高级计量经济学

Lecture 11: The Neyman-Rubin Causal Model

黄嘉平

工学博士 经济学博士
深圳大学中国经济特区研究中心 讲师

办公室 粤海校区汇文楼1510
E-mail huangjp@szu.edu.cn
Website https://huangjp.com

Neyman-Rubin Causal Model

Leamer 针对实证研究的批评以及可信性革命

The Leamer critique and the Credibility Revolution

Leamer (1983) 针对当时的经济学实证研究提出了如下批评：

- 控制变量、回归函数、样本的选择具有主观性。很多文章倾向于展示漂亮的（而非正确的）结果。
- 实验和观测数据都存在以上问题，只是程度有所不同。
- 回归分析对非现实假设的依赖性和结果的脆弱性降低了研究结论的可信度。

Leamer 提出的解决方法是敏感性分析（sensitivity analysis），简单的说就是针对同一问题用不同方法进行分析，并观察结果是否稳健。

1980 年代中期开始，随着因果推断方法在劳动经济学等领域的兴起，以因果推断为基础的研究设计逐渐在实证研究中占据主要地位，后来被 Angrist & Pischke (2010) 称为可信性革命（credibility revolution）。他们认为，造成实证研究可信性不足（无法揭示因果效应）的主要原因是缺乏正确的研究设计（research design）。

Leamer (1983). Let's Take the Con Out of Econometrics. *AER*, 73:1, 31-43.

Angrist & Pischke (2010). The Credibility Revolution in Empirical Economics: How Better Research Design is Taking the Con out of Econometrics. *JEP*, 24:2, 3-30.

因果推断：我们应该关注什么?

Focus of Causal Inference

Angrist & Pischke (2009) 中总结了四个关于因果推断研究的主要关注点（他们称为 FAQ）：

- 什么是因果关系：我们需要基于潜在结果定义因果关系
  - 清华大学毕业生的收入是否高于深圳大学毕业生?
  - 单纯的比较两校毕业生的平均收入差无法体现因果关系，我们应当比较同一个学生毕业于清华时的收入和毕业于深大时的收入，即基于事实和反事实的比较。
- 什么时候可以用实验方法：
  - 可行性：经费允许、不存在伦理和社会问题
  - 处理变量不影响实验结果：
    推迟小学入学年龄是否能提高学生成绩？小学生的学习成绩受智力发育程度和学习年限的影响，而智力发育程度和年龄相关。但是 年龄 = 入学年龄 + 学习年限，因此无法在同时控制年龄和学习年限的前提下，随机分配入学年龄。也就是说，处理变量（入学年龄）会影响结果（学习成绩）。
- 如何识别：在不同假设条件下的识别方法
- 如何进行统计推断：估计目标的性质，估计量的性质，标准误差的非偏估计等

Angrist & Pischke (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Neyman-Rubin 因果模型

The Neyman-Rubin Causal Model

现在学术界广泛认为 Neyman (1923) 首次针对完全随机实验明确定义了潜在结果的概念。Rubin (1974) 及其后续研究将此概念推广到其他类型的实验研究和观测性研究（observational studies）。

Neyman-Rubin 因果模型是建立在潜在结果概念上的因果推断框架。

- 处理变量（treatment variable）：W_i 代表个体 i 接受的处理（W_i 可以取离散值和连续值，最简单的设定是 $W_i \in \{0,1\}$，1 代表接受处理，0 代分没有接受处理）
- 潜在结果（potential outcome）：Y_{1i} 代表 i 接受处理时的结果, Y_{0i} 代表 i 没有接受处理时的结果
- 观测结果（observed outcome）：Y_i^{obs} 代表实际观测到的潜在结果，Y_i^{mis} 代表没有观测到的潜在结果
- 个体处理效应（unit-level treatment effect）：$Y_{1i} - Y_{0i}$ （或 Y_{1i}/Y_{0i}）

在个体层面无法同时观测所有的潜在结果，因此无法计算个体处理效应。这一性质被称为因果推断的基本问题（the fundamental problem of causal inference），是一种缺失值问题（missing value problem）。

Neyman (1923). On the application of probability theory to agricultural experiments. Essay on principles. Section 9. *Statistical Science*, 5:465–80.

Rubin (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66:5, 688-701.

