

§ 3.7 Sensitivity analysis.

Outline:

1. Understand the concept of sensitivity analysis.
2. Carrying out a sensitivity analysis in practice.

Subsection (Background)

Recall: The main goal of matching is to achieve balance on observed covariates.

Over bias: occur when balance on observed covariates is not achieved. We \times fully control for these variables.

目标: 我们希望能够识别出 overt bias.

手段: **Unobserved variables: matching** 无法控制未观测到的变量.

Recall 2: Randomized trials — achieve balance on observed and unobserved variables. (因此 treatment 的分配完全是随机的) 与所有的 X 都无关.

hidden bias: 存在未观测到的变量, 并且这些变量都是 confounders.

违背 ignorability assumption $Y^0, Y^1 \perp A | X$.

subsection { Sensitivity analysis }

Main idea: 如果存在 hidden bias, 考虑 hidden bias 多严重时会改变结论 (① 改变 causal effect 的显著性; ② 改变 effect 的方向)

符号说明: π_j : 第 j 个 subject 接受 treatment 的概率; π_k : 第 k 个 subject 接受 treatment 的概率.

• 如果第 j 个 subject 与第 k 个 subject **perfectly matched**, 那么 $X_j = X_k$.

• Inequality (odds ratio): 如果 $\pi_j = \pi_k$, no hidden bias.

$$\frac{1}{\Gamma} \leq \frac{\pi_j}{1-\pi_j} \leq \Gamma \text{ (odds ratio)}$$

$$\pi_k$$

$$\frac{1}{1-\pi_k}$$

• $P=1$, no overt bias.

• $P>1$, implies hidden bias.

• P 的大小可以检验 Ignorability assumption 是否被 violated, 以及被 violated 的程度.

3.8 Propensity Score. subsection {What's Pps?}

Propensity Score: probability of receiving treatment given covariates X .

在给定协变量 X 下接受treatment的概率。

表示: subject i 的 propensity score.

$$\pi_i = P(A=1|X_i)$$

例: Covariate X = age.

年龄大的人更可能接受治疗 (treatment).

故 $\pi_i = P(A=1|X_i) > \pi_j = P(A=1|X_j)$, if $X_i > X_j$.

如果 Pps 的直观含义: 若 $\pi_i = 0.3$, 表示第 i 个病人接受 treatment 的概率为 0.3.
可能性为 30%.

subsection {Balancing score}

Motivation: 2 subjects have the same value of the propensity score.

but they possibly have different covariate values X .

$\pi = P(A=1|X)$
Pps is the function of X .

说明这两个 subjects 有相同概率被分到 treatment group;

propensity score 是一种 balance score.

对于有相同 propensity score 的子群, 在两个 treatment 组中应该具有 balance.

Balance score: If you condition on it, you'll have balance.

考虑等式: $P(X=x|\pi(x)=p, A=1) = P(X=x|\pi(x)=p, A=0)$.

Match on the propensity score \Rightarrow achieve balance.

(替代 Match on 所有的 covariates 之间的 distance)

通过它们

在 Randomized trial 中, Pps 是已知的. $P(A=1|X) = P(A=1) = 0.5$.

分配 A 与 X 无关

在 observational study 中, Pps 未知的.

注意: 我们可以通过 observed data: $A, X \rightarrow$ 估计 Pps.

所以当我们提起 Pps 时, 通常是估计的 Pps.

subsection {Estimating Pps}

目的估计 $P(A=1|X)$

方法: logistic 回归.

步骤: Step 1: 拟合逻辑回归模型, A 作为 Outcome, X 协变量.

Step 2: 对每个 subject:

得到 Predicted value of Pps.



中国科学技术大学

UNIVERSITY OF SCIENCE AND TECHNOLOGY OF CHINA

地址: 中国安徽省合肥市 电话: 0551-63602184 传真: 0551-63631760 网址: <http://www.ustc.edu.cn>

§3.9. Pps matching

- ° Pps is a balancing score.
- ° Matching on the Pps achieve balance. should

° Pps 是一个度量, 每个 subject 都有一个 确定的 Pps.

° Match on Pps 的复杂度: 只需要匹配 1 个变量, matching 问题被 simplified.

对比 Match on covariates: 需要匹配 1 个变量集合.

\sub{section}{overlap}

- ° 比较 Pps 在 treated 组与 control 组 subjects 的分布. (通过画图).

Example 1:

Fig. {example}

① 观察峰值

② $\pi(\text{treated}) > \pi(\text{control})$

Overlap 是什么?

不管 Pps 的哪部分数值, 仍然有人接受处理或控制.

° 在 Fig {ex} 的最右端, 即使是 control 组, 也总有人有很高的概率接受 treated.

° 在 Fig {ex} 的最左端, 即使是 treated 组, 也总有人以很小的概率才能接受 treated.

✓ positivity assumption 成立.

Example 2: {反例} \Rightarrow Fig {Example 2}

消除有着极端 Pps 的个体.

\downarrow
trimming tails.

\sub{trimming tails} Fig. {Trimming tails}

使用: 在 lack of overlap 时适用.

目的: 去掉那些具有极端 Pps 值的 subjects. 极端 Pps 的含义: 对于使用数据的人而言, 接受某类处理的概率较高.

控制: 去掉那些 Pps 值小于 "treated group 中最小 Pps 的 subjects"

treated: 去掉那些 Pps 值大于 "control group 中最大 Pps 的 subjects."

作用: 保障 positivity assumption 合理性.

\sub{Matching}

思想: 计算 2 subjects Pps 的 distance.

求 min distance.

方法: greedy / optimal / 最近邻 matching.

★ 计算每一个 treated subject 与每一个 control 的 Pps.

取对数. $0 < Pps < 1$, 在 0-1 之间非常密集导致很多 Pps 都十分相似.

取 log 性质: ① 单调性, 保证 Pps 的 rank 不变.

② 无界性, 拉大了 Pps 的分布, 使其不再密集.

\Rightarrow 对 $\log(\pi)$ 作 Match, 而不用 π .

\sub{Caliper}

目的: 避免接受 bad matches.

设定: 最大距离

我们能接受的 PPs 间的.

一般设定: $0.2 \times Sd \Rightarrow$ 如果对 balance 的要求更高
就降到 $0.1 Sd \Rightarrow$ fewer matched pairs, better matches.
 $Sd = Sd(\log(\pi_i))$ 所有的 person?

流程: 省略 Fig{caliper}. treated 和 control 合在一起?

Small Caliper - less bias, more variance

trade-off in bias 和 variance. 方差得多了.

treatment effect estimate will have more variability

\sub{After matching}

同 Match based on distance of covariates.

Match 之后要检验 Match 是否达标. | Randomization test.

方法: 是否达到 balance. | GEE. 等.