# Sensitivity Analysis for Inverse Probability Weighting Estimators via the Percentile Bootstrap (Q. Zhao, D.S. Small, & B.B. Bhattacharya, 2017)

for "Observational Study Design and Causal Inference" (Statistics 260),

organized by S. Pimentel, Spring 2018, University of California, Berkeley

---

Nima Hejazi

12 April 2018

Group in Biostatistics,
University of California, Berkeley
https://statistics.berkeley.edu/~nhejazi

# Introduction

# Outline

# Preliminaries: Notation

- Data: $(A_1, \boldsymbol{X_1}, Y_1), \ldots, (A_n, \boldsymbol{X_n}, Y_n) \overset{\text{iid}}{\sim} F_0$, for a (binary) treatment $A_i$ and measured confounders $\boldsymbol{X_i}$.
- Outcome: $Y_i = A_i Y_i(1) + (1 - A_i) Y_i(0)$, using potential outcome notation.
- Estimand (Parameter): $\Delta := \mathbb{E}_0[Y(1)] - \mathbb{E}_0[Y(0)]$ (average treatment effect)

# Preliminaries: Identifiability and Assumptions

- Stable Unit Treatment Value Assumption (SUTVA)
- No Unmeasured Confounders (NUC): $(Y(0), Y(1)) \perp\!\!\!\perp A \mid \boldsymbol{X}$ (strong ignorability)
- Overlap: $e_0(x) := \mathbb{P}_0(A = 1 \mid X = x) \in (0, 1)$ (bounded propensity score)

# Preliminaries: IPW Estimators

- Inverse Probability of Weighting (IPW) Estimators:

$$\hat{\Delta}_{\mathrm{IPW}} = \frac{1}{n}\sum_{i=1}^{n} \frac{A_i Y_i}{\hat{e}(X_i)} - \frac{(1-A_i)Y_i}{1-\hat{e}(X_i)}$$

- $\hat{\Delta}$ consistently estimates $\Delta$ as long as $\hat{e}(\boldsymbol{X})$ converges to $e_0$.

# Outline

# The Objective: Sensitivity Analysis

- Under violations of the NUC assumption, $\hat{\Delta}_{\text{IPW}}$ is biased and has a confidence interval that doesn't cover $\Delta$ properly.
- The goal of a sensitivity analysis is to gauge the degree to which a statistical inference is incorrect under violations of the NUC assumption.
    - To what extent could the existence of potentially unmeasured confounders invalidate our findings?
    - ...

# The Objective: Sensitivity Analysis

- Under the NUC assumption, we have the following

$$e_a(x, y) := \mathbb{P}_0(A = 1 \mid \boldsymbol{X} = x, Y(a) = y) = e_a(x),$$

  for $a \in \{0, 1\}$.
  - $e_a(x, y)$ — "complete data" selection probability.
  - $e_a(x)$ — "observed data" selection probability.

- Thus, a sensitivity model might consider gauging whether $e_a(x, y) = e_a(x)$ holds in order to assess violations of NUC.

# Methodology

# Outline

# Sensitivity Parameter

- Marginal Sensitivity Model: let

$$\mathcal{E}(\Lambda) = \left\{ e(x,y) : \frac{1}{\Lambda} \leq \mathrm{OR}(e(x,y), e_0(x)) \leq \Lambda, x \in \mathcal{X}, y \in \mathbb{R} \right\}$$

- Then, for observational causal inference, let us assume that $e_a(x,y) \in \mathcal{E}(\Lambda)$, for $a \in \{0,1\}$.
- ...

# Sensitivity Parameter

- For convenience, we'll use a logistic representation of the sensitivity model:

$$h_0(x, y) = g_0(x) - g_0(x, y),$$

where

  - $g_0(x) = \text{logit}(e_0(x)) = \log \frac{e_0(x)}{1 - e_0(x)}$
  - Similarly, let $g_0(x, y) = \text{logit}(e_0(x, y))$.

- Then, we may express $\mathcal{E}(\Lambda)$ as

$$\mathcal{E}(\Lambda) = \{e^{(h)}(x, y) : h \in \mathcal{H}(\lambda)\},$$

where $\mathcal{H}(\lambda) = \{h : \mathcal{X} \times \mathbb{R} \to \mathbb{R} \text{ and } \|h\|_\infty \leq \lambda\}$.

