Sensitivity Analysis to Observed Confounding

January 31, 2013

Overview and Goals

- Review permutation inference and sensitivity analysis
- Extend permutation framework to facilitate the design and evaluation of observational research *in general* (i.e., design, pre-analysis, pre-matching, post-matching stages)
  1. Bound inferences due to imbalances on X, before or after conditioning
  2. Develop sensitivity analysis as a bias diagnostic in the data
  3. Improve interpretation of sensitivity analyses testing for unobserved confounding, on the scale of confounding on X

Model of an Observational Study

- For M units, with observed covariates $\mathbf{x}$, number the M units $j = 1, \ldots, M$, so $\mathbf{x}_{[j]}$ and $Z_{[j]}$ is the covariate and the treatment assignment for the jth unit.

Model of an Observational Study

- For M units, with observed covariates $\mathbf{x}$, number the M units $j = 1, \ldots, M$, so $\mathbf{x}_{[j]}$ and $Z_{[j]}$ is the covariate and the treatment assignment for the jth unit.
- Unit j is assigned to treatment with probability

$$\pi_j = \Pr(Z_{[j]} = 1)$$

and to control with probability

$$1 - \pi_j = \Pr(Z_{[j]} = 0)$$

Model of an Observational Study

- For M units, with observed covariates $\mathbf{x}$, number the M units $j = 1, \dots, M$, so $\mathbf{x}_{[j]}$ and $Z_{[j]}$ is the covariate and the treatment assignment for the jth unit.
- Unit j is assigned to treatment with probability

$$\pi_j = \Pr(Z_{[j]} = 1)$$

and to control with probability

$$1 - \pi_j = \Pr(Z_{[j]} = 0)$$

- Treatments are assigned by flipping biased coins (each unit might have a different biased coin):

$$\Pr(Z = z) = \prod_{j=1}^{M} \pi_{[j]}^{z} \{1 - \pi_{[j]}\}^{1 - z_j}$$

Permutation Inference

- Divide the M units into S strata, which are formed on the basis of pre-treatment characteristics.

Permutation Inference

- Divide the M units into S strata, which are formed on the basis of pre-treatment characteristics.
- There are m_s units in stratum s for $s = 1, \ldots, S$, so $M = \sum m_s$.

Permutation Inference

- Divide the M units into S strata, which are formed on the basis of pre-treatment characteristics.
- There are m_s units in stratum s for $s = 1, ..., S$, so $M = \sum m_s$.
- Write $Z_{si} = 1$ if the ith unit in stratum s receives the treatment and write $Z_{si} = 0$ if this unit receives control.

Permutation Inference

- Divide the M units into S strata, which are formed on the basis of pre-treatment characteristics.
- There are m_s units in stratum s for $s = 1, \ldots, S$, so $M = \sum m_s$.
- Write $Z_{si} = 1$ if the ith unit in stratum s receives the treatment and write $Z_{si} = 0$ if this unit receives control.
- Write n_s for the number of treated units in stratum s, so $n_s = \sum_{i=1}^{m_s} Z_{si}$ and $0 \leq n_s \leq m_s$.

Permutation Inference

- Divide the M units into S strata, which are formed on the basis of pre-treatment characteristics.
- There are m_s units in stratum s for $s = 1, ..., S$, so $M = \sum m_s$.
- Write $Z_{si} = 1$ if the ith unit in stratum s receives the treatment and write $Z_{si} = 0$ if this unit receives control.
- Write n_s for the number of treated units in stratum s, so $n_s = \sum_{i=1}^{m_s} Z_{si}$ and $0 \leq n_s \leq m_s$.
- Let Ω be the set containing $K = \Pi_{s=1}^{s} \binom{m_s}{n_s}$ possible treatment assigments $\mathbf{z}$.

Significance Test

- With known π_j, each of these K possible assignments is given the same probability, $\text{prob}(\mathbf{Z} = \mathbf{z}) = 1/K$ for all $\mathbf{z}$ in Ω.

Significance Test

- With known π_j, each of these K possible assignments is given the same probability, $\text{prob}(\mathbf{Z} = \mathbf{z}) = 1/K$ for all $\mathbf{z}$ in Ω.
- Permute each (or a large number of) $z \in \Omega$, using the known process in π_j, collect each test statistic $t(\mathbf{Z}, \mathbf{r})$, and determine what proportion of these are greater than T from our data.

