Covariate-Balancing Designs

Baseline covariates $X_i \in \mathcal{X}$

Block-randomized / Stratified design

Coarsen covariates into J strata

$$C : \mathcal{X} \to \{1, \dots, J\} = [J]$$

Then completely randomize subjects within each stratum independently & seperately

(if uneven # in a stratum then completely randomize leftovers)

$$\hat{\tau} = \frac{2}{n} \sum_{i=1}^{n} (-1)^{1+Z_i} Y_i =$$ weighted avg of within-stratum effect estimate

Same as drawing $Z_{1:n}$ uniformly at random from

$$BR_n = \left\{ Z_{1:n} : \sum_i Z_i = \frac{n}{2}, \ \forall j \in [J] \ \left| \sum_i (-1)^{Z_i} \mathbb{I}[C(X_i) = j] \right| \le 1 \right\}$$

Pair-matched design

Same as block-randomized except:

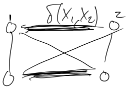
- each stratum has exactly two units
- Strata can depend on data

e.g. pair up the units into $n/2$ pairs & randomly split into control & treatment

Optimal pair matching:

Given a distance metric $\delta(x, x')$, put units into pairs to minimize the sum of within-pair distances (aka, nonbipartite matching)

poly-time solvable using Edmond's algo



"Rerandomization" (Morgan & Rubin '12)

Let $\Delta(Z_{1:n}) = \frac{2}{n} \sum_i (-1)^{1+Z_i} X_i$ $\quad (\mathcal{X} \subseteq \mathbb{R}^d)$

= diff in mean vectors of covars b/w gps

Fix $V \succeq 0$,

choose unif at random from

$$RR_n = \Big\{ z_{1:n} : \textstyle\sum_i z_i = n/2, \quad \#\Delta(z_{1:n})^T V \#\Delta(z_{1:n}) \le \underset{\text{quantile of } \chi^2}{\boxed{a}} \Big\}$$

will look like χ^2

Usually:
- Set $V = \hat{\Sigma}^{-1}$ inverse sample cov of $X_{1:n}$
- do the drawing using rejection sampling

Inference: $\hat{\tau} = \frac{2}{n} \sum_{i=1}^n (-1)^{z_i+1} y_i$

Exactly as before!

Just draw new assignments from design used at onset.

What is balance? all of these designs satisfy:

Def: A design a priori balancing if

a) $Z_{1:n} \perp\!\!\!\perp Y_{1:n}(0,1) \mid X_{1:n}$ ← only depends on $X_{1:n}$

b) $\mathbb{P}(Z_{1:n} = z_{1:n} \mid X) = \mathbb{P}(Z_{1:n} = 1 - z_{1:n} \mid X)$

(blinding the identity of treatment)

c) $\sum_{i=1}^n Z_i = n/2$ ← can be generalized

Properties: $\mathbb{E}[\hat{\tau} \mid X_{1:n}, Y_{1:n}(0,1)]$

$$= \frac{2}{n} \sum_{i=1}^n \mathbb{E}\left[Z_i Y_i(1) - (1 - Z_i) Y_i(0) \,\middle|\, \begin{matrix} X_{1:n} \\ Y_{1:n}(0,1) \end{matrix} \right]$$

$$\left(\mathbb{E}\left[Z_i \,\middle|\, \begin{matrix} X_{1:n} \\ Y_{1:n}(0,1) \end{matrix} \right] \overset{ⓐ}{=} \mathbb{E}[Z_i \mid X_{1:n}] \overset{ⓑ}{=} \frac{1}{2} \right)$$

$$\Rightarrow = \frac{2}{n} \sum_i \left(\tfrac{1}{2} Y_i(1) - \tfrac{1}{2} Y_i(0) \right) = SATE$$

$$\boxed{\operatorname{Var} \hat{\tau}} = \mathbb{E} \underbrace{\operatorname{Var}(\hat{\tau} \mid X_{1:n}, Y_{1:n}(0,1))}_{\mathbb{E}[(\hat{\tau} - SATE)^2 \mid X_{1:n}, Y_{1:n}(0,1)]} + \operatorname{Var} \underbrace{\mathbb{E}[\hat{\tau} \mid X_{1:n}, Y_{1:n}(0,1)]}_{SATE}$$

$$= \mathbb{E}(\hat{\tau} - SATE)^2 + \operatorname{Var}(SATE)$$

$$\hat{\tau} = \frac{2}{n} \sum_{i: Z_i = 1} Y_i(1) - \frac{2}{n} \sum_{i: Z_i = 0} Y_i(0)$$

$$SATE = \frac{1}{n} \sum_{i: Z_i = 1} (Y_i(1) - Y_i(0)) + \frac{1}{n} \sum_{i: Z_i = 0} (Y_i(1) - Y_i(0))$$

$$\hat{\tau} - SATE = \frac{2}{n}\sum_{i:Z_i=1}\left(\frac{Y_i(0)+Y_i(1)}{2}\right) - \frac{2}{n}\sum_{i:Z_i=0}\left(\frac{Y_i(0)+Y_i(1)}{2}\right)$$

$$= \frac{2}{n}\sum_{i=1}^{n}(-1)^{1+Z_i}\underbrace{\left(\frac{Y_i(0)+Y_i(1)}{2}\right)}_{\hat{Y}_i}$$

Is "balance" always good?

