

# 因果推断的潜在结果框架在实验性研究的应用

## ｜因果科学读书会第2期

### 【读书会相关信息】

“因果”并不是一个新概念，而是一个已经在多个学科中使用了数十年的分析技术。通过前两季的分  
享，我们主要梳理了因果科学在计算机领域的前沿进展。如要融会贯通，我们需要回顾数十年来在社  
会学、经济学、医学、生物学等多个领域中，都是使用了什么样的因果模型、以什么样的范式、解决  
了什么样的问题。我们还要尝试进行对比和创新，看能否以现在的眼光，用其他的模型，为这些研究  
提供新的解决思路。

如果大家对这个读书会感兴趣，欢迎报名：

[https://pattern.swarma.org/mobile/study\\_group/10?from=wechat](https://pattern.swarma.org/mobile/study_group/10?from=wechat)

【时间】2021年10月31日 09:00--11:00

### 【主讲人】

**李昊轩**，北京大学大数据科学研究中心博士研究生，导师为周晓华教授，专业为数据科学（统计  
学），研究兴趣为因果推断，推荐系统，强化学习。

### 【笔记小分队】

杨二茶 | 清华大学在读博士

旷野

## 【讲座笔记】

### 【内容简介】

本周日的讨论将围绕W. Imben和B. Rubin的著作Causal Inference for Statistics, Social, and  
Biomedical Sciences第一部分和第二部分进行讨论。将会介绍因果推断的潜在结果框架清晰定义因  
果，利用随机化实验的思想作为因果效应识别的基础，通过因果效应的识别与估计尝试回答上述问  
题。按分配机制是否已知，可以分为实验性研究和观察性研究，本次分享将重点讨论分配机制已知的  
实验性研究的因果效应估计方法。

记录人：旷野

## Part1:因果推断的潜在结果框架

1. 引子：诺贝尔经济学奖与因果推断，表彰他们为劳动力市场提供了新的见解。给出了从自然实验  
中得到因果关系的方法。

2. 三个关键概念: 潜在结果; 多个个体及稳定个体干预值假设 ; 分配机制

- 潜在结果：先验上，个体接受了相应的处理，就可以观察到；后验上，个体一旦进行了处理，最多只能观察到一个潜在的结果。
- 多个个体及稳定个体干预值假设：大多数情况下都使用稳定性假设来处理多个个体。
- 分配机制：至关重要，构成了潜在结果框架的架构模块。

3. 例子:

- 例 1: “因为我服用了阿司匹林, 所以我的头痛减弱了。”
- 例 2: “因为她上了大学, 所以她去年找到了一份好工作。”
- 例 3: “她留长发是因为她是女孩。”

· 潜在结果分析:

(1) 关联行为-个体 -潜在结果。 (2) 因果效应涉及对潜在结果的比较，但只有一个结果实现，其他的未实现，无法观测到。 (3) 处理在潜在结果出现之前。 (4) 三个例子的本质差别，可以通过实际处理/行动与未处理的潜在结果的反事实数值的差值得到。 (5) 要定义因果关系，必须先明确干预措施。（第三个例子没有明确定义行为/处理）

· 因果效应

Table 1.1. Example of Potential Outcomes and Causal Effect with One Unit			
Unit	Potential Outcomes		Causal Effect
	$Y(\text{Aspirin})$	$Y(\text{No Aspirin})$	
You	No Headache	Headache	Improvement due to Aspirin

图 1: 示例：单个个体的潜在结果和因果效应

- (1) 因果效应取决于潜在结果，但不取决于实际观察到的结果； (2) 因果效应：比较同一个体，同一时间的潜在结果。 (3) 特别指出：仅对比服/不服用阿司匹林前后的头疼状况不能得到阿司匹林与头疼的因果效应。（注：吃了不疼，不吃也可能不疼，因此仅靠吃了不疼不能推出因果关系）
- (4) “因果推断的基本问题” (Holland,1986, 第 947 页) 是最多只可以实现并观测到一种潜在结果（注：但要得到因果关系）的问题。 (5) 因果效应的估计要潜在结果实现出现差异的原因，即分配机制。

