Propensity scores

Discussion of papers

Danny Pfeffermann

Southampton, March 2018

The problem of observational studies

In **observational studies**, assignment of units to treatments (and possibly **the sample selection),** **not under control**.

The treatment and control groups may be very different in their **characteristics** (covariates), possibly resulting in **biased estimates** of the treatment effects.

Propensity scores

The conditional probability of treatment, given known covariates or interactions between them.

Once estimated (**e.g.,** via **logistic regression**) and under a strong ignorability assumption (see later), the treatment effects can be estimated **"unbiasedly"** by use of methods such as **pairwise matching, stratification, regression modelling** or **inverse probability weighting**.

Notation and definitions

Population: **U=1,…,*N*.** Every unit $i \in$**U** **potentially exposed** to each of ***m*** treatments with outcomes $y_i^t,\ t=1,\ldots,m$.

$\rightarrow$ **"Counterfactual approach".**

(Simple**)** **target parameters:**

$$\mu^{U,t} = \sum_{i=1}^{N} y_i^t / N,\quad \mu^t = \sum_{i=1}^{N} E(y_i^t \mid x_i)/N \quad \ldots t=1\ldots m$$

Or contrasts between the means**, e.g., (**$\mu^{U,1} - \mu^{U,2}$**).**

$\Downarrow$

Average Treatment Effect (ATE)

Notation and definitions (cont.)

Sample **S** of observational units obtained with **(known or unknown)** probabilities $\pi_i = \Pr(i \in S)$.

Every unit ***j*** $\in$ **S** exposed to **one** treatment with probability

$p_j^t = \Pr[T(j) = t \mid j \in S]$ **;** $\sum_{t=1}^{m} p_j^t = 1.$

$$\Downarrow$$

$$P(i \in S, T(i) = t) = \pi_i \times p_i^t = q_i^t$$

After assignment, $S = S^1 \cup, \ldots, \cup S^T$ **;** $S^t = \{i \mid i \in S, T(i) = t\}.$

Problem revisited

For an uncontrolled observational study,

$$f_{S^t}(y_i^t \mid x_i) = f(y_i^t \mid x_i, \boldsymbol{i} \in \boldsymbol{S^t}) \neq f_U(y_i^t \mid x_i) = f(y_i^t \mid x_i, \boldsymbol{i} \in \boldsymbol{U})$$

$f(y_i^t \mid x_i, i \in U)$**=** ***pdf*** of y_i^t if every unit $\boldsymbol{i} \in \mathbf{U}$ is exposed to treatment ***t*** **('population model').**

In particular,

$$E_{S^t}(y_i^t \mid i \in S^t) \neq E_U(y_i^t \mid i \in U)$$

unless under **"strong ignorability".**

Methods based on propensity scores

Assume the availability of **auxiliary variables (X)** that control the assignment bias**.** Suppose **m=2, 1= treatment, 0=control.**

Strong ignorability assumption: $P(T_i = 1 | y_i, x_i) = P(T_i = 1 | x_i)$ = $e(x_i)$ **(=PS).** The treatment assignment, **T,** and the response, **Y,** are **conditionally independent,** given the covariates**, X**.

Example of estimator:

$$A\hat{T}E = \sum_{i=1}^{n} \frac{T_i y_i}{\hat{e}(x_i)} / \sum_{i=1}^{n} \frac{T_i}{{}_i\hat{e}(x_i)} - \sum_{i=1}^{n} \frac{(1-T_i) y_i}{[1-\hat{e}(x_i)]} / \sum_{i=1}^{n} \frac{(1-T_i)}{[1-\hat{e}(x_i)]}.$$

(Assuming sampling with equal probabilities.)

Do propensity scores solve the bias problem?

Problem: $f_{S^t}(y_j^t \mid x_j) = f(y_j^t \mid x_j, j \in S^t) \neq f_U(y_j^t \mid x_j)$.

$$f_{S^t}(y_j^t \mid x_j) = \frac{\Pr(j \in S^t \mid y_j^t, x_j) f_U(y_j^t \mid x_j)}{\Pr(j \in S^t \mid x_j)}$$

$f_U(y_j^t \mid x_j)$ - ***pdf*** **under 'ignorable' assignment.**

$$f_{S^t}(y_j^t \mid x_j) = f_U(y_j^t \mid x_j) \Leftrightarrow \Pr(j \in S^t \mid y_j^t, x_j) = \Pr(j \in S^t \mid x_j)$$

$\downarrow$

propensity scores

Issues: Availability of all relevant **X's** not guaranteed.

Difficult (possible**?**) to test the existence of strong ignorability.

Discussion of Article by Stijn Vansteelandt

How to obtain valid tests and confidence intervals for treatment effects after confounder selection?

