Машинное обучение в экономике

Эффекты воздействия

Потанин Богдан Станиславович

доцент, научный сотрудник, кандидат экономических наук

2023–2024

Эффект воздействия

Определение

- Предположим, что одновременно существуют два гипотетически **режима** (counterfactual states) целевой переменной, обозначаемых Y_{0i} и Y_{1i}. Но в данных мы наблюдаем только один из них, в зависимости от значения бинарной **переменной воздействия** (treatment) T_i.

$$Y_i = \begin{cases} Y_{1i}, \text{ если } T_i = 1 \\ Y_{0i}, \text{ если } T_i = 0 \end{cases} = T_i Y_{1i} + (1 - T_i) Y_{0i}$$

- **Эффект воздействия** (treatment effect) T_i на Y_i определяется как:

$$\text{TE}_i = Y_{1i} - Y_{0i}$$

- Например, Y_i может отражать факт соверешения покупки клиентом, а T_i – факт наличия персонального предложения по бесплатной доставке товара на дом.
- На практике мы не можем одновременно наблюдать, как клиент мог бы повести себя и в случае наличия Y_{1i} и в случае отсутствия Y_{i0} предложения о доставке T_i, поскольку он либо получает это предложение (воздейсвтие) $T_i = 1$, либо нет $T_i = 0$.

Средний эффект воздействия
Определение

- На практике часто рассматривается **средний эффект воздействия**:

$$\text{ATE} = E(\text{TE}_i) = E(Y_{1i} - Y_{0i}) = E(Y_{1i}) - E(Y_{0i})$$

- Предположим, что Y_{1i} и Y_{0i} являются бинарными переменными, то есть $Y_{1i} \sim Ber(p_1)$ и $Y_{0i} \sim Ber(p_0)$.
- Обратим внимание, что:

$$E(Y_{1i}) = p_1 = P(Y_{1i} = 1) \qquad E(Y_{0i}) = p_0 = P(Y_{0i} = 1)$$

- Следовательно, в случае с бинарными целевыми переменными средний эффект воздействия можно интерпретировать как среднюю разность в вероятностях единицы в режимах, соответствующих наличию $T_i = 1$ и отсутствию $T_i = 0$ эффекта воздействия.
- Для простоты дальнейшего изложения будем придерживаться допущения о том, что Y_{ji} и Y_{jt} независимы при любых $i \neq t$.

Средний эффект воздействия

Допущение о независимости

- Предположим, что в данных n_1 наблюдений попали в **группу воздействия** $T_i = 1$, а n_0 наблюдений оказались в **контрольной группе** $T_i = 0$.
- Введем **допущение о независимости**, согласно которому $E(Y_{1i}|T_i = 1) = E(Y_{1i})$ и $E(Y_{0i}|T_i = 0) = E(Y_{0i})$.
- Это допущение обычно соблюдается в рамках контролируемых случайных экспериментов.
- Например, для измеренеия среднего эффекта воздействия вакцины на излечение от болезни, пациентов случайным образом распределеняют между группой воздействия, получающей лекарство, и контрольной группой, принимающей плацебо.
- При соблюдении допущения о независимости вследствие закона больших чисел получаем состоятельные оценки:

$$\hat{E}(Y_{1i}) = \frac{1}{n_1}\sum_{i:T_i=1} Y_{1i} \qquad \hat{E}(Y_{0i}) = \frac{1}{n_0}\sum_{i:T_i=0} Y_{0i}$$

- Тогда по теореме Слуцкого состоятельная оценка среднего эффекта воздействия может быть получена как:

$$\widehat{\text{ATE}} = \hat{E}(Y_{1i}) - \hat{E}(Y_{0i}) = \frac{1}{n_1}\sum_{i:T_i=1} Y_{1i} - \frac{1}{n_0}\sum_{i:T_i=0} Y_{0i}$$

Средний эффект воздействия

Технический комментарий о числе наблюдений в группе воздействия и контрольной группе

