Sensitivity analysis for observational studies: Past, Present, and Future

Qingyuan Zhao

Statistical Laboratory, University of Cambridge

February 13, 2024 @ University of East Anglia

Sensitivity analysis

Sensitivity analysis is widely used in many areas that use mathematical models.

The broader concept (Saltelli et al. 2004)

- "The study of how **the uncertainty in the output** of a mathematical model or system 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

- Making causal inference from observational studies require **untestable assumptions**.
- So a typical question is:

  How do the qualitative and/or quantitative conclusions if the identification assumptions (e.g. no unmeasured confounding) are violated?

Sensitivity analysis

Sensitivity analysis is widely used in many areas that use mathematical models.

The broader concept (Saltelli et al. 2004)

- "The study of how **the uncertainty in the output** of a mathematical model or system 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

- Making causal inference from observational studies require **untestable assumptions**.
- So a typical question is:

  How do the qualitative and/or quantitative conclusions if the identification assumptions (e.g. no unmeasured confounding) are violated?

Cornfield et al. (1959)

- The first (and probably most important) sensitivity analysis.

$$\begin{tikzcd}
 & U \text{ (Hormone X)} \arrow[dl] \arrow[dr] & \\
A \text{ (Smoking)} & & Y \text{ (Lung cancer)}
\end{tikzcd}$$

- Cornfield et al. (1959) showed that if $A \perp\!\!\!\perp Y \mid U$, then

$$\frac{\mathbb{P}(U = 1 \mid A = 1)}{\mathbb{P}(U = 1 \mid A = 0)} \geq \frac{\mathbb{P}(Y = 1 \mid A = 1)}{\mathbb{P}(Y = 1 \mid A = 0)}.$$

Rosenbaum's sensitivity model

- Limitations of the Cornfield bound: A, U, Y must be binary; no covariates X.
- Rosenbaum (1987) assume $A \perp\!\!\!\perp Y(a) \mid \boldsymbol{X}, U$ and for some $\Gamma \geq 1$,

$$\frac{1}{\Gamma} \leq \frac{\mathbb{P}(A=1 \mid \boldsymbol{X}=\boldsymbol{x}, U=u)/\mathbb{P}(A=0 \mid \boldsymbol{X}=\boldsymbol{x}, U=u)}{\mathbb{P}(A=1 \mid \boldsymbol{X}=\boldsymbol{x}, U=u')/\mathbb{P}(A=0 \mid \boldsymbol{X}=\boldsymbol{x}, U=u')} \leq \Gamma, \ \forall \boldsymbol{x}, u.$$

- In this and a subsequence series of work, Rosenbaum considered randomization inference under this model. Suppose units are pair matched so that for ith pair $\boldsymbol{X}_{i1} = \boldsymbol{X}_{i2}$ and $A_{i1} + A_{i2} = 1$. Then within a pair, the treatment is assigned by a possibly biased coin flip:

$$\frac{1}{1+\Gamma} \leq \mathbb{P}(A_{i1} = 1 \mid A_{i1} + A_{i2} = 1, \boldsymbol{X}_{i1}, \boldsymbol{X}_{i2}, U_{i1}, U_{i2}) \leq \frac{\Gamma}{1+\Gamma}.$$

- Rosenbaum used the FKG/Holley correlation inequality on distributive lattices to derive sharp upper and lower bounds on randomization p-value; see Rosenbaum (2002).

Marginal sensitivity models

- Denote $\pi(X) = \mathbb{P}(A = 1 \mid X)$, $\pi(X, U) = \mathbb{P}(A = 1 \mid X, U)$, and $h(X, U) = \pi(X)/\pi(X, U)$.
- Tan (2006) considered the following sensitivity model:

$$\Gamma^{-1} \leq \frac{\pi(X)/\{1 - \pi(X)\}}{\pi(X, U)/\{1 - \pi(X, U)\}} \leq \Gamma, \quad \Gamma \geq 1.$$

- Zhao, Small, and Bhattacharya (2019) rediscovered it and called it "marginal sensitivity model", in relation to the celebrated Rosenbaum sensitivity model.
- Dorn and Guo (2022) obtained sharp lower and upper bounds of the average treatment effect under this model.