$X_{1:n}$: $-\frac{n}{2}$, , , -1 , 1 , , , $\frac{n}{2}$ $\in \mathbb{R}$

$\hat{Y}_i = Y_i(0) = Y_i(1)$: 1 0 1 1 0

$$Var(\hat{\tau} \mid X_{1:n}, Y_{1:n}(0,1)) =$$

CR: $\frac{4}{n-1}$ Blocking: $\frac{4}{n-6}$

Opt pair matching: $8/n$ Rerand ($a=0$): 4^{*} ← w/ some perturbation

Thm: Among all a priori balancing designs CR minimizes $\max_{\|\hat{Y}\|_2 \le 1} Var(\hat{\tau} \mid X_{1:n}, Y_{1:n}(0,1))$

Cor: If $\mathbb{P}(Z_i = Z_j \mid X_{1:n}) \neq \frac{n-2}{2n}$ (as it is for CR) then $\exists\, Y_{1:n}(0,1)$ s.t. $Var(\hat{\tau} \mid X_{1:n}, Y_{1:n}(0,1)) > Var(\hat{\tau}^{CR} \mid X_{1:n}, Y_{1:n}(0,1))$

Reinterpreting balance: units are iid from a population

$$\hat{Y}_i = \frac{Y_i(0) + Y_i(1)}{2} \qquad \hat{f}(x) = \mathbb{E}[\hat{Y}_i \mid X_i = x]$$

$$\varepsilon_i = \hat{Y}_i - \hat{f}(X_i)$$

$$\hat{\tau} - SATE = \frac{2}{n}\sum_i (-1)^{1+Z_i}\hat{Y}_i$$

$$= \frac{2}{n}\sum_i (-1)^{1+Z_i}\hat{f}(X_i) + \frac{2}{n}\sum_i (-1)^{1+Z_i}\varepsilon_i$$

$$u_i = (-1)^{1+Z_i} \qquad B(Z_{1:n}; f) = \frac{2}{n}\sum_i u_i f(X_i)$$

$$\mathbb{E}[\varepsilon_i \mid X_{1:n}] = \mathbb{E}[\hat{Y}_i \mid X_i] - \mathbb{E}[\hat{f}(X_i) \mid X_i]$$

$$= \hat{f}(X_i) - \hat{f}(X_i) = 0$$

$$\mathbb{E}[\varepsilon_i h(X_i) \mid X_{1:n}] = h(X_i)\cancelto{0}{\mathbb{E}[\varepsilon_i \mid X_{1:n}]} = 0$$

$\Rightarrow$ ε_i is uncorrelated w/ any function of X_i

$$\mathbb{E}(\hat{\tau} - SATE)^2$$

$$\mathbb{E}[u_i u_j \varepsilon_i \hat{f}(x_j)] = \mathbb{E}[\hat{f}(x_j)\, \mathbb{E}[u_i u_j \varepsilon_i \mid X_{1:n}]]$$

$$= \mathbb{E}[\hat{f}(x_j) (\mathbb{E}[u_i u_j \hat{Y}_i \mid X_{1:n}] - \mathbb{E}[u_i u_j \hat{f}(x_j) \mid X_{1:n}])]$$

$$= \mathbb{E}[\hat{f}(x_j)(\mathbb{E}[u_i u_j \mid X_{1:n}] \cdot \mathbb{E}[\cancelto{0}{\varepsilon_i \mid X_{1:n}}]))$$

$$\mathbb{E}[u_i u_j \varepsilon_i \varepsilon_j] \overset{= 0}{=} \mathbb{E}[\mathbb{E}[u_i u_j \mid X_{1:n}] \cdot \mathbb{E}[\varepsilon_i \varepsilon_j \mid X_{1:n}]]$$

$$= \begin{cases} \mathbb{E}[\mathbb{E}[u_i u_j \mid X_{1:n}]\, \mathbb{E}[\cancelto{0}{\varepsilon_i \mid X_{1:n}}]\, \mathbb{E}[\cancelto{0}{\varepsilon_j \mid X_{1:n}}]] = 0 & i \neq j \\ \mathbb{E}[1\, \varepsilon_i^2] = \mathbb{E}[\varepsilon_1^2] = \sigma_\varepsilon^2 & i = j \end{cases}$$

$$\mathbb{E}(\hat{\tau} - SATE)^2 = \mathbb{E}\, B^2(Z_{1:n}; \hat{f}) + \frac{4}{n}\sigma_\varepsilon^2$$

$$\operatorname{Var}\hat{\tau} = \underbrace{\mathbb{E}[B^2(Z_{1:n}; \hat{f})]}_{\text{want this minimal}} + \frac{4}{n}\sigma_\varepsilon^2 + \operatorname{Var}(SATE)$$

What if $\mathbb{E}B^2(Z_{1:n}; \hat{f}) = 0$?