因果推断

Causal Inference

即然个体处理效应无法获得，我们就将注意力转移到总体中的处理效应。

估计目标（estimand）是研究中进行统计推断的目标量。估计目标是针对总体定义的，而估计量（estimator）是用样本定义的估计目标的近似量（或者说是一种估计方法），估计值（estimate）是估计量的观测值。在回归分析中，总体回归函数中的系数都是估计目标。

在确定一个总体后，我们可以定义不同种类的因果效应估计目标，例如均值、分位数等：

- 平均处理效应（average treatment effect, ATE）：

$$\tau_{\text{ATE}} = E[Y_{1i} - Y_{0i}] = \frac{1}{N}\sum_{i=1}^{N}(Y_{1i} - Y_{0i})$$

- 处理组的平均处理效应（average treatment effect on the treated, ATET）：

$$\tau_{\text{ATET}} = E[Y_{1i} - Y_{0i} \mid W_i = 1] = \frac{1}{N_1}\sum_{i:W_i=1}(Y_{1i} - Y_{0i})$$

平均处理效应也无法直接计算，因此需要对它们进行估计。

SUTVA 假设

The Stable Unit Treatment Value Assumption

稳定个体处理值（干预值）假设（stable unit treatment value assumption, SUTVA）：

1. 潜在结果 Y_{1i}, Y_{0i} 不受其他处理结果 $W_{j \neq i}$ 的影响（no interference）

   学生的学习成绩也会受到周围其他同学的影响（peer effects），因此在讨论某个变量对成绩的因果效应时，SUTVA 往往不成立。

   关于读研是否会增加工作收入这一问题，如果读研的人数非常多，则毕业后的就业市场会形成过度竞争（general equilibrium effects），从而影响工作收入，此时 SUTVA 也不成立。类似的例子还有接种疫苗的有效性（个体接种效果受整体接种率的影响）。

2. 对于每一种处理 k，潜在结果 Y_{ki} 是唯一的（no hidden variations of treatment）

   当读研与否是一种处理时，研究生教育的水平应该是一致的，而不应该出现水平不一的研究生院（现实中显然是不同的）。

   潜在结果也不应被接受处理的方式影响。例如推免生和普通考生在研究生阶段的学习效果应该一致（现实中可能不一样）。

选择偏差

Selection Bias

当 SUTVA 成立时，个体的观测结果 Y_i^{obs} 和反事实结果 Y_i^{mis} 可以表达为

$$Y_i^{\text{obs}} = \begin{cases} Y_{1i} & \text{if } W_i = 1 \\ Y_{0i} & \text{if } W_i = 0 \end{cases} \quad \Leftrightarrow \quad Y_i^{\text{obs}} = Y_{1i} W_i + Y_{0i}(1 - W_i)$$

$$Y_i^{\text{mis}} = \begin{cases} Y_{0i} & \text{if } W_i = 1 \\ Y_{1i} & \text{if } W_i = 0 \end{cases} \quad \Leftrightarrow \quad Y_i^{\text{mis}} = Y_{0i} W_i + Y_{1i}(1 - W_i)$$

在下面的讨论中通常只需要用到 Y_i^{obs}。当我们考虑用观测结果估计平均处理效应时，有

$$\begin{aligned} E[Y_i^{\text{obs}} \mid W_i = 1] - E[Y_i^{\text{obs}} \mid W_i = 0] &= E[Y_{1i} \mid W_i = 1] - E[Y_{0i} \mid W_i = 0] \\ &= (E[Y_{1i} \mid W_i = 1] - E[Y_{0i} \mid W_i = 1]) \\ &\quad + (E[Y_{0i} \mid W_i = 1] - E[Y_{0i} \mid W_i = 0]) \end{aligned}$$

处理组和对照组的平均观测结果之差 = τ_{ATET} + 选择偏差

如果选择偏差为零，则 τ_{ATET} 可以通过观测结果估计。实证研究中（尤其是观测性研究中）选择偏差不会自动为零，因此因果推断研究的主要目的是如何通过研究设计克服选择偏差。

随机分配可以消除选择偏差

Random Assignment Eliminates Selection Bias

分配机制（assignment mechanism）是指如何确定 W_i 的取值。最简单的分配方法是随机分配（random assignment），即对任意 i，$\Pr(W_i = 1 \mid Y_{1i}, Y_{0i}) = \Pr(W_i = 0 \mid Y_{1i}, Y_{0i}) = 1/2$。

