Primal-dual Covariate Balance and Minimal Double Robustness via Entropy Balancing

Qingyuan Zhao

(Joint work with Daniel Percival)

Department of Statistics, Stanford University

JSM, August 9, 2015

Outline

1 Background

2 Results for EB

3 Equivalence of PS and OR

Setting

Rubin's causal model

Consider an observational study:

- Treatment assignment: $T \in \{0, 1\}$;
- Potential outcomes: $Y(0)$, $Y(1)$;
- Pre-treatment covariates: X;
- No hidden bias: $(Y(0), Y(1)) \perp\!\!\!\perp T | X$.
- Overlap: $0 < \mathrm{P}(T = 1|X) = e(X) < 1$.

Covariate balance and propensity score

Covariate balance plays a crucial rule in observational study:

$$\mathrm{E}[c(X)|T=1] = \mathrm{E}\left[\frac{e(X)}{1-e(X)}c(X)\middle|T=0\right], \ \forall\ c(X).$$

Rosenbaum and Rubin (1983): any balancing score is a function of propensity score (PS).

Covariate balance and propensity score

Covariate balance plays a crucial rule in observational study:

$$\mathrm{E}[c(X)|T=1] = \mathrm{E}\left[\frac{e(X)}{1-e(X)}c(X)\middle|T=0\right], \ \forall\ c(X).$$

Rosenbaum and Rubin (1983): any balancing score is a function of propensity score (PS).
In practice, PS model is subject to misspecification.

Propensity score tautology (Imai et al., 2008)

Iterate between

1. Modeling propensity score
2. Checking covariate balance.

Entropy Balancing (Hainmueller, 2011)

Entropy balancing (EB) is a one-step solution of the tautology.

$$\begin{aligned} \underset{w}{\text{maximize}} \quad & -\sum_{T_i=0} w_i \log w_i \\ \text{subject to} \quad & \sum_{T_i=0} w_i c_j(X_i) = \bar{c}_j(1) = \frac{1}{n_1} \sum_{T_i=1} c_j(X_i),\ j = 1, \ldots, p, \\ & \sum_{T_i=0} w_i = 1, \\ & w_i > 0,\ i = 1, \ldots, n. \end{aligned}$$

EB estimates the average treatment effect on the treated

$$\gamma = \mathrm{E}[Y(1)|T = 1] - \mathrm{E}[Y(0)|T = 1]$$

by

$$\hat{\gamma}^{\mathrm{EB}} = \sum_{T_i=1} \frac{Y_i}{n_1} - \sum_{T_i=0} w_i^{\mathrm{EB}} Y_i.$$

This talk

EB was proposed purely from an applied perspective and is very easy to interpret, but is it actually safe to use EB?

We give theoretical justifications for entropy balancing:

- EB has a "minimal" double robustness property.
- Elegant correspondence between primal-dual optimization and double robustness.

Outline

1 Background

2 Results for EB

3 Equivalence of PS and OR

Heuristics

Let $m(x)$ be the density of X for the control population.

Minimum relative entropy principle

Estimate the density of the treatment population by

$$\underset{\tilde{m}}{\text{maximize}}\ H(\tilde{m}\|m) \quad \text{s.t.} \quad \mathrm{E}_{\tilde{m}}[c(X)] = \bar{c}(1). \tag{1}$$

where $H(\tilde{m}\|m) = \mathrm{E}_{\tilde{m}}[\log(\tilde{m}(X)/m(X))]$ is the relative entropy.

Heuristics

Let $m(x)$ be the density of X for the control population.

Minimum relative entropy principle

Estimate the density of the treatment population by

$$\underset{\tilde{m}}{\text{maximize}}\ H(\tilde{m}\|m) \quad \text{s.t.} \quad \mathrm{E}_{\tilde{m}}[c(X)] = \bar{c}(1). \tag{1}$$

where $H(\tilde{m}\|m) = \mathrm{E}_{\tilde{m}}[\log(\tilde{m}(X)/m(X))]$ is the relative entropy.

The optimization (1) is equivalent to

$$\underset{w}{\text{maximize}}\ \mathrm{E}_m[w(X)\log w(X)] \quad \text{s.t.} \quad \mathrm{E}_m[w(X)c(X)] = \bar{c}(1).$$

EB is the finite sample version of this problem.

Exponential tilting

The solution to (1) belongs to the family of exponential titled distributions of m (Cover and Thomas, 2012):

$$m_\theta(x) = m(x) \exp(\theta^T c(x) - \psi(\theta)).$$

By Bayes' formula, this implies a logistic PS model

$$\frac{\mathrm{P}(T=1|X=x)}{\mathrm{P}(T=0|X=x)} = w(x) = \exp(\alpha + \theta^T c(x))$$

Intuitively, EB solves the logistic regression by a criterion different than the MLE.

"Minimal" double robustness

Theorem (Zhao and Percival, 2015)

Assume there is no hidden bias, the expectation of $c(X)$ exists and $\mathrm{Var}(Y(0)) < \infty$. Let $e(X) = \mathrm{P}(T = 1|X)$ and $g_t(X) = \mathrm{E}[Y(t)|X]$. Then

1. If $\mathrm{logit}(e(X))$ or $g_0(X)$ is linear in $c_j(X)$, $j = 1, \ldots, p$, then $\hat{\gamma}^{\mathrm{EB}}$ is statistically consistent.
2. Moreover, if $\mathrm{logit}(e(X))$, $g_0(X)$ and $g_1(X)$ are all linear in $c_j(X)$, $j = 1, \ldots, p$, then $\hat{\gamma}^{\mathrm{EB}}$ reaches the semiparametric variance bound of γ derived in Hahn (1998).

