Inference with Covariate-Adaptive Randomization

by

Azeem M. Shaikh, University of Chicago

June 2017

amshaikh@uchicago.edu

Collaborators:
Federico Bugni, Duke University
Ivan Canay, Northwestern University

Introduction

Covariate-Adaptive Randomization

– In rand. controlled trials, want to avoid "imbalance" in baseline cov.

– First stratify according to baseline covariates ...

... then assign treatment status to achieve "balance" within strata.

– Stratified block randomization popular example, but many others.

Implications for Inference about Avg. Treatment Effect

– Leads to (strong) dependence *within* treatment status ...

... and *between* treatment status and baseline covariates.

– Methods for i.i.d. data (e.g., two-sample t-test) remain valid ...

... but may be very conservative.

– Easy-to-implement methods that are exact (and hence more powerful).

Outline of Talk

- Setup and Notation
- Assumptions
- Two-Sample t-Test
    - Generally conservative
    - Simple adj. leads to exact test
- t-Test with Strata Fixed Effects
    - Except in one important special case, generally conservative
    - Simple adj. leads to exact test
- Covariate-Adaptive Permutation Test
    - Important to Studentize correctly
- Simulation Study
- Extensions
- Conclusion

Setup and Notation

Observed data for ith unit:

$$\begin{aligned} Y_i &= \text{observed outcome of interest for } i\text{th unit} \\ A_i &= \text{indicator for whether } i\text{th unit is treated} \\ Z_i &= \text{observed, baseline covariates for } i\text{th unit} \end{aligned}$$

Stratum for ith unit given by $S_i = S(Z_i)$, where

$$S : \text{supp}(Z_i) \to \mathcal{S} \text{ with } \mathcal{S} \text{ a finite set}$$

As usual,

$$Y_i = Y_i(1)A_i + Y_i(0)(1 - A_i) \ ,$$

where

$$\begin{aligned} Y_i(1) &= \text{potential outcome for } i\text{th unit if treated} \\ Y_i(0) &= \text{potential outcome for } i\text{th unit if not treated} \end{aligned}$$

Setup and Notation (cont.)

Useful shorthand:

$$\begin{aligned}
Y^{(n)} &= (Y_i : 1 \le i \le n) \\
A^{(n)} &= (A_i : 1 \le i \le n) \\
Z^{(n)} &= (Z_i : 1 \le i \le n) \\
X^{(n)} &= ((Y_i, A_i, Z_i) : 1 \le i \le n)
\end{aligned}$$

Define $S^{(n)}$, $Y^{(n)}(1)$ and $Y^{(n)}(0)$ analogously.

Interested in testing

$$H_0 : E[Y_i(1) - Y_i(0)] = 0 \text{ vs. } H_1 : E[Y_i(1) - Y_i(0)] \neq 0$$

at level $\alpha \in (0, 1)$.

Assumptions

Assumptions stated in terms of following quantity:

For $s \in \mathcal{S}$, define

$$D_n(s) = n_1(s) - \pi n(s) ,$$

where

$$n_a(s) = |\{1 \le i \le n : A_i = a, S_i = s\}|$$
$$n(s) = |\{1 \le i \le n : S_i = s\}| ,$$

and $\pi \in (0,1)$ is target proportion of units to assign to treatment.

i.e., $D_n(s)$ is measure of imbalance in stratum s relative to π.

Remark 1 Results will vary according to whether $\pi = \frac{1}{2}$ or $\pi \neq \frac{1}{2}$.

Assumptions (cont.)

Assumption 1

(a) $(Y_i(1), Y_i(0), Z_i), i = 1, \ldots, n$ are i.i.d.

(b) $p(s) = P\{S_i = s\} > 0$ for all $s \in \mathcal{S}$

(c) $E[|Y_i(1)|^2] < \infty$ and $E[|Y_i(0)|^2] < \infty$

(d) $\text{Var}[Y_i(a)|S_i = s] > 0$ for some $s \in \mathcal{S}$ and $a \in \{0, 1\}$

Remark 2 Importantly, do *not* assume that $A_i, i = 1, \ldots, n$ are i.i.d.

Assumptions (cont.)

