Machine-learning a kauzálne ekonometrické modely

Lukáš Lafférs

Katedra matematiky, Univerzita Mateja Bela

Vedecké kolokvium FPV UMB

Machine-learning a kauzálne ekonometrické modely

Applied Statistics:

- Biostatistics
- Sociometrics
- Epidemiology
- Survey Research Methods
- Econometrics
- Actuarial Statistics
- Reliability statistics
- Environmental statistics
- ...

Machine-learning a kauzálne ekonometrické modely

Applied Statistics:

- Biostatistics
- Sociometrics
- Epidemiology
- Survey Research Methods
- Econometrics
- Actuarial Statistics
- Reliability statistics
- Environmental statistics
- ...

"Econometrics is what Econometricians do." (Arthur Goldberger)

Machine-learning a kauzálne ekonometrické modely

Algoritmy strojového učenia sa ukázali ako skvelý nástroj na predikciu!

Machine-learning a kauzálne ekonometrické modely

Vedia nám však pomôcť aj s odhadovaním kauzálnych vzťahov??

Kauzalita

Prečo sa to stane?

Vyžaduje hloboké porozumenie problému.

Náročný problém.

Predikcia

Čo sa stane?

Pokiaľ funguje, nevadí, že nerozumieme prečo.

Principiálne jednoduchšie.

Michal absolvoval rekvalifikačný kurz. Pomohlo to?

Michal absolvoval rekvalifikačný kurz. Pomohlo to?

Mohli by sme ho porovnať s Jozefom, ktorý ho neabsolvoval.

Michal

- 27 rokov
- ženatý
- z Brezna
- stredná škola
- vie anglicky
- zdravý
- vod. preukaz typu B
- býva s 3 ďalšími ľudmi
- ...

Jozef

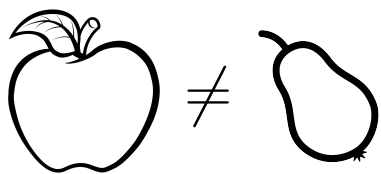
- 53 rokov
- slobodný
- z Myjavy
- VŠ
- vie nemecky
- mal autonehodu
- vod. preukazy typu B,C
- stará sa o rodičov
- ...

Prekvapivo. Michal a Jozef sú stále rôzni.

$$\text{Michal} \neq \text{Jozef}$$

Prekvapivo. Michal a Jozef sú stále rôzni.

$$\text{Michal} \neq \text{Jozef}$$



Michal

- 27 rokov
- ženatý
- z Brezna
- stredná škola
- anglicky
- zdravý
- vod. preukaz typu B
- býva v dome s 3 ďalšími ľudmi
- ...

Peter

- 29 rokov
- slobodný
- z Podbrezovej
- stredná škola
- anglicky, francúzsky
- zdravý
- vod. preukaz typu B
- býva s 2 ľudmi a so psom
- ...

Prekvapivo. Aj Michal a Peter sú rôzni.

Ale o dosť podobnejší.

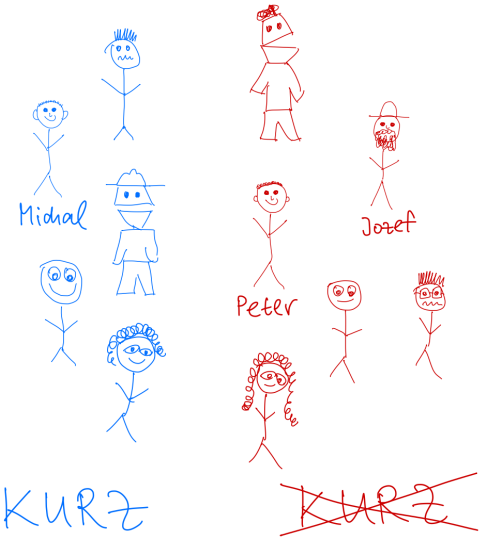
$$\text{Michal} \approx \text{Peter}$$

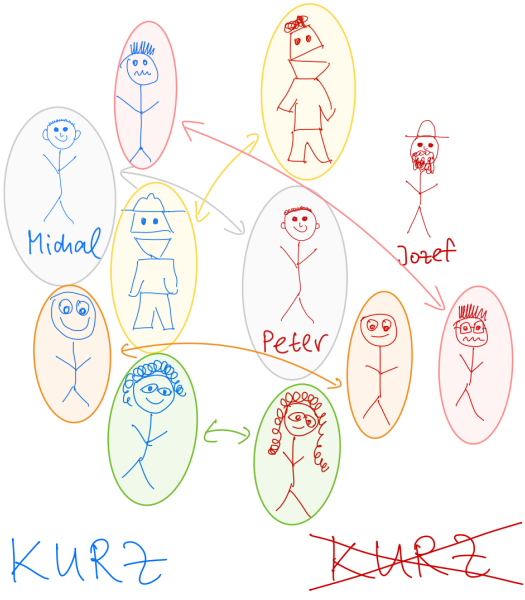
Prekvapivo. Aj Michal a Peter sú rôzni.

Ale o dosť podobnejší.

$$\text{Michal} \approx \text{Peter}$$

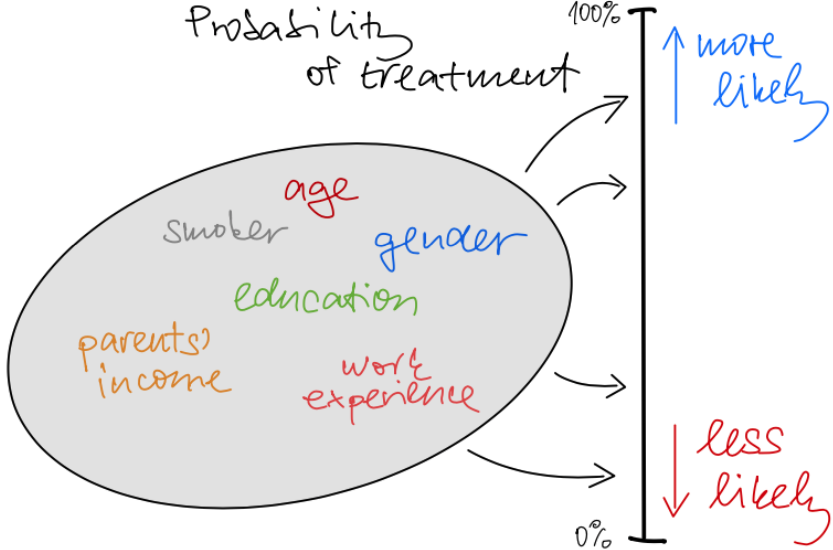
Párovanie (matching)

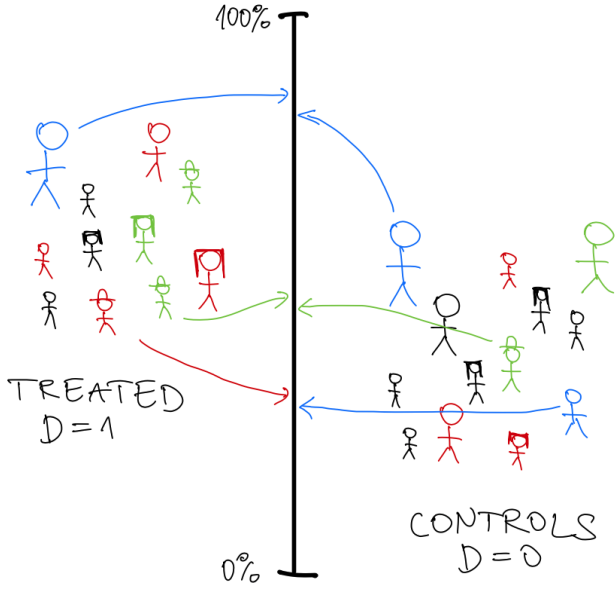



Michal
Peter
Jozef
KURZ
KURZ

Takýchto Petrov je však málo.