Significance Test

- With known π_j, each of these K possible assignments is given the same probability, $\text{prob}(\mathbf{Z} = \mathbf{z}) = 1/K$ for all $\mathbf{z}$ in Ω.
- Permute each (or a large number of) $z \in \Omega$, using the known process in π_j, collect each test statistic $t(\mathbf{Z}, \mathbf{r})$, and determine what proportion of these are greater than T from our data.
- What happens when:

$$\text{prob}(\mathbf{Z} = \mathbf{z}) \neq 1/K$$

Significance Test

- With known π_j, each of these K possible assignments is given the same probability, $\text{prob}(\mathbf{Z} = \mathbf{z}) = 1/K$ for all $\mathbf{z}$ in Ω.
- Permute each (or a large number of) $z \in \Omega$, using the known process in π_j, collect each test statistic $t(\mathbf{Z}, \mathbf{r})$, and determine what proportion of these are greater than T from our data.
- What happens when:

$$\text{prob}(\mathbf{Z} = \mathbf{z}) \neq 1/K$$

- Observational study with **Overt** and **Hidden** bias:

$$\text{prob}(\mathbf{Z} = \mathbf{1}) = f(\mathbf{x}, u)$$

Overt Bias

- An observational study is free of *hidden* bias if every $\pi_{[j]}$ (though unknown), only depend on the observed covariates:

$$\pi_{[j]} = \pi_{[k]} \quad \text{iff} \quad \mathbf{x}_{[j]} = \mathbf{x}_{[k]}$$

Overt Bias

- An observational study is free of *hidden* bias if every $\pi_{[j]}$ (though unknown), only depend on the observed covariates:

$$\pi_{[j]} = \pi_{[k]} \text{ iff } \mathbf{x}_{[j]} = \mathbf{x}_{[k]}$$

- Define the probability that j will be in treatment as some unknown function of $\mathbf{x}$:

$$\pi_{[j]} = \lambda(\mathbf{x}_{[j]})$$

Overt Bias

- An observational study is free of *hidden* bias if every $\pi_{[j]}$ (though unknown), only depend on the observed covariates:

$$\pi_{[j]} = \pi_{[k]} \quad \text{iff} \quad \mathbf{x}_{[j]} = \mathbf{x}_{[k]}$$

- Define the probability that j will be in treatment as some unknown function of $\mathbf{x}$:

$$\pi_{[j]} = \lambda(\mathbf{x}_{[j]})$$

- The probability of treatment assignment becomes:

$$\Pr(Z = z|\mathbf{x}) = \prod_{j=1}^{M} \lambda(\mathbf{x}_{[j]})^{z_j}\{1 - \lambda(\mathbf{x}_{[j]})\}^{1-z_j}$$

Overt Bias

- An observational study can have *overt* bias for two reasons:

Overt Bias

- An observational study can have *overt* bias for two reasons:
  1. No Overlap:
  $$\pi_{[j]} = \pi_{[k]} \text{ iff } \mathbf{x}_{[j]} = \mathbf{x}_{[k]}$$
  No units $\{j, k\}$ exist (in the study) for which $Z_k = 1 - Z_j$

Overt Bias

- An observational study can have *overt* bias for two reasons:

  1. No Overlap:
  $$\pi_{[j]} = \pi_{[k]} \text{ iff } \mathbf{x}_{[j]} = \mathbf{x}_{[k]}$$
  No units $\{j, k\}$ exist (in the study) for which $Z_k = 1 - Z_j$

  2. Imbalances on $\mathbf{x}$:
  For units $\{j, k\}$ in which $Z_k = 1 - Z_j$,
  $\mu(\mathbf{x}_{[j]}) \neq \mu(\mathbf{x}_{[k]})$

Overt Bias

- An observational study can have *overt* bias for two reasons:
  1. No Overlap:
     $$\pi_{[j]} = \pi_{[k]} \text{ iff } \mathbf{x}_{[j]} = \mathbf{x}_{[k]}$$
     No units $\{j, k\}$ exist (in the study) for which $Z_k = 1 - Z_j$
  2. Imbalances on $\mathbf{x}$:
     For units $\{j, k\}$ in which $Z_k = 1 - Z_j$,
     $\mu(\mathbf{x}_{[j]}) \neq \mu(\mathbf{x}_{[k]})$
- Balance diagnostics can evaluate the degree to which these are a problem in a design

Hidden Bias

- There is *hidden* bias if two units with the same observed covariates x have differing treatment probability:

$$\text{if } \mathbf{x}_{[j]} = \mathbf{x}_{[k]}, \text{ but } \pi_{[j]} \neq \pi_{[k]} \text{ for some } j \text{ and } k.$$

Hidden Bias

- There is *hidden* bias if two units with the same observed covariates x have differing treatment probability:

$$\text{if } \mathbf{x}_{[j]} = \mathbf{x}_{[k]}, \text{ but } \pi_{[j]} \neq \pi_{[k]} \text{ for some } j \text{ and } k.$$

- Unit j has an observed covariate $\mathbf{x}_{[j]}$ and an unobserved covariate $u_{[j]}$. The model links the probability of assignment to treatment as follows:

$$\log\left(\frac{\pi_{[j]}}{1-\pi_{[j]}}\right) = k(\mathbf{x}_{[j]}) + \gamma u_{[j]}$$

with $0 \leq u_{[j]} \leq 1$ and where $k(\cdot)$ is an unknown function and γ is an unknown parameter.

