Bootstrapping Sensitivity Analysis

Qingyuan Zhao

Department of Statistics, The Wharton School
University of Pennsylvania

June 2, 2019 @ OSU Bayesian Causal Inference Workshop

(Joint work with Bhaswar B. Battacharya and Dylan S. Small)

Why sensitivity analysis?

- Unless we have perfectly executed randomized experiment, causal inference is based on some **unverifiable assumptions**.
- In observational studies, the most commonly used assumption is **ignorability** or **no unmeasured confounding**:

$$A \perp\!\!\!\perp Y(0), Y(1) \,\big|\, \boldsymbol{X}.$$

  We can only say this assumption is "plausible".
- Sensitivity analysis asks: what if this assumption does not hold? Does our qualitative conclusion still hold?
- This question appears in many settings:
  1. Confounded observational studies.
  2. Survey sampling with missing not at random (MNAR).
  3. Longitudinal study with non-ignorable dropout.
- In general, this means that the target parameter (e.g. average treatment effect) is only **partially identified**.

Overview: Bootstrapping sensitivity analysis

Point-identified parameter: Efron's bootstrap

$$\text{Point estimator} \quad \xRightarrow{\textit{Bootstrap}} \quad \text{Confidence interval}$$

Partially identified parameter: An analogy

$$\underset{\textit{Optimization}}{\text{Extrema estimator}} \quad \xRightarrow{\textit{Percentile Bootstrap}} \quad \underset{\textit{Minimax inequality}}{\text{Confidence interval}}$$

Rest of the talk

Apply this idea to IPW estimators in a marginal sensitivity model.

Some existing sensitivity models

Generally, we need to specify how unconfoundedness is violated.

1. Y **models:** Consider **a specific difference** between the conditional distribution $Y(a) \mid \boldsymbol{X}, A$ and $Y(a) \mid \boldsymbol{X}$.
   - Commonly called "pattern mixture models".
   - Robins (1999, 2002); Birmingham et al. (2003); Vansteelandt et al. (2006); Daniels and Hogan (2008).
2. A **models:** Consider **a specific difference** between the conditional distribution $A \mid \boldsymbol{X}, Y(a)$ and $A \mid \boldsymbol{X}$.
   - Commonly called "selection models".
   - Scharfstein et al. (1999); Gilbert et al. (2003).
3. **Simultaneous models:** Consider **a range of** A models and/or Y models and report the "worst case" result.
   - Cornfield et al. (1959); Rosenbaum (2002); Ding and VanderWeele (2016).

Our sensitivity model—

A hybrid of 2nd and 3rd, similar to Rosenbaum's.

Rosenbaum's sensitivity model

- Imagine there is an unobserved confounder U that "summarizes" all confounding, so $A \perp\!\!\!\perp Y(0), Y(1) \mid \boldsymbol{X}, U$.
- Let $e_0(\boldsymbol{x}, u) = \mathbb{P}_0(A = 1 | \boldsymbol{X} = \boldsymbol{x}, U = u)$.

Rosenbaum's sensitivity model

$$\mathcal{R}(\Gamma) = \left\{ e(\boldsymbol{x}, u) : \frac{1}{\Gamma} \leq \mathrm{OR}(e(\boldsymbol{x}, u_1), e(\boldsymbol{x}, u_2)) \leq \Gamma, \forall \boldsymbol{x} \in \mathcal{X}, u_1, u_2 \right\},$$

where $\mathrm{OR}(p_1, p_2) := [p_1/(1 - p_1)]/[p_2/(1 - p_2)]$ is the *odds ratio*.

- Rosenbaum's question: can we reject the sharp null hypothesis $Y(0) \equiv Y(1)$ for every $e_0(\boldsymbol{x}, u) \in \mathcal{R}(\Gamma)$?
- Robins (2002): we don't need to assume the existence of U. Let $U = Y(1)$ when the goal is to estimate $\mathrm{E}[Y(1)]$.

Our sensitivity model

- Let $e_0(\boldsymbol{x}) = \mathbb{P}_0(A = 1|\boldsymbol{X} = \boldsymbol{x})$ be the propensity score.