The article proposes "**specific treatment effect estimators** in combination with **a specific selection strategy".**

The term **propensity scores** not mentioned even once ☹☹☹

Assumes parametric models for $E(A|L)=\pi(L;\gamma)$ (**A model**) and for $E(Y\,|\,L)=m(L;\beta)$ (**B model**).

A$\rightarrow$ **exposure** (treatment),

L$\rightarrow$ **pre-exposure** (background) characteristics.

Basic test statistic (known model parameters)

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{A_i-\pi(L_i;\gamma)\}\{Y_i-m(L_i;\beta)\}$$

"has mean zero under the null when either **model A** or **model B is correct**."

Question 1: what is the null? No exposure effect? Do you assume that **A and Y are uncorrelated given L?**

- In practice need to estimate the model parameters.

Variable selection carried out by penalizing the estimating equations with bridge penalties.

Proposed test statistic

$$T_n = \frac{\frac{1}{n}\sum_{i=1}^{n}\{A_i - \pi(L_i;\hat{\gamma})\}\{Y_i - m(L_i;\hat{\beta})\}}{\sqrt{\frac{1}{n}\{\frac{1}{n-1}\sum_{i=1}^{n}[\{A_i - \pi(L_i;\hat{\gamma})\}\{Y_i - m(L_i;\hat{\beta})\}]^2}}.$$

- In illustrations assumes $\pi(L_i;\hat{\gamma})$ **logistic**, $m(L_i;\hat{\beta})$ **linear**.

Properties

- When models **A** and **B** **are correct**, T_n converges **uniformly** to **standard normal distribution**.
- **Remains valid** when model A or B **misspecified**.
- Allows **arbitrary** conditional models A and B.

More questions

Question 2: I can see the logic of penalizing the estimating equations to account for the use of different number of parameters. **But do you actually select the variables L that should be included in the final model? How?**

Question 3: The title of the paper talks about **valid tests and confidence intervals** for **treatment effects**. Numerical results are also given. How do you set the **confidence intervals? Nothing said in the slides that I saw** ☹☹☹

Exp: you start the presentation by considering the linear model $E(Y \mid A, L) = \psi A + \beta' L$. Under this model we need a confidence interval for ψ. **Do you assume this model?**

Discussion of article by Rhian Daniel

Methods for dealing with measured confounding based on the propensity scores.

Propensity score $\boldsymbol{p(L)} = \Pr(A = 1 \mid L)$.

Rubin and Rosenbaum (1983): If ***Y*** **indep.** of ***A*** **given L** $\Rightarrow$

$\Rightarrow$ ***Y*** **indep.** of $\boldsymbol{p}(\mathrm{L})$ **given L** $\Rightarrow$ "Validity of methods based on propensity scores relies on correctly modelling **A|L**."

Conditions for proper use of propensity scores

“As in a randomised trial, an exposed and unexposed subject with the same ***p*****(L)** are exchangeable, unless,

(1) important confounders were not included in ***p*****(L)**

(2) the propensity score was incorrectly modelled.**”**

Question 1: How do we test that important confounders are **missing?** Modelling ***p*****(L)** with **available confounders,** **e.g.,** by logistic function unlikely to reveal this.

Question 2: What if for **available** **L**, **A** depends also on **Y**?

Standard methods based on propensity scores

The article compares empirical results obtained from 4 standard methods of using the estimated propensity scores:

Stratification, Matching, Adjusting, Weighting,

Followed by **nice discussions** of their virtues and limitations.

General conclusion: **"**These alternative methods, like traditional regression methods, are valid **only if L is sufficient** to control for **all confounding**.

Brings me back to my question:

Question 1: How do we test that important confounders are **missing?**

Alternative approach - Pfeffermann and Landsman (2011)

(*The Annals of Applied Statistics*)

Problem: $f_{S^t}(y_j^t \mid x_j) = f(y_j^t \mid x_j, j \in S^t) \neq f_U(y_j^t \mid x_j)$.

Solution: use the sample distribution for inference.

$$f_{S^t}(y_j^t \mid x_j) = \frac{\Pr(j \in S^t \mid y_j^t, x_j)\, f_U(y_j^t \mid x_j)}{\Pr(j \in S^t \mid x_j)}$$

$f_U(y_j^t \mid x_j)$ - ***pdf* under ‘ignorable’ assignment.**

- $f_{S^t}(y_j^t \mid x_j) = f_U(y_j^t \mid x_j) \Leftrightarrow \Pr(j \in S^t \mid y_j^t, x_j) = \Pr(j \in S^t \mid x_j)$

$\downarrow$

Propensity scores

Alternative approach (cont.)

$$f_{S^t}(y_j^t \mid x_j) = \frac{\Pr(j \in S^t \mid y_j^t, x_j)\, f_U(y_j^t \mid x_j)}{\Pr(j \in S^t \mid x_j)}$$