- При оценивании $\mathrm{E}\left(Y_{1i}\right)$ и $\mathrm{E}\left(Y_{0i}\right)$ для простоты изложения ранее и далее предполагается, что n_1 и n_0 являются константами.
- Однако, за пределами контролируемых экспериментов размеры группы воздействия и контрольной группы, как правило, являются случайными величинами:

$$n_1 = \sum_{i=1}^{n} T_i \qquad n_0 = \sum_{i=1}^{n} 1 - T_i$$

- При введеном ранее допущении о независимости для доказательства состоятельности необходимо воспользоваться законом больших чисел (в числителе и знаменателе) и теоремой Слуцкого:

$$\hat{\mathrm{E}}\left(Y_{1i}\right) = \frac{1}{n_1}\sum_{i:T_i=1} Y_{1i} = \frac{\frac{1}{n}\sum\limits_{i=1}^{n} T_i Y_{1i}}{\frac{1}{n}\sum\limits_{i=1}^{n} T_i} \xrightarrow{p} \frac{\mathrm{E}(T_i Y_{1i})}{E(T_i)} =$$

$$= \frac{E(Y_{1i}|T_i = 1)P(T_i = 1) + 0 \times P(T_i = 0)}{P(T_i = 1)} = E(Y_{1i}|T_i = 1) = E(Y_{1i})$$

- По аналогии нетрудно показать, что $\hat{\mathrm{E}}\left(Y_{0i}\right) \xrightarrow{p} E(Y_{0i})$.

- Как правило под **АВ-тестированием** понимается проверка гипотезы H_0 : ATE $= 0$ при допущении о независимости $E(Y_{ji}|T_i = j) = E(Y_{ji})$.
- Например, представим, что клиентская база продавца составляет 1000 человек. Из них он случайным образом отобрал 100 и предоставил им специальное предложение, согласно которому при покупке телефона они получат в подарок наушники.
- Из 100 человек, получивших предложение, покупку совершили 50, а из оставшихся 900 покупку осуществили 360 человек.
- Оценим средний эффект воздействия, то есть насколько, всреднем, возросла вероятность покупки благодаря предоставлению предложения:

$$\widehat{ATE} = 50/100 - 360/900 = 0.1$$

- Протестируем гипотезу H_0 : ATE $= 0$ против альтернативы H_1 : ATE > 0 с помощью теста о разнице долей, тестовая статистика которого, при верной нулевой гипотезе, в асимптотике (при стремящемся к бесконечности числе наблюдений) имеет стандартное нормальное распределение:

$$T = \frac{0.1}{\sqrt{0.410 \times (1 - 0.410)(1/100 + 1/900)}} \approx 1.93 \qquad \text{p-value} = 1 - \Phi(T) \approx 0.03$$

Средний эффект воздействия

Последствия нарушения допущения о независимости

- Если допущение о независимости не соблюдается, то $E(Y_{ji}) \neq E(Y_{ji}|T_i = j)$.
- Обычно оно нарушается в неконтролируемых экспериментах, например, когда имеется самоотбор в число тех, кто решил принять участие в новой программе лояльности магазина.
- Нарушение предпосылки о независимости приводит к смещению введенной ранее оценки среднего эффекта воздействия, что, в частности, не позволяет применять AB-тестирование (для простоты предположим n_1 и n_0 экзогенными):

$$E\left(\widehat{\text{ATE}}\right) = E\left(\frac{1}{n_1}\sum_{i:T_i=1}(Y_{1i}|T_i = 1) - \frac{1}{n_0}\sum_{i:T_i=0}(Y_{0i}|T_i = 0)\right) =$$

$$= \frac{1}{n_1}\sum_{i:T_i=1}E(Y_{1i}|T_i = 1) - \frac{1}{n_0}\sum_{i:T_i=0}E(Y_{0i}|T_i = 0) =$$

$$= E(Y_{1i}|T_i = 1) - E(Y_{0i}|T_i = 0) \neq E(Y_{1i}) - E(Y_{0i})$$

- Рассмотрим альтернативную оценку эффекта воздействия:

$$\widehat{\text{ATE}} = \frac{1}{n}\sum_{i=1}^{n} E(Y_{1i}|X_i) - E(Y_{0i}|X_i)$$

- Несмещенность этой оценки следует из закона чередующихся математических ожиданий.