Hidden Bias

- Define the probability that j will be in treatment as some unknown function of $\mathbf{x}$ and u:

$$\pi_{[j]} = \delta(\mathbf{x}_{[j]}, u_{[j]})$$

Hidden Bias

- Define the probability that j will be in treatment as some unknown function of $\mathbf{x}$ and u:

$$\pi_{[j]} = \delta(\mathbf{x}_{[j]}, u_{[j]})$$

- The probability of treatment assignment becomes:

$$\Pr(Z = z|\mathbf{x}, u) = \prod_{j=1}^{M} \delta(\mathbf{x}_{[j]}, u_{[j]})^{z_j}\{1 - \delta(\mathbf{x}_{[j]}, u_{[j]})\}^{1-z_j}$$

Hidden Bias

- Define the probability that j will be in treatment as some unknown function of $\mathbf{x}$ and u:

$$\pi_{[j]} = \delta(\mathbf{x}_{[j]}, u_{[j]})$$

- The probability of treatment assignment becomes:

$$\Pr(Z = z|\mathbf{x}, u) = \prod_{j=1}^{M} \delta(\mathbf{x}_{[j]}, u_{[j]})^{z_j}\{1 - \delta(\mathbf{x}_{[j]}, u_{[j]})\}^{1-z_j}$$

- Sensitivity analysis can test whether an inference is robust to hidden bias up to an arbitrary magnitude

Inference: Overt or Hidden Bias

- In the presence of hidden and overt bias, the issue is that the T permutation test distribution, under the sharp null of no effect, does not have uniform density over Ω

Inference: Overt or Hidden Bias

- In the presence of hidden and overt bias, the issue is that the T permutation test distribution, under the sharp null of no effect, does not have uniform density over Ω
- In permuting each $z \in \Omega$, we need to weight each realization of $t(\mathbf{z}, \mathbf{r})$ by $\Pr(Z = z|\mathbf{x}, u)$:

$$\text{prob}\{t(\mathbf{Z}, \mathbf{r}) \geq T\} = \sum_{z \in \Omega} I[t(\mathbf{Z}, \mathbf{r}) \geq T] \cdot \Pr(Z = z|\mathbf{x}, u)$$

Inference: Overt or Hidden Bias

- In the presence of hidden and overt bias, the issue is that the T permutation test distribution, under the sharp null of no effect, does not have uniform density over Ω
- In permuting each $z \in \Omega$, we need to weight each realization of $t(\mathbf{z}, \mathbf{r})$ by $\Pr(Z = z|\mathbf{x}, u)$:

$$\text{prob}\{t(\mathbf{Z}, \mathbf{r}) \geq T\} = \sum_{z \in \Omega} I[t(\mathbf{Z}, \mathbf{r}) \geq T] \cdot \Pr(Z = z|\mathbf{x}, u)$$

- Recall:

$$\Pr(Z = z|\mathbf{x}, u) = \prod_{j=1}^{M} \delta(\mathbf{x}_{[j]}, u_{[j]})^{z_j} \{1 - \delta(\mathbf{x}_{[j]}, u_{[j]})\}^{1-z_j}$$

Remember Γ

- How do we make progress?

Remember Γ

- How do we make progress?
- Recall the Γ parameter in a Rosenbaum-style sensitivity analysis: the treatment odds ratio

Remember Γ

- How do we make progress?
- Recall the Γ parameter in a Rosenbaum-style sensitivity analysis: the treatment odds ratio
- In a sensitivity analysis, we ask the effect of varying Γ on our inferences, where Γ is bounded as follows:

$$\frac{1}{\Gamma} \leq \frac{\pi_{[j]}(1-\pi_{[k]})}{\pi_{[k]}(1-\pi_{[j]})} \leq \Gamma$$

Remember Γ

- How do we make progress?
- Recall the Γ parameter in a Rosenbaum-style sensitivity analysis: the treatment odds ratio
- In a sensitivity analysis, we ask the effect of varying Γ on our inferences, where Γ is bounded as follows:

$$\frac{1}{\Gamma} \leq \frac{\pi_{[j]}(1-\pi_{[k]})}{\pi_{[k]}(1-\pi_{[j]})} \leq \Gamma$$