4. 稳定个体干预值假设 (SUTVA)

- 任何个体的潜在结果都不随分配给其他个体的处理水平而变化，而且处理水平对于所有个体都是相同的，这就导致了不同的潜在结果。
- 挑战一：互干预；挑战二：每个处理都要有多个观测；
- 解决方法：将子样本 (subpopulation) 视为一个个体。

5. 分配机制

- 分配每个个体接受处理的程度，并给出进行此种处理的原因。

· 例子

**Table 1.4. Medical Example with Two Treatments, Four Units, and SUTVA: Surgery (S) and Drug Treatment (D)**

Unit	Potential Outcomes		Causal Effect
	$Y_i(0)$	$Y_i(1)$	$Y_i(1) - Y_i(0)$
Patient #1	1	7	6
Patient #2	6	5	-1
Patient #3	1	5	4
Patient #4	8	7	-1
Average	4	6	2

**图 4:** 示例：SUTVA 假设下的四个个体医学案例

**Table 1.5. Ideal Medical Practice: Patients Assigned to the Individually Optimal Treatment; Example from Table 1.4**

Unit $i$	Treatment $W_i$	Observed Outcome $Y_i^{obs}$
Patient #1	1	7
Patient #2	0	6
Patient #3	1	5
Patient #4	0	8

**图 5:** 示例：四个个体的最优治疗情形

考虑平均因果效应，手术治疗比药物治疗会延长 2 年寿命；但是考虑医生对潜在结果有足够了解，采用最佳治疗方案，手术治疗的平均观测结果比药物治疗的平均观测结果少1年（PPT有误）；因此必须考虑分配机制。

## 6. 协变量

- 定义：个体特定性质（也称为预处理变量或协变量），用个体  $i$  的 含有  $K$  个分量的行向量  $X$  表示。
- 关键特征：先验性，不受处理分配机制的影响。
- 作用：协变量最重要的作用在于它们对分配机制的影响；协变量通常通过解释因变量/结果中的一些变化使得估计更精确；由协变量定义的子样本可能更有研究价值；关于分配机制及其可能不依赖于潜在结果的假设，在某些关于协变量同质的子样本中通常更为合理。

## 7. 因果估计量

- 个体因果效应：差值  $Y_i(1) - Y_i(0)$  或者比值  $Y_i(1)/Y_i(0)$
- 平均因果效应（ATE）：

$$\tau_{fs} = \frac{1}{N} \sum_{i=1}^N (Y_i(1) - Y_i(0)).$$

- 条件平均因果效应（CATE）：

$$\tau_{fs}(f) = \frac{1}{N(f)} \sum_{i: X_i=f} (Y_i(1) - Y_i(0))$$

- 其他形式的因果效应

$$\tau_{fs,t} = \frac{1}{N_t} \sum_{i: W_i=1} (Y_i(1) - Y_i(0))$$

$$\tau_{fs, pos} = \frac{1}{N_{pos}} \sum_{i: Y_i(0)>0, Y_i(1)>0} (Y_i(1) - Y_i(0))$$

## 8. 总体，超总体，样本

我们把观测到的集合视为有限样本（下标"fs"）；我们把所感兴趣的集合视为总体；把无限总体视为超总体（下标"sp"）

## 9. Lord悖论

- 假想案例：一所大学想考察其食堂膳食对学生体重的影响以及影响的性别差异,为此,收集了 9 月份入学时的体重,然后次年6月份又获得了学生在校一学年后的体重。
- 统计学家A对比了所有的男生和女生的体重,没有差异,所以认为膳食对学生体重无影响;统计学家B,用了回归分析,观察初始同样体重的男生,女生,发现1年后男生的体重增加更大,比女生平均高 7.3 磅,膳食对学生体重有影响。所以就有了悖论。
- 解释：悖论的根本原因是缺少没有去食堂吃饭的对照组；两者都可以是对的,但是依据不同的假设。第一个的假设是：如果学生不来食堂进餐,他们的体重将和入学前一样。第二个的假设是：不去食堂进餐时的潜在体重Y(0)是入学前体重X的线性函数且截距是b,即：Y0 = a + bX。（男生平均体重超过女生是事实,但不能得出膳食对学生体重有影响的因果效应,除非有前面的假设）