Scharfstein, Rotnitzky, and Robins (1999)

- A closely related problem is sensitivity analysis for **non-ignorable dropout**.
- Let Q be drop-out time and $\bar{\boldsymbol{V}}(t)$ be all variables collected at or before time $0 \le t \le T$. An outcome variable Y is collected at time T if $Q \ge T$. Let λ be the hazard function of Q:

$$\lambda(t) = \lim_{\epsilon \downarrow 0} \frac{\mathbb{P}(t \le Q \le t + \epsilon)}{\epsilon\, \mathbb{P}(Q \ge t)}.$$

- Scharfstein, Rotnitzky, and Robins (1999) considered the following sensitivity model for the conditional harzard function:

$$\lambda(t \mid \bar{\boldsymbol{V}}(T)) = \lambda(t \mid \bar{\boldsymbol{V}}(t)) \exp(\alpha Y),\ 0 \le t < T.$$

- By treating this as a semiparametric statistical model, they proposed estimators of $\mathbb{E}(Y)$ with any given α.

Omitted variable bias

- Many sensitivity analysis methods are motivated by the following simple observation, which dates back to at least Cochran (1938).

$$\begin{tikzcd} & U \arrow[dl] \arrow[dr] & \\ A \arrow[rr] & & Y \end{tikzcd}$$

- Let $\mathrm{coef}(Y \sim A + U)\$A$ denote the least squares coefficient of A in the linear regression of Y on A and U. Then

$$\mathrm{coef}(Y \sim A)\$A = \mathrm{coef}(Y \sim A + U)\$A + \mathrm{coef}(U \sim A)\$A \cdot \mathrm{coef}(Y \sim A + U)\$U.$$

- But this relies quite heavily on linearity.

Get ready for a slippery ride...

Three components of sensitivity analysis

1. **Model augmentation:** Extend the model used by primary analysis to allow for unmeasured confounding or other forms of violations.
2. **Statistical inference:** Vary the sensitivity parameter, estimate the causal effect, and control suitable statistical errors.
3. **Interpretation of the results:** Probe different "directions" of unmeasured confounding and "make sense" of the results.

Each step can be challenging!

Example: Child soldiering

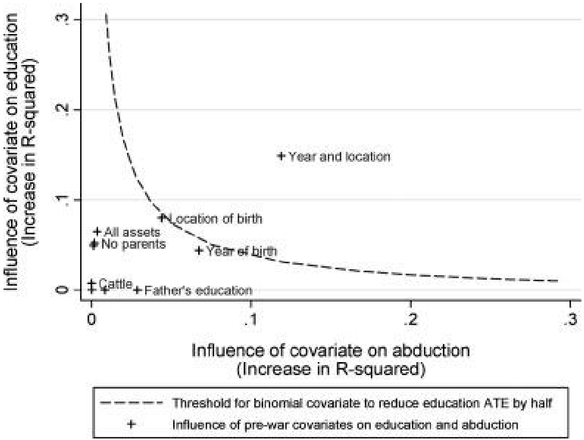
- About $60,000$ to $80,000$ youths were abducted in Uganda by a rebel force in 1995-2004.
- Question: What is the impact of child soldiering on the years of education?
- Blattman and Annan (2010) controlled for a variety of covariates $\boldsymbol{X}$ (age, household size, parental education, etc.) but were concerned about a unmeasured confounder U (ability to hide from the rebel).
- They used the following model proposed by Imbens (2003):

$$
\begin{aligned}
&A \perp\!\!\!\perp Y(a) \mid \boldsymbol{X}, U, \ \text{for } a = 0, 1, \\
&U \mid \boldsymbol{X} \sim \text{Bernoulli}(0.5), \\
&A \mid \boldsymbol{X}, U \sim \text{Bernoulli}(\text{expit}(\boldsymbol{\kappa}^T \boldsymbol{X} + \lambda U)), \\
&Y(a) \mid \boldsymbol{X}, U \sim \text{N}(\beta a + \boldsymbol{\nu}^T \boldsymbol{X} + \delta U, \sigma^2) \ \text{for } a = 0, 1,
\end{aligned}
$$

- A is the exposure and $Y(a)$ is the potential outcome.
- (λ, δ) are sensitivity parameters; $\lambda = \delta = 0$ corresponds to a primary analysis assuming no unmeasured confounding.