在随机分配机制下，处理变量和潜在结果相互独立，即 $(Y_{1i}, Y_{0i}) \perp\!\!\!\perp W_i$。此时可得，

$$\begin{aligned} E[Y_i^{\text{obs}} \mid W_i = 1] - E[Y_i^{\text{obs}} \mid W_i = 0] &= E[Y_{1i} \mid W_i = 1] - E[Y_{0i} \mid W_i = 0] \\ &= E[Y_{1i} \mid W_i = 1] - E[Y_{0i} \mid W_i = 1] \\ &= E[Y_{1i} - Y_{0i} \mid W_i = 1] \\ &= E[Y_{1i} - Y_{0i}] \end{aligned}$$

令 $\tau^{\text{diff}} = E[Y_i^{\text{obs}} \mid W_i = 1] - E[Y_i^{\text{obs}} \mid W_i = 0]$，则在随机分配机制下

$$\tau^{\text{diff}} = \tau_{\text{ATET}} = \tau_{\text{ATE}}$$

不难看出，为了消除选择偏差，我们只需满足独立性 $(Y_{1i}, Y_{0i}) \perp\!\!\!\perp W_i$。在观察性研究中，独立性假设不会自动被满足，而需要研究者对其进行讨论或提供其成立的证据。

关于 τ^{diff} 的其他统计性质，可参考

Imbens & Rubin (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge University Press.

用回归方法估计 ATE

Using Regression under Random Assignment

假设个体处理效应是恒定的（没有个体间差异），即 $Y_{1i} - Y_{0i} = \rho$。此时的估计目标就变成了参数 ρ。

考虑下面的回归模型

$$Y_i^{\text{obs}} = \alpha + \rho W_i + \varepsilon_i$$

其中 $\alpha = E[Y_{0i}], \varepsilon_i = Y_{0i} - E[Y_{0i}]$。由此模型可得

$$E[Y_i^{\text{obs}} \mid W_i = 1] = \alpha + \rho + E[\varepsilon_i \mid W_i = 1]$$
$$E[Y_i^{\text{obs}} \mid W_i = 0] = \alpha + E[\varepsilon_i \mid W_i = 0]$$

取两式之差可得

$$\tau^{\text{diff}} = \rho + \text{选择偏差}$$

在随机分配机制下，选择偏差为零，$\tau^{\text{diff}} = \rho = \tau_{\text{ATE}} = \tau_{\text{ATET}}$。平均处理效应可以用回归系数的 OLS 估计量来估计，即 $\hat{\tau}_{\text{ATE}} = \hat{\tau}_{\text{ATET}} = \hat{\rho}_{\text{OLS}}$。

回归方法的优点是可以方便地加入其他控制变量，并可以很容易的拓展到多值处理变量。缺点是需要假设处理效应不存在个体差异。

干扰

Confounding

和回归分析类似，在潜在结果模型中，处理变量和潜在结果变量也可能受到其他因素的影响，我们称其为干扰（confounding）。

如果干扰以变量的形式存在，我们将其称为干扰变量（confounding variables, confounders）或者协变量（covariates），记为 $\boldsymbol{X}_i$。协变量的存在可能威胁到独立性假设 $(Y_{1i}, Y_{0i}) \perp\!\!\!\perp W_i$ 的正当性，从而影响平均处理效应的估计。

当协变量可观测时，我们可以考虑下面的替代条件

- 条件独立假设（conditional independence assumption, CIA）：
$$(Y_{1i}, Y_{0i}) \perp\!\!\!\perp W_i \mid \boldsymbol{X}_i$$
- 条件均值独立假设（conditional mean independence, CMI）：
$$E[Y_{1i} \mid W_i, \boldsymbol{X}_i] = E[Y_{1i} \mid \boldsymbol{X}_i], \quad E[Y_{0i} \mid W_i, \boldsymbol{X}_i] = E[Y_{0i} \mid \boldsymbol{X}_i]$$

CMI 弱于 CIA，下面为了方便我们都假设 CIA。

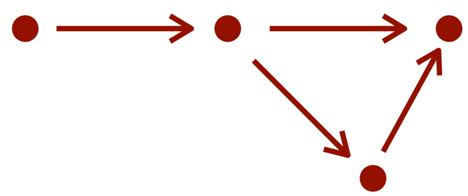
令 $\tau^{\text{diff}}(\boldsymbol{X}_i) = E[Y_i^{\text{obs}} \mid W_i = 1, \boldsymbol{X}_i] - E[Y_i^{\text{obs}} \mid W_i = 0, \boldsymbol{X}_i]$，则 $\tau^{\text{diff}}(\boldsymbol{X}_i)$ 是控制协变量后的平均处理效应。整体平均处理效应可以用 $\tau^{\text{diff}}(\boldsymbol{X}_i)$ 计算。