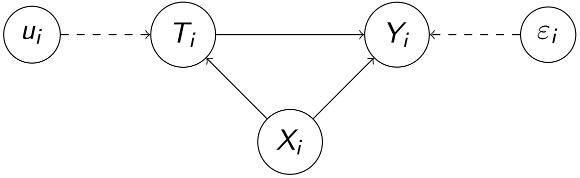
$$E\left(\widehat{\text{ATE}}\right) = E\left(E(Y_{1i}|X_i)\right) - E\left(E(Y_{0i}|X_i)\right) = E(Y_{1i}) - E(Y_{0i})$$

- **Проблема –** на практике мы не знаем $E(Y_{0i}|X_i)$ и $E(Y_{1i}|X_i)$.
- **Решение –** методами машинного обучения **при некоторых условиях** можно получить состоятельную оценку функций $\hat{E}(Y_{0i}|X_i)$ и $\hat{E}(Y_{1i}|X_i)$, а затем подставить их в формулу для оценки среднего эффекта воздействия.
- Отметим, что в задаче бинарной классификации $\hat{E}(Y_{ji}|X_i) = \hat{P}(Y_{ji} = 1|X_i)$.

Средний эффект воздействия

Допущение об условной независимости

- Введем **допущение об условной независимости**, при котором $E(Y_{ji}|X_i = x_i, T_i = j) = E(Y_{ji}|X_i = x_i)$ при любых $x_i \in \text{supp}(X_i)$ и $j \in \{0, 1\}$.
- Обычно это допущение соблюдается, когда X_i отражает все факторы, которые могут быть статистически связаны и с T_i, и с Y_{ji}.



- Предполагается, что связь между T_i и Y_i обусловлена наблюдаемыми в данных переменными X_i, именуемыми **смешивающими** (confounders).
- Прерывистыми линиями отображены связи T_i и Y_i с агрегированными ненаблюдаемыми переменными u_i и ε_i.

Средний эффект воздействия
Оценивание с помощью условных математических ожиданий

- При соблюдении допущения об условной независимости:

$$\begin{aligned}\text{ATE} = E(Y_{1i}) - E(Y_{0i}) = E\left(E(Y_{1i}|X_i) - E(Y_{0i}|X_i)\right) = \\ = E\left(E(Y_{1i}|X_i, T_i = 1) - E(Y_{0i}|X_i, T_i = 0)\right) = \\ = E\left(E(T_iY_{1i} + (1 - T_i)Y_{0i}|X_i, T_i = 1) - E(T_iY_{1i} + (1 - T_i)Y_{0i}|X_i, T_i = 0)\right) = \\ = E\left(E(Y_i|X_i, T_i = 1) - E(Y_i|X_i, T_i = 0)\right)\end{aligned}$$

- **Вывод** – достаточно найти состоятельную оценку функции $E(Y_i|X_i, T_i)$, что в большинстве случаев можно сделать **методами машинного обучения**, поскольку Y_i всегде наблюдается в данных.
- В итоге средний эффект воздействия оценивается как:

$$\widehat{\text{ATE}} = \frac{1}{n}\sum_{i=1}^{n}\hat{E}(Y_i|X_i, T_i = 1) - \hat{E}(Y_i|X_i, T_i = 0)$$

Средний эффект воздействия

Оценивание с помощью взвешивания на обратные вероятности

- Обратим внимание, что при соблюдении допущения об условной независимости:

$$\mathrm{E}\left(T_iY_i/P(T_i=1|X_i)|X_i\right)=\mathrm{E}\left(Y_{1i}/P(T_i=1|X_i)|X_i,T_i=1\right)P(T_i=1|X_i)=\mathrm{E}\left(Y_{1i}|X_i\right)\implies$$
$$\implies \mathrm{E}\left(T_iY_i/P(T_i=1|X_i)\right)=\mathrm{E}\left(\mathrm{E}\left(Y_{1i}|X_i\right)|X_i\right)=\mathrm{E}\left(\mathrm{E}\left(T_iY_i/P(T_i=1|X_i)\right)\right)=E(Y_{1i})$$

- По аналогии можно показать, что:

$$E\left((1-T_i)\,Y_i/\left(1-P(T_i=1|X_i)\right)\right)=\mathrm{E}\left(Y_{0i}\right)$$

- В итоге получаем:

$$\mathrm{ATE}=\mathrm{E}\left(T_iY_i/P(T_i=1|X_i)\right)-\mathrm{E}\left((1-T_i)\,Y_i/\left(1-P(T_i=1|X_i)\right)\right)$$

- Из полученны результатов следует альтернативный способ оценивания ATE, именуемый **оценкой с помощью взвешивания на обратные вероятности** (inverse probability weighting):

$$\widehat{\mathrm{ATE}}=\frac{1}{n}\sum_{i=1}^{n}\frac{T_iY_i}{\hat{P}(T_i=1|X_i)}-\frac{(1-T_i)\,Y_i}{1-\hat{P}(T_i=1|X_i)}$$

- **Преимущество** – достаточно с помощью методов машинного обучения оценить $\hat{P}(T_i=1|X_i)$.