Assumption 2

(a) $(Y^{(n)}(1), Y^{(n)}(0), Z^{(n)}) \perp\!\!\!\perp A^{(n)} | S^{(n)}$

(b) $\left(\frac{D_n(s)}{\sqrt{n}} : s \in \mathcal{S} \right) \Big| S^{(n)} \xrightarrow{d} N(0, \Sigma_D)$ a.s., where

$$\Sigma_D = \text{diag}(p(s)\tau(s) : s \in \mathcal{S}) \text{ with } 0 \leq \tau(s) \leq \pi(1 - \pi)$$

Remark 3 Assumption 2(b) requires

$$\frac{n_1(s)}{n(s)} \xrightarrow{P} \pi \ .$$

In particular, π does not depend on s.

Discuss implications of relaxing Assumption 2(b) later.

Remark 4 May interpret $\tau(s)$ as measure of degree of imbalance.

Assumptions (cont.)

Assumption 2 satisfied for many assignment mechanisms:

Example 1 *Simple Random Sampling*

Independently of $S^{(n)}$,

$$A^{(n)} \sim \text{Binomial}\,(n, \pi) \ .$$

Can verify Assumption 2 holds with $\tau(s) = \pi(1-\pi)$ for all $s \in \mathcal{S}$.

Assumptions (cont.)

Example 2 *Stratified Block Randomization*

In stratum s, assign

$$m(s) = \lfloor \pi n(s) \rfloor$$

units (uniformly) at random to treatment and remainder to control.

Can show Assumption 2 holds with $\tau(s) = 0$ for all $s \in \mathcal{S}$.

In fact, $|D_n(s)| \leq 1$ for all $s \in \mathcal{S}$.

Commonly used design in development economics.

See, e.g., Duflo et al. (2014), Callen et al. (2015), Dizon-Ross (2015).

Two-Sample t-Test

For $a \in \{0, 1\}$, define

$$\bar{Y}_{n,a} = \frac{1}{n_a} \sum_{1 \leq i \leq n : A_i = a} Y_i \quad \text{and} \quad \hat{\sigma}^2_{n,a} = \frac{1}{n_a} \sum_{1 \leq i \leq n : A_i = a} (Y_i - \bar{Y}_{n,a})^2 ,$$

where $n_a = |\{1 \leq i \leq n : A_i = a\}|$.

Consider

$$\phi_n^{t-\text{test}}(X^{(n)}) = I\{T_n^{t-\text{test}}(X^{(n)}) > z_{1-\frac{\alpha}{2}}\} ,$$

where $z_{1-\frac{\alpha}{2}} = \Phi^{-1}(1 - \frac{\alpha}{2})$ and

$$T_n^{t-\text{test}}(X^{(n)}) = \frac{|\bar{Y}_{n,1} - \bar{Y}_{n,0}|}{\sqrt{\frac{\hat{\sigma}^2_{n,1}}{n_1} + \frac{\hat{\sigma}^2_{n,0}}{n_0}}} .$$

For testing H_0, $\phi_n^{t-\text{test}}(X^{(n)})$ asymp. level α under simple rand. samp.

Is this true under covariate-adaptive rand. more generally?

Two-Sample t-Test (cont.)

Theorem 1 *If Assumptions 1 and 2 hold, then*

$$\lim_{n \to \infty} E[\phi_n^{t-\text{test}}(X^{(n)})] = P\{\sigma_{t-\text{test}}|Z| > z_{1-\frac{\alpha}{2}}\} \leq \alpha \ ,$$

where $Z \sim N(0,1)$*, whenever* $E[Y_i(1) - Y_i(0)] = 0$*.*

Furthermore, inequality strict unless

$$\Big(\tau(s) - \pi(1-\pi)\Big)\Big((E[Y_i(1)|S_i = s] - E[Y_i(1)]) + (E[Y_i(0)|S_i = s] - E[Y_i(0)])\Big)^2 = 0 \text{ for all } s \in \mathcal{S} \ .$$

Two-Sample t-Test (cont.)

Remark 5 Covariate-adaptive rand. $\Longrightarrow$ restricted variation in A_i.

Not accounted for in denom. of usual t-statistic $\Longrightarrow \sigma^2_{t-\text{test}} \leq 1$.