由于 CIA 和 CMI 基于协变量的可观测性，因此也被称为 selection on observables 条件。

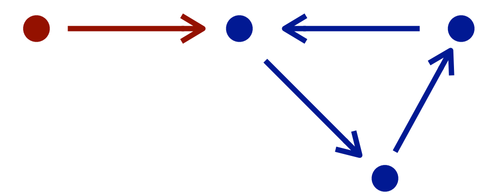
有向无环图

Directed Acyclic Graph (DAG) Representation

Pearl (2000) 提倡用有向无环图（directed acyclic graph, DAG）表达潜在结果模型中的因果关系。



有向图（directed graph）

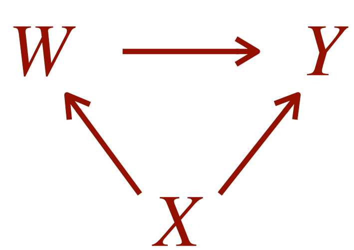


环（cycle）

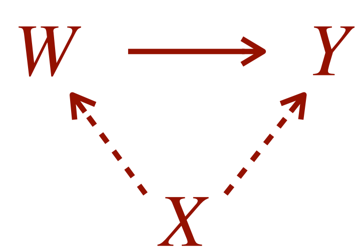
我们可以用图中的节点代表变量，用边代表因果关系的方向。例如 $W \rightarrow Y$ 代表 W 引起了 Y 的变化（W causes Y），W 为因 Y 为果。

$$W \longrightarrow Y$$

如果存在同时影响 W 和 Y 的协变量 X，则 DAG 可以表达为



X 可观测



X 不可观测

注意：如果两个节点间没有边连接，则代表二者间**不存在**直接因果关系。

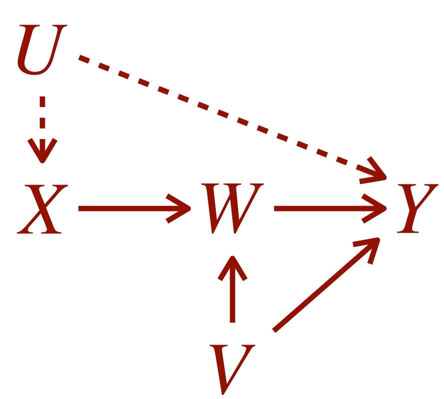
Pearl (2000, 2009). Causality: Models, Reasoning, and Inference, 1st & 2nd Edition. Cambridge University Press.

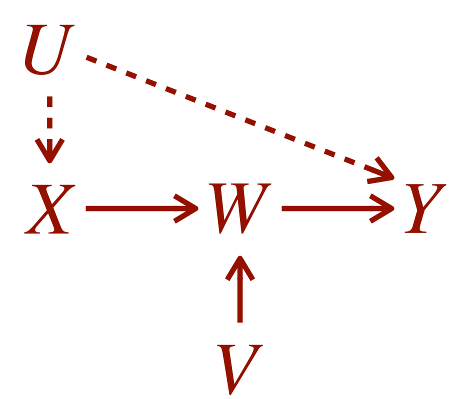
基于可观测协变量的识别

Identification under Selection on Observables

下面以奖学金和学习成绩间的因果效应问题为例。设处理变量 W 为奖学金获得情况，潜在结果 Y 为奖学金发放后的学生成绩，协变量包括 X（入学时的成绩），U（入学前的学术能力），V（学生的个人特征等）。

