观察性研究

① $Z \perp\!\!\!\perp (Y(1), Y(0)) \mid X$ — 技术性

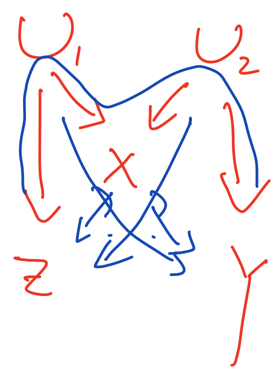
② $Z \not\perp\!\!\!\perp (Y(1), Y(0)) \mid X$ — 概念性

DE GRUYTER J. Causal Infer. 2015; 3(1): 41–57

Peng Ding* and Luke W. Miratrix

To Adjust or Not to Adjust? Sensitivity Analysis of *M*-Bias and Butterfly-Bias

Pearl 之启发



Abstract: "*M*-Bias," as it is called in the epidemiologic literature, is the bias introduced by conditioning on a pretreatment covariate due to a particular "*M*-Structure" between two latent factors, an observed treatment, an outcome, and a "collider." This potential source of bias, which can occur even when the treatment and the outcome are not confounded, has been a source of considerable controversy. We here present formulae for identifying under which circumstances biases are inflated or reduced. In particular, we show that the magnitude of *M*-Bias in linear structural equation models tends to be relatively small compared to confounding bias, suggesting that it is generally not a serious concern in many applied settings. These theoretical results are consistent with recent empirical findings from simulation studies. We also generalize the *M*-Bias setting (1) to allow for the correlation between the latent factors to be nonzero and (2) to allow for the collider to be a confounder between the treatment and the outcome. These results demonstrate that mild deviations from the *M*-Structure tend to increase confounding bias more rapidly than *M*-Bias, suggesting that choosing *to* condition on any given covariate is generally the superior choice. As an application, we re-examine a controversial example between Professors Donald Rubin and Judea Pearl.

Keywords: causality, collider, confounding, controversy, covariate

针对 Rosenbaum & Rubin

$e(x) = Pr(Z=1 \mid X)$

Biometrika (2017), **104**, 2, *pp.* 291–302
Printed in Great Britain
doi: 10.1093/biomet/asx009
Advance Access publication 17 April 2017

不局限于线性模型

Instrumental variables as bias amplifiers with general outcome and confounding

By P. DING

Department of Statistics, University of California, 425 Evans Hall, Berkeley, California 94720, U.S.A.
pengdingpku@berkeley.edu

T. J. VANDERWEELE and J. M. ROBINS

Department of Epidemiology, Harvard T. H. Chan School of Public Health, 677 Huntington Avenue, Boston, Massachusetts 02115, U.S.A.
tvanderw@hsph.harvard.edu robins@hsph.harvard.edu

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

Instrumental Variable (IV)

Summary

Drawing causal inference with observational studies is the central pillar of many disciplines. One sufficient condition for identifying the causal effect is that the treatment-outcome relationship is unconfounded conditional on the observed covariates. It is often believed that the more covariates we condition on, the more plausible this unconfoundedness assumption is. This belief has had a huge impact on practical causal inference, suggesting that we should adjust for all pretreatment covariates. However, when there is unmeasured confounding between the treatment and outcome, estimators adjusting for some pretreatment covariate might have greater bias than estimators that do not adjust for this covariate. This kind of covariate is called a bias amplifier, and includes instrumental variables that are independent of the confounder and affect the outcome only through the treatment. Previously, theoretical results for this phenomenon have been established only for linear models. We fill this gap in the literature by providing a general theory, showing that this phenomenon happens under a wide class of models satisfying certain monotonicity assumptions.

调整X增加偏差

Some key words: Causal inference; Directed acyclic graph; Interaction; Monotonicity; Potential outcome.

$$\begin{tikzcd} & X \arrow[dl] \\ Z & U \arrow[l] \\ & \varepsilon_Z \arrow[ul] \end{tikzcd}$$

The treatment assignment is a function of the instrumental variable, the unmeasured confounder and some other independent random error, which are the three sources of variation of the treatment. If we adjust for the instrumental variable, the treatment variation is driven more by the unmeasured confounder, which could result in increased bias due to this confounder. Seemingly paradoxically, without adjusting for the instrumental variable, the observational study is more like a randomized experiment, and the bias due to confounding is smaller. Although applied researchers (Myers et al., 2011; Walker, 2013; Brooks & Ohsfeldt, 2013; Ali et al., 2014) have confirmed through extensive simulation studies that this bias amplification phenomenon exists in a wide range of reasonable models, definite theoretical results have been established only for linear models. We fill this gap in the literature by showing that adjusting for an instrumental variable amplifies bias for estimating causal effects under a wide class of models satisfying certain monotonicity assumptions. However, we also show that there exist data-generating processes under which an instrumental variable is not a bias amplifier.

