

内容回顾

- ▶ 有时候分配的处理不等于实际接受的处理
- ▶ 如果出现了不顺从, 那么需要估计顺从者的比例和LATE
- ▶ 如果出现了干涉, 那么可以用两步法估计直接效应和间接效应
- ▶ 跟处理D互相独立的协变量可以被用来改进估计的效率
- ▶ 我们可以使用分块, 匹配, 林回归和随机森林来利用协变量包含的信息, 估计处理效应异质性

内生处理和分块实验

- ▶ 现在我们进一步放松假设, 允许处理跟协变量相关, 或者说给定协变量, 处理跟潜在结果互相独立:

$$D_i \perp\!\!\!\perp \{Y_i(1), Y_i(0)\} | X_i$$

- ▶ 这被称为条件独立性 (conditional independence)
- ▶ 条件独立性成立所必需的控制变量被称为混淆变量 (confounder)
- ▶ 从实验的角度来看, 这意味着在每个由协变量分隔出的区块里, 处理和潜在结果独立
- ▶ 我们仍然要求交叠性假设
- ▶ 条件独立和交叠性合称weak ignorability

从分块到观察性研究

- ▶ 基于设计的视角将任何观察性研究都看作一个分块实验
- ▶ 对于任何数据集, 我们都可以根据控制变量取值将观测分组, 并在每个组内计算处理效应, 最后加总
- ▶ 唯一不同的是, 观测在各个区块内接受处理的概率可以有所差异, 因而 X 会影响 D 的取值

从分块到观察性研究

- ▶ 基于设计的视角将任何观察性研究都看作一个分块实验
- ▶ 对于任何数据集, 我们都可以根据控制变量取值将观测分组, 并在每个组内计算处理效应, 最后加总
- ▶ 唯一不同的是, 观测在各个区块内接受处理的概率可以有所差异, 因而 X 会影响 D 的取值
- ▶ 看起来像分块实验, **不代表条件独立性真的成立!**
- ▶ 条件独立性是否成立, 取决于“自然”的分配机制

从分块到观察性研究

- ▶ 问题: 分块很多时候并不简单
- ▶ 维度诅咒: 如果有8个二值控制变量, 那么有 $2^8 = 256$ 个区块, 交叠性假设难以满足
- ▶ 如果存在连续控制变量, 更加难以处理
- ▶ 理想状况是处理组和控制组中协变量的联合分布相同,

$$f_t(\mathbf{X}_i) = f_c(\mathbf{X}_i)$$

但现实中难以实现

- ▶ 我们只能尽量接近这个目标
- ▶ 分块, 加权, 匹配都是常见的方法

最近邻居匹配 (Nearest Neighbor Matching)

- ▶ 对于每个处理组个体, 我们在控制组中寻找**M**个匹配
- ▶ 标准: 在协变量空间中, 寻找到该处理组个体距离最短的**M**个控制组个体
- ▶ 各个变量的计量单位不同, 所以要通过协方差矩阵加以调整:
$$d(X_i, X_j) = \sqrt{(X_i - X_j)' \mathbf{S}^{-1} (X_i - X_j)}$$
- ▶ 匹配完成之后, 我们可以估计每个处理组个体的处理效应, 再加总得到ATE

最近邻居匹配 (Nearest Neighbor Matching)

- ▶ 最近邻居匹配是有偏的, 因为对于连续变量的匹配不可能完美
- ▶ 误差随着匹配个数 M 增长, 且服从指数分布
- ▶ 由于这部分误差的存在, 不能用自助法 (bootstrap) 估计匹配的标准误
- ▶ Abadie and Imbens (2006) 给出了匹配估计量的渐进分布

模糊精确匹配 (Coarsened Exact Matching)

- ▶ 我们也可以把各个协变量模糊化, 即切分成几个更大的区间
- ▶ 例子: 把年龄变成 (0-30, 30-60, 60+)
- ▶ 这样部分解决了维度诅咒问题

倾向评分匹配 (Propensity Score Matching)

- ▶ Rosenbaum and Rubin (1983)指出, 所有混淆变量起到的作用可以被”坍缩”到一个维度上
- ▶ 倾向评分 $e(x) = P[D = 1|X = x]$
- ▶ 在分块实验中, 倾向评分即每个区块中个体接受处理的概率
- ▶ 倾向评分包含了混淆变量中的所有信息:

$$D_i \perp\!\!\!\perp X_i | e(X_i)$$

证明:

$$E[D_i|X_i, e(X_i)] = E[D_i|X_i] = e(X_i)$$

倾向评分匹配 (Propensity Score Matching)