- Can extend this parameterization to the general case:

$$\frac{1}{\Gamma} \leq \frac{\delta(\mathbf{x}_{[j]}, u_{[j]})(1-\delta(\mathbf{x}_{[k]}, u_{[k]}))}{\delta(\mathbf{x}_{[k]}, u_{[k]})(1-\delta(\mathbf{x}_{[j]}, u_{[j]}))} \leq \Gamma$$

Inference with Confounders

- For each $(\mathbf{x}, \gamma, \mathbf{u})$, statistic $t(\mathbf{Z}, \mathbf{r})$ is the sum of S independent RV, where the sth variable equals d_s with probability

$$p_s^+ = \frac{c_{s1} \cdot \exp(k(\mathbf{x}_{s1}) + \gamma u_{s1}) + c_{s2} \cdot \exp(k(\mathbf{x}_{s2}) + \gamma u_{s2})}{\exp(k(\mathbf{x}_{s1}) + \gamma u_{s1}) + \exp(k(\mathbf{x}_{s2}) + \gamma u_{s2})}$$

Inference with Confounders

- For each $(\mathbf{x}, \gamma, \mathbf{u})$, statistic $t(\mathbf{Z}, \mathbf{r})$ is the sum of S independent RV, where the sth variable equals d_s with probability

$$p_s^+ = \frac{c_{s1} \cdot \exp(k(\mathbf{x}_{s1}) + \gamma u_{s1}) + c_{s2} \cdot \exp(k(\mathbf{x}_{s2}) + \gamma u_{s2})}{\exp(k(\mathbf{x}_{s1}) + \gamma u_{s1}) + \exp(k(\mathbf{x}_{s2}) + \gamma u_{s2})}$$

- Recall c_{sj} is a function of the outcome R, such that
  $c_{s1} = 1, c_{s2} = 0$ if $r_{s1} > r_{s2}$;
  $c_{s1} = 0, c_{s2} = 1$ if $r_{s1} < r_{s2}$;
  $c_{s1} = c_{s2} = 0$ if $r_{s1} = r_{s2}$

Inference with Confounders

- For each $(\mathbf{x}, \gamma, \mathbf{u})$, statistic $t(\mathbf{Z}, \mathbf{r})$ is the sum of S independent RV, where the sth variable equals d_s with probability

$$p_s^+ = \frac{c_{s1} \cdot \exp(k(\mathbf{x}_{s1}) + \gamma u_{s1}) + c_{s2} \cdot \exp(k(\mathbf{x}_{s2}) + \gamma u_{s2})}{\exp(k(\mathbf{x}_{s1}) + \gamma u_{s1}) + \exp(k(\mathbf{x}_{s2}) + \gamma u_{s2})}$$

- Recall c_{sj} is a function of the outcome R, such that $c_{s1} = 1, c_{s2} = 0$ if $r_{s1} > r_{s2}$; $c_{s1} = 0, c_{s2} = 1$ if $r_{s1} < r_{s2}$; $c_{s1} = c_{s2} = 0$ if $r_{s1} = r_{s2}$
- Thus, p_s^+ is the probability the treated unit has the greater response: $p_s^+ = Pr[\sum_{i=1}^{m_s} c_{si} Z_{si} = 1]$

Bounded Inferences

- Draw bounded inferences using informative or sensible values of $k(\mathbf{x})$ and γ
- We can check for sensitivity to remaining overt bias after matching:
  - E.g., Set $\gamma = 0$, and $k(\mathbf{x}) = \mathbf{x}\beta$

$$p_s^+ = \frac{c_{s1} \cdot \exp(\beta \mathbf{x}_{s1}) + c_{s2} \cdot \exp(\beta \mathbf{x}_{s2})}{\exp(\beta \mathbf{x}_{s1}) + \exp(\beta \mathbf{x}_{s2})}$$

- Natural extension to combine tests of overt and hidden bias under a 'worst case'
  - E.g., Set $\gamma > 0$, $u_{si} = c_{si}$, and $k(\mathbf{x}) = \mathbf{x}\beta$

$$p_s^+ = \frac{c_{s1} \cdot \exp(\beta \mathbf{x}_{s1} + \gamma c_{s1}) + c_{s2} \cdot \exp(\beta \mathbf{x}_{s2} + \gamma c_{s2})}{\exp(\beta \mathbf{x}_{s1} + \gamma c_{s1}) + \exp(\beta \mathbf{x}_{s2} + \gamma c_{s2})}$$