→ Then: $\operatorname{Var}\hat{\tau} = \frac{4}{n}\sigma_\varepsilon^2 + \operatorname{Var}(SATE)$

→ Recall: $\operatorname{Var}\hat{\tau}^{CR} = \frac{4}{n}\sigma_{\hat{Y}}^2 + \operatorname{Var}(SATE)$ ← also in HW

define: $X_i = 0$ → know that $\hat{f}(x) = \mathbb{E}[\hat{Y}_i]$

$B^2(Z_{1:n}; \hat{f}) = 0$

$\varepsilon_i = \hat{Y}_i - \mathbb{E}[\hat{Y}_i]$

$\sigma_\varepsilon^2 = \sigma_{\hat{Y}}^2$

If X_i nontrivial → $\hat{f}$ nontrivial

What if $\operatorname{Var} SATE = 0$ (e.g. constant effect $Y_i(1) - Y_i(0) = \tau$)

$$1 - \frac{\operatorname{Var}\hat{\tau}}{\operatorname{Var}\hat{\tau}^{CR}} = 1 - \frac{\frac{4}{n}\sigma_\varepsilon^2 + \mathbb{E}[B^2(Z_{1:n}; \hat{f})]}{\frac{4}{n}\sigma_{\hat{Y}}^2}$$

$$\overset{\text{if } \mathbb{E}B^2 = 0}{=} 1 - \sigma_\varepsilon^2 / \sigma_{\hat{Y}}^2 = R^2$$ coeff of determination

$$\sigma_{\hat{Y}}^2 = \operatorname{Var}\hat{Y}_1 = \operatorname{Var}\mathbb{E}[\hat{Y}_1 \mid X_1] + \mathbb{E}\operatorname{Var}(\hat{Y}_1 \mid X_1)$$

$$= \mathrm{Var}\,\mathbb{E}[\hat{Y}_1 | X_1] + \mathbb{E}\left[(\hat{Y}_1 - \mathbb{E}[\hat{Y}_1 | X_1])^2\right]$$

$$= \mathrm{Var}\,\mathbb{E}[\hat{Y}_1 | X_1] + \mathbb{E}\,\varepsilon^2$$

$$= \mathrm{Var}\,\underbrace{\mathbb{E}[\hat{Y}_1 | X_1]}_{\hat{f}(X_1)} + \sigma_\varepsilon^2$$

How can we minimize $|B(Z_{1:n}; \hat{f})|$?

$B(Z_{1:n}; f)$ is linear in f:

$$B(Z_{1:n}; \alpha f + \beta g) = \alpha B(Z_{1:n}; f) + \beta B(Z_{1:n}; g)$$

So need to limit the magnitude of $\hat{f}$

O/W can just scale f up.

Let's consider min worst-case $|B(Z_{1:n}; \hat{f})|$ over f relative to its magnitude:

$$\mathbb{B}(Z_{1:n}; \|\cdot\|) = \sup_{\|f\| \le 1} B(Z_{1:n}; f) = \sup_{f} \frac{B(Z_{1:n}; f)}{\|f\|}$$ where $\frac{0}{0} = 0$

$\|f\| \in [0, \infty]$

a) $\|\alpha f\| = |\alpha|\,\|f\|$ b) $\|f + g\| \le \|f\| + \|g\|$

c) $|B(Z_{1:n}; f)| \le M_{Z_{1:n}} \cdot \|f\|$ for some $M_{Z_{1:n}}$

O/W $\mathbb{B}(Z_{1:n}; \|\cdot\|) = \infty$

Clearly: $|B(Z_{1:n}, \hat{f})| \le \|\hat{f}\|\, \mathbb{B}(Z_{1:n}; \|\cdot\|)$ ← $\|f\| = \infty \Rightarrow$ f that we don't guard against

So $\mathrm{Var}\,\hat{\tau} \le \|\hat{f}\|^2 \cdot \mathbb{E}\,\mathbb{B}^2(Z_{1:n}, \|\cdot\|) + \mathrm{Var\,SATE} + \frac{4}{n}\sigma_\varepsilon^2$

Some implications:

Def the equivalence $f \sim g \Leftrightarrow B(Z_{1:n}; f) = B(Z_{1:n}; g)\ \forall Z_{1:n} \in CR_n$