- ▶ 我们可以计算每个个体的倾向评分, 然后用该得分进行匹配
- ▶ 落在“共同支撑”之外的观测应该被删去
- ▶ 但是, 倾向评分匹配的效率较低
- ▶ **King and Nielsen (2019):** 尽可能使用最近邻居匹配
- ▶ 最近邻居匹配可以近似分块实验, 而倾向评分匹配近似的是完全随机实验

倾向评分加权 (IPW)

- ▶ 使用倾向评分的更常见方法是倾向评分加权 (Inverse Propensity Score Weighting):

$$\widehat{ATE} = \sum_i^N \frac{Y_i D_i}{P(D_i|X_i)} - \sum_i^N \frac{Y_i(1 - D_i)}{1 - P(D_i|X_i)}$$

- ▶ 这样得到的估计是无偏的:

$$\begin{aligned} \sum_i^N E \left[\frac{Y_i D_i}{P(D_i = 1|X_i)} \right] &= \sum_i^N E_X \left[E \left[\frac{Y_i D_i}{P(D_i = 1|X_i)} | X_i \right] \right] \\ &= \sum_i^N E_X \left[\frac{E[Y_i D_i | X_i]}{P(D_i = 1|X_i)} \right] \\ &= \sum_i^N E_X \left[\frac{E[Y_i | D_i = 1, X_i] P(D_i = 1|X_i)}{P(D_i = 1|X_i)} \right] \\ &= E[Y_i(1)] \end{aligned}$$

倾向评分加权的性质

- ▶ 显然IPW也是一个HT估计量
- ▶ 在观察性实验中, 我们不知道 $P(D_i = 1|X_i)$, 需要估计
- ▶ 问题: 在实验中我们知道真实的倾向评分, 那是使用真实值还是估计值?

倾向评分加权的性质

- ▶ 显然IPW也是一个HT估计量
- ▶ 在观察性实验中, 我们不知道 $P(D_i = 1|X_i)$, 需要估计
- ▶ 问题: 在实验中我们知道真实的倾向评分, 那是使用真实值还是估计值?
- ▶ Hahn (1998): 即使知道真实的倾向评分, 我们也应该使用估计的倾向评分
- ▶ 估计的倾向评分捕捉了实际数据里分配时的“噪音”
- ▶ 这在伯努利分配里已经体现得很明显

倾向评分加权的性质

- ▶ 倾向评分加权得到的估计是有效的 (Cramer-Rao lower bound)
- ▶ 倾向评分加权得到的估计收敛于正态分布
- ▶ 一个例外是使用了太多协变量 (Catteneo et al. 2019)
- ▶ 倾向评分加权的问题是, 有时候权重可能接近无穷大
- ▶ 此时可以使用稳健倾向评分加权

$$\widehat{ATE} = \sum_i^N \frac{Y_i D_i / P(D_i | X_i)}{D_i / P(D_i | X_i)} - \sum_i^N \frac{Y_i (1 - D_i) / (1 - P(D_i | X_i))}{(1 - D_i) / (1 - P(D_i | X_i))}$$

- ▶ 类似沃德估计量, 这个估计量是有偏的, 但更有效率
- ▶ 如果处理变量是连续的, 可以使用协变量平衡倾向评分 (CBPS)

熵平衡 (Entropy Balancing)

- ▶ Hainmueller (2011): 既然我们的最终目标是平衡控制变量, 那为什么不直接以此为目标来计算权重?
- ▶ 我们可以选择平衡协变量的K阶矩 (均值, 方差, 斜度...)

$$\min \sum_i w_i \log w_i$$

$$s.t. \quad w_i > 0$$

$$\sum_i w_i X_{t,i} = w_i X_{c,i}$$

- ▶ 相比于匹配, 熵平衡的优势是不会丢弃观测
- ▶ 但如果不平衡程度太高, 熵平衡未必能得到合适的权重

如何测量平衡性

- ▶ 如前所述, 我们的目标是

$$f_t(\mathbf{X}_i) = f_c(\mathbf{X}_i)$$

- ▶ 但如何确定我们已经达到了这个目标?