Bounded Inferences: $E[T^+]$

- Define T^+ to be the sum of S independent random variables, where the sth variable takes the value of d_s with probability p_s^+ and takes the value of 0 with probability $1 - p_s^+$.

$$T^+ = \sum_{s=1}^{S} d_s \sum_{i=1}^{m_s} c_{si} Z_{si}$$

$$E[T^+] = \sum_{s=1}^{S} d_s E[\sum_{i=1}^{m_s} c_{si} Z_{si}]$$

$$E[T^+] = \sum_{s=1}^{S} d_s p_s^+$$

Bounded Inferences: $V[T^+]$

- Similarly:

$$T^+ = \sum_{s=1}^{S} d_s \sum_{i=1}^{m_s} c_{si} Z_{si}$$

$$V[T^+] = \sum_{s=1}^{S} d_s^2 V[\sum_{i=1}^{m_s} c_{si} Z_{si}]$$

$$V[T^+] = \sum_{s=1}^{S} d_s^2 p_s^+ (1 - p_s^+)$$

Normal Approximations

- Deviate of a sum RV is asymptotically standard normal increasing in S

$$Deviate = \frac{T - E[T^+]}{\sqrt{V[T^+]}} \sim N(0,1)$$

1. Use this to approximate p-values for statistical inference given possible confounding in $\mathbf{x}$ – easily extended to inference without matching

More on Bounded Inferences

2. Can use p_s^+ to measure the degree of overt bias:

$$\frac{\pi_{[j]}(1-\pi_{[k]})}{\pi_{[k]}(1-\pi_{[j]})} = \frac{p_s^+}{1-p_s^+}$$

Descriptively useful in determining degree of selection problem –

More on Bounded Inferences

3. Can easily compare findings from overt bias to those for an unobserved factor. Set:

$$p_s^+ = \frac{\Gamma}{1+\Gamma}$$

Search over values of Γ until

$$\frac{T - E[T^+_{hidden}]}{\sqrt{V[T^+_{hidden}]}} \approx \frac{T - E[T^+_{overt}]}{\sqrt{V[T^+_{overt}]}}$$

Example: Empowering Effects of Coethnic Representatives

- Q: Do Hispanic incumbents increase Hispanic voter turnout and registration?

  Look at blocks of voters redistricted from white to Hispanic incumbents in CA – estimate differences in participation before (2000) and after (2002) being redistricted
- Define Γ to be a measure of the median treatment odds of treatment divided by the median treatment odds of control
- Use this empirical Γ, based on confounding in X, in a sensitivity analysis at various stages of conditioning in X

Example: Empowering Effects of Coethnic Representatives

	Γ	P-value
UNMATCHED		
Hispanic Registration	5.90	1.00
Hispanic Turnout	5.90	0.99
Non Hispanic Registration	5.90	1.00
Non Hispanic Turnout	5.90	1.00
TRIMMED		
Hispanic Registration	3.36	1.00
Hispanic Turnout	3.36	0.99
Non Hispanic Registration	3.36	1.00
Non Hispanic Turnout	3.36	1.00
MATCHED		
Hispanic Registration	1.39	0.51
Hispanic Turnout	1.39	0.99
Non Hispanic Registration	1.39	0.00
Non Hispanic Turnout	1.39	0.05

Additional Review Slides

The Sharp Null

- The most common hypothesis associated with randomization inference is the sharp null of no effect for all units.
- Under the null, the units' responses are *fixed* and the only random element is the meaningless rotation of labels.
- When testing the null hypothesis of no effect, the response of the ith unit in stratum s can be written r_{si} and the vector of responses is $\mathbf{r}$.

The Test Statistic

- A **test statistic** $t(\mathbf{Z}, r)$ is a quantity computed from the treatment assignment $\mathbf{Z}$ and the response r:

$$t(Z_{is}, r_{is}) = \sum_{s=1}^{S} d_s \sum_{i=1}^{n_s} c_{si} Z_{si}$$

- **Wilcoxon sign rank test**. In a stratified randomized experiment with S strata, the $|\mathbf{r_{1s}} - \mathbf{r_{0s}}|$ responses are ranked from smallest to largest to produce $\mathbf{d_s}$ (with average ranks for ties). Sum ranks for all strata for which $\mathbf{r_{1s}} > \mathbf{r_{0s}}$.
- **Wilcoxon rank sum test**. In an unstratified randomized experiment (S=1), the $\mathbf{r}$ responses are ranked from smallest to largest to produce $\mathbf{d}$ (with average ranks for ties). Sum ranks for all units assigned $\mathbf{Z} = 1$