Remark 6 Inequality is equality when, e.g.,

- $\tau(s) = \pi(1-\pi)$ for all $s \in S$

  (as in simple random sampling)

- $E[Y_i(a)|S_i] = E[Y_i(a)]$ for all $a \in \{0,1\}$

  (e.g., no stratification or when stratification irrelevant)

Two-Sample t-Test (cont.)

In proof of Theorem 1, we show that

$$\sqrt{n}(\bar{Y}_{n,1} - \bar{Y}_{n,0} - E[Y_i(1) - Y_i(0)]) \xrightarrow{d} N(0, \varsigma_1^2 + \varsigma_2^2 + \varsigma_3^2) \ ,$$

where

$$\begin{aligned}
\varsigma_1^2 &= \frac{\text{Var}[Y_i(1) - E[Y_i(1)|S_i]]}{\pi} + \frac{\text{Var}[Y_i(0) - E[Y_i(0)|S_i]]}{1-\pi} \\
\varsigma_2^2 &= E\left[\Big((E[Y_i(1)|S_i] - E[Y_i(1)]) - (E[Y_i(0)|S_i] - E[Y_i(0)])\Big)^2\right] \\
\varsigma_3^2 &= E\left[\tau(S_i)\left(\frac{E[Y_i(1)|S_i] - E[Y_i(1)]}{\pi} + \frac{E[Y_i(0)|S_i] - E[Y_i(0)]}{1-\pi}\right)^2\right] \ .
\end{aligned}$$

Two-Sample t-Test (cont.)

Let $\hat{\varsigma}_1^2$, $\hat{\varsigma}_2^2$ and $\hat{\varsigma}_3^2$ be "natural" estimators of ς_1^2, ς_2^2 and ς_3^2.

Consider "adjusted" test given by

$$\phi_n^{t-\text{test,adj}}(X^{(n)}) = I\{T_n^{t-\text{test,adj}}(X^{(n)}) > z_{1-\frac{\alpha}{2}}\} ,$$

where

$$T_n^{t-\text{test,adj}}(X^{(n)}) = \frac{|\bar{Y}_{n,1} - \bar{Y}_{n,0}|}{\sqrt{\frac{\hat{\varsigma}_1^2+\hat{\varsigma}_2^2+\hat{\varsigma}_3^2}{n}}} .$$

Theorem 2 *If Assumptions 1 and 2 hold, then,*

$$\lim_{n\to\infty} E[\phi_n^{t-\text{test,adj}}(X^{(n)})] = \alpha$$

whenever $E[Y_i(1) - Y_i(0)] = 0$.

t-Test with Strata Fixed Effects

Consider estimation of

$$Y_i = \beta A_i + \sum_{s \in \mathcal{S}} \delta_s I\{S_i = s\} + \epsilon_i$$

by ordinary least squares.

Let $\hat{\beta}_n$ be resulting estimator of β.

Consider

$$\phi_n^{\text{sfe}}(X^{(n)}) = I\{T_n^{\text{sfe}}(X^{(n)}) > z_{1-\frac{\alpha}{2}}\} \ ,$$

where

$$T_n^{\text{sfe}}(X^{(n)}) = \frac{|\hat{\beta}_n|}{\sqrt{\frac{\hat{V}_n}{n}}}$$

and $\hat{V}_n =$ usual heteroskedasticity-consistent estimator of variance.

t-Test with Strata Fixed Effects (cont.)

Theorem 3 *If Assumptions 1 and 2 hold, then*

$$\lim_{n\to\infty} E[\phi_n^{\text{sfe}}(X^{(n)})] = P\{\sigma_{\text{sfe}}|Z| > z_{1-\frac{\alpha}{2}}\} \leq \alpha \ ,$$

where $Z \sim N(0,1)$, *whenever* $E[Y_i(1) - Y_i(0)] = 0$.

Furthermore, inequality strict unless

$$\Big(1 - 2\pi\Big)\Big(\tau(s) - \pi(1-\pi)\Big)\Big((E[Y_i(1)|S_i = s] - E[Y_i(1)]) + (E[Y_i(0)|S_i = s] - E[Y_i(0)])\Big)^2 = 0 \text{ for all } s \in \mathcal{S} \ .$$

t-Test with Strata Fixed Effects (cont.)

Remark 7 Intuition for result same as before.

Remark 8 Inequality is equality in same situations as before,

i.e., simple random sampling and irrelevant stratification,

but additionally when $\pi = \frac{1}{2}$.

t-Test with Strata Fixed Effects (cont.)