Marginal sensitivity models

$$\mathcal{M}(\Gamma) = \Big\{e(\boldsymbol{x}, y) : \frac{1}{\Gamma} \le \mathrm{OR}(e(\boldsymbol{x}, y), e_0(\boldsymbol{x})) \le \Gamma, \forall \boldsymbol{x} \in \mathcal{X}, y\Big\}.$$

- Compare this to Rosenbaum's model:

$$\mathcal{R}(\Gamma) = \Big\{e(\boldsymbol{x}, u) : \frac{1}{\Gamma} \le \mathrm{OR}(e(\boldsymbol{x}, u_1), e(\boldsymbol{x}, u_2)) \le \Gamma, \forall \boldsymbol{x} \in \mathcal{X}, u_1, u_2\Big\}.$$

- Tan (2006) first considered this model, but he did not consider statistical inference in finite sample.
- Relationship between the two models: $\mathcal{M}(\sqrt{\Gamma}) \subseteq \mathcal{R}(\Gamma) \subseteq \mathcal{M}(\Gamma)$.[1]
- For observational studies, we assume both $\mathbb{P}_0(A = 1|\boldsymbol{X}, Y(1)), \mathbb{P}_0(A = 1|\boldsymbol{X}, Y(0)) \in \mathcal{M}(\Gamma)$.

[1] The second part needs "compatibility": $e(\boldsymbol{x}, y)$ marginalizes to $e_0(\boldsymbol{x})$.

Parametric extension

- In practice, the propensity score $e_0(\boldsymbol{X}) = \mathbb{P}_0(A = 1|\boldsymbol{X})$ is often estimated by a parametric model.

Definition (Parametric marginal sensitivity models)

$\mathcal{M}_{\boldsymbol{\beta}_0}(\Gamma) = \left\{ e(\boldsymbol{x}, y) : \frac{1}{\Gamma} \leq \mathrm{OR}(e(\boldsymbol{x}, y), e_{\boldsymbol{\beta}_0}(\boldsymbol{x})) \leq \Gamma, \forall \boldsymbol{x} \in \mathcal{X}, y \right\}$, where $e_{\boldsymbol{\beta}_0}(\boldsymbol{x})$ is the best parametric approximation of $e_0(\boldsymbol{x})$.

This sensitivity model covers both

1. **Model misspecification**, that is, $e_{\boldsymbol{\beta}_0}(\boldsymbol{x}) \neq e_0(\boldsymbol{x})$; and
2. **Missing not at random**, that is, $e_0(\boldsymbol{x}) \neq e_0(\boldsymbol{x}, y)$.

Logistic representations

1. **Rosenbaum's sensitivity model:**

$$\operatorname{logit}(e(\boldsymbol{x}, u)) = g(\boldsymbol{x}) + \gamma u,$$

   where $0 \leq U \leq 1$ and $\gamma = \log \Gamma$.

2. **Marginal sensitivity model:**

$$\operatorname{logit}(e^{(h)}(\boldsymbol{x}, y)) = \operatorname{logit}(e_0(\boldsymbol{x})) + h(\boldsymbol{x}, y),$$

   where $\|h\|_\infty = \sup |h(\boldsymbol{x}, y)| \leq \gamma$. Due to this representation, we also call it a **marginal L_∞-sensitivity model**.

3. **Parametric marginal sensitivity model:**

$$\operatorname{logit}(e^{(h)}(\boldsymbol{x}, y)) = \operatorname{logit}(e_{\boldsymbol{\beta}_0}(\boldsymbol{x})) + h(\boldsymbol{x}, y),$$

   where $\|h\|_\infty = \sup |h(\boldsymbol{x}, y)| \leq \gamma$.

Confidence interval I

- For simplicity, consider the "missing data" problem where $Y = Y(1)$ is only observed if $A = 1$.
- Observe i.i.d. samples $(A_i, \boldsymbol{X}_i, A_i Y_i)$, $i = 1, \ldots, n$.
- The estimand is $\mu_0 = \mathrm{E}_0[Y]$, however it is only partially identified under a simultaneous sensitivity model.

Goal 1 (Coverage of true parameter)

Construct a data-dependent interval $[L, U]$ such that

$$\mathbb{P}_0\big(\boldsymbol{\mu_0} \in [L, U]\big) \geq 1 - \alpha$$

whenever $e_0(\boldsymbol{X}, Y) = \mathbb{P}_0(A = 1 | \boldsymbol{X}, Y) \in \mathcal{M}(\Gamma)$.