Biometrika (2014), **101**, 4, *pp.* 971–977
Printed in Great Britain
doi: 10.1093/biomet/asu030
Advance Access publication 4 August 2014

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

流行病学家

Generalized Cornfield conditions for the risk difference

$RR_{ZU} \geqslant RR_{ZY}^{obs}$

BY PENG DING

Department of Statistics, Harvard University, One Oxford Street, Cambridge, Massachusetts 02138, U.S.A.
pengding@fas.harvard.edu

AND TYLER J. VANDERWEELE

Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts 02115, U.S.A.
tvanderw@hsph.harvard.edu

SUMMARY

A central question in causal inference with observational studies is the sensitivity of conclusions to unmeasured confounding. The classical Cornfield condition allows us to assess whether an unmeasured binary confounder can explain away the observed relative risk of the exposure on the outcome. It states that for an unmeasured confounder to explain away an observed relative risk, the association between the unmeasured confounder and the exposure and the association between the unmeasured confounder and the outcome must both be larger than the observed relative risk. In this paper, we extend the classical Cornfield condition in three directions. First, we consider analogous conditions for the risk difference and allow for a categorical, not just a binary, unmeasured confounder. Second, we provide more stringent thresholds that the maximum of the above-mentioned associations must satisfy, rather than weaker conditions that both must satisfy. Third, we show that all the earlier results on Cornfield conditions hold under weaker assumptions than previously used. We illustrate the potential applications by real examples, where our new conditions give more information than the classical ones.

Some key words: Causal inference; Confounding; Observational study; Sensitivity analysis.

RD 有类似结果
用处不大 因为
结果依赖于
U的取值

流行病学家

Lee (2011) obtained the above results (3) and (4) under Assumption 3, which can in fact be weakened to Assumption 2. Furthermore, in the Supplementary Material, we show that under Assumption 1, the following conditions must hold:

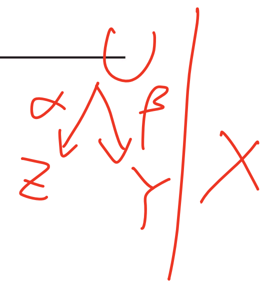
$$\min(U_E, U'_D) \geqslant \mathrm{RR}_{ED}, \quad \max(U_E, U'_D) \geqslant \{\mathrm{RR}_{ED}^{1/2} + (\mathrm{RR}_{ED} - 1)^{1/2}\}^2,$$

where $U'_D = \max(U_D, U^*_D)$ replaces U_D in conditions (3) and (4).

Reviewer 启发了

E-value

OPEN



Sensitivity Analysis Without Assumptions

Peng Ding[a] *and Tyler J. VanderWeele*[b]

Theorem 17.1 若 $Z \perp\!\!\!\perp Y \mid X, U$

则 $RR^{obs}_{ZY|X} \le \frac{\alpha\beta}{\alpha+\beta-1}$

$\Downarrow$

其中 $\alpha = RR_{ZU|X}$

定义 3

$\beta = RR_{UY|X}$

E-value

The claim that our technique is "without assumptions" requires some clarification. As we will see below, we will, without any assumptions, be able to make statements of the form: "For an observed association to be due solely to unmeasured confounding, two sensitivity analysis parameters must satisfy [a specific inequality]." We will also, without assumptions, be able to make statements of the form: "For unmeasured confounding alone to be able to reduce an observed association [to a given level], two sensitivity analysis parameters must satisfy [another specific inequality]." We believe the ability to make statements of this form without imposing any specific structure on the nature of the unmeasured confounder or confounders constitutes a major advance in the literature.

Chapter 16 HW

统计 Cochran's formula

计量 omitted-variable bias formula

$$\begin{cases} y = \beta_1^T x_1 + \beta_2^T x_2 + \varepsilon \\ y = \tilde{\beta}_1^T x_1 + e \end{cases}$$

$$\Rightarrow \tilde{\beta}_1 - \beta_1 = \delta \beta_2$$

$$\begin{tikzcd} & x_2 \arrow[dr, "\beta_2"] & \\ x_1 \arrow[ur, "\delta"] \arrow[rr, "\beta_1"'] & & y \end{tikzcd} \qquad x_2 = \delta^T x_1 + v$$