In proof of Theorem 3, we show that

$$\sqrt{n}(\hat{\beta}_n - E[Y_i(1) - Y_i(0)]) \xrightarrow{d} N(0, \varsigma_1^2 + \varsigma_2^2 + \varsigma_4^2) ,$$

where ς_1^2 and ς_2^2 are defined as before and

$$\varsigma_4^2 = \frac{(1-2\pi)^2}{\pi^2(1-\pi)^2} E\Bigg[\tau(S_i)\Big((E[Y_i(1)|S_i] - E[Y_i(1)]) - (E[Y_i(0)|S_i] - E[Y_i(0)])\Big)^2\Bigg] .$$

t-Test with Strata Fixed Effects (cont.)

Let $\hat{\varsigma}_1^2$, $\hat{\varsigma}_2^2$ and $\hat{\varsigma}_4^2$ be "natural" estimators of ς_1^2, ς_2^2 and ς_4^2.

Consider "adjusted" test given by

$$\phi_n^{\text{sfe,adj}}(X^{(n)}) = I\{T_n^{\text{sfe,adj}}(X^{(n)}) > z_{1-\frac{\alpha}{2}}\} ,$$

where

$$T_n^{\text{sfe,adj}}(X^{(n)}) = \frac{|\hat{\beta}_n|}{\sqrt{\frac{\hat{\varsigma}_1^2+\hat{\varsigma}_2^2+\hat{\varsigma}_4^2}{n}}}$$

Theorem 4 *If Assumptions 1 and 2 hold, then,*

$$\lim_{n\to\infty} E[\phi_n^{\text{sfe,adj}}(X^{(n)})] = \alpha$$

whenever $E[Y_i(1) - Y_i(0)] = 0$.

t-Test with Strata Fixed Effects (cont.)

Remark 9 In general, ς_3^2 and ς_4^2 unordered, but

- $\varsigma_3^2 = 0$ when $\tau(s) = 0$ for all $s \in \mathcal{S}$
- $\varsigma_4^2 = 0$ when $\pi = \frac{1}{2}$ or $\tau(s) = 0$ for all $s \in \mathcal{S}$

Covariate-Adaptive Permutation Test

Consider a restricted permutation test. Let

$$\mathbf{G}_n(S^{(n)}) = \{g \in \mathbf{G}_n : S_{g(i)} = S_i \text{ for all } 1 \le i \le n\} ,$$

where $\mathbf{G}_n$ = group of permutations of $\{1, \ldots, n\}$.

i.e., subgroup that only permutes *within* strata.

Define action of $g \in \mathbf{G}_n$ on $X^{(n)}$ by

$$gX^{(n)} = ((Y_i, A_{g(i)}, S_i) : 1 \le i \le n) .$$

For $T_n(X^{(n)}) = T_n^{t-\text{test,adj}}(X^{(n)})$, consider

$$\phi_n^{\text{cap}}(X^{(n)}) = I\{T_n(X^{(n)}) > \hat{c}_n^{\text{cap}}(1-\alpha)\} ,$$

where $\hat{c}_n^{\text{cap}}(1-\alpha) =$

$$\inf\left\{x \in \mathbf{R} : \frac{1}{|\mathbf{G}_n(S^{(n)})|} \sum_{g \in \mathbf{G}_n(S^{(n)})} I\{T_n(gX^{(n)}) \le x\} \ge 1-\alpha\right\} .$$

Covariate-Adaptive Permutation Test (cont.)

Theorem 5 *If Assumptions 1 and 2 hold with* $\tau(s) = 0$ *for all* $s \in \mathcal{S}$*, then,*

$$\lim_{n \to \infty} E[\phi_n^{cap}(X^{(n)})] = \alpha$$

whenever $E[Y_i(1) - Y_i(0)] = 0$.

Covariate-Adaptive Permutation Test (cont.)

Remark 10 Replacing $\mathbf{G}_n(S^{(n)})$ with $\mathbf{G}_n$ leads to classical perm. test.

Remark 11 Choice of test statistic important!

Using $T_n^{t-\text{test}}(X^{(n)})$ in place of $T_n^{t-\text{test,adj}}(X^{(n)})$ leads to test that is

– exact when $\pi = \frac{1}{2}$

– invalid when $\pi \neq \frac{1}{2}$.