A nie každý má svoju ”dvojičku” ako Michal.


Probability of treatment
100%
0%
more likely
less likely
age
smoker
gender
education
parents' income
work experience


100%
0%
TREATED
D=1
CONTROLS
D=0

O Michalovi vieme veľa informácií.

Ale absolventov kurzu je málo.

- vek
- slobodný/á
- počet detí
- mesto
- typ vzdelania
- oblasť vzdelania
- znalosť cudzích jazykov
- zdravotné znevýhodnenie
- vod. preukaz
- história zamestnania
- typ predošlej práce
- dĺžka predošlého zamestnania
- klasifikácia predošlého zamestnania
- počet členov v domácnosti
- bariéry zamestnanosti
- národnosť
- sociálne dávky
- zdravotné poistenie
- zdravotné znevýhodnenie
- osamelý občan
- ochota vzdelávať sa
- ochota dochádzať za prácou
- účasť na predošlom kurze
- záujem o prácu o neúplný úväzok
- záujem o prácu v zahraničí
- úroveň segregácie
- vzdialenosť od krajského mesta
- vzdialenosť od Bratislavy
- ...

Tradičné štatistické metódy nevedia pracovať s mnohorozmernými dátami.

Čo s tým?

Priemerný efekt absolvovania kurzu na mzdu

$$\Delta = \mathrm{E}\Big(\mathrm{MZDA}(\mathrm{kurz}) - \mathrm{MZDA}(\cancel{\mathrm{kurz}})\Big)$$

μ

Predikovať priamo

- vek
- pohlavie
- vzdelanie
- zručnosti
- ...

Dáta

$\mu \rightarrow$ mzda

μ

Predikovať priamo

- vek
- pohlavie
- vzdelanie
- zručnosti
- ...

Dáta

μ
$\rightarrow$
mzda

π

Preváhovanie

- vek
- pohlavie
- vzdelanie
- zručnosti
- ...

Dáta

π
$\rightarrow$
ide na kurz

μ

Predikovať priamo

- vek
- pohlavie
- vzdelanie
- zručnosti
- ...

Dáta

μ
$\rightarrow$
mzda

π

Preváhovanie

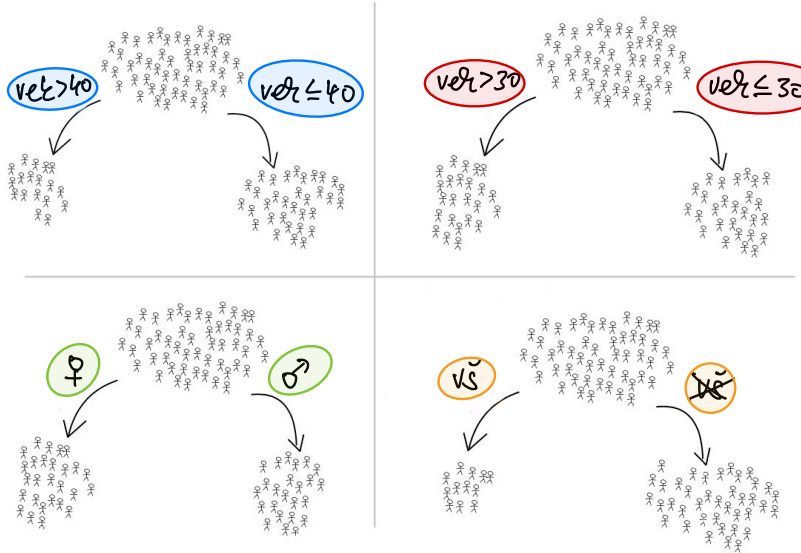
- vek
- pohlavie
- vzdelanie
- zručnosti
- ...

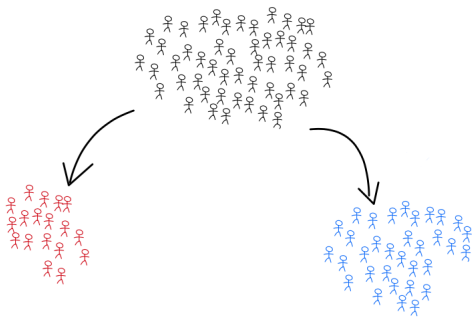
Dáta

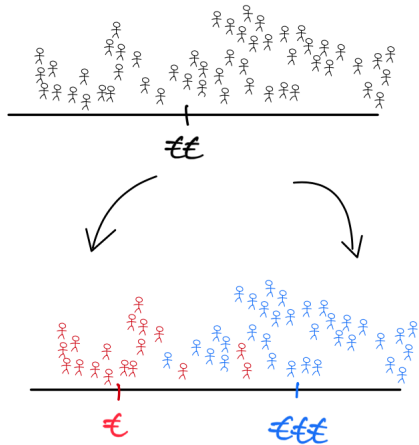
π
$\rightarrow$
ide na kurz

Aj μ aj π odhadneme pomocou MACHINE LEARNING metód strojového učenia.

Príklad algoritmu strojového učenia





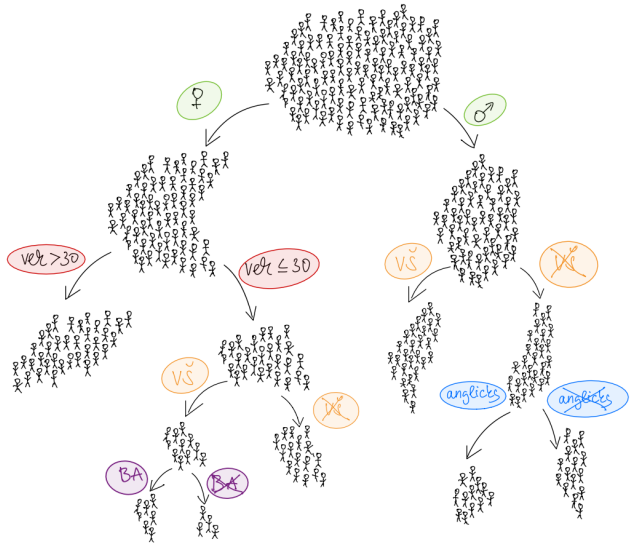


Ako narásť strom?

Dobrá áno/nie otázka?

Vieme to odmerať?

Ako veľký strom?