Main results of Blattman and Annan (2010)

- Their primary analysis found that the ATE is -0.76 (s.e. 0.17).
- Sensitivity analysis was summarized by a single calibration plot:



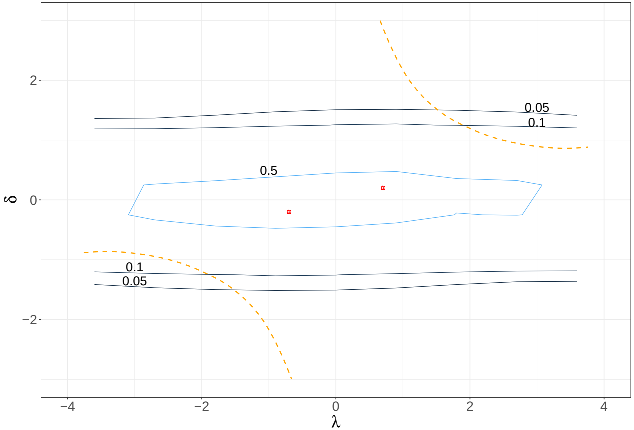
Some issues with the last analysis

Recall the model:

$$
\begin{aligned}
&A \perp\!\!\!\perp Y(a) \mid \boldsymbol{X}, U, \text{ for } a = 0, 1,\\
&U \mid \boldsymbol{X} \sim \text{Bernoulli}(0.5),\\
&A \mid \boldsymbol{X}, U \sim \text{Bernoulli}(\text{expit}(\boldsymbol{\kappa}^T \boldsymbol{X} + \lambda U)),\\
&Y(a) \mid \boldsymbol{X}, U \sim \text{N}(\beta a + \boldsymbol{\nu}^T \boldsymbol{X} + \delta U, \sigma^2) \text{ for } a = 0, 1,
\end{aligned}
$$

- Model augmentation: (λ, δ) are identifiable, so incoherent to vary them.
- Interpretation: In the calibration plot, partial R^2 for observed and unobserved confounders are not directly comparable because they use different reference models.
- Statistical inference: Cinelli and Hazlett (2020) proposed a nice solution to the last problem, but their heuristic sensitivity interval achieves no confidence guarantees.

Visualization the identifiability of (λ, δ)



Red dots: maximum likelihood estimate. Solid curves: contours of likelihood ratio test. Dashed curves: estimated ATE reduced by half.

Rest of the talk: Some recent work from my group

1. A general formulation of sensitivity analysis as a stochastic program.
2. Example 1: Linear models with no closed-form solutions.
3. Example 2: Nonparametric models with closed-form solutions.

Sensitivity analysis as a stochastic program

- Consider an i.i.d. sample $(V_i, U_i) \sim \mathsf{P}_{V,U}$ but only V_i is observed, $i = 1, \ldots, n$.
- In the last example, $V = (A, \boldsymbol{X}, Y)$ and $U = (Y(0), Y(1))$.
- Interested in a real-valued "causal parameter/functional" $\beta = \beta(\mathsf{P}_{V,U}) = \beta(\theta, \psi)$ where $\theta = \theta(\mathsf{P}_V)$ is identifiable and $\psi = \psi(\mathsf{P}_{V,U})$ is not.

