Bootstrapping Sensitivity analysis

Qingyuan Zhao

Statistical Laboratory, University of Cambridge

August 3, 2020 @ JSM

Sensitivity analysis

The broader concept [Saltelli et al., 2004]

- Sensitivity analysis is “the study of how **the uncertainty in the output** of a mathematical model or system (numerical or otherwise) can be apportioned to different sources of **uncertainty in its inputs**”.
- Model inputs may be any factor that “can be changed in a model prior to its execution”, including” “structural and epistemic sources of uncertainty”.

In observational studies

- The most typical question is:

  How do the qualitative and/or quantitative conclusions of the observational study change if the **no unmeasured confounding assumption** *is violated?*

Sensitivity analysis for observational studies

State of the art

- Gazillions of methods specifically designed for different problems.
- Various forms of statistical guarantees.
- Often not straightforward to interpret

Goals of this talk

1. What is the common structure behind various methods for sensitivity analysis?
2. Can we bootstrap sensitivity analysis?

What is a sensitivity model?

General setup

$$\text{Observed data } \boldsymbol{O} \overset{\textit{infer}}{\Longrightarrow} \text{Distribution of the full data } \boldsymbol{F}.$$

- Prototypical example: Observe iid copies of $\boldsymbol{O} = (\boldsymbol{X}, A, Y)$ from the underlying full data $\boldsymbol{F} = (\boldsymbol{X}, A, Y(0), Y(1))$, where A is a binary treatment, $\boldsymbol{X}$ is covariates, Y is outcome.

An abstraction

A *sensitivity model* is a family of distributions $\mathcal{F}_{\theta,\eta}$ of $\boldsymbol{F}$ that satisfies:

1. *Augmentation:* Setting $\eta = 0$ corresponds to a primary analysis assuming no unmeasured confounders.
2. *Model identifiability:* Given η, the implied marginal distribution $\mathcal{O}_{\theta,\eta}$ of the observed data $\boldsymbol{O}$ is identifiable.

Statistical problem

Given η (or the range of η), use the observed data to make inference about some causal parameter $\beta = \beta(\theta, \eta)$.

Understanding sensitivity models

Observational equivalence

- $\mathcal{F}_{\theta,\eta}$ and $\mathcal{F}_{\theta',\eta'}$ are said to be *observationally equivalent* if $\mathcal{O}_{\theta,\eta} = \mathcal{O}_{\theta',\eta'}$. We write this as $\mathcal{F}_{\theta,\eta} \simeq \mathcal{F}_{\theta',\eta'}$.
- Equivalence class $[\mathcal{F}_{\theta,\eta}] = \{\mathcal{F}_{\theta',\eta'} \mid \mathcal{F}_{\theta,\eta} \simeq \mathcal{F}_{\theta',\eta'}\}$.

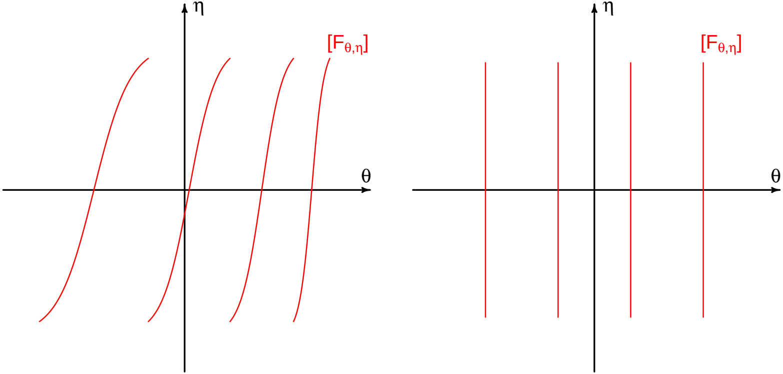
Types of sensitivity models

Testable models: When $\mathcal{F}_{\theta,\eta}$ is not rich enough, $[\mathcal{F}_{\theta,\eta}]$ is a singleton and η can be identified from the observed data (should be avoided in practice).