- $\Pr(j \in S^t \mid y_j^t, x_j) \neq q_j^t = \pi_i \times p_i^t = \Pr(j \in S^t)$.

q_j^t **may depend on** **unobservable** **variables that are possibly related to** y_j^t**. However,** **only need to model****,**

$$\Pr(j \in S^t \mid y_j^t, x_j) = E_U(\tilde{q}_j^t \mid y_j^t, x_j).$$

Inference based on the sample distribution

$$f_{S^t}(y_j^t \mid x_j) = \frac{\Pr(j \in S^t \mid y_j^t, x_j) f_U(y_j^t \mid x_j)}{\Pr(j \in S^t \mid x_j)}$$

Model $\Pr(j \in S^t \mid y_j^t, x_j; \alpha_t)$ and $f_U(y_j^t \mid x_j; \theta_t)$ **separately for each *t***

Maximum likelihood estimation

For given treatment *t*, the (full) likelihood is,

$$L[\{\theta^t, \alpha^t\}; (y_j^t, x_j), j \in S^t; x_i, i \notin S^t]$$

$$= \prod_{j \in S^t} f_{s^t}(y_j^t \mid x_j; \theta^t, \alpha^t) \prod_{j \in S^t} \Pr(j \in S^t \mid x_j; \theta^t, \alpha^t) \prod_{i \notin S^t} [1 - \Pr(i \in S^t \mid x_i; \theta^t, \alpha^t)]$$

Estimation of treatment effects

A- $\hat{\mu}^t = N^{-1}\sum_{j=1}^{N} \hat{E}_U(y_j^t \mid x_j) = N^{-1}\sum_{j=1}^{N} E_U(y_j^t \mid x_j; \hat{\theta}^t)$

B- $\hat{\mu}^{pt} = \sum_{j\in S^t}(y_j^t / \hat{q}_j^t) / \sum_{j\in S^t}(1/\hat{q}_j^t)$ **;** $\hat{q}_j^t = \Pr(j \in S^t \mid y_j^t, x_j; \hat{\alpha}^t)$ **or**

doubly robust estimator with probabilities $\hat{q}_j^t$ **instead of PS.**

- **Use of B does not require knowledge of *x* for units** $i \notin S$**.**

"Doubly robust estimate" (*Lunceford and Davidian*, 2005)

$$\hat{ATE} = \frac{1}{n}\sum_{i=1}^{n} \frac{T_i y_i - [T_i - \hat{e}(\mathrm{y}_i, x_i)]\hat{r}^1(x_i)}{\hat{e}(\mathrm{y}_i, x_i)} - \frac{1}{n}\sum_{i=1}^{n} \frac{(1-T_i)y_i + [T_i - \hat{e}(\mathrm{y}_i, x_i)]\hat{r}^0(x_i)}{[1-\hat{e}(\mathrm{y}_i, x_i)]}$$

$r^t(x) = E_U(y^t \mid x) = E_{S^t}(y^t \mid x)$.

- **Similar idea to "GREG" estimator in sampling.**

Properties of alternative procedure

1. Very flexible, possibly different models for different treatments.

2. Does not require to account of “all” potential confounders.

3. Identifiability of sample models needs to be verified.

4. It allows to test the assumptions underlying the use of propensity scores (and the use of instrumental variables).

5. The estimated model,

$$f_{S^t}(y_j^t \mid x_j; \hat{\alpha}, \hat{\theta}) = \frac{\Pr(j \in S^t \mid y_j^t, x_j; \hat{\alpha}) f_U(y_j^t \mid x_j; \hat{\theta})}{\Pr(j \in S^t \mid x_j; \hat{\alpha}, \hat{\theta})}$$ can be tested

based on the observations in S^t (**standard tests**).

Example: Are private schools better than public schools?

Programme for International Student Assessment (PISA).

- Collects information on children's proficiency in maths, science and reading **+** family and school characteristics.
- **32** countries**;** school children aged **15**.
- Survey waves every **3** years (first wave in 2000).
- In the present application we compare children's scores in Math between private and public schools in Ireland.

Sample sizes: private=1256 ; public=702

- **Same data analyzed by Vandenberghe and Robin (2004)**

Sample design

Stratified PPS sample of **150** schools**.**

- Equal probability sample of **35** pupils per school (or all the pupils if there are less).
- Unequal sampling weights (compensate for non-participating schools and nonresponse)**;**
- Present application uses data from year **2000.**

Covariates (confounders)

Gender, father education, socio-economic index, index of home educational resources, socio-economic index at school level.