Sensitivity analysis

- A **sensitivity model** constrains the non-identifiable parameter: $\psi \in \Psi(\theta)$.
- This leads to a **partially identified region** for β, which can be bounded by solving

$$\begin{aligned} \text{minimize/maximize} \quad & \beta(\theta(\mathsf{P}_V), \psi) \\ \text{subject to} \quad & \psi \in \Psi(\theta(\mathsf{P}_V)). \end{aligned}$$

Challenges

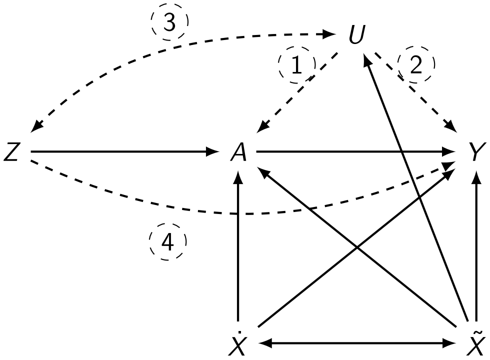
1. How to specify interpretable Ψ?
2. How to solve this optimization problem with a finite sample?
3. How to make inference for β and its partially identified region?

Example 1: Linear models with no closed-form solutions

Tobias Freidling and Qingyuan Zhao (2022). *Optimization-based Sensitivity Analysis for Unmeasured Confounding using Partial Correlations*. arXiv: 2301.00040 [stat.ME]

- Key idea: Specify an interpretable sensitivity model with the help of causal graphical models and the R^2-calculus.

Causal graph



- Regression analysis assumes no 1 and 2.
- Instrumental variable analysis assumes no 3 and 4.

Proposed method

- Objective function is the reparametrized omitted variable bias formula (Cinelli and Hazlett 2020; Hosman, Hansen, and Holland 2010)

$$\beta_{Y\sim A|X,Z,U} = \beta_{Y\sim A|X,Z} - R_{Y\sim U|X,Z,A}\, f_{A\sim U|X,Z}\frac{\sigma_{Y\sim X+Z+A}}{\sigma_{A\sim X+Z}}.$$

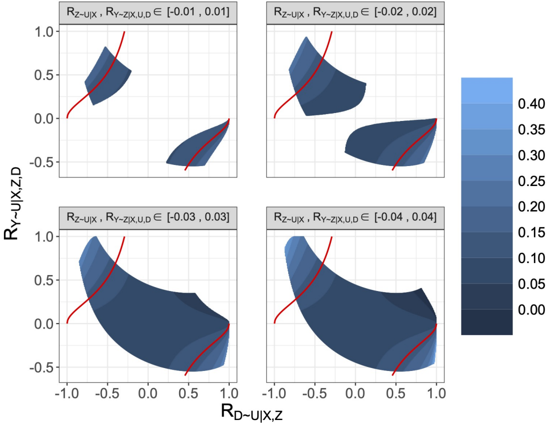
- Sensitivity models are formed by a mix-and-match of following constraints.

$$\begin{array}{lll}
(1)\ U \to A & 1. & R_{A\sim U|X,Z} \in [B^l_{UA}, B^u_{UA}] \\
 & 2. & R^2_{A\sim U|\tilde{X},\dot{X}_I,Z} \le b_{UA} R^2_{A\sim \dot{X}_J|\tilde{X},\dot{X}_I,Z} \\
(2)\ U \to Y & 1. & R_{Y\sim U|X,Z,A} \in [B^l_{UY}, B^u_{UY}] \\
 & 2. & R^2_{Y\sim U|\tilde{X},\dot{X}_I,Z} \le b_{UY} R^2_{Y\sim \dot{X}_J|\tilde{X},\dot{X}_I,Z} \\
 & 3. & R^2_{Y\sim U|\tilde{X},\dot{X}_I,Z,A} \le b_{UY} R^2_{Y\sim \dot{X}_J|\tilde{X},\dot{X}_I,Z,A} \\
(3)\ U \leftrightarrow Z & 1. & R_{Z\sim U|X} \in [B^l_{UZ}, B^u_{UZ}] \\
 & 2. & R^2_{Z\sim U|\tilde{X},\dot{X}_{-j}} \le b_{UZ} R^2_{Z\sim \dot{X}_j|\tilde{X},\dot{X}_{-j}} \\
(4)\ Z \to Y & 1. & R_{Y\sim Z|X,U,A} \in [B^l_{ZY}, B^u_{ZY}] \\
 & 2. & R^2_{Y\sim Z|X,U,A} \le b_{ZY} R^2_{Y\sim \dot{X}_j|\tilde{X},\dot{X}_{-j},Z,U,A}
\end{array}$$

- R partial correlation, $f = R/\sqrt{1-R^2}$, β least-squares coefficient, σ residual variance.