可以考虑右侧两种模型：

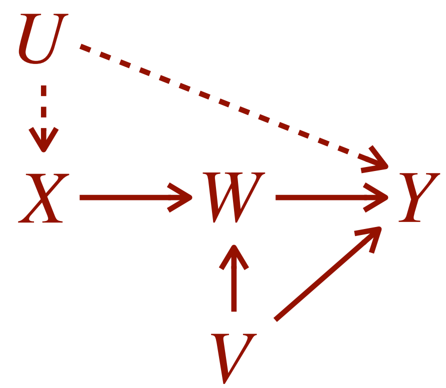


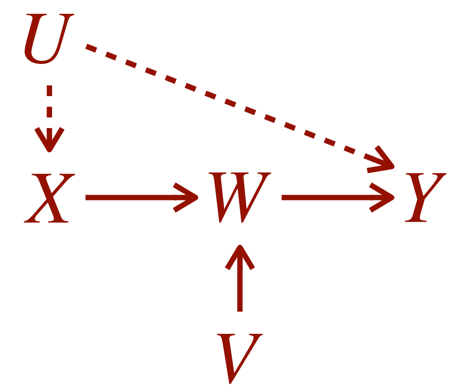


因果效应（直接或间接）可以用图中的路径（path）表达。例如 $X \to W \to Y$ 代表 X 通过 W 对 Y 产生因果效应。模型中是否存在干扰可以通过是否存在后门路径（backdoor path）判断

- 后门路径：连接处理变量 W 和结果变量 Y 的路径，其中连接 W 的边是指向 W 的。

图中的 $W \leftarrow X \leftarrow U \to Y$ 和 $W \leftarrow V \to Y$ 是后门路径。如果模型中存在后门路径，则可能存在干扰，使 τ^{diff} 包含选择偏差。





在 CIA 条件 $(Y_1, Y_0) \perp\!\!\!\perp W \mid X$ 成立时，我们可以通过控制协变量 X 阻断后门路径 $W \leftarrow X \leftarrow U \rightarrow Y$ 的发生。

另外，我们假设下面的条件成立：

- 共同支撑假设（common support assumption）或重叠假设（overlap）： $0 < \Pr(W = 1 \mid X) < 1$

此时，

$$
\begin{aligned}
E[Y_{1i} - Y_{0i} \mid X_i] &= E[Y_i^{\text{obs}} \mid W_i = 1, X_i] - E[Y_i^{\text{obs}} \mid W_i = 0, X_i] \\
&= \tau^{\text{diff}}(X_i)
\end{aligned}
$$

$$
\Rightarrow \quad
\begin{aligned}
\tau_{\text{ATE}} &= E_X\big[E[Y_{1i} - Y_{0i} \mid X_i]\big] = E[\tau^{\text{diff}}(X_i)] \\
\tau_{\text{ATET}} &= E_X\big[E[Y_{1i} - Y_{0i} \mid X_i] \mid W_i = 1\big] = E[\tau^{\text{diff}}(X_i) \mid W_i = 1]
\end{aligned}
$$

在上面的两个模型中，后门路径 $W \leftarrow X \leftarrow U \rightarrow Y$ 可以通过控制 X 或 U 阻断，但是 $W \leftarrow V \rightarrow Y$ 只能通过控制 V 阻断。

如果 V 不可观测，则上图左侧的模型无法对因果效应进行识别，右侧的则可以。

以上例子取自 Abadie & Catteneo (2018). Econometric Methods for Program Evaluation. *Annual Review of Economics*, 10, 465-503.

选择偏差的分解*

Decomposition of Selection Bias*

Heckman et al. (1998) 对选择偏差 $E[Y_{0i} \mid W_i = 1] - E[Y_{0i} \mid W_i = 0]$ 进行了分解。他们得出

$$\text{选择偏差} = B_1 + B_2 + B_3$$

- B_1：来源于弱重叠（weak overlap）的偏差。
  如果存在 $\Pr(W = 1 \mid X = x) = 0, \Pr(W = 0 \mid X = x) > 0$ 或 $\Pr(W = 0 \mid X = x) = 0, \Pr(W = 1 \mid X = x) > 0$ 的 x，此时针对这个 x 无法进行处理组与对照组间的比较。
- B_2：来源于弱平衡（weak balance）的偏差。
  如果存在 $\Pr(W = 1 \mid X = x) \neq \Pr(W = 0 \mid X = x)$ 的 x，则关于这个 x，处理组与对照组中的观测值数量不同，会造成估计偏差。
- B_3：来源于不可观测协变量的偏差（selection on unobservables）。
  如果存在不可观测的协变量，则 CIA 或 CMI 不成立。（类似于回归分析中的遗漏变量偏差）

Heckman, Ichimura, Smith, & Todd (1998). Characterizing Selection Bias Using Experimental Data. *Econometrica*, 66:5, 1017-1098.