如何测量平衡性

- ▶ 如前所述, 我们的目标是

$$f_t(\mathbf{X}_i) = f_c(\mathbf{X}_i)$$

- ▶ 但如何确定我们已经达到了这个目标?
- ▶ 常见做法是比较处理组和控制组协变量的组间均值之差; 如果均不显著, 则平衡

如何测量平衡性

- ▶ 如前所述, 我们的目标是

$$f_t(\mathbf{X}_i) = f_c(\mathbf{X}_i)$$

- ▶ 但如何确定我们已经达到了这个目标?
- ▶ 常见做法是比较处理组和控制组协变量的组间均值之差; 如果均不显著, 则平衡
- ▶ 但在样本量很大的时候, 很微小的差异可能也会统计显著

如何测量平衡性

- ▶ 如前所述, 我们的目标是

$$f_t(\mathbf{X}_i) = f_c(\mathbf{X}_i)$$

- ▶ 但如何确定我们已经达到了这个目标?
- ▶ 常见做法是比较处理组和控制组协变量的组间均值之差; 如果均不显著, 则平衡
- ▶ 但在样本量很大的时候, 很微小的差异可能也会统计显著
- ▶ 而且没有拒绝零假设, 不代表零假设正确

如何测量平衡性

- ▶ Hartman and Hidalgo (2018): 等价检验 (Equivalence test)

如何测量平衡性

- ▶ Hartman and Hidalgo (2018): 等价检验 (Equivalence test)
- ▶ 零假设: $\bar{X}_t - \bar{X}_c > \delta_1$ or $\bar{X}_t - \bar{X}_c < \delta_2$

如何测量平衡性

- ▶ Hartman and Hidalgo (2018): 等价检验 (Equivalence test)
- ▶ 零假设: $\bar{X}_t - \bar{X}_c > \delta_1$ or $\bar{X}_t - \bar{X}_c < \delta_2$
- ▶ 如果拒绝了零假设, 意味着两组协变量均值确实等价

如何测量平衡性

- ▶ Hartman and Hidalgo (2018): 等价检验 (Equivalence test)
- ▶ 零假设: $\bar{X}_t - \bar{X}_c > \delta_1$ or $\bar{X}_t - \bar{X}_c < \delta_2$
- ▶ 如果拒绝了零假设, 意味着两组协变量均值确实等价
- ▶ 有很多种检验方式, 比如两个单边t检验

如何测量平衡性

- ▶ Hartman and Hidalgo (2018): 等价检验 (Equivalence test)
- ▶ 零假设: $\bar{X}_t - \bar{X}_c > \delta_1$ or $\bar{X}_t - \bar{X}_c < \delta_2$
- ▶ 如果拒绝了零假设, 意味着两组协变量均值确实等价
- ▶ 有很多种检验方式, 比如两个单边t检验
- ▶ 等价检验中的一个关键问题是如何选择 δ_1 和 δ_2

如何测量平衡性

- ▶ Hartman and Hidalgo (2018): 等价检验 (Equivalence test)
- ▶ 零假设: $\bar{X}_t - \bar{X}_c > \delta_1$ or $\bar{X}_t - \bar{X}_c < \delta_2$
- ▶ 如果拒绝了零假设, 意味着两组协变量均值确实等价
- ▶ 有很多种检验方式, 比如两个单边t检验
- ▶ 等价检验中的一个关键问题是如何选择 δ_1 和 δ_2
- ▶ 最好是由理论决定, 否则可以选择 $\hat{\tau}$ 或者 $0.36\hat{SE}_c$

加权和回归: 双重稳健性

- ▶ 分块和匹配都相当于是加权估计量
- ▶ 加权估计量的偏误更小, 而回归的方差更小

加权和回归: 双重稳健性

- ▶ 分块和匹配都相当于是加权估计量
- ▶ 加权估计量的偏误更小, 而回归的方差更小
- ▶ 自然的想法: 加权回归
- ▶ 例子: 倾向评分加权回归, 先分块/匹配再回归

$$\widehat{ATE} = \sum_i^N \left(\frac{(Y_i - \mu_1(X_i))D_i}{P(D_i|X_i)} + \mu_1(X_i) \right) - \left(\sum_i^N \frac{(Y_i - \mu_0(X_i))(1 - D_i)}{1 - P(D_i|X_i)} + \mu_0(X_i) \right)$$

加权和回归: 双重稳健性

- ▶ 加权回归具有双重稳健性: 无论回归模型正确, 还是权重正确, 都能得到无偏的估计量
- ▶ 回归部分则可以提升估计的效率, 而加权部分可以处理控制变量的非线性影响

加权和回归: 双重稳健性

- ▶ 加权回归具有双重稳健性: 无论回归模型正确, 还是权重正确, 都能得到无偏的估计量
- ▶ 回归部分则可以提升估计的效率, 而加权部分可以处理控制变量的非线性影响
- ▶ 回归部分和权重都可以用机器学习方法来灵活估计, 这样得到的估计量被称为半参估计量

加权和回归: 双重稳健性

- ▶ 加权回归具有双重稳健性: 无论回归模型正确, 还是权重正确, 都能得到无偏的估计量
- ▶ 回归部分则可以提升估计的效率, 而加权部分可以处理控制变量的非线性影响
- ▶ 回归部分和权重都可以用机器学习方法来灵活估计, 这样得到的估计量被称为半参估计量
- ▶ 半参估计量本质上都是包含了无谓参数 (nuisance parameters) 的矩估计
- ▶ 在因果推断中, 矩条件即条件独立性

如何选择控制变量

- ▶ 基于理论和经验进行选择: 如果自然设计了这个实验, 那分配机制应当是怎样的?