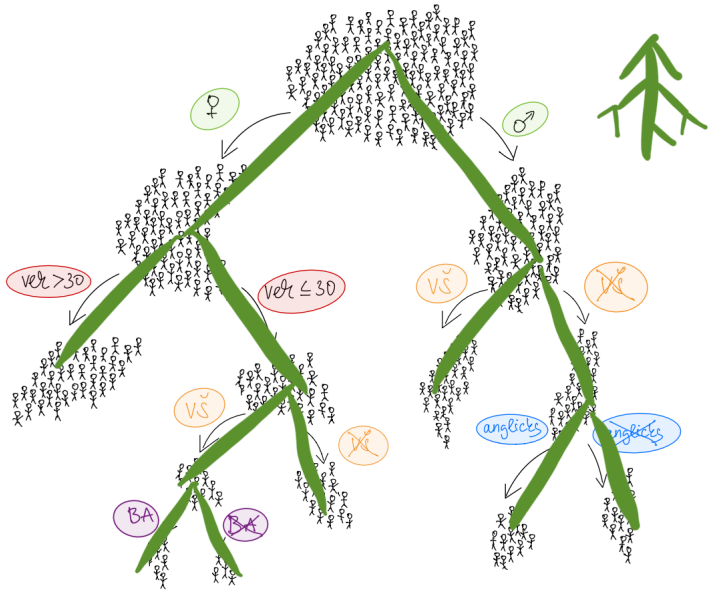
- Breiman, L., Friedman, J. H., Olshen, R., & Stone, C. J. (1984). Classification and Regression Trees.

Ako narásť strom?

Dobrá áno/nie otázka?

Vieme to odmerať?

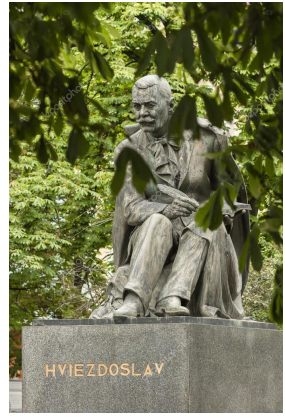
Ako veľký strom?



Náhodný les

Pozdravujem vás, lesy, hory,
z tej duše pozdravujem vás!

P. O. Hviezdoslav



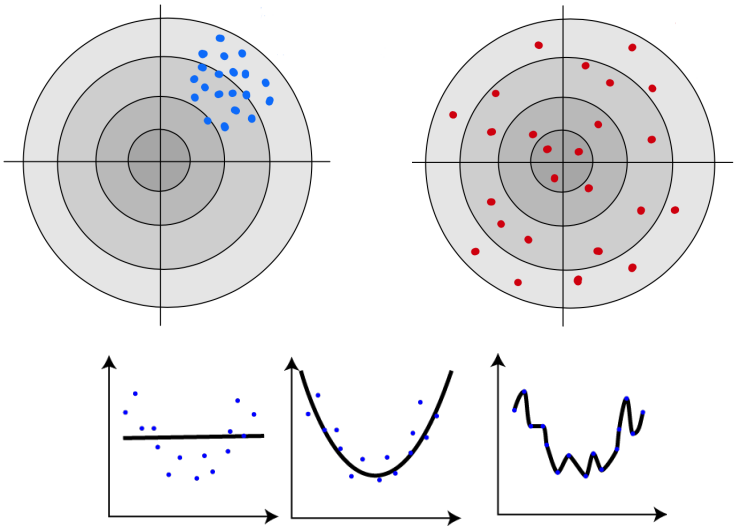
P.O.H. v lese.

(Nie až tak) náhodný les

(Už celkom) náhodný les

- Breiman, L. (2001). Random forests. Machine learning, 45, 5-32.

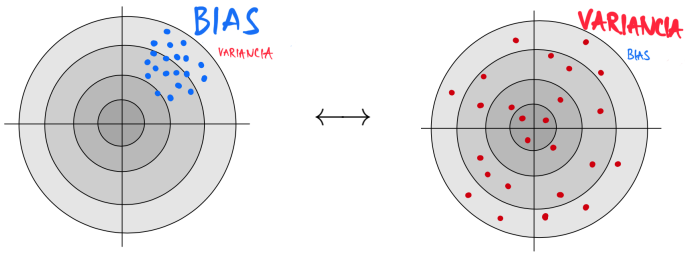
Nevýhody



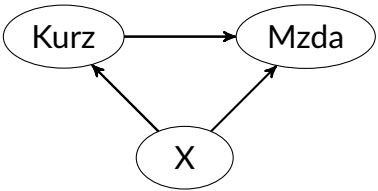
$$\text{PREDIKČNÁ CHYBA} = \text{NÁHODNÁ CHYBA} + \text{NENÁHODNÁ CHYBA}$$

$$\text{PREDIKČNÁ CHYBA} = \text{NÁHODNÁ CHYBA} + \text{NENÁHODNÁ CHYBA}$$

$$= \text{NÁHODNÁ CHYBA} + \underbrace{\text{BIAS}^2 + \text{VARIANCIA}}_{\text{NENÁHODNÁ CHYBA}}$$



Naspäť k príkladu

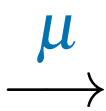


X - informácie o veku, pohlaví, vzdelaní, zručnostiach...

$$\Delta = \mathrm{E}\Big(\mathrm{MZDA}(\mathrm{kurz}) - \mathrm{MZDA}(\cancel{\mathrm{kurz}})\Big)$$

- Pearl, J. (2009). Causality. Cambridge university press.
- Imbens, G. W., & Rubin, D. B. (2015). Causal inference in statistics, social, and biomedical sciences. Cambridge university press.

- vek
- pohlavie
- vzdelanie
- zručnosti
- ...



Δ_μ

efekt na mzdu

- vek
- pohlavie
- vzdelanie
- zručnosti
- ...

$\mu \longrightarrow$

Δ_μ

efekt na mzdu

- vek
- pohlavie
- vzdelanie
- zručnosti
- ...

$\pi \longrightarrow$

Δ_π

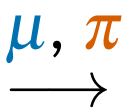
efekt na mzdu

Priama aplikácia μ alebo π (prevážením) na odhad Δ povedie k BIASu.

Obe Δ_μ aj Δ_π konvergujú pomaly: $\Delta_\mu \dashrightarrow \Delta$ a $\Delta_\pi \dashrightarrow \Delta$

Double Machine Learning

- vek
- pohlavie
- vzdelanie
- zručnosti
- ...



$\Delta_{\mu,\pi}$

efekt na mzdu

Double Machine Learning

- vek
- pohlavie
- vzdelanie
- zručnosti
- ...

$\mu, \pi \longrightarrow$

$\Delta_{\mu,\pi}$

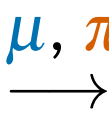
efekt na mzdu

$\Delta_{\mu,\pi} \rightarrow \Delta$

- Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., & Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters. Econometrics Journal, volume 21, pp. C1–C68.
- Robins, J. M., & Rotnitzky, A. (1995). Semiparametric efficiency in multivariate regression models with missing data. Journal of the American Statistical Association, 90(429), 122-129.
- Van Der Laan, M. J., & Rubin, D. (2006). Targeted maximum likelihood learning. The international journal of biostatistics, 2(1).

Double Machine Learning

- vek
- pohlavie
- vzdelanie
- zručnosti
- ...



$\Delta_{\mu,\pi}$

efekt na mzdu

Double Machine Learning

- vek
- pohlavie
- vzdelanie
- zručnosti
- ...