Proof: outcome regression $\longleftrightarrow$ primal problem

If the true OR model is linear: $Y_i(0) = \sum_{j=1}^{p} \beta_j c_j(X_i) + \epsilon_i$, then

$$\sum_{T_i=0} w_i Y_i - \mathrm{E}[Y(0)|T=1]$$
$$= \sum_{j=1}^{p} \beta_j \left[\sum_{T_i=0} w_i c_j(X_i) - \mathrm{E}[c_j(X)|T=1] \right] + \sum_{T_i=0} w_i \epsilon_i.$$

In the primal problem of EB, moment balancing constraints:

$$\sum_{i=1}^{n} w_i c_j(X_i) = \frac{1}{n_1} \sum_{T_i=1} c_j(X_i).$$

Proof: propensity scoring $\longleftrightarrow$ dual problem

The dual problem of EB is

$$\underset{\theta}{\text{minimize}} \quad \log\left(\sum_{T_i=0} \exp\left(\sum_{j=1}^{p} \theta_j c_j(X_i)\right)\right) - \sum_{j=1}^{p} \theta_j \bar{c}_j(1),$$

Intuitively, EB uses "exponential loss" instead of logistic loss. Consistency under logistic PS model can be rigorously proved by M-estimation theory.

Asymptotic efficiency of EB

A natural competitor is the inverse probability weighting estimator (PS model: logistic regression solved by MLE).

When the logistic PS model is correctly specified, Theorem 3 in our paper provides formulas for the asymptotic variance.

- When $Y(0)$ is correlated with $c(X)$, EB is more efficient than MLE
- When the true OR model is linear in $c(X)$, EB reaches the semiparametric variance bound.

Conclusion: EB should be preferred over IPW+MLE.

Outline

1 Background

2 Results for EB

3 Equivalence of PS and OR

Balancing PS weights $\longrightarrow$ OR model

Doubly robustify an OR estimator: given an OR model $\hat{g}_0(X)$,

$$\hat{\gamma}^{\mathrm{EB\text{-}DR}} = \sum_{T_i=1} \frac{1}{n_1}(Y_i - \hat{g}_0(X_i)) - \sum_{T_i=0} w_i^{\mathrm{EB}}(Y_i - \hat{g}_0(X_i)).$$

Theorem (the role of balancing PS weights)

If the fitted OR is $\hat{g}_0(X) = \sum_{j=1}^{p} \hat{\beta}_j c_j(X)$, whether or not this model is correctly specified, $\hat{\gamma}^{\mathrm{EB-DR}} = \hat{\gamma}^{\mathrm{EB}}$.

OR model $\longrightarrow$ balancing weights

Let X^t denote the matrix $X^t_{ij} = c_j(X_i)$ and Y^t denote the vector of outcomes for i in the group $t = 0$ or 1

For linear OR model $\mathrm{E}[Y(0)|X] = \sum_{j=1}^{p} \beta_j c_j(X)$, the OLS estimator of $\mathrm{E}[Y(0)|T = 1]$ is

$$\frac{1}{n_1} 1^T (X^1 \hat{\beta}) = \frac{1}{n_1} 1^T \left\{ X^1 [(X^0)^T X^0]^{-1} (X^0)^T \right\} Y^0.$$

This is a weighted average of Y^0!
Moreover, they are balancing weights:

$$\frac{1}{n_1} 1^T \left\{ X^1 [(X^0)^T X^0]^{-1} (X^0)^T \right\} X^0 = \frac{1}{n_1} 1^T X^1.$$

The role of covariate balance

Our analysis of EB reveals an interesting equivalence between PS and OR.

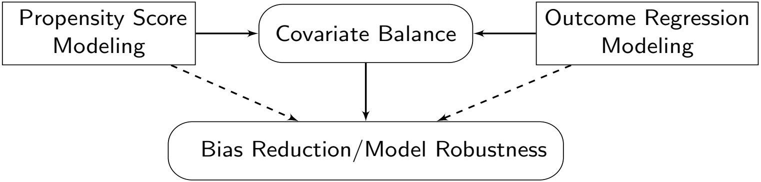

Figure : Dashed arrows: conventional understanding of double robustness. Solid arrows: our understanding of double robustness revealed by entropy balancing.

Thank you

References

Cover, T. M. and J. A. Thomas (2012). *Elements of information theory*. John Wiley & Sons.

Hahn, J. (1998). On the role of the propensity score in efficient semiparametric estimation of average treatment effects. *Econometrica 66*(2), 315–332.

Hainmueller, J. (2011). Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies. *Political Analysis*, mpr025.

Imai, K., G. King, and E. A. Stuart (2008). Misunderstandings between experimentalists and observationalists about causal inference. *Journal of the Royal Statistical Society: Series A (Statistics in Society) 171*(2), 481–502.

Rosenbaum, P. and D. Rubin (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika 70*, 41–55.

Zhao, Q. and D. Percival (2015). Primal-dual Covariate Balance and Minimal Double Robustness via Entropy Balancing. *ArXiv e-prints*.