Qualitatively similar to classical results about permutation tests.

Remark 12 Similar results hold for t-test with strata fixed effecsts.

Covariate-Adaptive Permutation Test (cont.)

Remark 13 Note that $D_n(s)$ is invariant w.r.t. $g \in \mathbf{G}_n(S^{(n)})$.

Thus, not surprising Theorem 5 requires

$$\tau(s) = 0 \text{ for all } s \in \mathcal{S} \ .$$

Holds, e.g., for Efron's biased coin design and stratified block rand.

Covariate-Adaptive Permutation Test (cont.)

Remark 14 For testing

$$H_0' : Y_i(1)|S_i \stackrel{d}{=} Y_i(0)|S_i \ ,$$

$\phi_n^{\text{cap}}(X^{(n)})$ level α whenever

$$gA^{(n)}|S^{(n)} \stackrel{d}{=} A^{(n)}|S^{(n)}$$

for $g \in \mathbf{G}_n(S^{(n)})$, which holds, e.g., under stratified block rand.

Follows closely arguments from classical theory of randomization tests.

See, e.g., Heckman et al. (2011) and Rosenberger & Lachin (2016).

In comparison, we use $\phi_n^{\text{cap}}(X^{(n)})$ for testing

$$H_0 : E[Y_i(1) - Y_i(0)] = 0 \ .$$

Simulation Study

For $i = 1, \ldots, n$ with $n = 200$,

$$Y_i(a) = \mu_a + m_a(Z_i) + \sigma_a(Z_i)\epsilon_{a,i} \ .$$

Model 1 $Z_i \sim$ "standardized" Beta$(2,2)$, $\epsilon_{a,i} \sim N(0,1)$,

$m_a(Z_i) = \gamma Z_i$, $\sigma_0(Z_i) = 1$, and $\sigma_1(Z_i) = \sigma_1$.

Model 2 As in Model 1, but $m_0(Z_i) = -\gamma \log(Z_i + 3) I\{Z_i \leq \frac{1}{2}\}$.

Model 3 $Z_i \sim$ Unif$(-2,2)$, $\epsilon_{a,i} \sim \frac{1}{3} t_3$,

$m_a(Z_i) = \gamma Z_i^2 I\{Z_i \in [-1,1]\} + \gamma(2 - Z_i^2) I\{Z_i \notin [-1,1]\}$,

$\sigma_0(Z_i) = Z_i^2$, and $\sigma_1(Z_i) = \sigma_1 Z_i^2$.

Model 4 As in Model 3, but

$m_0(Z_i) = \gamma Z_i^2 I\{Z_i \in [-1,1]\} + \gamma Z_i I\{Z_i \notin [-1,1]\}$

$m_1(Z_i) = \gamma Z_i I\{Z_i \in [-1,1]\} + \gamma Z_i^2 I\{Z_i \notin [-1,1]\}$

Simulation Study (cont.)

Strata S_i determined by dividing supp(Z_i) into four intervals of eq. length.

Consider $\gamma = 2$, $\sigma_1 = 1$, and $\alpha = .05$ with both $\pi = \frac{1}{2}$ and $\pi = \frac{7}{10}$.

Treatment assigned according to

SRS = simple random sampling

SBR = stratified block randomization

(Other parameters and treatment assignment mechanisms in paper.)

Seven tests:

***t*-Test** = usual/"adj." two-sample t-test

Reg = linear reg. of Y_i on A_i and Z_i (when $\pi = \frac{1}{2}$ only)

SFE = usual/"adj." t-test with strata fixed effects

CAP = cov.-adaptive perm. test using usual/"adj." two-sample t-stat.

Simulation Study (cont.)

$\pi = \frac{1}{2}$: Rej. prob. (%) of H_0 under $E[Y_i(1) - Y_i(0)] = 0$

		t-test	Reg	SFE	CAP
Model 1	SRS	5.58/5.29	5.18	5.08/5.49	5.19/5.20
	SBR	0.02/5.45	4.81	4.86/5.40	4.77/4.78
Model 2	SRS	5.50/5.31	5.30	4.89/5.84	5.26/5.20
	SBR	1.49/5.46	4.60	4.78/5.34	4.73/4.84
Model 3	SRS	5.25/5.25	5.08	4.92/5.93	4.98/4.97
	SBR	3.87/5.51	3.73	5.07/5.50	5.24/5.23
Model 4	SRS	5.53/5.41	5.51	5.26/5.80	5.36/5.36
	SBR	1.81/5.51	3.18	5.11/5.08	4.96/4.99

Simulation Study (cont.)