Global models: For any (θ, η) and η', there exists $\mathcal{F}_{\theta',\eta'} \simeq \mathcal{F}_{\theta,\eta}$.

Separable models: For any (θ, η), $\mathcal{F}_{\theta,\eta} \simeq \mathcal{F}_{\theta,0}$.

A visualization



Left: **Global** sensitivity models; Right: **Separable** sensitivity models.

Statistical inference

Modes of inference

1. **Point identified** sensitivity analysis is performed at a fixed η.
2. **Partially identified** sensitivity analysis is performed simultaneously over $\eta \in H$ for a given range H.

Statistical guarantees of interval estimators

1. **Confidence interval** $[C_L(\boldsymbol{O}_{1:n};\eta), C_U(\boldsymbol{O}_{1:n};\eta)]$ satisfies

$$\inf_{\theta_0,\eta_0} \mathbb{P}_{\theta_0,\eta_0}\big\{\beta(\theta_0,\eta_0) \in [C_L(\eta_0), C_U(\eta_0)]\big\} \geq 1-\alpha.$$

2. **Sensitivity interval** $[C_L(\boldsymbol{O}_{1:n};H), C_U(\boldsymbol{O}_{1:n};H)]$ satisfies

$$\inf_{\theta_0,\eta_0} \mathbb{P}_{\theta_0,\eta_0}\big\{\beta(\theta_0,\eta_0) \in [C_L(H), C_U(H)]\big\} \geq 1-\alpha. \tag{1}$$

They look almost the same, but because the latter interval only depends on H, (1) is actually equivalent to

$$\inf_{\theta_0,\eta_0} \inf_{\mathcal{F}_{\theta,\eta}\simeq\mathcal{F}_{\theta_0,\eta_0}} \mathbb{P}_{\theta_0,\eta_0}\big\{\beta(\theta,\eta) \in [C_L(H), C_U(H)]\big\} \geq 1-\alpha.$$

Approaches to sensitivity analysis

- **Point identified** sensitivity analysis is basically the same as primary analysis with known "offset" η.
- **Partially identified** sensitivity analysis is much harder. Let $\mathcal{F}_{\theta_0,\eta_0}$ be the truth. The fundamental problem is to make inference about

$$\inf_{\eta \in H} \{\beta(\theta,\eta) \mid \mathcal{F}_{\theta,\eta} \simeq \mathcal{F}_{\theta_0,\eta_0}\} \text{ and } \sup_{\eta \in H} \{\beta(\theta,\eta) \mid \mathcal{F}_{\theta,\eta} \simeq \mathcal{F}_{\theta_0,\eta_0}\}$$

Method 1 Solve the population optimization problems analytically.

- Not always feasible.

Method 2 Solve the sample approximation problem and use asymptotic normality.

- Central limit theorems not always true or established.

Method 3 Take the union of confidence intervals

$$[C_L(H), C_U(H)] = \bigcup_{\eta \in H} [C_L(\eta), C_U(\eta)].$$

- By the union bound, this is a $(1-\alpha)$-sensitivity interval if all $[C_L(\eta), C_U(\eta)]$ are $(1-\alpha)$-confidence intervals.

Computational challenges for Method 3

$$[C_L(H), C_U(H)] = \bigcup_{\eta \in H} [C_L(\eta), C_U(\eta)].$$

- Using asymptotic theory, it is often not difficult to construct asymptotic confidence intervals of the form

$$[C_L(\eta), C_U(\eta)] = \hat{\beta}(\eta) \mp z_{\frac{\alpha}{2}} \cdot \frac{\hat{\sigma}(\eta)}{\sqrt{n}}$$

- Unlike Method 2 that only needs to optimize $\hat{\beta}(\eta)$, Method 3 further needs to optimize the usually much more complicated $\hat{\sigma}(\eta)$ over $\eta \in H$.