Confidence interval II

- The inverse probability weighting (IPW) identity:

$$\mathbb{E}_0[Y] = \mathbb{E}\Big[\frac{AY}{e_0(\boldsymbol{X}, Y)}\Big] \overset{MAR}{=} \mathbb{E}\Big[\frac{AY}{e_0(\boldsymbol{X})}\Big].$$

- Define

$$\mu^{(h)} = \mathbb{E}_0\left[\frac{AY}{e^{(h)}(\boldsymbol{X}, Y)}\right]$$

- **Partially identified region:** $\{\mu^{(h)} : \ e^{(h)} \in \mathcal{M}(\Gamma)\}$.

Goal 2 (Coverage of partially identified region)

Construct a data-dependent interval $[L, U]$ such that

$$\mathbb{P}_0\Big(\{\boldsymbol{\mu}^{(h)} : \boldsymbol{e}^{(h)} \in \boldsymbol{\mathcal{M}}(\boldsymbol{\Gamma})\} \subseteq [L, U]\Big) \geq 1 - \alpha.$$

- Imbens and Manski (2004) have discussed the difference between these two Goals.

An intuitive idea: "The Union Method"

- Suppose for any h, we have a confidence interval $[L^{(h)}, U^{(h)}]$ such that
$$\lim_{n\to\infty}\inf \mathbb{P}_0(\mu^{(h)} \in [L^{(h)}, U^{(h)}]) \geq 1-\alpha$$
- Let $L = \inf_{\|h\|} L^{(h)}$ and $U = \sup_{\|h\|} U^{(h)}$, so $[L, U]$ is **the union interval**.

Theorem

1. *$[L, U]$ satisfies Goal 1 asymptotically.*
2. *Furthermore if the intervals are "congruent": $\exists\ \alpha' < \alpha$ such that*
$$\limsup_{n\to\infty} \mathbb{P}_0\left(\mu^{(h)} < L^{(h)}\right) \leq \alpha', \quad \limsup_{n\to\infty} \mathbb{P}_0\left(\mu^{(h)} > U^{(h)}\right) \leq \alpha - \alpha'.$$
Then $[L, U]$ satisfies Goal 2 asymptotically.

Practical challenge: How to take the union?

- Suppose $\hat{g}(\boldsymbol{x})$ is an estimate of $\operatorname{logit}(e_0(\boldsymbol{x}))$.
- For a specific difference h, we can estimate $e^{(h)}(\boldsymbol{x}, y)$ by

$$\hat{e}^{(\boldsymbol{h})}(\boldsymbol{x}, y) = \frac{1}{1 + e^{\boldsymbol{h}(\boldsymbol{x},\boldsymbol{y}) - \hat{g}(\boldsymbol{x}, y)}}.$$

- This leads to an (stabilized) IPW estimate of $\mu^{(h)}$:

$$\hat{\mu}^{(h)} = \left[\frac{1}{n}\sum_{i=1}^{n} \frac{A_i}{\hat{e}^{(h)}(\boldsymbol{X}_i, Y_i)}\right]^{-1} \left[\frac{1}{n}\sum_{i=1}^{n} \frac{A_i Y_i}{\hat{e}^{(h)}(\boldsymbol{X}_i, Y_i)}\right].$$

- Under regularity conditions, the Z-estimation theory tells us

$$\sqrt{n}\left(\hat{\mu}^{(h)} - \mu^{(h)}\right) \xrightarrow{d} \mathrm{N}(0, (\sigma^{(h)})^2)$$

- Therefore we can use $[L^{(h)}, U^{(h)}] = \hat{\mu}^{(h)} \mp z_{\frac{\alpha}{2}} \cdot \dfrac{\hat{\sigma}^{(h)}}{\sqrt{n}}$.
- However, computing the union interval requires solving a complicated optimization problem.

Bootstrapping sensitivity analysis

Point-identified parameter: Efron's bootstrap

$$\text{Point estimator} \quad \xRightarrow{\textit{Bootstrap}} \quad \text{Confidence interval}$$

Partially identified parameter: An analogy

$$\underset{\text{Extrema estimator}}{\textit{Optimization}} \quad \xRightarrow{\textit{Percentile Bootstrap}} \quad \underset{\text{Confidence interval}}{\textit{Minimax inequality}}$$

A simple procedure for simultaneous sensitivity analysis

1. Generate B random resamples of the data. For each resample, compute the extrema of IPW estimates under $\mathcal{M}_{\boldsymbol{\beta}_0}(\Gamma)$.
2. Construct the confidence interval using $L = Q_{\alpha/2}$ of the B minima and $U = Q_{1-\alpha/2}$ of the B maxima.

