The Role of the Propensity Score in Observational Studies with Complex Data Structures

Fabrizia Mealli
mealli@disia.unifi.it

Department of Statistics, Computer Science, Applications
University of Florence

TATE Talks, UNC, School of Social Work – May 22, 2017

Introduction and rationale of the talk

Arpino B. and Mealli F. (2011). The specification of the propensity score in multilevel observational studies. Computational Statistics and Data Analysis 55, 1770–1780

Forastiere L., Airoldi E. M., and Mealli F. (2016). Identification and estimation of treatment and interference effects in observational studies on networks. Arxiv working paper (http://arxiv.org/abs/1609.06245)

Papadogeorgou G., Mealli F., and Zigler C. (2017). Inverse probability weighted estimators under partial interference. (Work in progress, poster at ACIC 2017, *Causal Inference for interfering units under treatment regimes that incorporate covariate information in the counterfactual treatment assignment*)

Common feature of these papers is that they use the (generalized) propensity score to propose methods to adjust for covariates in complex settings under various form of unconfoundedness and SUTVA

Notation

Each unit (in a population of N) is characterized by a K-vector of characteristics, denoted by $\mathbf{X}_i$ for unit i, with $\mathbf{X}$ denoting the $N \times K$ matrix of characteristics

Let W_i denote the treatment, to which unit i is assigned : $W_i \in \mathbb{W} = \{0, 1\}$

Stable Unit Treatment Value Assumption (SUTVA)

- ✓ SUTVA: the potential outcomes for any unit do not vary with the treatments assigned to any other units, and there are no different versions of the treatment
- ✓ SUTVA is a form of exclusion restriction: assumptions that rely on outside information to rule out the possibility of any causal effect of a particular treatment

For each unit, let $Y_i(0)$ and $Y_i(1)$ denote the outcomes under the two values of the treatment

Potential outcomes $(\mathbf{Y}(0), \mathbf{Y}(1)) = [(Y_i(0), Y_i(1))]_{i=1}^N$ and assignments $\mathbf{W} = [W_i]_{i=1}^N$ jointly determine the values of the observed and missing outcomes:

$$Y_i^{obs} \equiv Y_i(W_i) = W_i \cdot Y_i(1) + (1 - W_i) \cdot Y_i(0)$$

$$Y_i^{mis} \equiv Y_i(1 - W_i) = (1 - W_i) \cdot Y_i(1) + W_i \cdot Y_i(0)$$

Basics of Propensity Scores

The assignment mechanism (AM) gives the conditional probability of each vector of assignments given the covariates and potential outcomes:

$$p(\mathbf{W}|\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1))$$

Given a population of N units, the AM defines the probability of receiving the treatment for each unit i as a function of the covariates and the potential outcomes:

$$p_i(\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) = \sum_{\mathbf{W}:W_i=1} p(\mathbf{W}|\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) \quad \forall\, i = 1, 2, \ldots, N$$

Restrictions on the AM: Individualistic, Probabilistic and Unconfounded

- ✓ Individualistic Assignment:

$$p_i(\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) = p(W_i = 1|\mathbf{X}_i, Y_i(0), Y_i(1)) \qquad \forall\, i = 1, 2, \ldots, N$$

- ✓ Probabilistic Assignment: For each possible $\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)$

$$0 < p_i(\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) < 1 \qquad \forall\, i = 1, 2, \ldots, N$$

- ✓ Unconfounded Assignment: An AM is unconfounded if it does not depend on the potential outcomes:

$$p(\mathbf{W}|\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) = p(\mathbf{W}|\mathbf{X})$$

Propensity scores

Propensity score for binary treatments. The propensity score at $\mathbf{x}$ is the average unit assignment probability for units with $\mathbf{X}_i = \mathbf{x}$

$$e(\mathbf{x}) = \frac{1}{N(\mathbf{x})} \sum_{i:\mathbf{X}_i=\mathbf{x}} p_i(\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1))$$

where $N(\mathbf{x}) = \sharp\{i = 1, \ldots, N \mid \mathbf{X}_i = \mathbf{x}\}$ is the number of units with $\mathbf{X}_i = \mathbf{x}$ ($e(\mathbf{x}) \equiv 0$ if $N(\mathbf{x}) = 0$)

Unconfoundedness and Individualistic assignment implies that the propensity score is the unit-level assignment probability of receiving the treatment

$$e(\mathbf{x}) = p(W_i = 1 | \mathbf{X}_i)$$

Observational studies: An assignment mechanism corresponds to an observational study if it is an unknown function of its arguments