如何选择控制变量

- ▶ 基于理论和经验进行选择: 如果自然设计了这个实验, 那分配机制应当是怎样的?
- ▶ 控制变量并不是越多越好

如何选择控制变量

- ▶ 基于理论和经验进行选择: 如果自然设计了这个实验, 那分配机制应当是怎样的?
- ▶ 控制变量并不是越多越好
 1. X 可能会吸收 D 中包含的信息, 降低 D 的显著性
 2. 如果 X 跟 Y 正相关, 某不可观测的混淆变量跟 Y 负相关, 那么不控制 X 反而可以抵消一部分误差

如何选择控制变量

- ▶ 基于理论和经验进行选择: 如果自然设计了这个实验, 那分配机制应当是怎样的?
- ▶ 控制变量并不是越多越好
 1. X可能会吸收D中包含的信息, 降低D的显著性
 2. 如果X跟Y正相关, 某不可观测的混淆变量跟Y负相关, 那么不控制X反而可以抵消一部分误差
- ▶ 控制变量一定不能被处理影响, 否则会导致预后偏误 (post-treatment variable bias)
- ▶ 被处理影响的控制变量相当于一个中介变量 (mediator), 会吸收部分处理效应, 导致偏误

如何选择控制变量

- ▶ 比如研究历史事件对今天政治态度的影响, 就不应该控制今天的经济发展水平
- ▶ 例子: 政治学里被引用数最高的研究
- ▶ **Fearon and Latin (2003)**: 控制了经济发展水平之后, 族群分裂程度并未导致更多内战
- ▶ 但问题是族群分裂会影响经济发展, 所以回归结果并未告诉我们太多信息 (Samii, 2016)

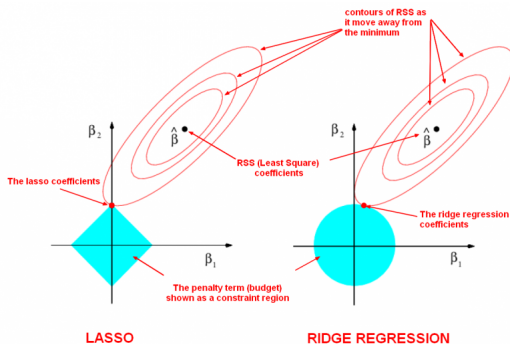
双重选择: 让机器选择控制变量

- ▶ 机器学习在变量选择方面也很有用处

双重选择: 让机器选择控制变量

- ▶ 机器学习在变量选择方面也很有用处
- ▶ 线性模型里最常见的方法是LASSO

$$Loss = (Y - X\beta)'(Y - X\beta) + \lambda * Penalty$$



双重选择: 让机器选择控制变量

- ▶ 但如果只让LASSO选择对Y解释力较强的变量, 条件独立性可能就不成立了

双重选择: 让机器选择控制变量

- ▶ 但如果只让LASSO选择对Y解释力较强的变量, 条件独立性可能就不成立了
- ▶ Belloni et al., 2013: 双重选择 (Double selection)
- ▶ 生成特征 (feature) 集: 所有控制变量, 它们的交互项和高阶项
- ▶ 先用LASSO选择对Y解释力较强的特征, X_1 , 然后用LASSO选择对D解释力较强的特征, X_2
- ▶ 实际分析的时候同时控制 X_1 和 X_2
- ▶ 进阶版本: 双重机器学习 (多次剖分样本, 每个观测既在训练集又在测试集)

检验条件独立性

- 无论是哪种估计方法,都不能让条件独立性变得更加可信

检验条件独立性

- ▶ 无论是哪种估计方法,都不能让条件独立性变得更加可信
- ▶ 条件独立性也没有直接的检验方法,其可信性依赖于研究设计和理论知识

检验条件独立性

- ▶ 无论是哪种估计方法,都不能让条件独立性变得更加可信
- ▶ 条件独立性也没有直接的检验方法,其可信性依赖于研究设计和理论知识
- ▶ 一个间接的检验方式是寻找安慰剂结果 (placebo outcome), 即理论上不会被处理影响的变量
- ▶ 朋友的体重可能影响你的体重,但应该不会影响你的身高...