Theorem

$[L, U]$ *achieves Goal 2 for* $\mathcal{M}_{\boldsymbol{\beta}_0}(\Gamma)$ *asymptotically.*

Proof of the Theorem

Partially identified parameter: Three ideas

$$\begin{array}{ccc} \textit{Optimization} & \textit{1. Percentile Bootstrap} & \textit{2. Minimax inequality} \\ \text{Extrema estimator} & \Longrightarrow & \text{Confidence interval} \end{array}$$

1. The sampling variability of $\hat{\mu}^{(h)}$ can be captured by bootstrap. The percentile bootstrap CI is given by

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

2. Generalized minimax inequality:

$$\overbrace{Q_{\frac{\alpha}{2}}\left(\inf_h \hat{\hat{\mu}}_b^{(h)}\right) \leq \underbrace{\inf_h Q_{\frac{\alpha}{2}}\left(\hat{\hat{\mu}}_b^{(h)}\right) \leq \sup_h Q_{1-\frac{\alpha}{2}}\left(\hat{\hat{\mu}}_b^{(h)}\right)}_{\textbf{Union CI}} \leq Q_{1-\frac{\alpha}{2}}\left(\sup_h \hat{\hat{\mu}}_b^{(h)}\right)}^{\textbf{Percentile Bootstrap CI}}.$$

Computation

Partially identified parameter: Three ideas

3. Optimization	*Percentile Bootstrap*	*Minimax inequality*
Extrema estimator	$\Longrightarrow$	Confidence interval

3. Computing extrema of $\hat{\mu}^{(h)}$ is a **linear fractional programming**:
   Let $z_i = e^{h(\boldsymbol{X}_i, Y_i)}$, we just need to solve

$$\begin{aligned} \text{max or min} \quad & \frac{\sum_{i=1}^n A_i Y_i \left(1 + z_i e^{-\hat{g}(\boldsymbol{X}_i)}\right)}{\sum_{i=1}^n A_i \left(1 + z_i e^{-\hat{g}(\boldsymbol{X}_i)}\right)}, \\ \text{subject to} \quad & z_i \in [\Gamma^{-1}, \Gamma], \; i = 1, \dots, n. \end{aligned}$$

   - This can be converted to a linear programming.
   - Moreover, the solution $\boldsymbol{z}$ must have the same/opposite order as $\boldsymbol{Y}$, so the time complexity can be reduced to $O(n)$ (optimal).

The role of Bootstrap

Comapred to the union method, the workflow is greatly simplified:

1. No need to **derive** $\sigma^{(h)}$ analytically (though we could).
2. No need to **optimize** $\sigma^{(h)}$ (which is very challenging).

Comparison with Rosenbaum's sensitivity analysis

	Rosenbaum's paradigm	New bootstrap approach
Population	Sample	Super-population
Design	Matching	Weighting
Sensitivity model	$\mathcal{M}(\sqrt{\Gamma}) \subseteq \mathcal{R}(\Gamma) \subseteq \mathcal{M}(\Gamma)$	
Inference	Bounding p-value	CI for ATE/ATT
Effect modification	Constant effect	Allow for heterogeneity
Extension	Carefully developed for observational studies	Can be applied to missing data problems

Example

Fish consumption and blood mercury

- 873 controls: ≤ 1 serving of fish per month.
- 234 treated: ≥ 12 servings of fish per month.
- Covariates: gender, age, income (very imblanced), race, education, ever smoked, # cigarettes.

Implementation details

- Rosenbaum's method: 1-1 matching, CI constructed by Hodges-Lehmann (assuming causal effect is constant).
- Our method (percentile Bootstrap): stabilized IPW for ATT w/wo augmentation by outcome linear regression.

Results

- Recall that $\mathcal{M}(\sqrt{\Gamma}) \subseteq \mathcal{R}(\Gamma) \subseteq \mathcal{M}(\Gamma)$.