Method 4: Percentile bootstrap

1. For fixed η, use the percentile bootstrap confidence interval (b is an index for data resample)

$$[C_L(\eta), C_U(\eta)] = \left[Q_{\frac{\alpha}{2}} \left(\hat{\hat{\beta}}_b(\eta) \right), Q_{1-\frac{\alpha}{2}} \left(\hat{\hat{\beta}}_b(\eta) \right) \right].$$

2. Use the generalized minimax inequality to interchange quantile and infimum/supremum:

Percentile bootstrap sensitivity interval

$$Q_{\frac{\alpha}{2}} \left(\inf_{\eta} \hat{\hat{\beta}}_b(\eta) \right) \leq \inf_{\eta} Q_{\frac{\alpha}{2}} \left(\hat{\hat{\beta}}_b(\eta) \right) \leq \sup_{\eta} Q_{1-\frac{\alpha}{2}} \left(\hat{\hat{\beta}}_b(\eta) \right) \leq Q_{1-\frac{\alpha}{2}} \left(\sup_{\eta} \hat{\hat{\beta}}_b(\eta) \right).$$

Union sensitivity interval

Advantages

- Computation is reduced to repeating Method 2 over data resamples.
- Only need coverage guarantee for $[C_L(\eta), C_U(\eta)]$ for **fixed** η.

Bootstrapping sensitivity analysis

Point-identified parameter: Efron's bootstrap

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

Partially identified parameter: Three ideas

Optimization	*Percentile Bootstrap*	*Minimax inequality*
Extrema estimator	$\Longrightarrow$	Sensitivity interval

Rest of the talk

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

Our sensitivity model

- Consider the prototypical example: A is a binary treatment, $\boldsymbol{X}$ is covariates, Y is outcome.
- U "summarizes" unmeasured confounding, so $A \perp\!\!\!\perp Y(0), Y(1) \mid \boldsymbol{X}, U$.
- Let $e_0(\boldsymbol{x}) = \mathbb{P}_0(A = 1 \mid \boldsymbol{X} = \boldsymbol{x}), e(\boldsymbol{x}, u) = \mathbb{P}(A = 1 \mid \boldsymbol{X} = \boldsymbol{x}, U = u)$.

Marginal sensitivity models

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

- Compare this to the Rosenbaum [2002] model:

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

- Tan [2006] first considered the marginal model, but he did not consider statistical inference in finite sample.
- Relationship between the two models: $E_M(\sqrt{\Gamma}) \subseteq E_R(\Gamma) \subseteq E_M(\Gamma)$.[1]

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

Parametric extension

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

Parametric marginal sensitivity models

$$E_M(\Gamma, \beta_0) = \left\{ e(\boldsymbol{x}, u) : \frac{1}{\Gamma} \leq \mathrm{OR}(e(\boldsymbol{x}, u), e_{\boldsymbol{\beta}_0}(\boldsymbol{x})) \leq \Gamma, \forall \boldsymbol{x} \in \mathcal{X}, y \right\}$$

- $e_{\boldsymbol{\beta}_0}(\boldsymbol{x})$ is the best parametric approximation to $e_0(\boldsymbol{x})$.

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, $e_0(\boldsymbol{x}) \neq e(\boldsymbol{x}, u)$.

Logistic representations

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

$$\mathrm{logit}(e(\boldsymbol{x}, u)) = g(\boldsymbol{x}) + u \log \Gamma,$$

   where $0 \leq U \leq 1$.

2. **Marginal sensitivity model:**

$$\mathrm{logit}(e_\eta(\boldsymbol{x}, u)) = \mathrm{logit}(e_0(\boldsymbol{x})) + \eta(\boldsymbol{x}, u),$$

   where $\eta \in H_\Gamma = \{\eta(\boldsymbol{x}, u) \mid \|\eta\|_\infty = \sup |\eta(\boldsymbol{x}, u)| \leq \log \Gamma\}$.

3. **Parametric marginal sensitivity model:**

$$\mathrm{logit}(e_\eta(\boldsymbol{x}, u)) = \mathrm{logit}(e_{\boldsymbol{\beta}_0}(\boldsymbol{x})) + \eta(\boldsymbol{x}, u),$$

   where $\eta \in H_\Gamma$.