Средний эффект воздействия

Двойная устойчивость

- Мы рассмотрели два способа оценивания ATE при допущении об условной независимости, первый из которых опирается на оценки $\mathrm{E}(Y_i|X_i, T_i)$, а второй – на оценки $P(T_i = 1|X_i)$.
- **Проблема** – точность оценок каждого из этих способов зависит от точности оценок соответствующих условных математических ожиданий или вероятностей. Если они оценены неточно, то и итоговая оценка ATE также будет неточной.
- **Решение** – обеспечить **двойную устойчивость**, то есть совместить оба способа, чтобы оценка ATE оказывалась состоятельной, если по крайней мере один из них дает состоятельную оценку.

$$\widehat{\mathrm{ATE}} =$$

$$= \underbrace{\frac{1}{n}\sum_{i=1}^{n}\hat{g}_{1i} - \hat{g}_{0i}}_{\text{стремится к ATE если } \hat{\mathrm{E}}(Y_i|X_i,T_i) \text{ состоятельна}} + \underbrace{\frac{1}{n}\sum_{i=1}^{n}\frac{T_i(Y_i - \hat{g}_{1i})}{\hat{g}_{Ti}} - \frac{(1-T_i)(Y_i - \hat{g}_{0i})}{1-\hat{g}_{Ti}}}_{\text{стремится к 0 если } \hat{\mathrm{E}}(Y_i|X_i,T_i) \text{ состоятельна}} =$$

$$= \underbrace{\frac{1}{n}\sum_{i=1}^{n}\frac{T_i Y_i}{\hat{g}_{Ti}} - \frac{(1-T_i)Y_i}{1-\hat{g}_{Ti}}}_{\text{стремится к ATE если } \hat{P}(T_i=1|X_i) \text{ состоятельна}} + \underbrace{\frac{1}{n}\sum_{i=1}^{n}\frac{(T_i - \hat{g}_{Ti})\,\hat{g}_{0i}}{1-\hat{g}_{Ti}} - \frac{(T_i - \hat{g}_{Ti})\,\hat{g}_{1i}}{\hat{g}_{Ti}}}_{\text{стремится к 0 если } \hat{P}(T_i=1|X_i) \text{ состоятельна}}$$

$$\hat{g}_{1i} = \hat{\mathrm{E}}(Y_i|X_i, T_i = 1) \qquad \hat{g}_{0i} = \hat{\mathrm{E}}(Y_i|X_i, T_i = 0) \qquad \hat{g}_{Ti} = \hat{\mathrm{P}}(T_i = 1|X_i)$$

Средний эффект воздействия

Двойное машинное обучение

- Средний эффект воздействия можно также оценить с помощью DML метода, рассмотрев уравнение:

$$Y_i = g(T_i, X_i) + \varepsilon_i \qquad \text{ATE} = \mathrm{E}\left(g(1, X_i) - g(0, X_i)\right)$$

- Допущение об условной независимости можно сформулировать как $E(\varepsilon_i | X_i, T_i) = 0$.
- Рассмотрим вклад, удовлетворяющий условию ортогональности по Нейману:

$$\psi = \frac{T_i(Y_i - g_1(X_i))}{g_T(X_i)} - \frac{(1 - T_i)(Y_i - g_0(X_i))}{1 - g_T(X_i)} + g_1(X_i) - g_0(X_i) - \text{ATE}$$

$$g_1(X_i) = E(Y_i | X = x_i, T_i = 1), \quad g_0(X_i) = E(Y_i | X = x_i, T_i = 0), \quad g_T(X_i) = P(T_i = 1 | X_i)$$

- Решая $E(\psi) = 0$ для ATE, а также подставляя оценки (полученные с помощью машинного обучения) неизвестных функций и применяя кросс-фиттинг, получаем оценку, обладающую свойством двойной устойчивости:

$$\widehat{ATE} = \frac{1}{n} \sum_{i=1}^{n} \frac{T_i(Y_i - \hat{g}_1^{(q_i)}(X_i))}{\hat{g}_T^{(q_i)}(X_i)} - \frac{(1 - T_i)(Y_i - \hat{g}_0^{(q_i)}(X_i))}{1 - \hat{g}_T^{(q_i)}(X_i)} + \hat{g}_1^{(q_i)}(X_i) - \hat{g}_0^{(q_i)}(X_i)$$