Methods under
Selection on Observables

匹配估计量

Matching Estimator

假设只存在一个可观测协变量 X_i，且 X_i 取离散值 $x_1, \ldots, x_m$。同时假设 CIA 和共同支撑假设成立。
前面定义了 $\tau^{\text{diff}}(X_i) = E[Y_i^{\text{obs}} \mid W_i = 1, X_i] - E[Y_i^{\text{obs}} \mid W_i = 0, X_i]$，因此 ATE 和 ATET 可以表达为

$$\begin{aligned}
\tau_{\text{ATE}} &= E[Y_{1i} - Y_{0i}] = E\big[E[Y_{1i} - Y_{0i} \mid X_i]\big] \\
&= E[\tau^{\text{diff}}(X_i)] = \sum_{k=1}^{m} \tau^{\text{diff}}(x_k)\Pr(X_i = x_k), \\
\tau_{\text{ATET}} &= E[Y_{1i} - Y_{0i} \mid W_i = 1] = E\big[E[Y_{1i} - Y_{0i} \mid X_i] \mid W_i = 1\big] \\
&= E[\tau^{\text{diff}}(X_i) \mid W_i = 1] = \sum_{k=1}^{m} \tau^{\text{diff}}(x_k)\Pr(X_i = x_k \mid W_i = 1)
\end{aligned}$$

根据等式最右侧的表达式可知，平均处理效应是条件均值之差 $\tau^{\text{diff}}(X_i)$ 的加权平均。

样本中对应 $\tau^{\text{diff}}(x_k)$ 的是 $\hat{\tau}^{\text{diff}}(x_k) = \frac{1}{n_{1k}}\sum_{j=1}^{n_{1k}} y_{1j}^{\text{obs}} - \frac{1}{n_{0k}}\sum_{j=1}^{n_{0k}} y_{0j}^{\text{obs}}$，而用作权重的概率也可以由样本提供的经验分布计算。因此平均处理效应的估计量是

$$\hat{\tau}_{\text{ATE}} = \sum_{k=1}^{m} \hat{\tau}^{\text{diff}}(x_k)\, p_k, \quad \hat{\tau}_{\text{ATET}} = \sum_{k=1}^{m} \hat{\tau}^{\text{diff}}(x_k)\, p_{k|W_i=1}$$

文献中称这种估计量为匹配估计量（matching estimator）或 subclassification。

回归调整
Regression Adjustment

令 D_{ik} 为 $X_i = x_k$ 时取值为 1 的虚拟变量。如果假设个体处理效应恒定，则可以考虑下面的回归模型

$$Y_i^{\text{obs}} = \rho W_i + \sum_{k=1}^{m} \beta_k D_{ik} + \varepsilon_i$$

Angrist & Pischke (2009) 应用 FWL 定理得出 ρ_{OLS} 可以表达为

$$\rho_{\text{OLS}} = \sum_{k=1}^{m} \tau^{\text{diff}}(x_k)\,\omega_k, \quad \omega_k = \frac{\text{Var}[W_i \mid X_i = x_k]\,\Pr(X_i = x_k)}{\sum_{r=1}^{m} \text{Var}[W_i \mid X_i = x_k]\,\Pr(X_i = x_r)}$$

因此，$\hat{\rho}_{\text{OLS}}$ 是用条件方差进行加权的匹配估计量。只有当 $\tau^{\text{diff}}(x_k) = \tau^{\text{diff}}$ （不随 x_k 的取值而改变），或 $\omega_k = p_k$ 时，$\hat{\rho}_{\text{OLS}}$ 才能正确估计 ATE 和 ATET。然而实践中，这两个条件都很难成立。

因为 $W_i \in \{0,1\}$，其条件方差是

$$\text{Var}[W_i \mid X_i = x_k] = \Pr(W_i = 1 \mid X_i = x_k)\big(1 - \Pr(W_i = 1 \mid X_i = x_k)\big)$$

因此当 $\Pr(W_i = 1 \mid X_i = x_k) = \frac{1}{2}$ 时取最大值，此时回归赋予 $\tau^{\text{diff}}(x_k)$ 的权重最大。