$\pi = \frac{1}{2}$: Rej. prob. (%) of H_0 under $E[Y_i(1) - Y_i(0)] = \frac{1}{2}$

		t-test	Reg	SFE	CAP
Model 1	SRS	36.00/35.98	93.95	85.04/85.95	60.83/60.91
	SBR	24.68/86.09	93.96	85.42/86.12	84.50/84.54
Model 2	SRS	48.33/48.50	69.22	65.96/68.00	54.90/55.12
	SBR	46.94/67.31	68.71	66.20/68.09	64.99/65.53
Model 3	SRS	52.67/52.82	51.96	57.43/59.41	53.14/53.25
	SBR	53.42/59.98	52.89	58.73/59.13	58.43/58.42
Model 4	SRS	29.69/29.54	32.70	38.52/41.67	34.04/34.02
	SBR	25.52/42.40	32.53	40.90/41.47	40.49/40.56

Simulation Study (cont.)

$\pi = \frac{7}{10}$: Rej. prob. (%) of H_0 under $E[Y_i(1) - Y_i(0)] = 0$

		t-test	SFE	CAP
Model 1	SRS	5.43/4.94	5.30/6.35	4.85/4.82
	SBR	0.03/5.56	5.20/5.32	4.90/5.07
Model 2	SRS	5.00/4.92	5.12/5.98	6.94/7.06
	SBR	2.46/5.57	3.34/5.49	10.25/4.98
Model 3	SRS	5.45/5.54	4.80/5.61	4.89/4.89
	SBR	3.83/5.64	5.21/5.64	5.00/5.06
Model 4	SRS	5.51/5.07	5.09/6.07	7.01/6.79
	SBR	1.20/5.34	1.65/5.30	5.08/4.63

Simulation Study (cont.)

$\pi = \frac{7}{10}$: Rej. prob. (%) of H_0 under $E[Y_i(1) - Y_i(0)] = \frac{1}{2}$

		t-test	SFE	CAP
Model 1	SRS	31.30/31.38	77.61/78.85	54.73/54.97
	SBR	17.38/79.02	79.13/79.63	77.58/77.42
Model 2	SRS	50.14/50.38	53.65/55.73	58.53/58.91
	SBR	49.55/62.61	54.76/63.31	73.11/60.62
Model 3	SRS	45.25/45.59	50.67/52.90	45.78/45.93
	SBR	46.13/52.69	51.67/52.93	51.09/50.85
Model 4	SRS	27.70/27.73	27.03/28.51	33.11/33.18
	SBR	22.66/43.12	24.98/43.09	41.99/40.52

Extensions

What if Assumption 2(b) is not satisfied, i.e.,

$$\frac{n_1(s)}{n(s)} \xrightarrow{P} \pi_s \neq \pi \ ?$$

Subject of Bugni, Canay & Shaikh (2017).

Of interest, e.g., in Project STAR.

Then,

- None of previously described methods valid.
- Indeed, diff. in means not even consistent for avg. treatment effect.
- Can be consistently estimated with a fully saturated linear reg.
- Exact inference, however, requires "adjusting" std. errors!

Also develops results for commonly encountered multiple treatments.

Conclusion

Covariate-adaptive rand. has important implications for inference!

Methods for i.i.d. data (e.g., two-sample t-test) remain valid ...

... but may be conservative.

Easy to "adjust" tests so that they are exact (and thus more powerful).

Covariate-adaptive perm. test also exact in many cases ...

... when correctly implemented!

Covariate-adaptive perm. test also has some finite-sample validity.

Conclusion (cont.)

Natural questions:

- Group-level treatment
- Non-compliance
- Multiple testing
- What if number of strata large relative to sample size?

  (relevant, e.g., when size of strata is "small")

$\vdots$

Subject of on going work with ...

... V. Kamat, J. P. Romano & M. Tabord-Meehan.