Theoretical Mathematics & Applications, vol.2, no.3, 2012, 75-86
ISSN: 1792-9687 (print), 1792-9709 (online)
Scienpress Ltd, 2012

Strategy of Bayesian Propensity Score Estimation Approach in Observational Study

O. Shojaee[1], M. Babanezhad*[2], N. Jalali[3] and F. Yaghmaei[4]

Abstract

Estimating causal effect of a treatment on an outcome is often complicated. This is because the treatment effect may be deviated by the confounding variables. These variables affect treatment and outcome simultaneously, and the causal effect estimation thus depends upon these variables. Several methods have been proposed to reduce the attribute bias of confounding effect. In this article, we compare the traditional method of Propensity score through Stratification; and recent method of Propensity score through Bayesian in observational studies. Comparison is constructed on Mont Carlo simulation of the hypothetical binary treatment.

Mathematics Subject Classification: 68T05, 62B10

[1] Department of Statistics, Faculty of Sciences, Golestan University, Gorgan, Golestan, Iran
[2] Department of Statistics, Faculty of Sciences, Golestan University, Gorgan, Golestan, Iran. *Corresponding author, e-mail: m.babanezhad@gu.ac.ir.
[3] Department of Statistics, Faculty of Sciences, Golestan University, Gorgan, Golestan, Iran
[4] Department of Statistics, Faculty of Sciences, Golestan University, Gorgan, Golestan, Iran

Article Info: *Received* : August 12, 2012. *Revised* : October 23, 2012
Published online : December 30, 2012

Keywords: Causal effect, Observational studies, Confounding variables, Stratification propensity scores, Bayesian propensity scores

1 Introduction

In many fields of researchers are faced with the problem of estimating causal effects from non-experimental data. This interest of using study in fact increases gradually in recent decades [1, 2, 3]. This might be because the lack of time or consuming money to perform Randomized study [3, 4, 5]. One of the ubiquitous issues in observation studies is the impact of confounding variables [6, 7]. The simultaneous effect of these variables on treatment and outcome make causal effect estimation invalid. Propensity score is one method to decline the effect of measured confounding variables when we confront with a binary outcome in observational studies [7, 8]. This method allocates a value which is called score to each person in a population. Then the causal effect is estimated based on new population which is stratified through propensity scores [9]. Recently, an alternative method is designed which combines propensity score with Bayesian techniques [7, 8]. In this method, propensity score as a latent variable is integrated from the posterior distribution for the treatment effect. Bayesian propensity score adjustment fits regression models for treatment and outcome simultaneously rather than one at a time. When propensity score is estimated, Bayesian propensity score adjustment incorporates prior information about relationship between outcome and propensity score within treatment groups [9, 10]. In contrast, standard analytic methods estimate propensity scores from the marginal model for treatment given measured confounding variables [9, 11].

In this paper, we use Stratification and Bayesian propensity score adjustment based on propensity score method to estimate the causal effect of a treatment. In Section 2, we review the content of potential outcomes which formalizes the notion of causal effect. In addition, the propensity score method is surveyed in this Section. In Section 3 Bayesian propensity score adjustment is explained. We report on extensive comparative simulations in Section 4. Conclusion is summarized in Section 5.

2 Strategy of Causal Effect Estimation

2.1 Potential outcomes

Without loss of generality, we throughout consider X is a binary treatment ($X = 1$ treated, $X = 0$ control); and C is a vector of measured covariates prior to receipt of treatment (baseline) or, if measured post treatment, not affected by either treatment. Each individual is assumed to have an associated random vector $(Y_0; Y_1)$, where Y_0 and Y_1 are the values for outcome that would be observed if, possibly contrary to the fact of what actually happened, she/he were to receive control or treatment, respectively. Consequently, Y_0 and Y_1 are referred to as potential outcomes and may be viewed as inherent characteristics of an individual. We can formalize the observed outcome Y which would be observed under the exposure actually received, as follow;