μ, π

$\longrightarrow$

$\Delta_{\mu,\pi}$

efekt na mzdu

$\Delta_{\mu,\pi} \to \Delta$

Dve dôležité myšlienky

Model je definovaný pomocou:

$$\mathrm{E}\left[\psi(\text{dáta}, \underbrace{\mu, \pi}_{\equiv \eta}, \Delta)\right] = 0$$

(1) Neyman orthogonality

- Funkcia ψ je lokálne necitlivá na zmeny v $\eta = (\mu, \pi)$

(2) Sample splitting

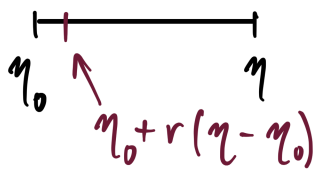
- Rozdelíme dáta a rôzne časti použijeme na odhad $\eta = (\mu, \pi)$ a na odhad Δ

(1) Lokálna necitlivosť (Neyman orthogonality)

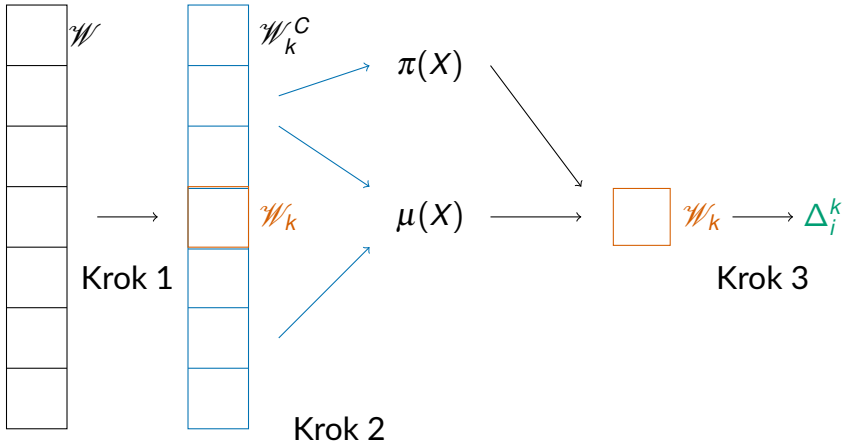
$$\mathrm{E}\big[\psi(\text{dáta}, \underbrace{\mu, \pi}_{\equiv \eta}, \Delta)\big] = 0$$

V blízkom okolí η_0 sa ψ príliš nemení v η:

$$\frac{\partial}{\partial r}\mathrm{E}[\psi(\text{dáta}, \eta_0 + r(\eta - \eta_0), \Delta)]\Big|_{r=0} = 0$$



(2) Vyhnúť sa overfittingu (Sample splitting)



$$\Delta = \frac{1}{n} \sum_{k=1}^{K} \sum_{i=1}^{n_k} \Delta_i^k$$

$$\sqrt{n}(\hat{\Delta} - \Delta) = \underbrace{a^*}_{\text{Approx. Gaussovský}} + \underbrace{b^*}_{\text{Regularizačný bias}} + \underbrace{c^*}_{\text{Overfitting bias}}$$

$$= \underbrace{a^*}_{\text{Approx. Gaussovský}} + \cancel{\underbrace{b^*}_{\text{Neyman-orthogonal}}} + \cancel{\underbrace{c^*}_{\text{Sample splitting}}}$$

$$\sim N(0, \sigma^2)$$

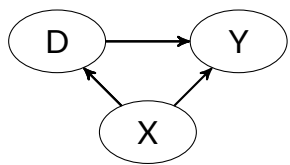
Príklad

Uvažujme nasledujúci parciálne lineárny model.
θ je parameter, o ktorý nám ide.
$g(X)$ a $m(X)$ sú flexibilné funkcie, o ktoré nemáme záujem

$$Y = \theta D + g(X) + U, \qquad E[U \mid D, X] = 0,$$
$$D = m(X) + V, \qquad E[V \mid X] = 0.$$

$$\psi = V \cdot U = (D - m(X)) \cdot (Y - g(X) - \theta D)$$

$$\mathrm{E}\big[\psi(\underbrace{Y, X, D}_{\equiv W \sim \text{dáta}}, \underbrace{m, g}_{\equiv \eta}, \theta)\big] = 0$$



Detaily

Double machine learning - zhrnutie

- Estimátor $\hat{\Delta}$ založený na Neyman-ortogonálnej funkcii ψ
- Použijúc sample splitting
- Estimátory μ and π sú ”dostatočne dobré” (e.g. konvergujú rýchlosťou aspoň $n^{-1/4}$)

Double machine learning - zhrnutie

- Estimátor $\hat{\Delta}$ založený na Neyman-ortogonálnej funkcii ψ
- Použijúc sample splitting
- Estimátory μ and π sú "dostatočne dobré" (e.g. konvergujú rýchlosťou aspoň $n^{-1/4}$)

Veta 1 (Chernozhukov et al. 2018):

$$\sqrt{n}(\hat{\Delta} - \Delta) \to N(0, \sigma^2)$$

Estimátor $\hat{\Delta}$ je asymptoticky normálne rozdelený a je $\sqrt{n}$-konzistentný.

Double machine learning - zhrnutie

DML nám poskytuje odhadovacie prístupy, ktoré:

Double machine learning - zhrnutie

DML nám poskytuje odhadovacie prístupy, ktoré:

- zvládajú veľadimenzionálne dáta

Double machine learning - zhrnutie

DML nám poskytuje odhadovacie prístupy, ktoré:

- zvládajú veľadimenzionálne dáta
- sú flexibilné

Double machine learning - zhrnutie

DML nám poskytuje odhadovacie prístupy, ktoré:

- zvládajú veľadimenzionálne dáta
- sú flexibilné
- vedia využiť prediktívne vlastnosti ML

Double machine learning - zhrnutie

DML nám poskytuje odhadovacie prístupy, ktoré:

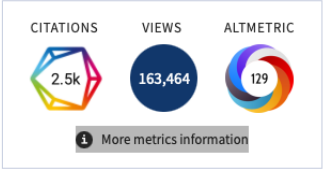
- zvládajú veľadimenzionálne dáta
- sú flexibilné
- vedia využiť prediktívne vlastnosti ML
- majú dobré štatistické vlastnosti

Double/debiased **machine learning** for treatment and structural parameters [PDF] oup.com

V Chernozhukov, D Chetverikov, M Demirer, E Duflo… - 2018 - academic.oup.com

… To estimate η 0 , we consider the use of statistical or **machine learning** (ML) methods, which are … We call the resulting set of methods **double** or debiased ML (DML). We verify that DML …

☆ Save 99 Cite Cited by 4471 Related articles All 30 versions



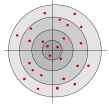
Limitácie - ”Kitchen sink” regresia

Hünermund, Beyers and Caspi (2023)

Zatiaľ

- Kauzalita je náročná
- O to viac v modeloch s veľa informáciami

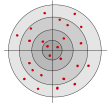
BIAS vs VARIANCIA



Zatiaľ

- Kauzalita je náročná
- O to viac v modeloch s veľa informáciami

BIAS vs VARIANCIA



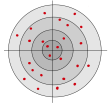
- ML metódy sú skvelé na predikciu
- Na odhadovanie parametrov už nie

$$\Delta_{\mu} \dashrightarrow \Delta \quad \text{a} \quad \Delta_{\pi} \dashrightarrow \Delta$$