Properties of the propensity score

Balancing property of the propensity score: The probability of receiving the active treatment given the covariates is free of dependence on the covariates given the propensity score

$$W_i \perp\!\!\!\perp \mathbf{X}_i \mid e(\mathbf{X}_i)$$

Unconfoundedness given the popensity score: Suppose assignment to treatment is unconfounded. Then assignment is unconfounded given the propensity score only:

$$\text{If} \quad W_i \perp\!\!\!\perp Y_i(0), Y_i(1) \mid \mathbf{X}_i) \quad \text{then} \quad W_i \perp\!\!\!\perp Y_i(0), Y_i(1) \mid e(\mathbf{X}_i)$$

Unconfoundedness given the propensity score has generated methods of adjusting based on the propensity score: weighting, regression, subclassification, matching

Propensity score with multilevel data

Clustered data: individual- and cluster-level covariates

Treatment assignment at cluster level
(Keele and Zubizarreta, 2017; Pimentel at al., 2017)

Treatment assignment at individual level
(Kim and Seltzer, 2007; Rosenbaum et al., 2007; Aussems, 2008; Su, 2008; Li et al., 2013; Arpino and Mealli, 2012)

The specification of the propensity score in multilevel studies

- Assignment mechanism may depend on individual- and cluster-level covariates
- Mimic block randomized experiments or multi-site experiments
- Arpino and Mealli (2012) consider cases of omitted variable bias due to unobserved cluster-level covariates
- Matching within clusters achieves perfect balance in cluster-level covariates but often not feasible and leading to poor balance in individual-level covariates

The specification of the propensity score in multilevel studies

(Arpino and Mealli, 2012)

Different specification of the propensity score (logit link):

- ✓ Random-effect multilevel models
- ✓ Fixed-effect models
- ✓ Models that ignore clustering

Simulations showing bias/efficiency of nearest-neighbour PS matching estimators

Motivating example: analyzing the effects of childbearing events on economic wellbeing in Vietnam, where community characteristics play important roles

Overall results and implications

Fixed-effect specification of the PS outperforms in terms of bias and efficiency

- ✓ Robust to different distribution of cluster-level covariates
- ✓ Good even with small and/or imbalanced cluster size
- ✓ Still good when irrelevant variables included

The inclusion of fixed-effects specifies a model for the PS more general than the ideal if cluster-level variables were available

When conducting PS analysis it is safer to specify a more general model than pursuing model parsimony

Interference

So far, we have assummed **SUTVA**, according to which the potential outcomes for any unit do not vary with the treatments assigned to any other units

SUTVA allows us to write that for a unit i there are two potential outcomes $Y_i(0), Y_i(1)$

In the presence of interference, a unit's outcome depends on the individual treatment, but also on the treatment of others

- ✓ For example, neighbor's vaccination status can affect an individual's outcome

Under interference, the set of potential outcomes is $\{Y_i(\mathbf{w}), \mathbf{w} \in \{0,1\}^n\}$

- ✓ This allows for 2^n potential outcomes for every unit, where n is the number of observations
- ✓ The treatment of any other observation can affect the outcome of unit i

Partial interference

Units can be clustered in groups within which there is interference, but not among them

Denote $k \in \{1, 2, \ldots, K\}$ to be a cluster with n_k individuals.

$\mathbb{W}(n) = \{0, 1\}^n$: set of vectors of possible treatment allocations of length n

Let W_{ki} to be the treatment indicator of unit i in cluster k, and write $\mathbf{W}_k = (W_{k1}, \ldots, W_{kn_k})$, and $\mathbf{W}_{k,-i} = (W_{k1}, \ldots, W_{kj-1}, W_{kj+1}, \ldots, W_{kn_k})$

Partial interference. Let $k(i) \in \{1, \ldots, K\}$ denote the class to which unit i belongs, and decompose $\mathbf{W} = (\mathbf{W}_1 \ldots, \mathbf{W}_K)$. For all $\mathbf{W}$ and $\mathbf{W}'$ such that $\mathbf{W}_{k(i)} = \mathbf{W}'_{k(i)}$ we have $Y_i(\mathbf{W}) = Y_i(\mathbf{W}')$

- ✓ Then, unit's i potential outcomes are $\{Y(\mathbf{w}_k), \mathbf{w}_k \in \mathbb{W}(n)\}$

$\mathbf{X}_{ki}$ = vector of fixed individual and group-level covariates; $\mathbf{X}_k, \mathbf{X}_{k,-i}$ similarly to $\mathbf{W}_k, \mathbf{W}_{k,-i}$