## 10. 因果推断的形式化体系

- 总体由N个个体组成,其协变量矩阵记为 X, 是一个  $N \times K$  矩阵, 个体协变量为  $X_i$ , 是X的某一行。
- 以治疗为例：潜在结果是一个标量,用Y(0) 和 Y(1), Y (0) 对照治疗, Y (1) 主动治疗; 治疗分配由 W 表示; 潜在结果框架只允许我们针对每一个个体观测到一个特定的潜在结果  $Y_i^{obs}$ , 另一个潜在结果则无法被观测  $Y_i^{mis}$

$$Y_i^{obs} = Y_i(W_i) = \begin{cases} Y_i(0) & \text{if } W_i = 0 \\ Y_i(1) & \text{if } W_i = 1 \end{cases}, Y_i^{mis} = Y_i(1 - W_i) = \begin{cases} Y_i(1) & \text{if } W_i = 0 \\ Y_i(0) & \text{if } W_i = 1 \end{cases}$$

给定  $N$  个单位构成的总体, 分配机制 (Assignment Mechanisms) 是一个行可交换函数  $Pr(W | X, Y(0), Y(1))$ , 取值为  $[0, 1]$ , 满足:

$$\sum_{\mathbf{W} \in \{0,1\}^N} Pr(\mathbf{W} | \mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) = 1.$$

样本有限时,  $W$  有随机性,  $X, Y(0), Y(1)$  固定; 超总体样本中, 认为  $W, X, Y(0), Y(1)$  均具有随机性

- 倾向性得分: 具有共同协变量值的子样本集的个体水平分配概率的平均值
  - 分配机制的个体化分配假设: 限制个体  $i$  的治疗分配对其他个体的结果和分配的依赖性。
  - 分配机制的非混淆性假设: 限制了分配机制对潜在结果的依赖性。
11. 观察性研究: 称未知分配机制的设计称为观察性研究。若该分配机制同时是个体化的, 概率的, 以及具有不可混淆性的, 则分配机制时正规的 (Regular)。知道分配机制的函数形式则成为经典随机实验。
12. 随机化实验的分类: 伯努利实验, 完全随机实验, 分层随机实验, 配对随机实验
- 伯努利实验: 干预状态可以投硬币决定,  $N$  个个体有  $2^N$  分配向量; (矩阵是  $N \times 2^N$ )
  - 完全随机实验: 若固定数量的受试者被分配接受积极治疗, 则称该分配机制是完全随机试验。分配机制是:

$$Pr(\mathbf{W} | \mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) = \begin{cases} \left(\frac{N}{N_t}\right)^{-1} & \text{if } \sum_{i=1}^N W_i = N_t \\ 0 & \text{otherwise} \end{cases}$$

其中

$$\binom{N}{N_t} = \frac{N!}{N_t!(N - N_t)!}$$

- 分层随机实验: 某些个体特征或变量  $X$  对潜在结果有重要影响是, 先按照  $X$  进行分层, 在层内随机。
- 配对随机实验: 根据某个特征, 将所有个体分成  $N/2$  组 ( $N$  为偶数), 每组只有 2 个个体, 并且两个个体随机化分配, 1 个在干预组, 1 个在控制组。

## Part2: 完全随机实验

### 1. 费希尔精确 P 值方法

- 简介: 对于一个完全随机实验的数据, 采用零假设的情况下, 实验中的每个个体, 两个潜在结果的值是相同的。基于此费希尔开发了计算“P 值”的方法, 称为费希尔精确 P 值 (FEPs)。可以用于除了个体治疗效果均为零的尖锐零假设之外的零假设, 好处是可以插值出缺失的很多数据。此处  $Y_i(1)$  和  $Y_i(0)$  相差或相除为一个常数都行。