Stochastic optimization is not simple

- The optimization problem is generally nonlinear and non-convex (example below). A grid-search algorithm tailored for this problem was developed in the paper.



- Asymptotic normality is expected but requires further theoretical development.

Example 2: Nonparametric models with closed-form solutions.

Yao Zhang and Qingyuan Zhao (2022). "Sharp Bounds and Semiparametric Inference in L^∞- and L^2-sensitivity Analysis for Observational Studies". In: arXiv: `2211.04697 [stat.ME]`

- Key idea: Identify variational optimization problems that admit closed-form solutions.

Marginal sensitivity analysis (L^∞)

- Recall the L^∞ marginal sensitivity model
$$\Gamma^{-1} \leq \frac{\pi(X)/\{1-\pi(X)\}}{\pi(X,U)/\{1-\pi(X,U)\}} \leq \Gamma, \quad \Gamma \geq 1.$$
- $\pi(X) = \mathbb{P}(A = 1 \mid X)$, $\pi(X,U) = \mathbb{P}(A = 1 \mid X, U)$, and $h(X,U) = \pi(X)/\pi(X,U)$.
- For simplicity, let $U = Y(1)$. Bounds for $\beta(X) = \mathrm{E}\{Y(1) \mid X\}$ can be obtained by solving
$$\begin{aligned} \text{minimize/maximize} \quad & \mathrm{E}\{h(X,Y)Y \mid X, A=1\} \\ \text{subject to} \quad & (1-\Gamma^{-1})\pi(X) + \Gamma^{-1} \leq h(X,Y) \leq (1-\Gamma)\pi(X) + \Gamma, \\ & \mathrm{E}\{h(X,Y) \mid X, A=1\} = 1. \end{aligned}$$
- Zhao, Small, and Bhattacharya (2019) ignored the equality constraint and proposed a percentile bootstrap method.
- Dorn and Guo (2022) obtained exact solutions to this linear program by using the Neyman-Pearson lemma.

Marginal sensitivity analysis (L^2)

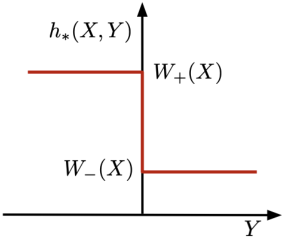
- Key idea: One can also constrain the degree of unmeasured confounding by $\mathrm{Var}(h(X, U))$.
- This leads to a quadratic program:

$$\begin{aligned} \text{minimize/maximize} \quad & \mathrm{E}\{h(X,Y)Y \mid X, A=1\} \\ \text{subject to} \quad & \mathrm{E}\{h(X,Y)^2 \mid X, A=1\} \le \Gamma, \\ & \mathrm{E}\{h(X,Y) \mid X, A=1\} = 1, \\ & h(X,Y) \ge \pi(X). \end{aligned}$$

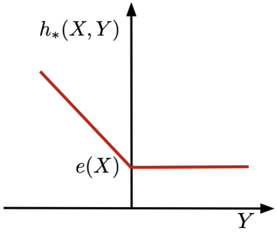
- Closed-form solutions are found in the paper by considering the Lagrangian problem.
- The paper also derived the influence function for the bounds of the L^∞- and L^2-sensitivity problems and discussed how to estimate them.[1]
- Calibration using observed confounders is much easier with the L^2 model than the L^∞ model.

[1] An alternative form of the influence function for the L^∞ problem was obtained by Dorn, Guo, and Kallus (2021).

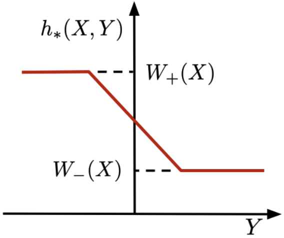
Schematic for the solution



(a) L^∞ constraint (linear program).



