Randomization Inference

Kosuke Imai

Harvard University

STAT186/GOV2002 CAUSAL INFERENCE

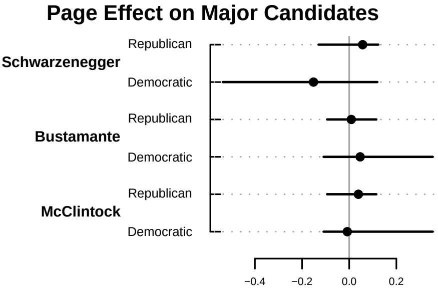
Fall 2018

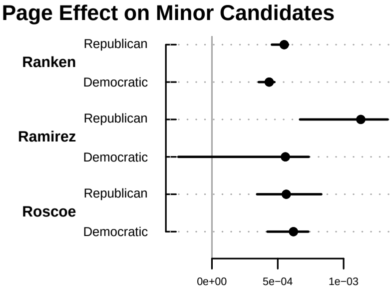
Randomization-based Confidence Intervals

- Permutation tests are useful but tell us nothing about effect size
- Invert a α-level test generates $(1-\alpha) \times 100\%$ confidence set
- Inverting a permutation test:
  1. Consider the constant additive effect model
  $$Y_i(1) - Y_i(0) = \tau_0 \quad \text{for all} \quad i$$
  2. Collect all null values τ_0 that cannot be rejected by α-level test
  $$A_\alpha \ = \ \{\tau_0 : \Pr(f(\{Y_i, T_i^{obs}\}_{i=1}^n, T_i^{obs}, \tau_0) \leq f(\{Y_i, T_i^{obs}\}_{i=1}^n, T_i, \tau_0)) \geq \alpha\}$$
  where A_α is the $(1-\alpha) \times 100\%$ confidence set

Revising the California Alphabet Lottery

- For each candidate, we test $H_0 : Y_i(1) - Y_i(0) = \tau_0$ for a range of τ_0 values using the permutation test at the 0.05 level
- We collect all τ_0 values where we cannot reject the null





Heterogenous Treatment Effects

- Most randomization inference assumes a homogeneous treatment effect, e.g., constant additive effect model
- In some cases, we can allow for heterogenous treatment effects under an alternative assumption

- Rank-sum test for continuous outcome with no ties
  - The reference distribution does not depend on unit index:

$$\Pr(\text{each set of ranks}) = \frac{1}{\binom{n}{n_1}}$$

  - Non-sharp null hypothesis:

$$H_0 : \Pr(Y(1) \leq y) = \Pr(Y(0) \leq y) \text{ for all } y$$

  - Population inference is identical to sample inference

Point Estimation under the Population Shift Model

- The model:

$$H_0 : \Pr(Y(0) \leq y) \; = \; \Pr(Y(1) - \tau_0 \leq y)$$

- Inverting the Wilcoxon's rank-sum test $\rightsquigarrow$ a confidence interval
- Hodges and Lehmann estimator (Hodges & Lehmann. 1963. *Ann. Math. Stat.*)
  - Idea: choose $\hat{\tau}$ to remove the treatment effect treatment effect s.t.

$$\Pr(Y_i - \hat{\tau} \leq y \mid T_i = 1) = \Pr(Y_i \leq y \mid T_i = 0) \quad \text{for all } y$$

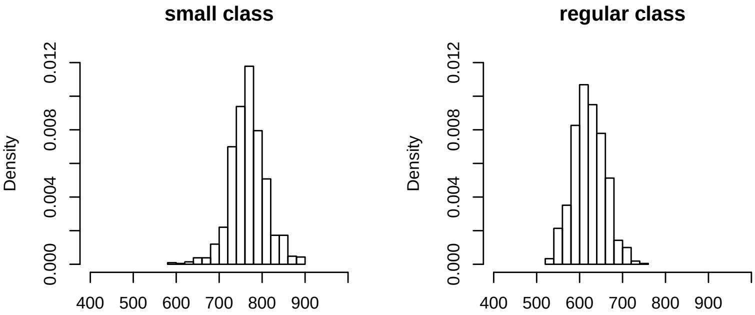
  - Wilcoxon's rank-sum test: $S_\tau = \sum_{i=1}^n T_i R_i(Y_i - \tau T_i)$
  - Estimator:

$$S_{\hat{\tau}} \; = \; \frac{n_1(n+1)}{2}$$

  - Computation: S_τ is non-increasing in τ
  - No exact solution $\rightsquigarrow$ take the average of the smallest τ that is too large and the largest τ that is too small