### 2. 统计量

统计量  $T^{dif}$ ,  $T^{rank}$  可以用于评估因果效应的显著性。

$$T^{\text{dif}} = \left| \bar{Y}_t^{\text{obs}} - \bar{Y}_c^{\text{obs}} \right| = \left| \frac{\sum_{i:W_i=1} Y_i^{\text{obs}}}{N_t} - \frac{\sum_{i:W_i=0} Y_i^{\text{obs}}}{N_c} \right|$$

$$R_i = R_i(Y_1^{\text{obs}}, \dots, Y_N^{\text{obs}}) = \sum_{j=1}^N \mathbf{1}_{Y_j^{\text{obs}} < Y_i^{\text{obs}}} + \frac{1}{2} \left( 1 + \sum_{j=1}^N \mathbf{1}_{Y_j^{\text{obs}} = Y_i^{\text{obs}}} \right) - \frac{N+1}{2}.$$

$$T^{\text{rank}} = |\bar{R}_t - \bar{R}_c| = \left| \frac{\sum_{i:W_i=1} R_i}{N_t} - \frac{\sum_{i:W_i=0} R_i}{N_c} \right|$$

- 有多个感兴趣的潜在结果时，可以把它看成是一个向量。可以将测试统计量合并成一个统计量，然后针对该统计量计算得到 P 值。
- 实验者也可以使用霍特林统计量，用于衡量治疗组和对照组平均值之间的马氏平方距离。
- 当样本量很大时，具有正概率的分配机制向量的选择会剧增，计算会很困难，此时可以用蒙特卡洛模拟来做。

$$\hat{p} = \frac{1}{K} \sum_{k=1}^N \mathbf{1}_{T^{\text{dif},k} \geq T^{\text{dif},\text{obs}}}.$$

- 例子中样本很小时，很难推出一个可信的推断。

## 2. 奈曼重复抽样方法

- 简介：奈曼专注于平均治疗效果的估计和推断方法；试图找到无偏的点估计量，以及在大样本中的区间估计量。用该方法，我们感兴趣的估计量是这些平均值之间的差异。我们的目标是为因果估计构建一个区间估计量，且该估计量是平均治疗效果估计量的样本方差的无偏估计量。
- 样本方差：

平均治疗效果的估计量：

$$\hat{\tau}^{\text{dif}} = \bar{Y}_t^{\text{obs}} - \bar{Y}_c^{\text{obs}}$$

样本方差：

$$\mathbb{V}_W(\bar{Y}_t^{\text{obs}} - \bar{Y}_c^{\text{obs}}) = \frac{S_c^2}{N_c} + \frac{S_t^2}{N_t} - \frac{S_{tc}^2}{N}$$

$S_c^2$ ， $S_t^2$  分别是  $Y_i(0)$  和  $Y_i(1)$  和样本方差，是不可知的（因为潜在结果不可知），需要估算

样本方差的另一种表示：

$$\mathbb{V}_W \left( \bar{Y}_t^{\text{obs}} - \bar{Y}_c^{\text{obs}} \right) = \frac{N_t}{N \cdot N_c} \cdot S_c^2 + \frac{N_c}{N \cdot N_t} \cdot S_t^2 + \frac{2}{N} \cdot \rho_{tc} \cdot S_c \cdot S_t$$

相关系数  $\rho_{tc} = 1$  时, 估计量  $\hat{\tau}^{dif}$  的样本方差达到最大值.

$$\begin{aligned} \mathbb{V}_W \left( \bar{Y}_t^{\text{obs}} - \bar{Y}_c^{\text{obs}} \mid \rho_{tc} = 1 \right) &= \frac{N_t}{N \cdot N_c} \cdot S_c^2 + \frac{N_c}{N \cdot N_t} \cdot S_t^2 + \frac{2}{N} \cdot S_c \cdot S_t \\ &= \frac{S_c^2}{N_c} + \frac{S_t^2}{N_t} - \frac{(S_c - S_t)^2}{N} \end{aligned}$$