Angrist & Pischke (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

近邻匹配

Nearest-Neighbor Matching

前面介绍的匹配估计量是组间的匹配（控制协变量 $X_i = x_k$ 时，处理组与对照组间的均值比较）。匹配也可以在个体层面实现。

个体处理效应可以表达为

$$
\begin{aligned}
\tau_i = Y_{1i} - Y_{0i} &= \begin{cases} Y_i^{\text{obs}} - Y_i^{\text{mis}} \text{ if } W_i = 1 \\ Y_i^{\text{mis}} - Y_i^{\text{obs}} \text{ if } W_i = 0 \end{cases} \\
&= W_i(Y_i^{\text{obs}} - Y_i^{\text{mis}}) + (1 - W_i)(Y_i^{\text{mis}} - Y_i^{\text{obs}})
\end{aligned}
$$

我们可以用下面的办法对 Y_i^{mis} 进行估计：

1. 针对每个处理组中的 i，从对照组中找到使 $\|X_i - X_j\|$ 最小的 j，并令 $\hat{Y}_i^{\text{mis}} = Y_{j(i)}^{\text{obs}}$；
2. 针对每个对照组中的 i，从处理组中找到使 $\|X_i - X_j\|$ 最小的 j，并令 $\hat{Y}_i^{\text{mis}} = Y_{j(i)}^{\text{obs}}$；
3. 计算个体处理效应的估计量 $\hat{\tau}_i = W_i(Y_i^{\text{obs}} - \hat{Y}_i^{\text{mis}}) + (1 - W_i)(\hat{Y}_i^{\text{mis}} - Y_i^{\text{obs}})$；
4. $\hat{\tau}_{\text{ATET}} = \frac{1}{N_1}\sum_{i:W_i=1}\hat{\tau}_i$，$\hat{\tau}_{\text{ATEC}} = \frac{1}{N_0}\sum_{i:W_i=0}\hat{\tau}_i$，$\hat{\tau}_{\text{ATE}} = \frac{1}{N}\sum_{i=1}^{N}\hat{\tau}_i = \frac{1}{N}(N_1\hat{\tau}_{\text{ATET}} + N_0\hat{\tau}_{\text{ATEC}})$。

这种匹配方法被称为近邻匹配（nearest-neighbor matching）。其优点是可以直接对应协变量为连续值或多变量的情况。近邻匹配有很多变种，比如考虑可重复或不可重复抽样，用多个匹配值计算 $\hat{Y}_i^{\text{mis}}$（取其均值），用不同的距离函数等。

倾向得分匹配

Matching using Propensity Scores

倾向得分（propensity score）是个体接受处理的条件概率，即

$$p(X_i) = \Pr(W_i = 1 \mid X_i)$$

Rosenbaum & Rubin (1983) 证明，在 CIA 和共同支撑假设成立的情况下，我们可以将假设中基于协变量 X_i 的部分置换成倾向得分 $p(X_i)$，置换后的假设仍然成立。这说明，控制倾向得分和控制协变量都能够起到阻断后门路径的作用。

$$\begin{tikzcd} W \arrow[r] & Y \\ p(X) \arrow[u] & X \arrow[l] \arrow[u] \end{tikzcd}$$

因此，我们可以将倾向得分运用到匹配中：

1. 针对 X_i 的每个取值 x_k，估计倾向得分 $\hat{p}(x_k)$；
2. 用 $\hat{p}(x_k)$ 替代 x_k 进行匹配。

Rosenbaum & Rubin (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70:1, 41-55.

逆概率加权法

Inverse Probability Weighting

逆概率加权法（inverse probability weighting）是另一种利用倾向得分估计平均处理效应的方法。

Hirano et al. (2003) 分析了下面的加权估计量

$$\hat{\tau}_{\text{ATE}}^{\text{ipw}} = \frac{1}{N}\sum_{i=1}^{N}\frac{W_i Y_i^{\text{obs}}}{\hat{p}(X_i)} - \frac{1}{N}\sum_{i=1}^{N}\frac{(1-W_i)Y_i^{\text{obs}}}{1-\hat{p}(X_i)}$$

当某个 $\hat{p}(x_k)$ 的取值不等于 1/2 时，说明该 x_k 在处理组和对照组中存在弱平衡，可能导致估计偏差。加权估计量的目的就是调整这种弱平衡带来的偏差。

用 $1/\hat{p}(X_i)$ 做权重的一个缺点是权重之和不等于 1。如果将权重之和调整为 1，则对应的 $\hat{\tau}_{\text{ATE}}^{\text{ipw}}$ 可以通过加权回归求得。

Hirano, Imbens, & Ridder (2003). Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score. Econometrica, 71:4, 1161-1189.