Zatiaľ

- Kauzalita je náročná
- O to viac v modeloch s veľa informáciami

BIAS vs VARIANCIA



- ML metódy sú skvelé na predikciu
- Na odhadovanie parametrov už nie

$$\Delta_{\mu} \dashrightarrow \Delta \quad \text{a} \quad \Delta_{\pi} \dashrightarrow \Delta$$

- Skombinujeme dve slabé ML metódy μ, π
- Dostaneme odhad s dobrými vlastnosťami

$$\Delta_{\mu,\pi} \to \Delta$$

Rozšírenia

μ, π

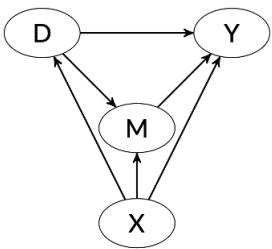
$\Delta_{\mu,\pi}$

efekt

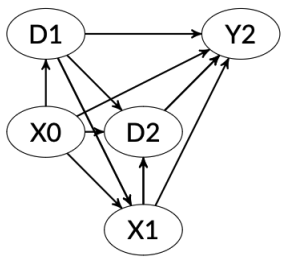
???

$\Delta_{\mu,\pi} \rightarrow \Delta$

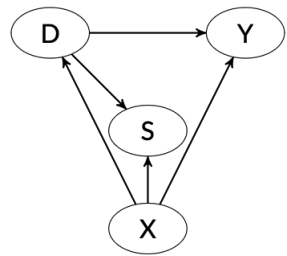
Mediačná analýza



Dynamické efekty



Výberová vzorka



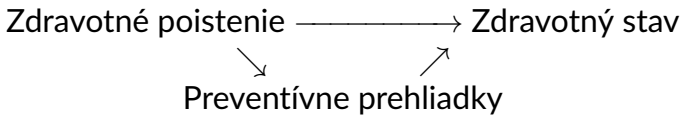
- Mediačná analýza (H. Farbmacher, M. Huber, H. Langen, LL, M. Spindler)
- Dynamické efekty (H. Bodory, M. Huber, LL)
- Výberová výchylka (M. Bia, M. Huber, LL)

Prvé rozšírenie

DML and mediation analysis

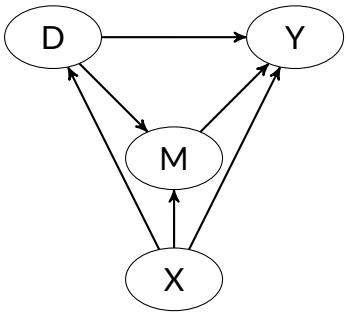
Helmut Farbmacher, Martin Huber, Lukáš Lafférs, Henrika Langen and Martin Spindler: Causal mediation analysis with double machine learning (Econometrics Journal, 2022, 25 (2), 277—300)

Príklad



Detaily | Príklady na mediačnú analýzu

DML a mediačná analýza



Objekt záujmu:

Nepriamy efekt: $E[Y(d, M(1)) - Y(d, M(0))]$
Priamy efekt: $E[Y(1, M(d)) - Y(0, M(d))]$

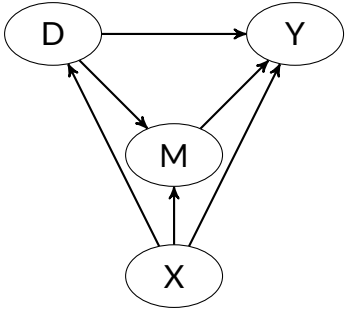
Indentifikačné predpoklady:

1) Podmienená nezávislosť D
$\{Y(d', m),\ M(d)\} \perp D \mid X$

2) Podmienená nezávislosť M
$Y(d', m) \perp M \mid D = d,\ X = x$

3) Spoločný nosič
$\Pr(D = d | M = m, X = x) > 0$

DML a mediačná analýza



Model:

$$
\begin{aligned}
\psi(W;\theta_0,\eta) &= \frac{I\{D=d\}(1-p_d(M,X))}{p_{dm}(M,X)\cdot 1-p_d(X)}\cdot[Y-\mu(d,M,X)] \\
&+ \frac{I\{D=1-d\}}{1-p_d(X)}\cdot\Big[\mu(d,M,X)-\omega(1-d,X)\Big] \\
&+ E\Big[\mu(d,M,X)\Big|D=1-d,X\Big]-\theta_0. \\
E\Big[\psi(W;\theta_0,\eta)\Big] &= E\Big[Y(d,M(1-d))\Big]-\theta_0=0
\end{aligned}
$$

Dáta: $W=(Y,D,M,X)$

Parametre vedľajšieho záujmu: $\eta=(p_d,p_{dm},\mu,\omega)$

- $p_d(X)=Pr(D=d|X)$
- $p_{dm}(M,X)=Pr(D=d|M,X)$
- $\mu(D,M,X)=E(Y|D,M,X)$
- $\omega(1-d,X)=E[\mu(d,M,X)|D=1-d,X]$

Aplikácia

Výsledky:

direct indirect

	$\hat{\Delta}$	$\hat{\theta}(1)$	$\hat{\theta}(0)$	$\hat{\delta}(1)$	$\hat{\delta}(0)$
	Modified score using Bayes' rule				
effect	-0.05	-0.07	-0.05	0.00	0.02
se	0.03	0.03	0.03	0.01	0.01
p-value	0.10	0.03	0.10	0.89	0.07

- Zdravotné poistenie mierne zlepšuje všeobecné zdravie v krátkodobom horizonte u mladých dospelých v USA mechanizmami inými než bežné preventívne prehliadky.

Detaily

Druhé rozšírenie

DML and dynamic treatment effects

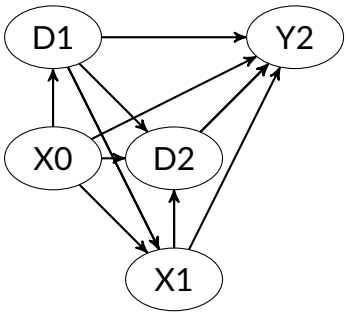
Hugo Bodory, Martin Huber and Lukáš Lafférs: Evaluating (weighted) dynamic treatment effects by double machine learning (The Econometrics Journal 25.3 (2022): 628—648

Príklad

Rekvalifikačné (akademické/odborné) kurzy $\longrightarrow$ Zamestnanosť

Detaily

DML a dynamické efekty



Objekt záujmu:

$$E[Y(\underline{d}_2)] - E[Y(\underline{d}_2^*)]$$

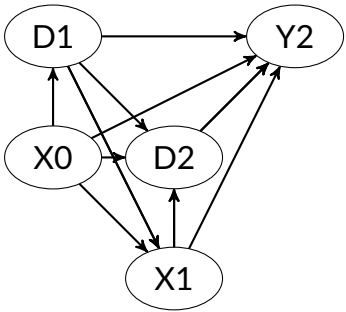
Indentifikačné predpoklady:

1) Podmienená nez. prvej intervencie
$Y_2(\underline{d}_2) \perp D_1 | X_0$, for $\underline{d}_2 \in \{0, 1, ..., Q\}^2$

2) Podmienená nez. druhej intervencie
$Y_2(\underline{d}_2) \perp D_2 | D_1, X_0, X_1$, for $\underline{d}_2 \in \{0, 1, ..., Q\}^2$.