- Функции $\hat{g}_1^{(k)}$, $\hat{g}_0^{(k)}$ и $\hat{g}_T^{(k)}$ оцениваются с помощью машинного обучения на данных, **не** вошедших в k-ю из K выборок. Также, $q_i = k$, если наблюдение i **не** вошло в k-ю выборку.

- Средний эффект воздействия не всегда достаточно информативен для принятия конкретных решений. Поэтому в качестве альтернативы часто оценивают **условный средний эффект воздействия**, что в маркетинге именуется **uplift** моделированием.

$$\text{CATE}_i = E\left(Y_{1i}|X_i\right) - E\left(Y_{0i}|X_i\right) = E(Y_i|X_i, T_i = 1) - E(Y_i|X_i, T_i = 0)$$

- При допущении об условной независимости нетрудно получить оценку условного среднего эффекта воздействия:

$$\widehat{\text{CATE}}_i = \hat{E}(Y_i|X_i, T_i = 1) - \hat{E}(Y_i|X_i, T_i = 0)$$

- Можно либо оценить $\hat{E}(Y_i|X_i, T_i)$ по всей выборке (**Single-learner/S-learner**), либо отдельно $\hat{E}(Y_i|X_i, T_i = 1)$ и $\hat{E}(Y_i|X_i, T_i = 0)$ по группе воздействия и контрольной группе соответственно (**Two-learner/T-learner**).
- Однако существуют и иные, менее очевидные подходы к оцениванию CATE_i.

- **Проблема** – иногда группа воздействия может включать малое число наблюдений, что осложняет оценивание $\hat{E}(Y_i|X_i, T_i = 1)$ по данным группы воздействия.
- **Решение** – рассмотрим вспомогательную переменную:

$$D_{1i} = Y_{1i} - E(Y_i|X_i, T_i = 0)$$

- Обратим внимание, что:

$$E(D_{1i}|X_i) = E(Y_{1i}|X_i) - E(Y_i|X_i, T_i = 0) = \text{CATE}_i$$

- Следовательно, $\hat{E}(D_{1i}|X)$ можно рассматривать как оценку CATE_i, полученную с помощью Y_{1i} и $\hat{E}(Y_i|X_i, T_i = 0)$, то есть без использования неэффективно оцениваемого по малому числу наблюдений группы воздействия $E(Y_i|X_i, T_i = 1)$.
- **Проблема** – в данных отсутствуют наблюдения по D_{1i}.
- **Решение** – рассмотреть $\hat{D}_{1i} = Y_{1i} - \hat{E}(Y_i|X_i, T_i = 0)$ и оценить $\widehat{\text{CATE}}_i = \hat{E}\left(\hat{D}_{1i}|X_i\right)$.

- **Первый шаг** – по аналогии с T-learner оцениваются условные математические ожидания $E(Y_i|X_i, T_i = 1)$ и $E(Y_i|X_i, T_i = 0)$.
- **Второгой шаг** – рассчитываются вспомогательные переменные:

$$\hat{D}_{1i} = Y_{1i} - \hat{E}(Y_i|X_i, T_i = 0) \qquad \hat{D}_{0i} = \hat{E}(Y_i|X_i, T_i = 1) - Y_{0i}$$

- **Третий шаг** – оцениваются $E\left(\hat{D}_{1i}|X_i\right)$ и $E\left(\hat{D}_{0i}|X_i\right)$.
- **Четвертый шаг** – оцениваются условные вероятности попадания в группу воздействия $P(T_i = 1|X_i)$.
- **Пятый шаг** – с помощью взвешивания оценивается условный эффект воздействия:

$$\widehat{\text{CATE}}_i = \left(1 - \hat{P}(T_i = 1|X_i)\right)\hat{E}\left(\hat{D}_{1i}|X_i\right) + \hat{P}(T_i = 1|X_i)\hat{E}\left(\hat{D}_{0i}|X_i\right)$$

- **Интуиция взвешивания** – чем меньше наблюдений в группе воздействия, тем больший вес присваивается $\hat{E}\left(\hat{D}_{1i}|X_i\right)$, который не зависит от $\hat{E}\left(Y_i|X_i, T_i = 1\right)$.