信息论也常见

Data processing inequality

Lihua Lei

Bin Yu

若 $Z \perp\!\!\!\perp Y \mid U$

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

则

$$I(Z, Y) \le I(Z, U)$$

$$I(Z, Y) \le I(U, Y)$$

其中 $I(A, B) = \iint p(a,b) \log_2 \frac{p(a,b)}{p(a)p(b)} \, da \, db$

mutual information

定理17.1 的初等证明:

Z, Y, U 二值

X 给定: 省略

$$RR^{obs}_{ZY} \overset{\text{定义}}{=} \frac{Pr(Y=1 \mid Z=1)}{Pr(Y=1 \mid Z=0)}$$

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

$$\overset{\text{全概公式}}{=} \frac{Pr(Y=1 \mid \cancel{Z=1}, U=1)\, Pr(U=1 \mid Z=1) + Pr(Y=1 \mid \cancel{Z=1}, U=0)\, Pr(U=0 \mid Z=1)}{Pr(Y=1 \mid \cancel{Z=0}, U=1)\, Pr(U=1 \mid Z=0) + Pr(Y=1 \mid \cancel{Z=0}, U=0)\, Pr(U=0 \mid Z=0)}$$

$$\overset{Z \perp\!\!\!\perp Y \mid U}{=\!=} \frac{Pr(Y=1 \mid U=1)\, \overbrace{Pr(U=1 \mid Z=1)}^{f_1} + \underline{\underline{Pr(Y=1 \mid U=0)}}\, \overbrace{Pr(U=0 \mid Z=1)}^{1-f_1}}{Pr(Y=1 \mid U=1)\, \underbrace{Pr(U=0 \mid Z=0)}_{f_0} + \underbrace{\underline{\underline{Pr(Y=1 \mid U=0)}}}_{\text{除此项}}\, \underbrace{Pr(U=0 \mid Z=0)}_{1-f_0}}$$

依然精确

$$= \frac{RR_{UY}\, f_1 + 1 - f_1}{RR_{UY}\, f_0 + 1 - f_0}$$

$$= \frac{(RR_{UY} - 1)\, f_1 + 1}{\frac{RR_{UY} - 1}{RR_{ZU}}\, f_1 + 1}$$

重新参数化

关于 f_1

若 $RR_{UY} > 1$

$RR_{ZU} > 1$

$f_1 \to 1$

$$\frac{RR_{UY} \cancel{-1} \cancel{+1}}{\frac{RR_{UY} - 1}{RR_{ZU}} + 1}$$

$$= \frac{RR_{ZU} \cdot RR_{UY}}{RR_{ZU} + RR_{UY} - 1} \quad \square$$

$$\Rightarrow \quad RR_{ZU} \geq RR_{ZY}^{obs}$$

$$RR_{UY} \geq RR_{ZY}^{obs}$$

$$\max\left(RR_{ZU},\ RR_{UY}\right) \geq E\text{-value}$$

$$= RR_{ZY}^{obs} + \sqrt{RR_{ZY}^{obs}\left(RR_{ZY}^{obs} - 1\right)}$$

流行病学. 病例－对照研究

case - control study

(case - referent study)

A 2.6.3) 之

$$\Pr\left(S=1 \mid x, y\right) = \Pr(S=1 \mid y)$$

x ↓ 协变量　y ↓ 结果

不依

$$\left(x_i, y_i - S_i = 1\right)_{i=1}^{n} : \cancel{iid}$$

Cornfield

Ross Prentice : logistic regression

结论: logistic 对 $\begin{cases} \text{iid} \\ \text{case-control} \end{cases}$

都对

$$\Pr(Y=1 \mid Z, X) = \frac{e^{\beta_0+\beta_1 Z+\beta_2^T X}}{1+e^{\beta_0+\beta_1 Z+\beta_2^T X}}$$

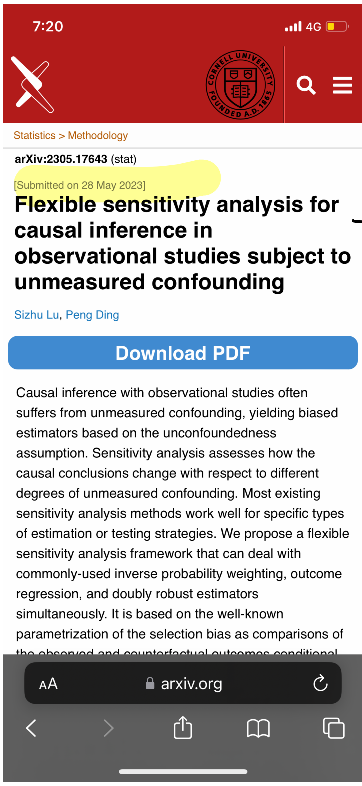
$\Rightarrow \beta_1 =$ conditional $OR_{ZY|X}$

rare disease

$\approx$ conditional $RR_{ZY|X}$

关于 U 的敏感性分析
sensitivity analysis
(Rosenbaum & Rubin
(1983 JRSSB)
比较参数化的做法

Chapter 18 关于 U
非常新的一章
疫情第一年 教的
目的：一个极其简单的方法
我推荐


7:20 4G

Statistics > Methodology

arXiv:2305.17643 (stat)