Observed and counterfactual treatment allocation

(Papadogeorgou, Mealli, Zigler, 2017)

Observed treatment allocation

- ✓ The mechanism that has assigned the observed treatment
- ✓ Clinical trials (randomization), observational studies (covariates)

Counterfactual treatment allocation

- ✓ What is the intervention that we are imagining?
- ✓ In what hypothesized world are we estimating the effect of interest?
- ✓ Interpretation of the effects requires a hypothesized treatment allocation that is applicable

Previous literature on interference has considered

- ✓ Randomized observed treatment allocation
- ✓ Covariate-dependent observed treatment allocation
- ✓ Randomization-based counterfactual treatment allocation

We propose the estimation of causal effects in the presence of interference under realistic interventions

Counterfactual treatment allocation

We consider counterfactual treatment allocation that

- ✓ Incorporate covariates as treatment predictors
- ✓ Allow for dependence of treatments within the cluster
- ✓ Intervention takes place at the cluster level

Denote $p_k(\mathbf{W}_k; \mathbf{X}_k, \alpha)$ to be the probability of allocating treatment $\mathbf{W}_k$ to cluster k, when the cluster average propensity of treatment is equal to α

The individual average potential outcome under $w \in \{0, 1\}$ is defined as

$$\overline{Y}_{ki}(w; \mathbf{X}_k, \alpha) = \sum Y_{ki}(W_{ki} = w, \mathbf{W}_{k,-i} = \boldsymbol{w}_{k,-i}) p_k(\mathbf{W}_{k,-i} = \boldsymbol{w}_{k,-i}; W_{ki} = w, \mathbf{X}_k, \alpha)$$

where the summation is over $\mathbf{w}_{k,-i} \in \mathbb{W}(n_k - 1)$

Group average potential outcome: $\overline{Y}_k(w; \mathbf{X}_k, \alpha) = \frac{1}{n_k} \overline{Y}_{ki}(w; \mathbf{X}_k, \alpha)$

Population average potential outcome: $\overline{Y}(w; \mathbf{X}, \alpha) = \frac{1}{K} \overline{Y}_k(w; \mathbf{X}_k, \alpha)$

Direct and indirect effects

If the percentage of units in the cluster that are treated is equal to α, what is the effect of treatment?

- ✓ Individual direct effect: $\overline{DE}_{ki}(\mathbf{X}_k, \alpha) = \overline{Y}_{ki}(0; \mathbf{X}_k, \alpha) - \overline{Y}_{ki}(1; \mathbf{X}_k, \alpha)$
- ✓ Group average direct effect: $\overline{DE}_k(\mathbf{X}_k, \alpha) = \frac{1}{n_k} \sum_{i=1}^{n_k} \overline{DE}_{ki}(\mathbf{X}_k, \alpha)$
- ✓ Population average direct effect: $\overline{DE}(\mathbf{X}, \alpha) = \frac{1}{K} \sum_{k=1}^{K} \overline{DE}_k(\mathbf{X}_k, \alpha)$

Among control units, what is the effect of changing α from α_1 to α_2?

- ✓ Individual indirect effect: $\overline{IE}_{ki}(\alpha_1, \alpha_2; \mathbf{X}_k) = \overline{Y}_{ki}(0; \mathbf{X}_k, \alpha_1) - \overline{Y}_{ki}(0; \mathbf{X}_k, \alpha_2)$
- ✓ Group and population level estimands $\overline{IE}_k(\alpha_1, \alpha_2; \mathbf{X}_k), \overline{IE}(\alpha_1, \alpha_2; \mathbf{X})$ are defined similarly

Allocation average potential outcome and effects

What if we are interested in evaluating the effect of interventions that shift the distribution of observed α from F_α to F'_α

For example, federal regulations could target the increase of state-specific vaccination rates

- ✓ Each state's compliance could be different
- ✓ We cannot know in advance which α each state/city will accept

Allocation average individual potential outcome under F_α

$$\overline{Y}_{ki}(w; \mathbf{X}_k, F_\alpha) = \int \overline{Y}_{ki}(w; \mathbf{X}_k, \alpha) \, \mathrm{d}F_\alpha(\alpha)$$

Allocation average population direct effect

$$\overline{DE}(\mathbf{X}, F_\alpha) = \int \overline{DE}(\mathbf{X}, a) \, \mathrm{d}F\alpha(\alpha)$$