# Parametric Model for Sensitivity Analysis

- Ideally, $e_0(x)$ would be estimated nonparametrically, but, in many cases, we restrict ourselves to parametric models:

$$
\begin{aligned}
e_{\beta_0}(x) &= \arg\min_{\beta\in\Theta} \mathrm{KL}(\mathbb{P}_0(A=1 \mid \boldsymbol{X}=x) \| \mathbb{P}_\beta(A=1 \mid \boldsymbol{X}=x)) \\
&= \arg\max_{\beta\in\Theta} \mathbb{E}_0[A \cdot \log e_\beta(X) + (1-A)\cdot \log(1-e_\beta(x)) \mid \boldsymbol{X}=x]
\end{aligned}
$$

- As before, now have $e_0(x,y) \in \mathcal{E}_{\beta_0}(\Lambda)$, where

$$
\mathcal{E}_{\beta_0}(\Lambda) := \left\{ e(x,y) : \frac{1}{\Lambda} \leq \mathrm{OR}(e(x,y), e_{\beta_0}(x,y)) \leq \Lambda, x \in \mathcal{X}, y \in \mathbb{R} \right\}
$$

# Confidence Intervals in Sensitivity Analysis

- $(1-\alpha)$–CI: $\mathbb{P}_0(\Delta \in [\mathrm{L}, \mathrm{U}]) \geq 1-\alpha$ is true for any $F_0$ s.t. $h_0 \in \mathcal{H}(\lambda)$, a collection of sensitivity models.
- Asymptotic $(1-\alpha)$–CI if $\liminf_{n\to\infty} \mathbb{P}_0(\Delta \in [\mathrm{L}, \mathrm{U}]) \geq 1-\alpha$

## Outline

# IPW Point Estimates

- Note that the ATE may be expressed $\hat{\Delta}^{(h_0,h_1)} := \hat{\mu}^{(h_1)}(1) - \hat{\mu}^{h_0}(0)$ for $h_0, h_1 \in \mathcal{H}(\lambda)$.

- Recall that the IPW estimator for a the mean $\mu$ in a missing data problem may be expressed:

$$\hat{\mu}^{(h)}_{\text{IPW}} = \frac{1}{n}\sum_{i=1}^{n} \frac{A_i Y_i}{\hat{e}^( h)(\boldsymbol{X_i}, Y_i)}$$

- The *stabilized* IPW (SIPW) estimator is often used instead

$$\hat{\mu}^{(h)}_{\text{SIPW}} = \left[\frac{1}{n}\sum_{i=1}^{n} \frac{A_i}{\hat{e}^( h)(\boldsymbol{X_i}, Y_i)}\right]^{-1} \left[\frac{1}{n}\sum_{i=1}^{n} \frac{A_i Y_i}{\hat{e}^( h)(\boldsymbol{X_i}, Y_i)}\right]$$

# Interval Construction for SIPW Estimators

- **Problem:** need to estimate variance computationally (e.g., via bootstrap) while considering all violations of NUC in $\mathcal{H}(\lambda)$.

- Given $(1-\alpha)$–CIs for all $h \in \mathcal{H}(\lambda)$:

$$\liminf_{n\to\infty} \mathbb{P}_0(\mu^{(h)} \in [L^{(h)}, U^{(h)}]) \geq 1-\alpha$$

- Then, we have that the following is an asymptotic $(1-\alpha)$–CI under the collection of sensitivity models $\mathcal{H}(\lambda)$:

$$L = \inf_{h\in\mathcal{H}(\lambda)} L^{(h)}, U = \sup_{h\in\mathcal{H}(\lambda)} U^{(h)}$$

# Outline

# Percentile Bootstrap

- Finally, the centerpiece! But first, we need still more notation.
- Let $\mathbb{P}_n$ be the empirical measure on the sample $\boldsymbol{T_1}, \ldots, \boldsymbol{T_n}$, where $\boldsymbol{T_i} = (A_i, \boldsymbol{X_i'}, A_i Y_i)$.
- Further, let $\boldsymbol{\hat{T}_1}, \ldots, \boldsymbol{\hat{T}_n}$ be i.i.d. re-samples from $\mathbb{P}_n$.
- Then, the SIPW estimate $\hat{\hat{\mu}}^{(h)}$ may be computed over the bootstrap re-samples $\{\boldsymbol{\hat{T}_i}\}_{i \in [n]}$.
- Now, for $h \in \mathcal{H}(\lambda)$, percentile bootstrap CI:

$$[L^{(h)}, U^{(h)}] = \left[Q_{\frac{\alpha}{2}}(\hat{\hat{\mu}}^{(h)}), Q_{1-\frac{\alpha}{2}}(\hat{\hat{\mu}}^{(h)})\right],$$

where $Q_\alpha(\hat{\hat{\mu}}) := \inf\{t : \hat{\mathbb{P}}_n(\hat{\hat{\mu}} \leq t) \geq \alpha\}$

# Percentile Bootstrap

- The percentile bootstrap interval $[L^{(h)}, U^{(h)}]$ is an asymptotically valid CI of the target estimate for the parametric sensitivity model $e^{(h)} \in \mathcal{E}_{\beta_0}(\Lambda)$.
- The bootstrap is not valid if the missingness probability is modeled nonparametrically (Abadie & Imbens, 2008).
- Percentile bootstrap CI under collection of sensitivity models:

$$[L, U] = \left[ \left( Q_{\frac{\alpha}{2}}\Big( \inf_{h \in \mathcal{H}(\lambda)} \hat{\hat{\mu}}^{(h)} \Big) \right), \left( Q_{1-\frac{\alpha}{2}}\Big( \sup_{h \in \mathcal{H}(\lambda)} \hat{\hat{\mu}}^{(h)} \Big) \right) \right]$$

  - Since infimum/supremum is inside the quantile function, the problem is efficiently solved by linear programming.
  - Exchange of quantile and infimum/supremum is justified by a generalized (von Neumann's) minimax/maximin inequality.