[Submitted on 28 May 2023]

Flexible sensitivity analysis for causal inference in observational studies subject to unmeasured confounding

Sizhu Lu, Peng Ding

Download PDF

Causal inference with observational studies often suffers from unmeasured confounding, yielding biased estimators based on the unconfoundedness assumption. Sensitivity analysis assesses how the causal conclusions change with respect to different degrees of unmeasured confounding. Most existing sensitivity analysis methods work well for specific types of estimation or testing strategies. We propose a flexible sensitivity analysis framework that can deal with commonly-used inverse probability weighting, outcome regression, and doubly robust estimators simultaneously. It is based on the well-known parametrization of the selection bias as comparisons of

AA arxiv.org

理论基础

Chapter 18 核心结果

找中间点

ignorability ? ? ? ? ? bounds

$$E(Y(1)\mid Z=1,X) = E(Y(1)\mid Z=0,X)$$

$$E(Y(0)\mid Z=1,X) = E(Y(0)\mid Z=0,X)$$

Manski

想法

上下界

Manski

$$\tau = E\left(Y(1) - Y(0)\right)$$

$$= \left[E(\overset{(1)}{Y}|Z=1)Pr(Z=1) + E(Y(1)|Z=0)Pr(Z=0)\right]$$ 忽略 X

$$- \left[E(Y(0)|Z=1)Pr(Z=1) + E(\overset{(0)}{Y}|Z=0)Pr(Z=0)\right]$$

反事实

容易

若 $Y \in [l, u]$，则 τ 有上下界

理想状态:

$$E(Y(1)|Z=0) = E\left\{E(Y|Z=1, X) \,\middle|\, Z=0\right\}$$

$$E(Y(0)|Z=1) = E\left\{E(Y|Z=0, X) \,\middle|\, Z=1\right\}$$

定义 敏感性参数

$$\frac{\overbrace{E(Y(1) \mid Z=1, X)}^{=\mu_1(X)}}{E(Y(1) \mid Z=0, X)} = \varepsilon_1(X)$$

$$\frac{E(Y(0) \mid Z=1, X)}{\underbrace{E(Y(0) \mid Z=0, X)}_{=\mu_0(X)}} = \varepsilon_0(X)$$

$\Rightarrow$ 推导出一系列结果

比 $Z \perp\!\!\!\perp (Y(1), Y(0)) \mid X$ 下之结果要广

当 $\varepsilon_1(X) = \varepsilon_0(X) = 1$ 时，所有结果

和 Part III 一样

定理 18.1 固定 $\varepsilon_1(x)$, $\varepsilon_0(x)$.

$$\tau = E\left\{ Z\mu_1(x) + (1-Z)\mu_1(x)/\varepsilon_1(x) \right\} - E\left\{ Z\mu_0(x)\varepsilon_0(x) + (1-Z)\mu_0(x) \right\}$$

其中 $\mu_1(x) = E(Y \mid Z=1, X)$

$\mu_0(x) = E(Y \mid Z=0, X)$

$$= E\left(w_1(X) \frac{ZY}{e(X)} \right) - E\left(w_0(X) \frac{(1-Z)Y}{1-e(X)} \right)$$

其中 $w_1(X) = e(X) + \dfrac{1-e(X)}{\varepsilon_1(X)}$

$w_0(X) = e(X)\varepsilon_0(X) + 1 - e(X)$

= 双稳健的理论也存在

见 Lu & Ding (2023)

$\hat{\tau}^{dr}$ 公式

$$\hat{\tau}^{ht} = \frac{1}{n}\sum_{i=1}^{n} \frac{\left(\hat{e}(X_i)\,\varepsilon_1(X_i) + 1 - \hat{e}(X_i)\right) Z_i Y_i}{\varepsilon_1(X_i)\,\hat{e}(X_i)}$$

$$- \frac{1}{n}\sum_{i=1}^{n} \frac{\left(\hat{e}(X_i)\,\varepsilon_0(X_i) + 1 - \hat{e}(X_i)\right)(1 - Z_i) Y_i}{1 - \hat{e}(X_i)}$$

$$\hat{\tau}^{dr} = \hat{\tau}^{ht} - \frac{1}{n}\sum_{i=1}^{n} \left(Z_i - \hat{e}(X_i)\right)\left(\frac{\hat{\mu}_1(X_i)}{\hat{e}(X_i)\,\varepsilon_1(X_i)} + \frac{\hat{\mu}_0(X_i)\,\varepsilon_0(X_i)}{1 - \hat{e}(X_i)}\right)$$