Allocation average population indirect effect

$$\overline{IE}(\mathbf{X}, F_\alpha, F'_\alpha) = \int \overline{Y}(0; \mathbf{X}, \alpha) \, \mathrm{d}F_\alpha(\alpha) - \int \overline{Y}(0; \mathbf{X}, \alpha) \, \mathrm{d}F'_\alpha(\alpha)$$

Assumptions, estimator, and asymptotic results

Positivity: $p(\mathbf{W}_k = \boldsymbol{w}_k \mid \mathbf{X}_k) > \delta_0 > 0$ for all $\boldsymbol{w}_k \in \mathbb{W}(n_k)$

Unconfoudedness: $\mathbf{W}_k \perp\!\!\!\perp \mathbf{Y}_k(\cdot) \mid \mathbf{X}_k$

Estimators

✓ $$\widehat{Y}_k(w; \mathbf{X}_k, \alpha) = \frac{1}{n_k} \sum_{i=1}^{n_k} \frac{p(\mathbf{W}_{k,-i} \mid W_{ki} = w, \mathbf{X}_k, \alpha)}{p(\mathbf{W}_k \mid \mathbf{X}_k)} \mathbb{I}\{W_{ki} = w\} Y_{ki}^{obs}$$

✓ $$\widehat{Y}^K(w, \mathbf{X}, \alpha) = \frac{1}{K} \sum_{k=1}^{K} \widehat{Y}_k(w, \mathbf{X}_k, \alpha)$$

Theorem 1 (Unbiasedness). Unbiased for $\overline{Y}_k(w; \mathbf{X}_k, \alpha)$, $\overline{Y}(w; \mathbf{X}, \alpha)$

Theorem 2 (Consistency). Let F_0 be the distribution of $(\mathbf{Y}_k, \mathbf{X}_k, \mathbf{W}_k)$ in the whole population.

$$\lim_{K \longrightarrow \infty} \widehat{Y}^K(w, \mathbf{X}', \alpha) = \mathbb{E}_{F_0}[\overline{Y}_k(w, \mathbf{X}, \alpha)] \quad a.s. \text{ and so in probability}$$

Theorem 3 (Asymptotic normality). If positivity and unconfoudedness hold, the propensity scores are known or estimated from the correctly specified propensity score model, and outcome and cluster size are bounded, then $\widehat{Y}^K(w, \mathbf{X}, \alpha)$ is asymptotically normal $\rightarrow$ standard errors!

The use of the propensity score

When the observed treatment allocation is not known (most times), $p(\mathbf{W}_k|\mathbf{X}_k)$ needs to be estimated

We model $W_{ki} \sim Bern(p_{ki})$ where

$$\text{logit}(p_{ki}) = b_k + \mathbf{X}_{ki}, \quad b_k \sim N(0, \sigma_b^2),$$

and $\mathbf{X}_{ki}$ includes both individual and cluster level covariates

Then $p(\mathbf{W}_k|\mathbf{X}_k) = \int \left[\prod_{i=1}^{n_k} p(W_{ki}|b_k, X_{ki})\right] f(b_i|\sigma_b^2)\ \mathrm{d}b_i$

The numerator is set equal to $\prod_{j\neq i} p(W_{kj}|X_{kj}, \alpha)$ where b_k has been set equal to b_k^α, for the b_k^α that satisfies $\dfrac{1}{n_k - 1}\sum_{j\neq i} p(W_{kj}|b_k = b_k^\alpha, X_{kj}) = \alpha$

The propensity score has been used to capture the covariate-treatment relationship in the observed (denominator) and counterfactual (numerator) treatment allocation

Estimating treatment and spillover effects in observational social network data using GPSs

(Forastiere, Airoldi, Mealli, 2016)

$\mathcal{N} = (V, E)$: Social Network

- ✓ $i = 1, \ldots, N = |V|$: Node (Unit)
- ✓ $\mathcal{N}_i = \{j \in V : e_{ij} = 1\}$: Neighborhood of unit i
- ✓ $N_i = |\mathcal{N}_i|$: Degree of unit i

$\mathbf{W} \in \{0, 1\}^N$: Treatment Vector $\quad$ $Y_i(\mathbf{W})$: Potential Outcomes

Under SUTVA: $Y_i(W_i, \mathbf{W}_{N_i}, \mathbf{W}_{-N_i}) = Y_i(W_i, \mathbf{W}'_{N_i}, \mathbf{W}'_{-N_i}) \; \forall \, \mathbf{W}_{N_i}, \mathbf{W}'_{N_i}, \mathbf{W}_{-N_i}, \mathbf{W}'_{-N_i}$