常数治疗效果假设情形：

当治疗效果是常数以及具有可加性时, 即  $Y_i(1) - Y_i(0) = \tau$ , 自动满足相关系数  $\rho_{tc} = 1$  的条件, 此时有：

$$\mathbb{V}^{\text{const}} = \mathbb{V}_W \left( \bar{Y}_t^{\text{obs}} - \bar{Y}_c^{\text{obs}} \mid \rho_{tc} = 1, S_c^2 = S_t^2 \right) = \frac{S_c^2}{N_c} + \frac{S_t^2}{N_t}$$

•  $\hat{V}^{\text{neyman}}$  估计量样本方差估计：

若  $Y_i(1) - Y_i(0) = C$ , 则  $\hat{\tau}^{dif}$  的样本方差的一个无偏估计量是：

$$\hat{V}^{\text{neyman}} = \frac{s_c^2}{N_c} + \frac{s_t^2}{N_t}$$

其中

$$s_c^2 = \frac{1}{N_c - 1} \sum_{i: W_i=0} \left( Y_i(0) - \bar{Y}_c^{\text{obs}} \right)^2 = \frac{1}{N_c - 1} \sum_{i: W_i=0} \left( Y_i^{\text{obs}} - \bar{Y}_c^{\text{obs}} \right)^2$$

$$s_t^2 = \frac{1}{N_t - 1} \sum_{i: W_i=1} \left( Y_i(1) - \bar{Y}_t^{\text{obs}} \right)^2 = \frac{1}{N_t - 1} \sum_{i: W_i=1} \left( Y_i^{\text{obs}} - \bar{Y}_t^{\text{obs}} \right)^2$$

与平均治疗效果的估计量  $\hat{\tau}^{dif}$  的真实样本方差一样大, 因此是统计保守的。此外, 它作为无限超总体的平均治疗效果  $\hat{\tau}^{dif}$  的样本方差估计量时, 它是一个无偏估计。

• 另一种保守估计  $\hat{V}^{\rho_{tc}=1}$ ：

对于另一个统计保守的样本方差估计量



$$\mathbb{V}_W(\bar{Y}_t^{\text{obs}} - \bar{Y}_c^{\text{obs}} \mid \rho_{tc} = 1)$$

也可以通过治疗组和对照组的样本方差来实现保守估计：

$$\begin{aligned}\hat{\mathbb{V}}^{\rho_{tc}=1} &= s_c^2 \cdot \frac{N_t}{N \cdot N_c} + s_t^2 \cdot \frac{N_c}{N \cdot N_t} + s_c \cdot s_t \cdot \frac{2}{N} \\ &= \frac{s_c^2}{N_c} + \frac{s_t^2}{N_t} - \frac{(s_t - s_c)^2}{N}\end{aligned}$$

该统计量在治疗组和对照组的样本方差不一致时, 将同样实现保守估计, 且比对应的  $\hat{V}^{neyman}$  置信区间更紧。(但很多时候相差不大)

· 常数治疗效果假设情形下的估计  $\hat{\mathbb{V}}^{const}$  :

考虑在治疗效果不变的附加假设下, 即:

$$Y_i(1) - Y_i(0) = \tau_{fs}.$$

这个治疗的无效性假设隐含了, 两个潜在结果的总体方差相等的事实

$$S^2 \equiv S_c^2 = S_t^2$$

该共同方差的估计量为

$$\begin{aligned}s^2 &= \frac{1}{N-2} \cdot (s_c^2 \cdot (N_c - 1) + s_t^2 \cdot (N_t - 1)) \\ &= \frac{1}{N-2} \cdot \left( \sum_{i:W_i=0} (Y_i^{\text{obs}} - \bar{Y}_c^{\text{obs}})^2 + \sum_{i:W_i=1} (Y_i^{\text{obs}} - \bar{Y}_t^{\text{obs}})^2 \right)\end{aligned}$$