Computation

Bootstrapping partially identified sensitivity analysis

$$\text{Extrema estimator} \xrightarrow[]{} \text{Sensitivity interval}$$

Optimization — *Percentile Bootstrap* — *Minimax inequality*

- Stabilized inverse-probability weighted (IPW) estimator for $\beta = \mathbb{E}[Y(1)]$:

$$\hat{\beta}(\eta) = \left[\frac{1}{n}\sum_{i=1}^{n} \frac{A_i}{\hat{e}_\eta(\boldsymbol{X}_i, U_i)}\right]^{-1} \left[\frac{1}{n}\sum_{i=1}^{n} \frac{A_i Y_i}{\hat{e}_\eta(\boldsymbol{X}_i, U_i)}\right],$$

  where $\hat{e}_\eta$ can be obtained by plugging in an estimator of $\boldsymbol{\beta}_0$.

- Computing extrema of $\hat{\beta}(\eta)$ is a **linear fractional programming**: Let $h_i = \exp\{-\eta(\boldsymbol{X}_i, U_i)\}$ and $g_i = 1/e_{\hat{\boldsymbol{\beta}}_0}(\boldsymbol{X}_i)$,

$$\begin{aligned} \text{max or min} \quad & \frac{\sum_{i=1}^{n} A_i Y_i [1 + h_i(g_i - 1)]}{\sum_{i=1}^{n} A_i [1 + h_i(g_i - 1)]}, \\ \text{subject to} \quad & h_i \in [\Gamma^{-1}, \Gamma],\ i = 1, \dots, n. \end{aligned}$$

  This can be converted to a linear programming and can in fact be solved in $O(n)$ time (optimal rate).

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 $E_M(\sqrt{\Gamma}) \subseteq E_R(\Gamma) \subseteq E_M(\Gamma)$.

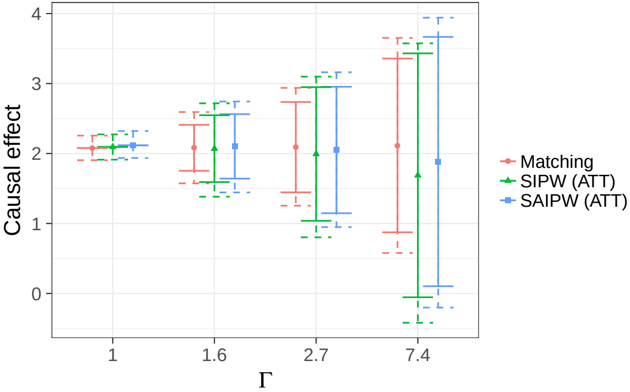


Figure: The solid error bars are the range of point estimates and the dashed error bars (together with the solid bars) are the confidence intervals. The circles/triangles/squares are the mid-points of the solid bars.

Recap

- **Sensitivity model** = Overparameterizing the full data distribution.
- Understand sensitivity models by visualizing their **observational equivalence** classes.
- **Point identified** versus **partially identified** inference.
- Percentile bootstrap can greatly simplify the problem.
- Example: Marginal sensitivity model & the IPW estimator.

References

1. Sensitivity analysis for inverse probability weighting estimators via the percentile bootstrap. *J Roy Stat Soc B*, 81(4) 735–761, 2019.
   - Joint work with Dylan Small and Bhaswar Bhattacharya.
   - `R` package: `https://github.com/qingyuanzhao/bootsens`.
2. Sensitivity analysis for observational studies: Principles, models, methods, and practice.
   - Ongoing work with Bo Zhang, Ting Ye, Joe Hogan, Dylan Small.

Further references

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

A. Saltelli, S. Tarantola, F. Campolongo, and M. Ratto. *Sensitivity analysis in practice: A guide to assessing scientific models*. John Wiley & Sons, Ltd, 2004.

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

Thank you!