SUTVA is untenable in the presence of network data

Neighborhood-Level SUTVA:

$$Y_i(W_i, \mathbf{W}_{N_i}, \mathbf{W}_{-N_i}) = Y_i(W_i, \mathbf{W}_{N_i}, \mathbf{W}'_{-N_i}) \quad \forall \, \mathbf{W}_{-N_i}, \mathbf{W}'_{-N_i}$$

G-Neighborhood-Level SUTVA (SUTNVA). Let $g_i(\cdot)$ be a function $g_i : \{0, 1\}^{N_i} \to \mathcal{G}_i \subset \mathbb{R}$

$$Y_i(W_i, \mathbf{W}_{N_i}, \mathbf{W}_{-N_i}) = Y_i(W_i, \mathbf{W}'_{N_i}, \mathbf{W}'_{-N_i})$$

$\forall \, \mathbf{W}_{-N_i}, \mathbf{W}'_{-N_i}$ and $\forall \, \mathbf{W}_{N_i}, \mathbf{W}'_{N_i} : g_i(\mathbf{W}_{N_i}) = g_i(\mathbf{W}'_{N_i})$

Main Effects and Spillover Effects

A potential outcome $Y_i(w, g)$ is defined only for a subset of nodes $V_g = \{i \in V : g \in \mathcal{G}_i\}$

Main Effect: Average effect of the individual treatment, when the neighborhood treatment is set to g

$$\tau(g) = E\left[\, Y_i(W_i = 1, G_i = g) - Y_i(W_i = 0, G_i = g) | \, i \in V_g \,\right]$$

Overall Main Effect: Average effect of the individual treatment, averaged over the neighborhood treatment distribution

$$\tau = \sum_{g \in \mathcal{G}} \tau(g) P(G_i = g) \qquad \mathcal{G} = \bigcup_i \mathcal{G}_i$$

Spillover Effect: average effect of having the neighborhood treatment at level g versus 0, when the individual treatment is set to z

$$\delta(g; w) = E\left[\, Y_i(W_i = w, G_i = g) - Y_i(W_i = w, G_i = 0) | \, i \in V_g \,\right]$$

Overall Spillover Effect $\Delta(w)$: average of spillover effects $\delta(g; w)$ over the neighborhood treatment distribution

$$\Delta(w) = \sum_{g \in \mathcal{G}} \delta(g; w) P(G_i = g) \qquad \mathcal{G} = \bigcup_i \mathcal{G}_i$$

Joint Treatment and Identifying Assumptions

(Z_i, G_i): Joint Treatment

Observational study: The assignment mechanism

$$Pr(\mathbf{W}, \mathbf{G} | \mathbf{X}, \{\mathbf{Y}(w, g), w = 0, 1; g \in \mathcal{G}\})$$

is unknown and depends on covariates,

$$\mathbf{X} = [\mathbf{X}_i]_i = [\mathbf{X}_i^{ind}, \mathbf{X}_i^{neig}]_i$$

where $\mathbf{X}_i^{ind}$ = Individual characteristics and $\mathbf{X}_i^{neig}$ = Summary of individual characteristics in neighboring units + Neighborhood structure (N_i, Shared friends,...)

Unconfoudedness Assumption of Joint Treatment

$$Y_i(w, g) \perp\!\!\!\perp W_i, G_i \mid \mathbf{X}_i$$

Under SUTNVA and unconfoudedness of the joint treatment an unbiased estimator for the adjusted average of Y_i^{obs}, conditional on the joint treatment

$$\overline{Y}_{w,g,\mathbf{X}}^{obs} := \mathbb{E}_{\mathbf{X}}[\mathbb{E}[Y_i^{obs} \mid W_i = w, G_i = g, \mathbf{X}_i, i \in V_g] \mid W_i = w, G_i = g, i \in V_g]$$

is unbiased for the marginal mean $\mathbb{E}[Y_i(w, g) \mid i \in V_g]$:

$$\overline{Y}_{w,g,\mathbf{X}_i}^{obs} = \mathbb{E}[Y_i(w, g) \mid i \in V_g]$$

Bias for main effects and overall main effects when SUTVA is wrongly assumed

Observed adjusted mean difference

$$\tau^{obs}_{\mathbf{X}^\star} = \sum_{\mathbf{x}\in\mathcal{X}^\star} \mathbb{E}\left[Y_i^{obs} \mid W_i = 1, \mathbf{X}^\star = \mathbf{x}\right] - \mathbb{E}\left[Y_i^{obs} \mid W_i = 0, \mathbf{X}^\star = \mathbf{x}\right] p(\mathbf{X}^\star = \mathbf{x})$$