当治疗效果为常数时, 样本量越大, 统计量  $\bar{Y}_t^{\text{obs}} - \bar{Y}_c^{\text{obs}}$  样本方差的估计越精确:

$$\hat{\mathbb{V}}^{const} = s^2 \cdot \left( \frac{1}{N_c} + \frac{1}{N_t} \right)$$

· 三种估计量的总结

若个体的治疗效果无差异, 为常数,  $\hat{\mathbb{V}}^{const}$  优于  $\hat{V}^{neyman}$  和  $\hat{\mathbb{V}}^{\rho_{tc}=1}$ , 否则统计量容易失效。

$\hat{V}^{neyman}$  和  $\hat{\mathbb{V}}^{\rho_{tc}=1}$  通常有效且保守, 故而实验经常采用。

### 3. 例子：教师工资激励计划数据



亮点：ATE=-0.03，随机化做的比较好；有些统计学者在关注再随机化，在数据收集的时候做一些调整，统计上会造成一些截断。

## Part3 总结

- 费希尔关心有限样本情形下，尖锐零假设的假设检验，在零假设成立时，可以精确求出统计量的分布，蕴含了bootstrap的思想。
- 奈曼在完全随机实验中关心治疗的平均效果，提出了有限样本中ATE的一个无偏估计量。
- 为抽样方差找到一个本身无偏的估计量是不可能的，相反，奈曼证明了这个估计量的抽样方差的标准估计量是正偏的，除非处理效果是常数和可加的，在这种情况下它是无偏的。

## 【精彩问答】

【记录人】杨二茶 | 清华大学在读博士

1. 倾向得分分配机制很多都是指协变量有关，这里做法是  $Y(0)$  和  $Y(1)$  与前面结果有关，请问考虑带来了什么好处，显然数学公式变得复杂了。

李昊轩：其实不是，这里本质上来讲也是协变量函数  $e(x)$ ，这里的  $x$  是协变量，这里会有一个分配机制。倾向性得分定理是接受treatment的平均概率，既有协变量  $e(x) = x$  这样的个体，如果要考虑接受treatment的概率的话，肯定会depends on  $x$  或者depends on  $Y(0)$  或者  $Y(1)$ ，至少depends on  $x$ ，分配机制和协变量来讲是相关的。

2. 右边写了潜在结果有关，实际上还是与协变量有关的函数？

李昊轩：对。是协变量的函数。

3. 如果和协变量有关， $Y(0)$ ， $Y(1)$  能不能当成selection bias的考虑呀？

李昊轩：可以这样理解。assignment mechanism我们认为，一个人是否接受treatment是depends on  $x$  或者depends on  $Y(0)$  或者  $Y(1)$ 。对分配机制来说是协变量求和，所以可以理解成是selection bias。

4.  $x$ 是所有可观测的协变量，这个就是说这机器学习里面，数据是什么？

李昊轩：数据就是 $x$ 。但是 $x$ 来讲现实意义，一定会有不可观测的混杂和协变量。 $x$ 可以看作是纳入研究的所有协变量，他必须是可观测的。比如某些treatment或者outcomes就可以不纳入。这个在倾向性匹配匹配得分很重要，很重要的是怎么选择协变量。

5. 解释selection bias的时候举过电影的例子，然后感觉解释selection bias产生sample的原因是因为存在一个confounding影响了他会不会打分和看电影，想问confounding会引起selection bias吗？