The Project STAR (Mosteller. 1997. *Bull. Am. Acad. Arts Sci.*)

- The Student-Teacher Achievement Ratio Project (1985–1989)
  - More than 10,000 students involved with the cost of $12 million
  - Effects of class size in early grade levels
  - 3 arms: Small class, Regular-sized class, Regular class with aid
- Effect of kindergraden class size on 8th grade reading score:



- Wilcoxon's rank-sum test (there are some ties):
  p-value < 0.001, 95% conf. int. $= [72.3, 80.0]$, est. $= 77.0$

Causal Heterogeneity for Binary Outcome

- Back to Fisher's exact test for 2×2 table
- Long-term impact of class size:

	Small class	Regular-sized class
Graduate	754	892
Not graduate	148	189
Total	902	1081

- Can't assume a constant additive model for binary outcome
- Monotonicity assumption: $Y_i(1) \geq Y_i(0)$ for all i
- Attributable effects $=$ number of successes in the treatment group caused by the treatment:

$$A = \sum_{i=1}^{n} T_i(Y_i(1) - Y_i(0))$$

Attributable Effects (Rosenbaum. 2001. *Biometrika*)

- Attributable effects as a pivot:

$$S - A = \sum_{i=1}^{n} T_i Y_i(0), \text{ where } S = \sum_{i=1}^{n} T_i Y_i(1)$$

- Adjusted 2×2 table:

	Treated ($T = 1$)	Control ($T = 0$)
Success ($Y = 1$)	$\sum_{i=1}^{n} T_i Y_i(0)$	$\sum_{i=1}^{n}(1 - T_i) Y_i(0)$
Failure ($Y = 0$)	$\sum_{i=1}^{n} T_i(1 - Y_i(0))$	$\sum_{i=1}^{n}(1 - T_i)(1 - Y_i(0))$
Total	n_1	n_0

- The reference distribution $\rightsquigarrow$ hypergeometric
- Allows for heterogeneity, Inversion gives a confidence interval
- STAR example: p-value = 0.55, 95% conf. int. $= [0, 41]$

Testing Spillover Effects (Aronow. 2012. *Sociol. Methods Res.*)

- So far, we have assumed no spillover effect
- Sharp null hypothesis of no effect implies no spillover effect as well as no effect of one's own treatment
- Choose a set of *focal units* $\mathcal{F}$
- Perform a permutation test using *conditional* randomization distribution given the treatment assignments of focal units Ω
- Sharp null hypothesis of no spillover effect among focal units:

$$H_0 : Y_i(T_i, \mathbf{t}_{-i}) = Y_i(T_i, \mathbf{t}'_{-i}) \quad \text{where } i \in \mathcal{F}, \text{ and } \mathbf{t}_{-i} \in \Omega$$

- Any dependence between the outcomes of forcal units and the treatment assignments of others is evidence of spillover effects
- Many choices: focal units, test statistics
- Extensions (Athey et al. 2017. *J. Amer. Stat. Assoc.*)

Spillover of Voter Mobilization (Nickerson. 2008. *Am. Political Sci. Rev.*)

- Randomized GOTV experiment in 2002 Congressional Primary (Denver and Minneapolis)
- Do interpersonal relationships affect turnout decision?
- Observational studies $\rightsquigarrow$ selection bias
- Placebo controlled design with two voter households:

	GOTV	Recycling	Control
Number of households	1285	1289	1286
Contacted	37.7%	36.5%	0%
Turnout	34.0%	31.7%	31.2%
contacted	39.3%	29.8%	
cohabitant of contacted	34.7%	28.9%	
noncontacted households	32.1%	33.1%	31.2%

- Focal units $=$ cohabitants of contacted voters
- Fisher's exact test among contacted households: p-value $= 0.061$

Summary

- Inference about effect size, going beyond hypothesis test
- Inverting a permutation test gives a randomization-based confidence interval $\rightsquigarrow$ no asymptotic approximation

- Overcoming the limitations of randomization inference
  1. heterogenous treatment effects
  2. spillover effects
- They often require additional assumptions