Under SUTVA and if $Y_i(w) \perp\!\!\!\perp W_i \mid \mathbf{X}_i^\star$, unbiased covariate-adjusted estimators of $\tau^{obs}_{\mathbf{X}^\star}$ are unbiased for $\mathbb{E}[Y_i(1)] - \mathbb{E}[Y_i(0)]$

Theorem 1. If $Y_i(w,g) \perp\!\!\!\perp W_i, G_i \mid \mathbf{X}_i^\star$, then unbiased estimators of $\tau^{obs}_{\mathbf{X}^\star}$ are biased for τ

Corollary 1. If $Y_i(w,g) \perp\!\!\!\perp W_i, G_i \mid \mathbf{X}_i^\star$ and $W_i \perp\!\!\!\perp G_i \mid \mathbf{X}_i^\star$, an unbiased estimator of $\tau^{obs}_{\mathbf{X}^\star}$ is unbiased for τ, even in the presence of interference: $\tau^{obs}_{\mathbf{X}^\star} = \tau$

Corollary 2. If $Y_i(w,g) \perp\!\!\!\perp W_i, G_i \mid \mathbf{X}_i^\star$ but $W_i \not\perp\!\!\!\perp G_i \mid \mathbf{X}_i^\star$, then an unbiased estimator of $\tau^{obs}_{\mathbf{X}^\star}$ is biased for τ

The bias depends on the level of interference and the association between W_i and G_i conditional on $\mathbf{X}_i^\star$

Theorem 2. If $Y_i(w,g) \not\perp\!\!\!\perp W_i, G_i \mid \mathbf{X}_i^\star$, this bias due to interference is combined with the bias due to unmeasured confounders ($U = \mathbf{X}_i \backslash \mathbf{X}_i^\star$)

Joint propensity score

Joint propensity score

$$\psi(w;g;\mathbf{x}) := Pr(W_i = w, G_i = g \mid \mathbf{X}_i = \mathbf{x}) = Pr(G_i = g \mid W_i = w, \mathbf{X}_i^g = \mathbf{x}^g) \times Pr(W_i = w | \mathbf{X}_i^w = \mathbf{x}^w)$$

$\lambda(g;w;\mathbf{x}^g)$ — Neighborhood Propensity Score

$\phi(w;\mathbf{x}^w)$ — Individual Propensity Score

The cardinality of the neighborhood treatment depends of the function g_i

Continuous treatments and the generalized propensity score

Continuous treatment: Let $\mathcal{G} \subseteq \mathbb{R}$ be the set of values for the treatment

Average dose-response function: $\mu(g) = \mathbb{E}[Y_i(g)]$

Weak unconfoundedness: $G_i \perp\!\!\!\perp Y_i(g) \mid \mathbf{X}_i^g \quad$ for all $g \in \mathcal{G}$

Let $\lambda(g, \mathbf{x}) = f_{G|\mathbf{X}^g}(g \mid \mathbf{x}^g)$ be the conditional density of the treatment given the covariates

The GPS for a continuous treatments is $\Lambda_i = \lambda(G_i, \mathbf{X}_i^g)$

Properties of the GPS

- ✓ The GPS is a balancing score: $f_G(g \mid \mathbf{X}_i^g) = f_G(g \mid \mathbf{X}_i^g, \lambda(g, \mathbf{X}_i)) = f_G(g \mid \lambda(g, \mathbf{X}_i^g))$
- ✓ Weak unconfoundedness given the GPS. If the assignment to the treatment is weakly unconfounded given pretreatment variables $\mathbf{X}^g$, then, for every g,

$$f_G(g \mid \lambda(g, \mathbf{X}_i^g), Y_i(g)) = f_G(g \mid \lambda(g, \mathbf{X}_i^g))$$

- ✓ Bias Removal with GPS. If the assignment to the treatment is weakly unconfounded given pretreatment variables $\mathbf{X}$, then,

$$\beta(g, \lambda) = \mathbb{E}\left[Y_i(g) \mid \lambda(g, \mathbf{X}_i^g) = \lambda\right] = \mathbb{E}\left[Y_i^{obs} \mid G_i = g, \Lambda_i = \lambda\right]$$
$$\mu(g) = \mathbb{E}\left[\beta(g, \lambda(g, \mathbf{X}_i^g))\right]$$

(e.g., Hirano and Imbens, 2004; Imai and Van Dyk, 2004; Bia and Mattei, 2008, 2012; Flores et al., 2012; Kluve et al., 2012; Zhao et al., 2013, Bia et al., 2014)