李昊轩：对，在Rubin框架里，没有定义confounding bias或者selection bias，但是这又是一个特别重要的问题。在刚才电影评分例子给大家讲什么是selection bias，推荐的场景里，我们需要去定义一下confounding bias和selection bias，这里的selection bias本质上可以理解为数据收集时候产生的bias。但是实际上，selection bias和confounding bias是区分开的，暴露的这批人不能体现群体，这个叫selection bias。confounding bias可能是违背了假设，given  $x$  按照confounding的假设不成立。很多时候sample, given  $IE[Y(0) | Y = 0]$  是不能等于  $E[Y(0)]$  也就是说你用observation数据没有办法反映群体，这个叫confounding bias。大家可以想一想有没有更合理的定义方式。

confounding:

$$IE[Y(0) | Y = 0] \neq E[Y(0)]$$

袁冰：Selection bias在现实应该很多，高校里做调查，看看特长生和学习成绩之间负相关，往往特长生成绩都低于平均，给人一个假象，是不是要锻炼身体，就一定学习成绩不好，这就是典型的selection bias。这种调查只在高校学生样本进行，但是高校入学机制就是selection bias，特长生的招生机制，本身没有考虑整体的学生，这个招生机制使得样本产生了虚假相关。selection bias用因果图的话，会更加清晰一些。

李昊轩：对。大家也可以参考吴鹏师兄之前说的，数据生成机制中导致的和数据收集机制中导致的。你收集到数据，没有的假设，你很  $IP = IP_0$  难在往下做一些推断。现在来讲，很多人都是基于这样的假设考虑confounding bias，这样假设很多时候是很必要的。

李奉治：心理学实验本身就有限制，之前很多心理学结论的适用范围就是美国大学白人男性学生，这就是整个学科的一个selection bias。

李昊轩：所以你怎么定义population很重要，这个假设本身也是很必要。

6. SUTVA和no inference是等价的吗？

李昊轩：SUTVA里有一条是no inference，另外一条是no alternative formal treatment，这两条合起来叫SUTVA。目前讨论都是基于SUTVA 假设，measure confounder也是大家关注的一个问题。

7. Selection bias有时候会被model成collider bias，不知道causal inference是不是也有文章这么做？

李昊轩：在pearl的体系里，collider bias是一种产生confounder的方式，是一种confounding bias。这是在Pearl框架里，Rubin框架里是违背了unconfounder的假设，这个其实是我们想去说明的。Rubin里的confounding不只是confounding，还有可能是其他的导致的。

8. Assignment mechanism是只针对于treatment而言吗？不针对covariate或outcome variable？

李昊轩： $w$  是treatment。分配机制的定义是一个probability  $w$ ，因为这final sample你只认为 $w$  的随机性，给定协变量和潜在结果情况下，分配概率是多少，其实是一个概率分布。从随机性角度来讲，只有 $w$ 是随机性的。

9. 关于第二部分，完全随机实验的两个，Fisher和Neyman之间的分歧，怎么直观上理解，在其他统计模型上是不是也有这样的分歧？

李昊轩：从因果角度来讲，Fisher研究零假设，每个人的Potential outcome在接受治疗和不接受都是一样的，一个药对所有人来讲都无效。你可以接受或者拒绝这个假设，在实际应用中，没有办法给一个confidence interval，也不能做统计。他非常直接到讨论所有个体来说，都是无效的。Neyman就是研究Average effect，所以要根据对应问题的情景，产生两种学派和分析问题的模式。Neyman关心怎么给一个整体的治疗效果。感兴趣的 $1/n * Y(I=1)$ 和 $Y(I=0)$ 怎么去estimate，怎么给出渐进的理论，怎么构造confidence interval，构建hypothesis test。一个test个体是不是0，另一个test整体的ATE是0.这是整个学派分歧所在。

10. 我们主要用Neyman，Fisher怎么发展的，bootstrap应该很常用？

李昊轩：蕴含了bootstrap的思想。从时间来讲，Fisher的东西要更早，原来就有，也存在大家说的

11. Bayesian和Fisher其实没有什么关系，这两者的话？和Fisher Lehman的分歧？

李昊轩：是有的。下一次会讲model-based imputation，就是用模型去插补，会对 $Y(0)$ 和 $Y(1)$ ，有一个prior distribution，通过你的observation data去调整distribution，得到posterior distribution。和Fisher-Lehman没有关系。

12. 对因果模型的研究基本上潜在结果模型和因果图墨香两个主要方面，潜在结果模型的假设，随机性是很强的条件，实际应用中很难满足，包括刚刚举到的工具变量方法，混杂偏差的影响，要使