敏感性检验 (Sensitivity test)

- ▶ 敏感性检验的基本思想: 假定存在不可观测的混淆变量 U , 改变 U 的“大小”, 观察估计值如何随之改变

敏感性检验 (Sensitivity test)

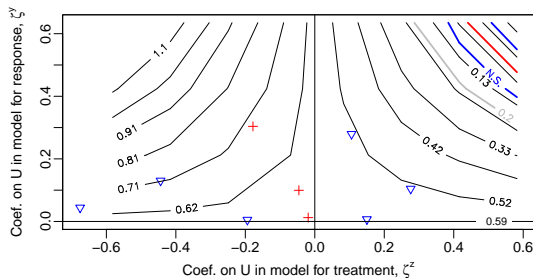
- ▶ 敏感性检验的基本思想: 假定存在不可观测的混淆变量 U , 改变 U 的“大小”, 观察估计值如何随之改变
- ▶ 如果 U 的取值变化了一点, 估计值就不再显著, 那么我们就应当对其可信性有所怀疑

敏感性检验 (Sensitivity test)

- ▶ 敏感性检验的基本思想: 假定存在不可观测的混淆变量 U , 改变 U 的“大小”, 观察估计值如何随之改变
- ▶ 如果 U 的取值变化了一点, 估计值就不再显著, 那么我们就应当对其可信性有所怀疑
- ▶ 但 U 同时影响结果 Y 和处理 D , 所以我们可以在 U 跟 (Y, D) 相关系数的空间里画出处理效应的“等势线”, 并考察何时效应消失为零
- ▶ 我们还可以将 U 的大小跟其他控制变量的大小作对比

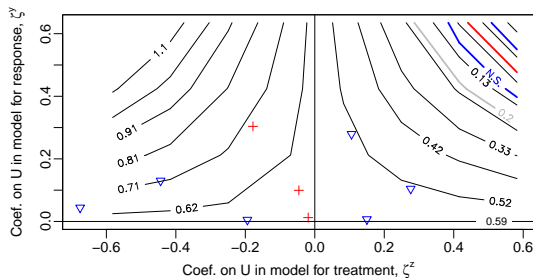
敏感性检验 (Sensitivity test)

- ▶ 在红色曲线处, 处理效应消失; 在蓝色曲线处, 其统计显著性消失
- ▶ 红色十字和蓝色三角标出了其他协变量跟(Y, D)的相关系数
- ▶ 该图说明处理效应比较显著



敏感性检验 (Sensitivity test)

- ▶ 在红色曲线处, 处理效应消失; 在蓝色曲线处, 其统计显著性消失
- ▶ 红色十字和蓝色三角标出了其他协变量跟(Y, D)的相关系数
- ▶ 该图说明处理效应比较显著



- ▶ 这只是一进行敏感性检验的方法

怎么寻找自然实验

- ▶ 因果推断并不教你怎么找到因果关系,也不能从任意的数据集里创造出因果关系

怎么寻找自然实验

- ▶ 因果推断并不教你怎么找到因果关系,也不能从任意的数据集里创造出因果关系
- ▶ 为了得到因果关系,最直接的办法是做实验,其次是寻找自然实验

怎么寻找自然实验

- ▶ 因果推断并不教你怎么找到因果关系,也不能从任意的数据集里创造出因果关系
- ▶ 为了得到因果关系,最直接的办法是做实验,其次是寻找自然实验
- ▶ 寻找自然实验要依靠理论

怎么寻找自然实验

- ▶ 因果推断并不教你怎么找到因果关系,也不能从任意的数据集里创造出因果关系
- ▶ 为了得到因果关系,最直接的办法是做实验,其次是寻找自然实验
- ▶ 寻找自然实验要依靠理论
- ▶ 理论意味着一个系统,系统内部的变量互相作用

怎么寻找自然实验

- ▶ 因果推断并不教你怎么找到因果关系, 也不能从任意的数据集里创造出因果关系
- ▶ 为了得到因果关系, 最直接的办法是做实验, 其次是寻找自然实验
- ▶ 寻找自然实验要依靠理论
- ▶ 理论意味着一个系统, 系统内部的变量互相作用
- ▶ 一般来说, 我们不能同时检验系统内部的所有相互作用, 只能关注其中的一种作用渠道

怎么寻找自然实验

- ▶ 因果推断并不教你怎么找到因果关系,也不能从任意的数据集里创造出因果关系
- ▶ 为了得到因果关系,最直接的办法是做实验,其次是寻找自然实验
- ▶ 寻找自然实验要依靠理论
- ▶ 理论意味着一个系统,系统内部的变量互相作用
- ▶ 一般来说,我们不能同时检验系统内部的所有相互作用,只能关注其中的一种作用渠道
- ▶ 所谓自然实验,即系统之外的某些因素,引发了系统内部某个变量超出预期之外的变化