(b) L^2 constraint (quadratic program).



(c) L^∞ and L^2 constraints (quadratic program).

Future directions

- Recent work have revealed connections between sensitivity analysis in statistics and stochastic optimization.
- There is a wide spectrum of problems between the last two examples:
  1. Linear models with no closed-form solutions.
  2. Nonparametric models with closed-form solutions.
- Many difficult challenges:
  1. General asymptotic (and effiency?) theory for stochastic programs with possibly infinite-dimensional parameters.
  2. Semiparametric estimators with nuisance parameters/functions obtained by machine-learning.
  3. Theory for bootstrap confidence intervals.
  4. Numerical solutions to the nonlinear and nonconvex stochastic programs.
  5. Toolbox for specifying interpretable sensitivity parameters and models.
  6. Integration with graphical models and existing algorithms for causal identification.

References I

Blattman, Christopher and Jeannie Annan (2010). “The Consequences of Child Soldiering”. In: *The Review of Economics and Statistics* 92.4, pp. 882–898.

Cinelli, Carlos and Chad Hazlett (2020). “Making sense of sensitivity: Extending omitted variable bias”. In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 82.1, pp. 39–67.

Cochran, William G (1938). “The omission or addition of an independent variate in multiple linear regression”. In: *Supplement to the Journal of the Royal Statistical Society* 5.2, pp. 171–176.

Cornfield, J. et al. (1959). “Smoking and lung cancer”. In: *Journal of the National Cancer Institute* 22, pp. 173–203.

Dorn, Jacob and Kevin Guo (2022). “Sharp Sensitivity Analysis for Inverse Propensity Weighting Via Quantile Balancing”. In: *Journal of the American Statistical Association*, in press. DOI: 10.1080/01621459.2022.2069572.

Dorn, Jacob, Kevin Guo, and Nathan Kallus (2021). “Doubly-Valid/doubly-Sharp Sensitivity Analysis for Causal Inference With Unmeasured Confounding”. In: arXiv: 2112.11449 [stat.ME].

References II

- Freidling, Tobias and Qingyuan Zhao (2022). *Optimization-based Sensitivity Analysis for Unmeasured Confounding using Partial Correlations*. arXiv: 2301.00040 [stat.ME].
- Hosman, Carrie A., Ben B. Hansen, and Paul W. Holland (2010). "The sensitivity of linear regression coefficients' confidence limits to the omission of a confounder". In: *The Annals of Applied Statistics* 4.2. ISSN: 1932-6157. (Visited on 11/29/2021).
- Imbens, G. W. (2003). "Sensitivity to exogeneity assumptions in program evaluation". In: *American Economic Review* 93, pp. 126–132.
- Rosenbaum, P. R. (1987). "Sensitivity analysis for certain permutation inferences in matched observational studies". In: *Biometrika* 74, pp. 13–26.
- — (2002). *Observational Studies*. Springer.
- Saltelli, Andrea et al. (2004). *Sensitivity analysis in practice: A guide to assessing scientific models*. John Wiley & Sons, Ltd.
- Scharfstein, Daniel O, Andrea Rotnitzky, and James M Robins (1999). "Adjusting for nonignorable drop-out using semiparametric nonresponse models". In: *Journal of the American Statistical Association* 94.448, pp. 1096–1120.
- Tan, Zhiqiang (2006). "A distributional approach for causal inference using propensity scores". In: *Journal of the American Statistical Association* 101.476, pp. 1619–1637.

References III

Zhang, Yao and Qingyuan Zhao (2022). "Sharp Bounds and Semiparametric Inference in L^∞- and L^2-sensitivity Analysis for Observational Studies". In: arXiv: 2211.04697 [stat.ME].

Zhao, Qingyuan, Dylan S. Small, and Bhaswar B. Bhattacharya (2019). "Sensitivity Analysis for Inverse Probability Weighting Estimators Via the Percentile Bootstrap". In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 81.4, pp. 735–761. DOI: 10.1111/rssb.12327.