得通过随机性的办法，理论上的结果，实际上非参是满足的。这个被认为潜在结果模型作用缺陷之一。Pearl提出因果图模型，反事实和干预的方法，能不能消除潜在结果模型本身就达不到随机实验带来的弊端，能不能通过图模型解决，或者结构方程模型解决吗？

李昊轩：因果推断本身没有additional assumption是不能做可信推断的。在两个框架下，都有很强的假设的。我个人认为rubin假设更弱，因果图模型中pearl的框架包含两个部分，第一部分是causal discovery，第二部分是causal effect estimate。首先要有因果图，这个图的构建蕴含的假设非常多。最强的假设不在于链接边，而在于没有链接的边。两者面临的挑战是共同的，比如confounding假设和unmeasured confounders。Rubin的假设会更弱一些，你可以认为我去相信我现在拿到的协变量信息，以至于给定协变量信息后，我都分配是随机的，独立于潜在结果的。这个比直接给因果图的假设，Rubin假设更弱些。很多文章做Pearl框架都有因果图假设，通过因果图学习模型。包括反事实框架，是上一期主题。Counterfactual总有建模的过程，这个用什么去建，很多人是用神经网络建，实际上你本质上还是捕捉相关性。

### 13. 对于SCM模型，我们是否可以把随机性假设放宽，直接通过结构方程模型估计因果？

吴鹏：POM是没有图结构的，SCM模型本身蕴含了因果图。所以POM条件要弱一些。POM要不要随机化，那是后一阶段的过程。目前我们用的只是符号体系用来清晰定义因果关系。

李昊轩：潜在结果框架，我们希望通过一些表示学习的方法，BNN，BOR模型，本质上通过因果表征，达到一种拟的随机化。刚才我们讲unconfoundness假设，控制协变量后，我们当做一个随机实验，或者condition一个balance score，可以理解为是一个随机的实验。即使是SCM，我不考虑因果图，构造X对Y，问题是什么呢？是问因果项存在与否，是要捕捉causal discovery还是问X对Y的因果效应有多少，如果因果图建立错误了，你是永远得不到因果效应的。

宋：实验所带来的信息远比我们观察到的信息要大很多，因为不是一个层次的信息，所以没有这个假设的话，会有很多问题。也有观察性研究的，不一定必须要随机化来完成的。

袁冰：您提这个就是因果发现的问题。Pearl对POM方法的批判主要是不透明，因果图是可以把形式化假设都有描述，剩下的都是数学问题，就可以用计算机解决的自动化问题。他认为POM假设太过零散和分散，大家交流起来缺乏协同效应。Pearl的意思是用统一语言，用严格的形式化推理过程。

李昊轩：因果发现可以根据半监督学习，去得出有没有因果效应，在faithful假设下进行发现。在有图后，我们讨论causal effect estimate，我们还是要假设这个图是对的。图就是这些独立性。

### 14. 关于Social Network违背SUTVA假设的问题？

李昊轩：Social Network 样本和个体都是嵌入网络中，包括推荐里的知识图谱的方式，都可以违背基本都SUTVA假设，主要还是统计上关注的问题，即使做社交网络和因果腿短，还是要遵守iid假设。这个问题在推荐和社会网络都是具有挑战的，怎么和机器学习做结合，也是未来的因果效应。你的估计量肯定是有偏的，现在是统计要求比较高。推荐elements of causal inference，这个讲因果发现的问题，大家还是要把因果发现和因果效应分开。

- Peters, J., Janzing, D., & Schölkopf, B. (2017). *Elements of causal inference: foundations and learning algorithms* (p. 288). The MIT Press.

## 【参考文献】

1. Peters, J., Janzing, D., & Schölkopf, B. (2017). *Elements of causal inference: foundations and learning algorithms* (p. 288). The MIT Press.

## 【读书会PPT】



20211031 【李昊轩】Causal\_Talk .pdf  
2.24MB