3) Spoločný nosič
$\Pr(D_1 = d_1 | X_0) > 0$, $\Pr(D_2 = d_2 | D_1, \underline{X}_1) > 0$

DML a dynamické efekty



Model:

$$\begin{aligned}\psi(W;\theta_0,\eta) &= \frac{I\{D_1=d_1\}\cdot I\{D_2=d_2\}\cdot[Y_2-\mu^{Y_2}(\underline{d}_2,\underline{X}_1)]}{p^{d_1}(X_0)\cdot p^{d_2}(d_1,\underline{X}_1)} \\ &+ \frac{I\{D_1=d_1\}\cdot[\mu^{Y_2}(\underline{d}_2,\underline{X}_1)-\nu^{Y_2}(\underline{d}_2,X_0)]}{p^{d_1}(X_0)}+\nu^{Y_2}(\underline{d}_2,X_0)-\theta_0. \\ E\Big[\psi(W;\theta_0,\eta)\Big] &= E\Big[Y_2(\underline{d}_2)\Big]-\theta_0=0\end{aligned}$$

Dáta: $W=(Y_2,D_1,D_2,X_0,X_1)$

Parametre vedľajšieho záujmu:

$\eta=(p^{d_1},p^{d_2},\mu^{Y_2},\nu^{Y_2})$

- $p^{d_1}(X_0)\equiv \Pr(D_1=d_1|X_0)$
- $p^{d_2}(D_1,\underline{X}_1)\equiv \Pr(D_2=d_2|D_1,\underline{X}_1)$
- $\mu^{Y_2}(\underline{D}_2,\underline{X}_1)\equiv E[Y_2|\underline{D}_2,X_0,X_1]$
- $\nu^{Y_2}(\underline{D}_2,X_0)\equiv E[E[Y_2|\underline{D}_2,X_0,X_1]|D_1,X_0]$,

DML a dynamické efekty: Aplikácia

Výsledky (zamestnanosť po 4 rokoch):

3 = vocational 2=academic

10% ↑ employment after 4years

d_1	d_2^*	$\hat{E}[Y_2(d_2^*)\|S=1]$	$\hat{\Delta}(d_2, d_2^*, S=1)$	SE	p-value	observations	trimmed
33	22	0.76	0.1	0.06	0.11	3783	507
33	21	0.82	0.05	0.03	0.07	3783	43
33	11	0.81	0.08	0.03	0.02	2346	22

1= no training

Detaily

Tretie rozšírenie

DML and sample selection models

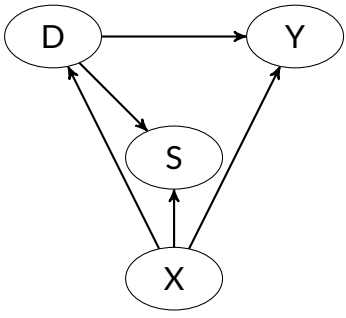
Michela Bia, Martin Huber and Lukáš Lafférs: Double machine learning for sample selection models. Journal of Business & Economic Statistics, 2024, 42 (3), 958-969

Príklad

Rekvalifikačné (akademické/odborné) kurzy $\longrightarrow$ Mzdy

Detaily

DML a modely s výberovou výchylkou



Objekt záujmu:

$$E[Y(d)] - E[Y(d^*)]$$

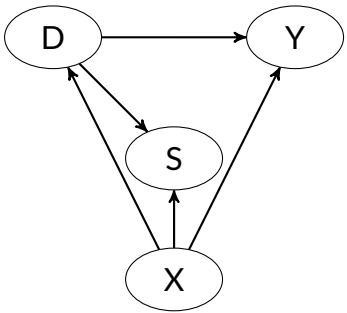
Indentifikačné predpoklady

1) Podmienená nez. intervencie:
$Y(d) \perp D | X = x$

2) Podmienená nez. selekcie
$Y \perp S | D = d, X = x$

3) Spoločný nosič
(a) $\Pr(D = d | X = x) > 0$ and (b)
$\Pr(S = 1 | D = d, X = x) > 0$

DML a modely s výberovou výchylkou



Model:

$$\psi(W;\theta_0,\eta) = \frac{I\{D=d\}\cdot S\cdot[Y-\mu(d,1,X)]}{p_d(X)\cdot\pi(d,X)} + \mu(d,1,X) - \theta_0.$$

$$E\Big[\psi(W;\theta_0,\eta)\Big] = E\Big[Y(d)\Big] - \theta_0 = 0$$

Dáta: $W = (Y.S, S, D, X)$

Parametre vedľajšieho záujmu: $\eta = (p^d, \pi, \mu)$

- $p^d(X) = \Pr(D = d|X)$
- $\pi(D,X) = \Pr(S = 1|D,X)$
- $\mu(D,S,X) = E[Y|D,S,X]$

DML a modely s výberovou výchylkou: Aplikácia

$D=1$	$D=0$	ATE	štandardná odchýlka	p-hodnota
Veta 1 (MAR)				
akademické	bez kurzu	-0.683	1.073	0.524
odborné	bez kurzu	0.611	0.629	0.331
Veta 3 (IV)				
akademické	bez kurzu	-0.631	1.052	0.549
odborné	bez kurzu	0.586	0.645	0.364
Veta 4 (sekvenčné)				
akademické	bez kurzu	0.149	0.199	0.454
odborné	bez kurzu	0.567	0.208	0.007

Pozorujeme mierne zvýšenie dlhodobejších hodinových miezd.

Detaily

Rekapitulácia

Rekapitulácia

DML je užitočný rámec pre odhadovanie v prostredí s vysokou dimenzionalitou.

Rekapitulácia

DML je užitočný rámec pre odhadovanie v prostredí s vysokou dimenzionalitou.

Dokáže automaticky vyberať medzi mnohými premennými a vyhnúť sa

regularizačnému biasu (cez Neymanovu ortogonalitu) a
overfitting biasu (cez cross-fitting) a

poskytnúť odhad, ktorý je root-n konzistentný a asymptoticky normálny.

Rekapitulácia

DML je užitočný rámec pre odhadovanie v prostredí s vysokou dimenzionalitou.

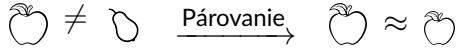
Dokáže automaticky vyberať medzi mnohými premennými a vyhnúť sa

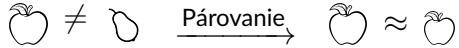
regularizačnému biasu (cez Neymanovu ortogonalitu) a
overfitting biasu (cez cross-fitting) a

poskytnúť odhad, ktorý je root-n konzistentný a asymptoticky normálny.

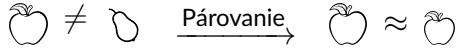
Ukázal som tri rozšírenia DML, ktoré sa javia ako empiricky relevantné a užitočné.