Instrumental variable (for existing methods)

z **= School location: 1** if school in big city

0 otherwise

- Used in other studies as **instrumental variable.**

Model fitted

Population model (normal):

$$y_i^t \sim \mathrm{N}\left(\beta_0^t + x_i'\beta^t, \sigma_t^2\right),\ t\text{=0,1}$$

Covariates (x): all from previous slide

Assignment probabilities (logistic):

$$P(i \in S^t \mid y_i^t, v_i) = \frac{\exp(\gamma_0^t + \boldsymbol{\delta}^{\mathbf{t}} y_i^t + v_i'\gamma^t)}{1+\exp(\gamma_0^t + \boldsymbol{\delta}^{\mathbf{t}} y_i^t + v_i'\gamma^t)},\ t\text{=0,1;}$$

Estimation of model parameters in private schools

Assignment (logistic)

Coefficient	γ_0	δ	Gen.	S.E.I	H.E.R	S.E.S	S.loc
Estimate	-2.95	0.49	0.77	-0.12	3.16	0.09	1.13
Std error	1.30	0.21	0.13	0.07	0.20	0.07	0.13

Population (normal)

Parameter	σ	β_0	Gen.	M.E	S.E.I	H.E.R	S.E.S	S.loc
Estimate	0.83	6.09	-0.20	0.18	0.16	0.39	0.21	- 0.09
Std error	0.02	0.07	0.05	0.05	0.03	0.09	0.02	0.06

(*) Supports use of propensity scores (?) and IVM approach.

Estimation of model parameters in public schools

Assignment (logistic)

Coefficient	γ_0	δ	Gen.	S.E.I	H.E.R	S.E.S	S.loc
Estimate	13.88	-2.02	-0.76	0.40	-2.57	0.27	-1.63
Std error	2.90	0.39	0.18	0.12	0.30	0.11	0.24

Population (normal)

Parameter	σ	β_0	Gen.	M.E	S.E.I	H.E.R	S.E.S	S.loc
Estimate	1.10	6.89	0.17	0.11	0.16	1.35	0.30	0.23
Std error	0.07	0.14	0.08	0.07	0.04	0.20	0.04	0.15

Supports use of IVM but propensity scores questionable.

Estimation of population means by type of school

	Private School		Public School	
	$\hat{\mu}^1 = \overline{x}'\hat{\beta}^1$	$\hat{\mu}^1_{DR}$	$\hat{\mu}^0 = \overline{x}'\hat{\beta}^0$	$\hat{\mu}^0_{DR}$
Estimate	**6.91**	**7.05**	**6.09**	**6.10**
Std error	**0.12**	**0.15**	**0.06**	**0.05**

Estimation of ATE for Ireland

Alternative method

Method	$\hat{\Delta} = \hat{\mu}^1 - \hat{\mu}^0$	$\hat{\Delta}_{DR}$
Estimator	**- 0.82**	**- 0.95**
Std error	**0.13**	**0.16**

Model diagnostics

Goodness of fit test statistics (p-values)

Private schools			Public schools		
KS	CM	AD	KS	CM	AD
0.023 (0.12)	**0.089** (0.18)	**0.62** (0.11)	**0.027** (0.17)	**0.062** (0.32)	**0.45** (0.15)

Max$_{ps}$ $=\max_{j\in S}|1-\sum_{t=1}^{m}\hat{\Pr}(j\in S^t\,|\,x_j,v_j)|$=**0.0564**

$$\Pr(j\in S^t\,|\,x_j,v_j)=\int \Pr(j\in S^t\,|\,y_j^t,v_j)\,f_p(y_j^t\,|\,x_j)\,dy_j^t\,.$$

ATE estimates when data generated from Ireland model

Method	Ireland	Simulated Data
$\bar{y}^1 - \bar{y}^0$	0.36 (0.05)	0.37 (0.04)
Regress.	0.12 (0.05)	0.11 (0.05)
PS Match	0.21 (0.05)	0.16 (0.09)
Hajek	0.16 (0.05)	0.18 (0.05)
Doub.Rob.	0.17 (0.05)	0.17 (0.05)
IVM	-0.76 (0.26)	-0.72 (0.26)
LVM	-0.60 (0.20)	-0.60 (0.20)
$\hat{\mu}^1 - \hat{\mu}^0$	-0.95 (0.13)	-0.93 (0.13)
$\hat{\mu}^1_{DR} - \hat{\mu}^0_{DR}$	-0.83 (0.25)	-0.85 (0.21)

Estimation of ATE by alternative and existing methods

Alternative method

Method	$\hat{\Delta} = \hat{\mu}^1 - \hat{\mu}^0$	$\hat{\Delta}_{DR}$
Estimator	- 0.82	- 0.95
Std error	0.13	0.16

Existing methods

Method	$\bar{y}^1 - \bar{y}^0$	Reg.	PS Match	PS Hajek	Doub. Rob.	IVM	LVM
Estimate	0.36	0.12	0.21	0.16	0.17	- 0.61	- 0.49
Std error	0.05	0.05	0.05	0.05	0.05	0.24	0.19