- Скронструируем **псевдоисход**:

$$Y_i^* = Y_i\left(\frac{T_i}{P(T_i=1|X_i)} + \frac{1-T_i}{1-P(T_i=1|X_i)}\right)$$

- Обозначим через $\hat{Y}_i$ величину Y_i, посчитанную с использованием $\hat{P}(T_i = 1|X_i)$ вместо $P(T_i = 1|X_i)$.
- При соблюдении допущения об условной независимости можно показать, по аналогии с тем, как это было сделано с оцениванием ATE с помощью взвешивания на обратные вероятности, что:

$$E(Y_i^*|X_i) = \text{CATE}_i$$

- Следовательно, для оценивания CATE_i можно воспользоваться двухшаговой процедурой, часто именуемой **методом трансформации классов**.
- **Первый шаг** – вычислить $\hat{P}(T_i = 1|X_i)$ и $\hat{Y}_i^*$.
- **Второй шаг** – оценить $E(\hat{Y}_i^*|X_i)$ и получить $\widehat{\text{CATE}}_i = \hat{E}(\hat{Y}_i^*|X_i)$.

- Для измерения качества прогнозирования условных эффектов воздействия хотелось бы использовать метрику:

$$\text{MSE}_0 = \frac{1}{n}\sum_{i=1}^{n}\left(\text{CATE}_i - \widehat{CATE}_i\right)^2$$

- **Проблема** – на практике CATE_i неизвестно.
- **Решение** – в качестве косвенной метрики качества прогнорзов условных эффектов воздействия использовать метрику качества точности прогнозов исхода:

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{E}\left(Y_i|X_i, T_i\right)\right)^2$$

- **Проблема** – модель, точнее оценивающая Y_i, не обязательно будет точнее оценивать CATE_i. Например, даже если оценки $\hat{E}\left(Y_i|X_i, T_i = 1\right)$ и $\hat{E}\left(Y_i|X_i, T_i = 0\right)$ обладают очень большим, но примерно одинаковым смещением, то их разница, то есть $\widehat{\text{CATE}}_i$, может оказаться весьма точной оценкой, поскольку смещения сократятся.

Качество оценивания условных средних эффектов воздействия

Использование псевдоисходов

- Рассмотрим псевдоисход Y_i^*, такой, что $E(Y_i^*|X_i) = \text{CATE}_i$ и соблюдены некоторые дополнительные, технические условия. Например, можно использовать псевдоисход метода трансформации классов:

$$Y_i^* = Y_i\left(\frac{T_i}{P(T_i = 1|X_i)} + \frac{1 - T_i}{1 - P(T_i = 1|X_i)}\right)$$

- Рассмотрим следующую метрику качества:

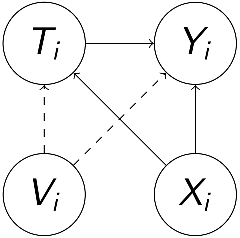
$$\text{MSE}^* = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{Y}_i^* - \widehat{\text{CATE}}_i\right)^2$$

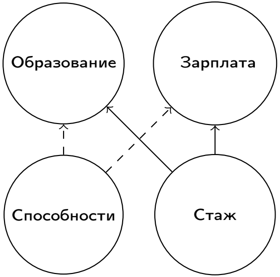
- Можно показать, что при определенных условиях, в частности, больших объемах выборки, MSE^* является достаточно точной аппроксимацией MSE_0.
- Существует множество иных подходов к аппроксимации MSE_0 с помощью метрик, зависящх от различных псевдоисходов.

Локальный средний эффект воздействия

Постановка проблемы

- **Проблема** – на практике допущение об условной независимости $E(Y_{ji}|X_i, T_i) = E(Y_{ji}|X_i)$ часто нарушается вследствие **эндогенности**, из-за чего описанные ранее методы обычно дают несостоятельные оценки эффектов воздействия.





- Эндогенность обычно возникает из-за наличия ненаблюдаемых характеристик V_i, одновременено влияющих и на целевую переменную Y_i, и на вероятность воздействия $T_i = 1$.
- **Решение** – адаптировать метод инструментальных переменных.

Локальный средний эффект воздействия

Определение

- 123
-