Implementované v balíku `causalweight` pre R (Bodory a Huber 2018)

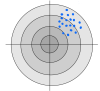

≠
Párovanie
≈


Párovanie

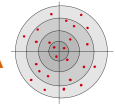

BIAS vs VARIANCIA



:

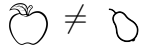


BIAS vs VARIANCIA

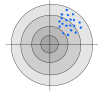


$\Delta_\mu \dashrightarrow \Delta$

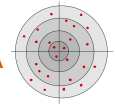
$\Delta_\pi \dashrightarrow \Delta$



:



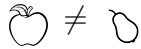
BIAS vs VARIANCIA



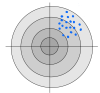
$\Delta_\mu \dashrightarrow \Delta$

$\Delta_\pi \dashrightarrow \Delta$

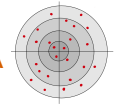
Double ML $\longrightarrow$ $\Delta_{\mu,\pi} \to \Delta$



:



BIAS vs VARIANCIA

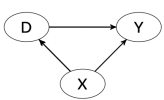


$\Delta_\mu \dashrightarrow \Delta$

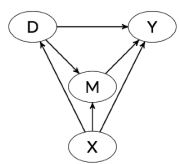
$\Delta_\pi \dashrightarrow \Delta$

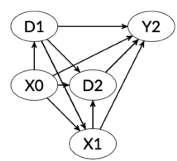
Double ML

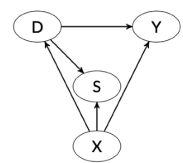
$\Delta_{\mu,\pi} \rightarrow \Delta$

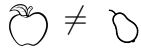


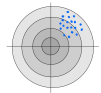
Rozšírenia



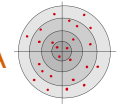








BIAS vs VARIANCIA

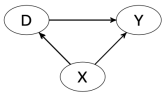


$\Delta_\mu \dashrightarrow \Delta$

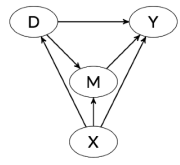
$\Delta_\pi \dashrightarrow \Delta$

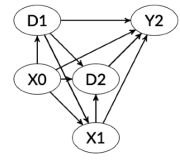
Double ML

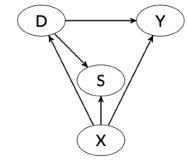
$\Delta_{\mu,\pi} \rightarrow \Delta$



Rozšírenia







Ďakujem za pozornosť. `lukaslaffers.github.io` `#UMBmath`

References

- Double machine learning framework: Chernozhukov, Victor, et al. "Double/debiased machine learning for treatment and structural parameters." The Econometrics Journal 21.1 (2018): C1-C68.
- DoubleML package in R `https://cran.r-project.org/web/packages/DoubleML/DoubleML.pdf`
- Bach, Philipp, et al. "DoubleML–An Object-Oriented Implementation of Double Machine Learning in R." arXiv preprint arXiv:2103.09603 (2021).
- Knaus, M. C. (2022). Double machine learning-based programme evaluation under unconfoundedness. The Econometrics Journal, 25(3), 602-627.
- Bang, Heejung, and James M. Robins. "Doubly robust estimation in missing data and causal inference models." Biometrics 61.4 (2005): 962-973.
- Hünermund, P., Louw, B., and Caspi, I. (2023). Double machine learning and automated confounder selection: A cautionary tale. Journal of Causal Inference, 11(1), 20220078.
- Farbmacher, Helmut, et al. "Causal mediation analysis with double machine learning." The Econometrics Journal 25.2 (2022): 277-300.
- Bodory H., Huber M. and Lafférs L. "Evaluating (weighted) dynamic treatment effects by double machine learning." The Econometrics Journal 25.3 (2022): 628—648.
- Bia, M., Huber, M., and Lafférs, L. (2023). Double machine learning for sample selection models. Journal of Business & Economic Statistics, 1-12.
- Bodory, Hugo, and Martin Huber. "The causalweight package for causal inference in R." (2018).

Neyman-ortogonalita pre ψ

Overíme, že ψ spĺňa podmienku Neymanovej ortogonality.
Označenie:

- $\eta = (m, g)$ je vektor vedľajších (nuisance) parametrov, $\eta_0 = (m_0, g_0)$ je skutočný vektor,
- $\eta_r = \eta_0 + r(\eta - \eta_0)$.

Neyman-ortogonalita pre ψ

$$
\begin{aligned}
\psi(W;\eta_r,\theta_0) &= (D-m_0(X)-r(m(X)-m_0(X)))\cdot(Y-g_0(X)-r(g(x)-g_0(X))-D\theta_0) \\
&= (D-m_0(X))\cdot(Y-g_0(X)-D\theta_0)+ \\
&\quad -r(D-m_0(X))\cdot(g(x)-g_0(X)) \\
&\quad -r(m(X)-m_0(X))\cdot(Y-g_0(X)-D\theta_0) \\
&\quad +r^2(m(X)-m_0(X))\cdot(g(x)-g_0(X))
\end{aligned}
$$

Neyman-ortogonalita pre ψ

$$
\begin{aligned}
\psi(W;\eta_r,\theta_0) &= (D-m_0(X)-r(m(X)-m_0(X)))\cdot(Y-g_0(X)-r(g(x)-g_0(X))-D\theta_0) \\
&= (D-m_0(X))\cdot(Y-g_0(X)-D\theta_0)+ \\
&\quad -r(D-m_0(X))\cdot(g(x)-g_0(X)) \\
&\quad -r(m(X)-m_0(X))\cdot(Y-g_0(X)-D\theta_0) \\
&\quad +r^2(m(X)-m_0(X))\cdot(g(x)-g_0(X)) \\
\frac{\partial}{\partial r}\mathsf{E}[\psi(W;\eta_r,\theta_0)] &= -\mathsf{E}[(D-m_0(X))\cdot(g(x)-g_0(X))] \\
&\quad -\mathsf{E}[(m(X)-m_0(X))\cdot(Y-g_0(X)-D\theta_0)] \\
&\quad +2\cdot r\cdot\mathsf{E}[(m(X)-m_0(X))\cdot(g(x)-g_0(X))]
\end{aligned}
$$

Neyman-ortogonalita pre ψ

$$
\begin{aligned}
\psi(W;\eta_r,\theta_0) &= (D-m_0(X)-r(m(X)-m_0(X)))\cdot(Y-g_0(X)-r(g(x)-g_0(X))-D\theta_0) \\
&= (D-m_0(X))\cdot(Y-g_0(X)-D\theta_0)+ \\
&\quad -r(D-m_0(X))\cdot(g(x)-g_0(X)) \\
&\quad -r(m(X)-m_0(X))\cdot(Y-g_0(X)-D\theta_0) \\
&\quad +r^2(m(X)-m_0(X))\cdot(g(x)-g_0(X)) \\
\frac{\partial}{\partial r}\mathsf{E}[\psi(W;\eta_r,\theta_0)] &= -\mathsf{E}[(D-m_0(X))\cdot(g(x)-g_0(X))] \\
&\quad -\mathsf{E}[(m(X)-m_0(X))\cdot(Y-g_0(X)-D\theta_0)] \\
&\quad +2\cdot r\cdot\mathsf{E}[(m(X)-m_0(X))\cdot(g(x)-g_0(X))] \\
\frac{\partial}{\partial r}\mathsf{E}[\psi(W;\theta_0,\eta_r)]\Big|_{r=0} &= -\mathsf{E}[(D-m_0(X))\cdot(g(x)-g_0(X))] \\
&\quad -\mathsf{E}[(m(X)-m_0(X))\cdot(Y-g_0(X)-D\theta_0)]
\end{aligned}
$$