怎么寻找自然实验

- ▶ 因此到底什么是“外生冲击”，取决于你对于“内生变量”的定义

怎么寻找自然实验

- ▶ 因此到底什么是“外生冲击”, 取决于你对于“内生变量”的定义
- ▶ 某些变量在研究A里可能是内生的, 但在研究B里则是外生的

怎么寻找自然实验

- ▶ 因此到底什么是“外生冲击”，取决于你对于“内生变量”的定义
- ▶ 某些变量在研究A里可能是内生的，但在研究B里则是外生的
- ▶ 例子：美联储加息

怎么寻找自然实验

- ▶ 因此到底什么是“外生冲击”，取决于你对于“内生变量”的定义
- ▶ 某些变量在研究A里可能是内生的，但在研究B里则是外生的
- ▶ 例子：美联储加息
- ▶ 如果我们想研究美国政府内部的博弈，以及货币政策的制定，那么加息当然是内生的决策

怎么寻找自然实验

- ▶ 因此到底什么是“外生冲击”，取决于你对于“内生变量”的定义
- ▶ 某些变量在研究A里可能是内生的，但在研究B里则是外生的
- ▶ 例子：美联储加息
- ▶ 如果我们想研究美国政府内部的博弈，以及货币政策的制定，那么加息当然是内生的决策
- ▶ 但如果我们感兴趣的是加息怎么影响个体消费水平，那么可以将其视为外生变量

怎么寻找自然实验

- ▶ 精妙的自然实验往往发生在常识之外, 是普通人未能预计之事

怎么寻找自然实验

- ▶ 精妙的自然实验往往发生在常识之外, 是普通人未能预计之事
- ▶ 例子: 卢旺达政府的宣传, 是否推动了种族屠杀的发展?

怎么寻找自然实验

- ▶ 精妙的自然实验往往发生在常识之外, 是普通人未能预计之事
- ▶ 例子: 卢旺达政府的宣传, 是否推动了种族屠杀的发展?
- ▶ 从政府的角度来看, 宣传策略是既定的

怎么寻找自然实验

- ▶ 精妙的自然实验往往发生在常识之外, 是普通人未能预计之事
- ▶ 例子: 卢旺达政府的宣传, 是否推动了种族屠杀的发展?
- ▶ 从政府的角度来看, 宣传策略是既定的
- ▶ 然而卢旺达是多山国家, 政府广播的信号受到地形阻碍, 不同地区收到信号的概率有差异

怎么寻找自然实验

- ▶ 精妙的自然实验往往发生在常识之外, 是普通人未能预计之事
- ▶ 例子: 卢旺达政府的宣传, 是否推动了种族屠杀的发展?
- ▶ 从政府的角度来看, 宣传策略是既定的
- ▶ 然而卢旺达是多山国家, 政府广播的信号受到地形阻碍, 不同地区收到信号的概率有差异
- ▶ 无论是政府还是屠杀者, 应该都没能预计到这种情况的出现

怎么寻找自然实验

- ▶ 精妙的自然实验往往发生在常识之外, 是普通人未能预计之事
- ▶ 例子: 卢旺达政府的宣传, 是否推动了种族屠杀的发展?
- ▶ 从政府的角度来看, 宣传策略是既定的
- ▶ 然而卢旺达是多山国家, 政府广播的信号受到地形阻碍, 不同地区收到信号的概率有差异
- ▶ 无论是政府还是屠杀者, 应该都没能预计到这种情况的出现
- ▶ 想找到一个好的自然实验, 需要像人类学家和历史学家一样, 对实际情况有充分地了解, 才能分辨哪些是预料之中, 哪些是预料之外

怎么寻找自然实验

- ▶ 精妙的自然实验往往发生在常识之外, 是普通人未能预计之事
- ▶ 例子: 卢旺达政府的宣传, 是否推动了种族屠杀的发展?
- ▶ 从政府的角度来看, 宣传策略是既定的
- ▶ 然而卢旺达是多山国家, 政府广播的信号受到地形阻碍, 不同地区收到信号的概率有差异
- ▶ 无论是政府还是屠杀者, 应该都没能预计到这种情况的出现
- ▶ 想找到一个好的自然实验, 需要像人类学家和历史学家一样, 对实际情况有充分地了解, 才能分辨哪些是预料之中, 哪些是预料之外
- ▶ 当我们用理论解释现实的时候, 那些不太好解释的部分, 可能就是自然实验的潜在来源