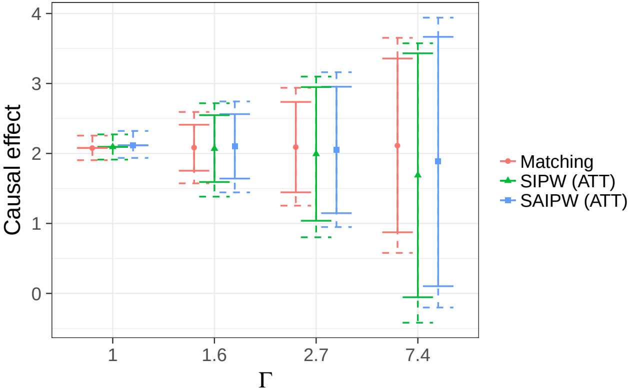

Figure: The solid error bars are the range of point estimates and the dashed error bars (together with the solid bars) are the confidence intervals. The circles/triangles/squares are the mid-points of the solid bars.

Discussion: The general sensitivity analysis problem

$$\underset{\text{Extrema estimator}}{\textit{Optimization}} \quad \xRightarrow{\textit{Percentile Bootstrap}} \quad \underset{\text{Confidence interval}}{\textit{Minimax inequality}}$$

The percentile bootstrap idea can be extended to the following problem:

$$\begin{aligned} \text{max or min} \quad & \mathbb{E}[f(\boldsymbol{X}, \theta, h(\boldsymbol{X}))], \\ \text{subject to} \quad & \|h(\boldsymbol{x})\|_\infty \leq \gamma, \end{aligned}$$

where f is a functional of the observed data $\boldsymbol{X}$, some finite-dimensional nuisance parameter θ, and a sensitivity function $h(\boldsymbol{X})$, as long as

- θ is **"estimable"** given $\boldsymbol{X}$ and h;
- Bootstrap **"works"** for $\mathbb{E}_n[f(\boldsymbol{X}, \hat{\theta}, h(\boldsymbol{X}))]$, **given** *h*.

The challenges...

1. How to solve the sample verision of the optimization problem?
2. Can we allow infinite-dimensional θ?
3. Can we include additional constraints such as $\mathbb{E}[g(\boldsymbol{X}, \theta, h(\boldsymbol{X}))] \leq 0$?

References I

Reference for this talk

- "Sensitivity analysis for inverse probability weighting estimators via the percentile bootstrap." To appear in *JRSSB*.
- R package: `https://github.com/qingyuanzhao/bootsens`.

Further references

J. Birmingham, A. Rotnitzky, and G. M. Fitzmaurice. Pattern-mixture and selection models for analysing longitudinal data with monotone missing patterns. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(1):275–297, 2003.

J. Cornfield, W. Haenszel, E. C. Hammond, A. M. Lilienfeld, M. B. Shimkin, and E. L. Wynder. Smoking and lung cancer: recent evidence and a discussion of some questions. *Journal of the National Cancer Institute*, 22(1):173–203, 1959.

M. J. Daniels and J. W. Hogan. *Missing data in longitudinal studies: Strategies for Bayesian modeling and sensitivity analysis*. CRC Press, 2008.

P. Ding and T. J. VanderWeele. Sensitivity analysis without assumptions. *Epidemiology*, 27(3): 368, 2016.

P. B. Gilbert, R. J. Bosch, and M. G. Hudgens. Sensitivity analysis for the assessment of causal vaccine effects on viral load in hiv vaccine trials. *Biometrics*, 59(3):531–541, 2003.

G. W. Imbens and C. F. Manski. Confidence intervals for partially identified parameters. *Econometrica*, 72(6):1845–1857, 2004.

References II

J. M. Robins. Association, causation, and marginal structural models. *Synthese*, 121(1):151–179, 1999.

J. M. Robins. Comment on "covariance adjustment in randomized experiments and observational studies". *Statistical Science*, 17(3):309–321, 2002.

P. R. Rosenbaum. *Observational Studies*. Springer New York, 2002.

D. O. Scharfstein, A. Rotnitzky, and J. M. Robins. Adjusting for nonignorable drop-out using semiparametric nonresponse models. *Journal of the American Statistical Association*, 94(448): 1096–1120, 1999.

Z. Tan. A distributional approach for causal inference using propensity scores. *Journal of the American Statistical Association*, 101(476):1619–1637, 2006.

S. Vansteelandt, E. Goetghebeur, M. G. Kenward, and G. Molenberghs. Ignorance and uncertainty regions as inferential tools in a sensitivity analysis. *Statistica Sinica*, 16(3):953–979, 2006.