Neyman-ortogonalita pre ψ

$$\begin{aligned}\frac{\partial}{\partial r}\mathrm{E}[\psi(W;\theta_0,\eta_r)]\Big|_{r=0} &= -\mathrm{E}[(D-m_0(X))\cdot(g(x)-g_0(X))] \\ &\quad -\mathrm{E}[(m(X)-m_0(X))\cdot(Y-g_0(X)-D\theta_0)] \\ &= 0\end{aligned}$$

Neyman-ortogonalita pre ψ

$$\begin{aligned}\frac{\partial}{\partial r}\mathrm{E}[\psi(W;\theta_0,\eta_r)]\Big|_{r=0} &= -\mathrm{E}[(D-m_0(X))\cdot(g(x)-g_0(X))] \\ &\quad -\mathrm{E}[(m(X)-m_0(X))\cdot(Y-g_0(X)-D\theta_0)] \\ &= 0\end{aligned}$$

nakoľko

Neyman-ortogonalita pre ψ

$$\begin{aligned}\frac{\partial}{\partial r}\mathrm{E}[\psi(W;\theta_0,\eta_r)]\Big|_{r=0} &= -\mathrm{E}[(D-m_0(X))\cdot(g(x)-g_0(X))] \\ &\quad -\mathrm{E}[(m(X)-m_0(X))\cdot(Y-g_0(X)-D\theta_0)] \\ &= 0\end{aligned}$$

nakoľko

$$\mathrm{E}[(D-m_0(X))\cdot(g(X)-g_0(X))] = \mathrm{E}[(g(X)-g_0(X))\cdot\underbrace{\mathrm{E}[D-m_0(X)|X]}_{\mathrm{E}[V|X]=0}] = 0$$

Neyman-ortogonalita pre ψ

$$\begin{aligned}\frac{\partial}{\partial r}\mathrm{E}[\psi(W;\theta_0,\eta_r)]\Big|_{r=0} &= -\mathrm{E}[(D-m_0(X))\cdot(g(x)-g_0(X))] \\ &\quad -\mathrm{E}[(m(X)-m_0(X))\cdot(Y-g_0(X)-D\theta_0)] \\ &= 0\end{aligned}$$

nakoľko

$$\begin{aligned}\mathrm{E}[(D-m_0(X))\cdot(g(X)-g_0(X))] &= \mathrm{E}[(g(X)-g_0(X))\cdot\underbrace{\mathrm{E}[D-m_0(X)|X]}_{\mathrm{E}[V|X]=0}]=0 \\ \mathrm{E}[(m(X)-m_0(X))\cdot(Y-g_0(X)-D\theta_0)] &= \mathrm{E}[(m(X)-m_0(X))\cdot\underbrace{\mathrm{E}[Y-g_0(X)-D\theta_0|X]}_{\mathrm{E}[U|X,D]=0}]=0\end{aligned}$$

a teda ψ je vskutku Neyman-ortogonálna.

Späť

Súvisiace s hodnotením pracovného tréningu:

- Priamy efekt na mzdy programu Job Corps s využitím pracovných skúseností ako mediátora (Flores a Flores-Lagunes (2009))
- Efekt Perry Preschool Program na zdravé správanie sprostredkovaný osobnostnými črtami (Conti, Heckman a Pinto (2016))
- Aký je efekt dôkladnejších poradcov v procese poradenstva na zamestnanosť sprostredkovaný zaradením do programu trhu práce (Huber, Lechner a Mellace (2017))

Súvisiace so vzdelaním a mzdami:

- Ako rast v chudobe ovplyvňuje ekonomické výsledky v dospelosti s využitím vzdelania ako mediátora (Bellani a Bia (2018))
- Decompozícia mzdového rozdielu (pohlavie, sociálno-ekonomické premenné, mzda) (Huber (2015))
- Efekt vzdelania na úmrtnosť sprostredkovaný kognitívnymi schopnosťami (Bijwaard a Jones (2018))

Na základe nástrojových premenných:

- Efekt vzdelania na spokojnosť so životom s využitím príjmu ako mediátora (Powdthavee, Lekfuangfu a Wooden (2013))
- Efekt vzdelania na zdravie sprostredkovaný zdravotným správaním (Brunello, Fort, Schneeweis a Winter-Ebmer (2016))
- Efekt zloženia rodiny na vzdelanie prvého dieťaťa s využitím veľkosti rodiny ako mediátora (Chen, Chen a Liu (2017))

Späť

Mediáčná analýza príklad - detaily

- zdravotné poistenie $\rightarrow$ zdravotný stav
- zdravotné poistenie $\rightarrow$ pravidelné preventívne prehliadky $\rightarrow$ zdravotný stav
- X - demografia, rodinné zázemie, vzdelanie, trh práce, charakteristiky domácnosti, duševné zdravie, výživa, fyzická aktivita....
  (755 kontrolných premenných, z roku 2005)

- Národný longitudinálny prieskum mládeže 1997 (NLSY97), prieskum amerického Ministerstva práce (2019) (n $\approx$ 7500)
- väčšina štúdií zistila významný efekt na konkrétny typ skríningu (rakovina, mŕtvica...)
- máme mladších respondentov a krátkodobé efekty (2006 $\rightarrow$ 2007 $\rightarrow$ 2008)
- zdravie - „výborné“ až „zlé“, negatívny ATE $\approx$ zlepšenie

Späť

Dynamický príklad - detaily

- Tréning $\rightarrow$ Zamestnanie (po 4 rokoch)
- X - 1184 premenných ($X_0 - 814$, $X_1 - 374$) socio-ekonomické charakteristiky, vzdelanie a tréning pred zásahom, histórie trhu práce, aktivity hľadania zamestnania, prijímanie sociálnych dávok, zdravie, kriminalita...

- Job Corps poskytuje odborné školenie a akademické vyučovanie pre znevýhodnené osoby vo veku 16 až 24 rokov
- V súčasnosti približne 50 000 účastníkov každý rok.
- Vzorka pochádza z experimentálnej štúdie Job Corps realizovanej v polovici 90. rokov, Schochet et al. (2008): 11313 mladých osôb, s ukončenými rozhovormi štyri roky po randomizácii (6828 zaradených do Job Corps, 4485 randomizovaných mimo).
- Sekvencie zásahov sú založené na účasti v akademickom alebo odbornom tréningu v prvom alebo druhom roku po randomizácii medzi tými, ktorí boli randomizovaní do programu. Späť

Príklad výberu vzorky - detaily

- Tréning $\rightarrow$ Hodinová mzda
- Stovky počiatočných kovariátov X (socio-ekonomické premenné, história trhu práce, kriminalita, zdravie...).

- Job Corps poskytuje odborné školenie a akademické vyučovanie pre znevýhodnené osoby vo veku 16 až 24 rokov
- V súčasnosti približne 50 000 účastníkov každý rok.
- Vzorka pochádza z experimentálnej štúdie Job Corps ($n \approx 3600$)
- Výsledok Y je **hodinová mzda** v poslednom týždni prvého roka alebo štyri roky po randomizácii, pozorovaný pod podmienkou zamestnania S.
- Zásah D je účasť v akademickom alebo odbornom **školení** v prvom roku po randomizácii medzi tými, ktorí boli randomizovaní do programu.

Späť