Define $[f] = \{g : f \sim g\} = \{f + g : B(Z_{1:n}; g) = 0\ \forall Z_{1:n} \in CR_n\}$

$$\|[f]\| = \inf_{g \in [f]} \|g\| = \inf_{\substack{g : B(Z_{1:n}; g) = 0 \\ \forall Z_{1:n} \in CR_n}} \|f + g\|$$

Then: ⓐ $\|[\cdot]\|$ is a norm for the vector space

$$\mathcal{F} = \{[f] : \|[f]\| < \infty\}$$

$$\left(\begin{aligned} \|[f]\| = 0 &\Rightarrow \exists g \sim f \text{ s.t. } \|g\| = 0 \\ &\Rightarrow B(Z_{1:n}; f) = B(Z_{1:n}; g) = 0 = B(Z_{1:n}; 0) \\ &\Rightarrow f \sim 0 \Rightarrow [f] = [0] \end{aligned} \right)$$

ⓑ $B(Z_{1:n}; [f])$ well defined

ⓒ $B(Z_{1:n}; [f])$ is continuous in $[f]$

$$\begin{aligned} &|B(Z_{1:n}; [f]) - B(Z_{1:n}; [g])| \\ &\quad = |B(Z_{1:n}; [f-g])| \le M_{Z_{1:n}} \|[f-g]\| \\ &\quad = M_{Z_{1:n}} \|[f]-[g]\| \end{aligned}$$

Given a normed vector space $(V, \|\cdot\|)$ the dual space $(V^*, \|\cdot\|_*)$ is

$$\begin{aligned} V^* &= \text{all bounded linear operators} \\ &= \left\{ A : V \to \mathbb{R} : \begin{array}{l} A(\alpha v + \beta u) = \alpha A v + \beta A u \\ \exists M : |Av| \le M \|v\| \end{array} \right\} \\ &= \text{all continuous linear operators} \end{aligned}$$

$$\|A\|_* = \sup_{\|v\| \le 1} Av$$

(Note: if A is linear & not bounded then $\sup_{\|v\| \le 1} Av = \infty$)

$$\mathbb{B}(Z_{1:n}; \|\cdot\|) = \|B(Z_{1:n}; \cdot)\|_*$$

Def: Given $\|\cdot\|$, then the pure-strategy optimal design (PSOD) chooses unif at random from

$$\operatorname*{argmin}_{Z_{1:n} \in CR_n} \mathbb{B}(Z_{1:n} ; \|\cdot\|)$$

(Also possible mixed-strategy designs)

Block-randomized design:

f is piecewise constant on strata

$$\text{Let } \|f\|_{L_\infty(c)} = \begin{cases} \sup_{x \in X} f(x) & \overbrace{|f(c^{-1}(j))| = 1 \ \forall j} \\ \infty & o/w \end{cases}$$

Thm: The PSOD wrt is the block randomized design.

Pf: $\mathbb{B}(Z_{1:n} ; \|\cdot\|_{L_\infty(c)})$

$$= \frac{2}{n} \sup_{\substack{v \in \mathbb{R}^J \\ \|v\|_\infty \le 1}} \sum_i u_i \sum_j \mathbb{I}[c(x_i) = j] v_j$$

$$= \frac{2}{n} \sup_{\|v\|_\infty \le 1} v^T \Big(\sum_i u_i e_{c(x_i)} \Big)$$

$$= \frac{2}{n} \Big\| \sum_i u_i e_{c(x_i)} \Big\|_1$$

$$= \frac{2}{n} \sum_j \Big| \sum_i u_i \mathbb{I}[c(x_i) = j] \Big|$$

$$= \frac{2}{n} \sum_j \Big| \sum_{i : c(x_i) = j} u_i \Big|$$

$\delta(x, x') = 0 \quad x \neq x' \qquad \delta$ can be ∞

Thm: Fix a pseudo-metric δ on X

Let $\|f\|_{lip(\delta)} = \sup_{x \neq x'} \frac{f(x) - f(x')}{\delta(x, x')}$

is it a norm?

$\|f\| = 0 \Rightarrow ? \ f = 0$

need to quotient

Then the PSOD is opt pair matching

Pf: Let $D_{ij} = \delta(X_i, X_j)$

$$\mathbb{B}(Z_{1:n}; \|\cdot\|_{\mathrm{Lip}(\delta)}) = \frac{2}{n} \sup_{v : v_i - v_j \le D_{ij} \ \forall ij} u^T v$$

Need to show: set of optimal solns u to $\min_u$)
is equal to the set of
± 1 assignments to pairs in a
min-weight nonbipartite match

Let $M = \{\{i_1, j_1\}, \dots, \{i_{n/2}, j_{n/2}\}\}$
be a matching (aka indep edge set)