# Linear Fractional Programming of SIPW Point Estimates

- Bootstrap Intervals of the range of SIPW point estimates:

$$[L_B, U_B] = \left[ \left( Q_{\frac{\alpha}{2}} \big( \inf_{h \in \mathcal{H}(\lambda)} \hat{\hat{\mu}}^{(h)} \big)_{b \in [B]} \right), \left( Q_{1-\frac{\alpha}{2}} \big( \sup_{h \in \mathcal{H}(\lambda)} \hat{\hat{\mu}}^{(h)} \big)_{b \in [B]} \right) \right]$$

- In the linear programming problem, the optimization variable is merely $z_i = e^{h(\boldsymbol{X}_i, Y_i)}$, as all other relevant quantities may be readily estimated from the observed data.

- Computation is extremely efficient, with complexity $O(nB + n\log n)$.

# CI for the ATE in the Sensitivity Model

- Using the approach we've been discussing, we readily obtain an asymptotically valid CI for the ATE via the percentile bootstrap:

$$\left[Q_{\frac{\alpha}{2}}(\hat{\hat{\Delta}}^{(h_0,h_1)}), Q_{1-\frac{\alpha}{2}}(\hat{\hat{\Delta}}^{(h_0,h_1)})\right],$$

  where $Q_{\frac{\alpha}{2}}(\hat{\hat{\Delta}}^{(h_0,h_1)})$ is the $\alpha$-th bootstrap quantile of the SIPW estimates.

- From this, we obtain CIs

# Extensions: Augmented IPW Estimators

- AIPW estimators are *"double robust"*, incorporating an extra nuisance parameter: $f_0(\boldsymbol{x}) = \mathbb{E}_0[Y \mid A = 1, \boldsymbol{X} = \boldsymbol{x}]$.
- The AIPW estimator $\hat{\Delta}_{\text{AIPW}}$ is consistent for $\Delta$ as long as one of $\hat{e}(\boldsymbol{x})$ and $\hat{f}(\boldsymbol{x})$ is consistent.
- **Limitation:** In order to compute asymptotically valid confidence intervals, the outcome regression model $\hat{f}(\boldsymbol{X}_i)$ must be parametric.

# Extensions: Augmented IPW Estimators

- AIPW estimators are *"double robust"*, incorporating an extra nuisance parameter: $f_0(\boldsymbol{x}) = \mathbb{E}_0[Y \mid A = 1, \boldsymbol{X} = \boldsymbol{x}]$.
- The AIPW estimator $\hat{\Delta}_{\mathrm{AIPW}}$ is consistent for $\Delta$ as long as one of $\hat{e}(\boldsymbol{x})$ and $\hat{f}(\boldsymbol{x})$ is consistent.
- **Limitation:** In order to compute asymptotically valid confidence intervals, the outcome regression model $\hat{f}(\boldsymbol{X}_i)$ must be parametric.
  *Perhaps this could be loosened.*

# Discussion

# Outline

# Comparison with Rosenbaum Sensitivity

- Rosenbaum's method obtains point estimates and CIs under the collection of models $\mathcal{R}(\Gamma)$.
  - Here, classically, we assume that the causal effect is additive and constant.
  - This means the Fisher null can be used to determine whether an effect $\Delta$ ought to be included in the $(1-\alpha)$–CI.
- The present approach is a hybrid of existing approaches in the sense that it considers a range of differencs between $A \mid \boldsymbol{X}$ and $A \mid \boldsymbol{X}, Y(a)$.

# Comparison with Rosenbaum Sensitivity

- Generally, Rosenbaum's method treats the sample as the population, the present approach treats the observations as i.i.d. samples from a super-population.
- The present method uses IPW-type estimators — importantly, this makes exact matching completely unnecessary.
- The present approach is natural for IPW-type estimators while Rosenbaum's is natural for matched designs.
- Most methods use randomization inference based on Fisher's sharp null, the present approach takes a point estimation perspective.
- Heterogeneous treatment effects; applicability to missing data problems.

# Outline

# What do I think?

# What do you think?

# References

Qingyuan Zhao, Dylan S Small, and Bhaswar B Bhattacharya. Sensitivity analysis for inverse probability weighting estimators via the percentile bootstrap. *arXiv preprint arXiv:1711.11286*. URL `https://arxiv.org/abs/1711.11286v1`.