我们学到了什么

- ▶ 因果关系建立在反事实的基础上, 可以用**Rubin**模型描述

我们学到了什么

- ▶ 因果关系建立在反事实的基础上, 可以用**Rubin**模型描述
- ▶ 实验是检验因果关系的黄金标准, 任何试图得到因果关系的观察性研究都应该被看作自然设计的实验

我们学到了什么

- ▶ 因果关系建立在反事实的基础上, 可以用**Rubin**模型描述
- ▶ 实验是检验因果关系的黄金标准, 任何试图得到因果关系的观察性研究都应该被看作自然设计的实验
- ▶ 简单随机实验的估计用**H-T**估计量, 推断可以用**Fisher**的随机推断, 或者**Neyman**的方差分析

我们学到了什么

- ▶ 因果关系建立在反事实的基础上, 可以用**Rubin**模型描述
- ▶ 实验是检验因果关系的黄金标准, 任何试图得到因果关系的观察性研究都应该被看作自然设计的实验
- ▶ 简单随机实验的估计用**H-T**估计量, 推断可以用**Fisher**的随机推断, 或者**Neyman**的方差分析
- ▶ 如果存在不顺从, 使用沃德估计量; 不要滥用工具变量

我们学到了什么

- ▶ 因果关系建立在反事实的基础上, 可以用**Rubin**模型描述
- ▶ 实验是检验因果关系的黄金标准, 任何试图得到因果关系的观察性研究都应该被看作自然设计的实验
- ▶ 简单随机实验的估计用**H-T**估计量, 推断可以用**Fisher**的随机推断, 或者**Neyman**的方差分析
- ▶ 如果存在不顺从, 使用沃德估计量; 不要滥用工具变量
- ▶ 巧妙的实验设计可以解决干涉的问题, 并得到处理的扩散效应

我们学到了什么

- ▶ 控制变量不但能够减少偏误, 还能增加估计的效率; 但不要控制被处理影响的变量

我们学到了什么

- ▶ 控制变量不但能够减少偏误, 还能增加估计的效率; 但不要控制被处理影响的变量
- ▶ 时刻留心交叠性假设是否成立!

我们学到了什么

- ▶ 控制变量不但能够减少偏误, 还能增加估计的效率; 但不要控制被处理影响的变量
- ▶ 时刻留心交叠性假设是否成立!
- ▶ 线性回归包含了模型假设, 只有在简单随机实验中才是无偏的, 但使用Lin提出的方法能提高估计的效率

我们学到了什么

- ▶ 控制变量不但能够减少偏误, 还能增加估计的效率; 但不要控制被处理影响的变量
- ▶ 时刻留心交叠性假设是否成立!
- ▶ 线性回归包含了模型假设, 只有在简单随机实验中才是无偏的, 但使用Lin提出的方法能提高估计的效率
- ▶ 分块, 匹配和加权都能帮助你获得协变量平衡的样本, 但无法带来条件独立性

我们学到了什么

- ▶ 控制变量不但能够减少偏误, 还能增加估计的效率; 但不要控制被处理影响的变量
- ▶ 时刻留心交叠性假设是否成立!
- ▶ 线性回归包含了模型假设, 只有在简单随机实验中才是无偏的, 但使用Lin提出的方法能提高估计的效率
- ▶ 分块, 匹配和加权都能帮助你获得协变量平衡的样本, 但无法带来条件独立性
- ▶ 机器学习在变量筛选和估计异质性方面十分有用

我们学到了什么

- ▶ 控制变量不但能够减少偏误, 还能增加估计的效率; 但不要控制被处理影响的变量
- ▶ 时刻留心交叠性假设是否成立!
- ▶ 线性回归包含了模型假设, 只有在简单随机实验中才是无偏的, 但使用Lin提出的方法能提高估计的效率
- ▶ 分块, 匹配和加权都能帮助你获得协变量平衡的样本, 但无法带来条件独立性
- ▶ 机器学习在变量筛选和估计异质性方面十分有用
- ▶ 结合加权和回归的双重稳健估计量值得你拥有!

我们学到了什么

- ▶ 控制变量不但能够减少偏误, 还能增加估计的效率; 但不要控制被处理影响的变量
- ▶ 时刻留心交叠性假设是否成立!
- ▶ 线性回归包含了模型假设, 只有在简单随机实验中才是无偏的, 但使用Lin提出的方法能提高估计的效率
- ▶ 分块, 匹配和加权都能帮助你获得协变量平衡的样本, 但无法带来条件独立性
- ▶ 机器学习在变量筛选和估计异质性方面十分有用
- ▶ 结合加权和回归的双重稳健估计量值得你拥有!
- ▶ 可以用安慰剂检验或者敏感性检验评估识别假设的可信性

我们没学到什么

- ▶ 存在时间维度时的因果推断 (Y, D, X, T)

我们没学到什么

- ▶ 存在时间维度时的因果推断 (Y, D, X, T)
- ▶ 双重差分 (Difference-in-differences), 合成控制法 (Synthetic control), 因子模型 (Factor models), 矩阵填充 (Matrix completion)...

