T = 0$: no matching donation
- Outcome:
  $Y =$ donation amount

- Covariate: $X = 1($ "Red State" $)$
- In the field experiment: T randomly assigned, $Pr(T = 1) = 2/3$.
- We suppose this is the first of two stages of an experiment (with $\kappa = .5$).
- How would we want to assign treatment in second stage to best estimate average treatment effect?

Example 1: Karlan-List

Table: Karlan-List Experiment

	$\hat{\mu}_0$	$\hat{\sigma}_0^2$	$\hat{\mu}_1$	$\hat{\sigma}_1^2$	π^*
Blue State ($X = 0$)	0.90	73.44	0.89	67.74	0.49
Red State ($X = 1$)	0.69	57.01	1.06	97.67	0.57

- Variance using adaptive rule: 291
- Variance using nonadaptive rule: 320
- Can achieve same precision with 4558 fewer observations.

Example 2: Progresa

- Large-scale randomized experiment in Mexico
- Randomly allocated cash and nutritional supplements to families (with conditions)
- Similar experiments conducted or begun in Colombia, Ecuador, Honduras, Nicaragua
- Gertler, Martinez, and Rubio-Codina (2006) report conditional means and variances and sample sizes
- So we can apply our procedure without access to raw data.

Table: Progresa Experiment, Number of Draft Animals

	$\hat{\mu}_0$	$\hat{\sigma}_0^2$	$\hat{\mu}_1$	$\hat{\sigma}_1^2$	p_{orig}	π^*
NoAgAssets ($X = 0$)	0.41	0.34	0.34	0.07	0.55	0.69
Landless ($X = 1$)	0.49	0.79	0.44	0.37	0.67	0.59
SmallerFarm ($X = 2$)	0.68	1.3	0.58	0.63	0.68	0.59
BiggerFarm ($X = 3$)	0.83	1.2	0.87	1.83	0.62	0.45

Recommended probabilities differ from those used, but reduction in variance is quite small.

Karlan and Wood (2017)

As discussed in Class 1.

First Wave: 2/3 control, 1/3 treatment

Second Wave: 2 treatments and 1 control arm.

- Prob. of treatment conditional on prior donation, etc.
- Overall 1/3 proportions in each arm.

Discussion

- These approach requires discrete X_i.
- If X_i is continuous (or discrete and taking many values), could stratify, but it's not clear how best to choose stratification scheme.
- See Tabord-Meehan (2021) for one data-driven stratification procedure.
- Analysis depends on a specific objective (estimation of ATE); other objectives may lead to quite different solutions.