The generalized propensity score

Matching usually unfeasible

GPS allows avoiding to specify a model for the relationship between potential outcomes and covariates

How to use GPS

- ✓ Estimate the GPS, e.g., using a flexible parametric approach: Let $\hat{\Lambda}_i$ be the estimated GPS
- ✓ Estimate the conditional expectation function of Y_i^{obs} given G_i and Λ_i as a flexible function of its two arguments: $\widehat{\mathbb{E}}\left[Y_i^{obs} \mid G_i = g, \Lambda_i = \lambda\right]$
- ✓ Estimate the average dose-response function at treatment level w averaging $\widehat{\mathbb{E}}\left[Y_i^{obs} \mid G_i = g, \hat{\lambda}(g, \mathbf{X}_i^g)\right]$ over $\hat{\lambda}(g, \mathbf{X}_i^g)$

Joint propensity score

Joint propensity score

$$\psi(w;g;\mathbf{x}) := Pr(W_i = w, G_i = g \mid \mathbf{X}_i = \mathbf{x}) = \underbrace{Pr(G_i = g \mid W_i = w, \mathbf{X}_i^g = \mathbf{x}^g)}_{\lambda(g;w;\mathbf{x}^g)} \times \underbrace{Pr(W_i = w | \mathbf{X}_i^w = \mathbf{x}^w)}_{\phi(w;\mathbf{x}^w)}$$

Neighborhood Propensity Score

Individual Propensity Score

The joint propensity score is a balancing score:

$$p(W_i = w, G_i = g \mid \mathbf{X}_i, \psi(w;g;\mathbf{X}_i)) = p(W_i = w, G_i = g \mid \psi(w;g;\mathbf{X}_i))$$

Conditional unconfoundedness of W_i and G_i given the joint / (individual + neighborhood) PS:

If $Y_i(w,g) \perp\!\!\!\perp W_i, G_i \mid \mathbf{X}_i$ then

$$Y_i(w,g) \perp\!\!\!\perp W_i, G_i \mid \psi(w;g;\mathbf{X}_i)$$
$$Y_i(w,g) \perp\!\!\!\perp W_i, G_i \mid \lambda(g;w;\mathbf{X}_i^g), \phi(w;\mathbf{X}_i^w)$$

Propensity Score-Based-Estimator (Subclassification + GPS)

Subclassification on $\phi(1;\mathbf{X}_i^w)$

1. Estimate $\phi(1;\mathbf{X}_i^w)$ (logistic regression for W_i conditional on covariates $\mathbf{X}_i^w$)
2. Predict $\phi(1;\mathbf{X}_i^w)$ for each unit
3. Identify J subclasses B_j, with $j = 1,\ldots,J$, where $\mathbf{X}_i^w \perp\!\!\!\perp W_i \mid i \in B_j$

Within each subclass B_j estimate $\mu_j(w,g) = \mathbb{E}\left[Y_i(w,g) \mid i \in B_j^g\right]$, where $B_j^g = V_g \cap B_j$:

1. Estimate a model for the neighborhood propensity score $\lambda(g;w;\mathbf{X}_i^g)$.
2. Use the observed data ($Y_i^{obs}, W_i, G_i, \mathbf{X}_i^g$ and $\widehat{\Lambda} = \lambda(W_i;G_i;\mathbf{X}_i^g)$) to estimate a model

$$Y_i(w,g) \mid \lambda(w;g;\mathbf{X}_i^g) \sim f(w,g,\lambda(w;g;\mathbf{X}_i^g))$$

3. For each unit $i \in B_j^g$, predict $\lambda(w;g;\mathbf{X}_i^g)$, and use it to predict $Y_i(w,g)$
4. Estimate the dose-response function averaging the conditional potential outcomes over $\lambda(w;g;\mathbf{X}_i^g)$:

$$\widehat{\mu}_j(w,g) = \frac{\sum_{i\in B_j^g} \widehat{Y}_i(w,g)}{|B_j^g|}$$

Derive the average dose-response function

$$\widehat{\mu}(w,g) = \sum_{j=1}^{J} \widehat{\mu}_j(w,g)\pi_j^g \qquad \pi_j^g = \frac{\sum_{i\in V_g} 1(\phi(1;\mathbf{X}_i^z) \in B_j)}{v_g}$$

Standard errors and confidence intervals are derived using bootstrap methods

Some concluding remarks

Propensity scores are powerful tools

Must however be used with care

Underlying assumptions are crucial

They determine how the propensity score should be specified and estimated

References

Achy-Brou A.C., Frangakis C.E. and Griswold M. (2010). Estimating treatment effects of longitudinal designs using regression models on propensity scores. Biometrics 66, 824-833.