我们没学到什么

- ▶ 存在时间维度时的因果推断 (Y, D, X, T)
- ▶ 双重差分 (Difference-in-differences), 合成控制法 (Synthetic control), 因子模型 (Factor models), 矩阵填充 (Matrix completion)...
- ▶ 这里的本质问题是处理在时间维度上的干涉
- ▶ 否则我们可以按照横截面数据加以处理

我们没学到什么

- ▶ 存在时间维度时的因果推断 (Y, D, X, T)
- ▶ 双重差分 (Difference-in-differences), 合成控制法 (Synthetic control), 因子模型 (Factor models), 矩阵填充 (Matrix completion)...
- ▶ 这里的本质问题是处理在时间维度上的干涉
- ▶ 否则我们可以按照横截面数据加以处理
- ▶ 断点回归 (Y, D, X)
- ▶ 但这里的 X 不是协变量, 而是流动变量 (running variable)
- ▶ 我们只能估计某一点的处理效应
- ▶ 本质上也是一种加权估计量, 关键的问题是如何选择权重

我们没学到什么

- ▶ 统计学: 如何估计 Y 和 X 之间的关系?

我们没学到什么

- ▶ 统计学: 如何估计 Y 和 X 之间的关系?
- ▶ 各种非参估计方法: 局部回归 (local regression) 和样条曲线 (splines)

我们没学到什么

- ▶ 统计学: 如何估计 Y 和 X 之间的关系?
- ▶ 各种非参估计方法: 局部回归 (local regression) 和样条曲线 (splines)
- ▶ 以及其他机器学习方法: 支持向量机 (SVM), 岭回归 (ridge regression), boosting...

我们没学到什么

- ▶ 统计学: 如何估计 Y 和 X 之间的关系?
- ▶ 各种非参估计方法: 局部回归 (local regression) 和样条曲线 (splines)
- ▶ 以及其他机器学习方法: 支持向量机 (SVM), 岭回归 (ridge regression), boosting...
- ▶ 渐进理论: 在 N 足够大的时候, 统计量的分布是什么?
- ▶ 在nuisance parameters维度有限的时候, 可以利用极大似然估计的渐进分布
- ▶ 在nuisance parameters维度有限的时候, 需要证明函数空间上的收敛性

总结

- ▶ 但我们始终是社会科学家, 而不是统计学家

总结

- ▶ 但我们始终是社会科学家, 而不是统计学家
- ▶ 任何实验设计都应该以坚实的理论作为支撑, 而不是炫目的技巧

总结

- ▶ 但我们始终是社会科学家, 而不是统计学家
- ▶ 任何实验设计都应该以坚实的理论作为支撑, 而不是炫目的技巧
- ▶ 只有设计和理论能够保证因果性, 统计方法不能

总结

- ▶ 但我们始终是社会科学家, 而不是统计学家
- ▶ 任何实验设计都应该以坚实的理论作为支撑, 而不是炫目的技巧
- ▶ 只有设计和理论能够保证因果性, 统计方法不能
- ▶ 如果得不到因果性, 不要强求

总结

- ▶ 但我们始终是社会科学家, 而不是统计学家
- ▶ 任何实验设计都应该以坚实的理论作为支撑, 而不是炫目的技巧
- ▶ 只有设计和理论能够保证因果性, 统计方法不能
- ▶ 如果得不到因果性, 不要强求
- ▶ 问题足够重要的话, 相关性也很好

总结

- ▶ 但我们始终是社会科学家, 而不是统计学家
- ▶ 任何实验设计都应该以坚实的理论作为支撑, 而不是炫目的技巧
- ▶ 只有设计和理论能够保证因果性, 统计方法不能
- ▶ 如果得不到因果性, 不要强求
- ▶ 问题足够重要的话, 相关性也很好
- ▶ 如果要做出因果论断, 一定要想明白其背后的假设到底是什么

总结

- ▶ 但我们始终是社会科学家, 而不是统计学家
- ▶ 任何实验设计都应该以坚实的理论作为支撑, 而不是炫目的技巧
- ▶ 只有设计和理论能够保证因果性, 统计方法不能
- ▶ 如果得不到因果性, 不要强求
- ▶ 问题足够重要的话, 相关性也很好
- ▶ 如果要做出因果论断, 一定要想明白其背后的假设到底是什么
- ▶ 这才是基于设计的视角真正的含义

谢谢!