对于 ATT ： 平行理论

Chapter 19 Rosenbaum 的敏感性分析

观察性研究

1—1 匹配

精确匹配（有问题!!）

① 无混杂 $\Rightarrow$ MPE

$\Rightarrow$ FRT

② 有混杂？ $\not\Rightarrow$ MPE

$X_{i1} = X_{i2}$

$U_{i1} \neq U_{i2}$

$$\underset{\substack{\downarrow \\ i\text{对}}}{e}_{\underset{\substack{\downarrow \\ j=1,2}}{ij}} = \Pr\Big(Z_{ij}=1 \,\Big|\, \underset{\substack{\downarrow \\ \text{共同}}}{X_i}, \underbrace{Y_{ij}(1), Y_{ij}(0)}_{\substack{\text{潜在结果未知} \\ \text{也混杂因素}}}\Big)$$

$$\pi_{i1} = \Pr\big(Z_{i1}=1 \mid X_i, \mathcal{S}_i, Z_{i1}+Z_{i2}=1\big)$$

其中 $\mathcal{S}_i = \{Y_{i1}(1), Y_{i1}(0), Y_{i2}(1), Y_{i2}(0)\}$

$$= \frac{\Pr(Z_{i1}=1, Z_{i2}=0 \mid X_i, \mathcal{S}_i)}{\Pr(Z_{i1}+Z_{i2}=1 \mid X_i, \mathcal{S}_i)}$$

分母有两项

$$= \frac{e_{i1}(1-e_{i2})}{e_{i1}(1-e_{02}) + (1-e_{i1})e_{i2}}$$

定义 $= \dfrac{O_{i1}}{O_{i1}+O_{i2}}$

$O_{ij} = \dfrac{e_{ij}}{1-e_{ij}}$

理想case $O_{i1} = O_{i2}$

$\Rightarrow = \dfrac{1}{2}$

假设 $\Gamma^{-1} \le \dfrac{O_{i1}}{O_{i2}} \le \Gamma$ 与 logistic 回归系数有关

$$\Rightarrow \frac{1}{1+\Gamma} \le \pi_{i1} \le \frac{\Gamma}{1+\Gamma}$$

有: $p(T) =$ 最"差"情况的 p 值

$$T = \sum_{i=1}^{n} S_i Z_i$$

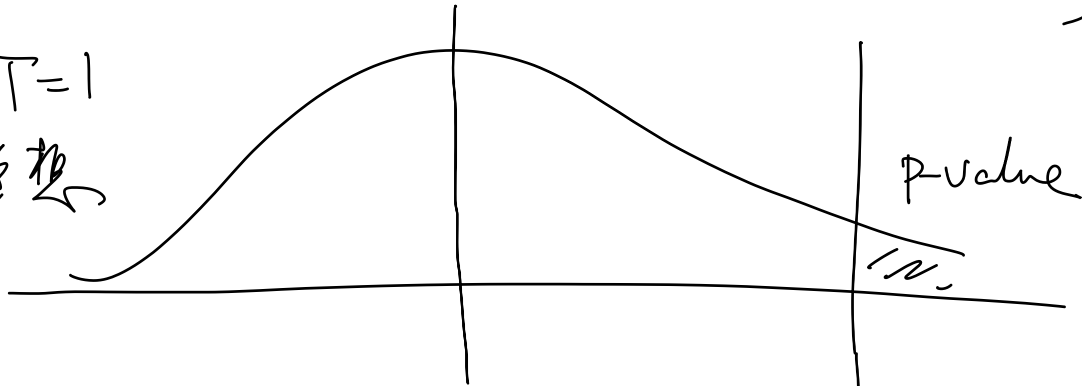
$S_i = 1\left(\dfrac{\Delta}{c_i} > 0\right)$

$Z_i =$ 是 $\left|\dfrac{\Delta}{c_1}\right| \cdots \left|\dfrac{\Delta}{c_n}\right|$ 的秩数

最差情况: $S_i \overset{iid}{\sim} \mathrm{Bern}\left(\frac{\Gamma}{1+\Gamma}\right)$

$\Gamma = 1$

理想



$\Gamma > 1$

不理想