Arpino B. and Mealli F. (2011). The specification of the propensity score in multilevel observational studies. Computational Statistics and Data Analysis 55, 1770-1780

Aussems M.E. (2008). Multilevel data and propensity scores: an application to a virtual Y after-school program. Mimeo.

Bia M. and Mattei A. (2008). A STATA Package for the estimation of the dose-response function through adjustment for the generalized propensity score. The STATA Journal 8, 354-373.

Bia M. and Mattei A. (2012). Assessing the effect of the amount of financial aids to Piedmont firms using the generalized propensity score. Statistical Methods and Applications, 21, 485-516.

Bia M., Flores A. C., Flores-Lagunes A., Mattei A. (2014). A Stata package for the application of semiparametric estimators of dose-response functions. The STATA Journal 14, 580-604.

Cattaneo M. (2010). Efficient semiparametric estimation of multi-valued treatment effects under ignorability. Journal of Econometrics 155, 138-154.

Flores C. A., Flores-Lagunes A. and Neumann T. (2012). Estimating the effects of length of exposure to instruction in a training program: The case of Job Corps. The Review of Economics and Statistics 94, 153-171.

Forastiere L., Airoldi E. M., and Mealli F. (2016). Identification and estimation of treatment and interference effects in observational studies on networks. Arxiv working paper (http://arxiv.org/abs/1609.06245)

Hirano K. and Imbens G.W. (2004). The propensity score with continuous treatments. In Applied Bayesian Modelling and Causal Inference from Missing Data Perspectives.

Imai K. and Van Dyk D. (2004). Causal inference with general treatment regimes: generalizing the propensity score. JASA 99, 854-866.

References

Imai K. and Ratkovic M. (2014). Covariate balancing propensity score. JRSS-B 76, 243-263.

Imai K. and Ratkovic M. (2015). Robust Estimation of Inverse Probability Weights for Marginal Structural Models. JASA 110, 1013-1023.

Imbens G.W. (2000). The role of the propensity score in estimating dose-response functions. Biometrika 87, 706-710.

Kim J. and Seltzer M. (2007). Causal inference in multilevel settings in which selection process vary across schools. Working Paper 708. Center for the Study of Evaluation (CSE), Los Angeles.

Keele L.J. and Zubizarreta J. (2017). Optimal Multilevel Matching in Clustered Observational Studies: A Case Study of the School Voucher System in Chile. JASA. Forthcoming.

Kluve J., Schneider H., Uhlendorff A. and Zhao Z. (2012) Evaluating continuous training programmes by using the generalized propensity score. JRSS-A 175, 587-617

Lechner M. (2001) Identification and estimation of causal effects of multiple treatments under the conditional independence assumption. Econometric Evaluations of Active Labor Market Policies in Europe, 43-58

Li F., Zaslavsky A.M and Landrum M.B. (2013). Propensity score weighting with multilevel data. Statistics in Medicine 32, 3373-3387.

Mattei A. and Mealli F. (2015) Commentary on “On Bayesian estimation of marginal structural models” by Saarela O., Stephens D. A., Moodie E. E. M., Klein M. B. Biometrics 71, 293-296.

Papadogeorgou G., Zigler C. and Mealli F. (2017). Inverse probability weighted estimators under partial interference. (Work in progress)

Pimentel S., Page L., Matthew L., and Lindsay and Keele, L. J. (2015). Optimal Multilevel Matching Using Network Flows: An Application to a Summer Reading Intervention. Mimeo (http://lukekeele.com/wp-content/uploads/2016/03/myon.pdf)

Robins J. M., Hernan M. A. and Brumback B. (2000). Marginal structural models and causal inference in epidemiology. Epidemiology 11, 550-560.

Su, Y., 2008. Causal inference of repeated observations: a synthesis of matching method and multilevel modeling. Paper Presented at the Annual Meeting of the APSA 2008, Boston, Massachusetts.

Yang S., Imbens G.W., Cui Z., Faries D. E. and Kadziola Z. (2016). Propensity score matching and subclassification in observational studies with multi-level treatments. Biometrics 72,1055-1065.

Zhao S., Van Dyk D. and Imai K. (2013) Propensity-score based methods for causal inference in observational studies with fixed non-binary treatments. Mimeo (http://imai.princeton.edu/research/files/gpscore.pdf)