$$Y = XY_1 + (1 - X)Y_0.$$

Thus (Y, X, C) are our observed variables for each individual. It is important to distinguish between the observed outcome Y and potential outcomes Y_0 and Y_1. The latter are hypothetical and may never be observed simultaneously. However, they are a convenient construct allowing precise statement of questions of interest, as we now describe. The distributions of Y_0 and Y_1 may be thought as representing the hypothetical distributions of outcome for the population of individuals were all individuals receive control or treatment, respectively. So the means of these distributions correspond to the mean outcome if all individuals were to receive each treatment. Hence, a difference in these means would be attributable to, or caused by, the treatments. Formally, we would have

$$\psi = \mu_1 - \mu_0 = E(Y_1) - E(Y_0)$$

and we define no unmeasured confounding variable assumption [9, 10]

$$(Y_1, Y_0) \perp X | C \tag{1}$$

where $A \perp B | D$ means A is independent from B given D. This assumption states if we can measure the confounding variables then potential outcomes are independent from treatment. That means there are no more confounding variables which affect on treatment causal effect.

2.2 Propensity score

The propensity score $e(C) = P(X = 1|C)$; $0 < e(C) < 1$, is the probability of treatment given the observed covariates, in which $C \perp X|e(C)$. Therefore individuals from either treatment group with the same propensity score are balanced; in distribution of C regardless of treatment status.

If (1) holds, then $(Y_1, Y_0) \perp X|e(C)$ [9], that is, treatment X is independent from the potential outcomes for individuals whom sharing the same propensity score. In practice, the propensity score is unlikely to be known, so it is common to estimate it from observed data (X_i, C_i), $i = 1, ..., n$ by assuming that $e(C)$ follows a parametric model; e.g., a logistic regression model [5, 6], $e(C, \beta) = \left\{1 + \exp(-C^T\beta)\right\}^{-1}$, where β is a $p \times 1$ vector parameter. Interaction and higher-order terms may also be included. Here, β may be estimated by the maximum likelihood (ML) estimator $\hat{\beta}$, by solving the following expression [5, 6]

$$\sum_{i=1}^{n} \psi_\beta(X_i, C_i, \beta) = \sum_{i=1}^{n} \frac{X_i - e(C_i, \beta)}{e(C_i, \beta)(1 - e(C_i, \beta))} \times \frac{\partial}{\partial \beta} e(C_i, \beta) = 0. \qquad (2)$$

We assume that the analyst is proficient at modeling $e(C, \beta)$, so that it is correctly specified.

2.3 Estimation of ψ based on stratification

The popular approach using stratification on estimated propensity scores to estimate ψ involves the following steps [5]:

(i) Estimate β as in (2) and calculate estimated propensity scores $\hat{e}_i = e(C_i, \hat{\beta})$ for all i.

(ii) Form K strata according to the sample quantiles of the $\hat{e}_i$, where the j^{th} sample quantile $\hat{q}_j$, $j = 1, ..., K$, is such that the proportion of $\hat{e}_i \leq \hat{q}_j$ is roughly j/k, $\hat{q}_0 = 0$ and $\hat{q}_1 = 1$.

(iii) Within each stratum, calculate the difference of sample means of the Y_i for each treatment.

(iv) Estimate ψ by a weighted sum of the differences of sample means across strata, where weighting is by the proportion of observations falling in each stratum.

Defining $\hat{Q}_j = [\hat{q}_{j-1}, \hat{q}_j]$; and $n_j = \sum_{i=1}^{n} I(\hat{e}_i \in \hat{Q}_j)$, the number of individuals in stratum j; and $n_1 j = \sum_{i=1}^{n} X_i I(\hat{e}_i \in \widehat{Q}_j)$ is the number of these who are treated, the estimator using a weighted sum is

$$\hat{\psi}_S = \sum_{i=1}^{n} (\frac{n_j}{n}) \left\{ n_{1j}^{-1} \sum_{i=1}^{n} X_i Y_i I(\hat{e}_i \in \hat{Q}_j) - (n_j - n_1 j)^{-1} \sum_{i=1}^{n} (1 - X_i) Y_i I(\hat{e}_i \in \hat{Q}_j) \right\} \quad (3)$$

As the weights $n_j/n \approx k^{-1}$, they may be replaced by k^{-1} to yield an average across strata.

3 Bayesian Propensity Score Adjustment (BPSA)

Suppose that (Y_i, X_i, C_i) for $i = 1, ..., n$ are identically distributed observations drawn from A population with probability density function $P(Y, X, C)$. Based on the Bayesian propensity score analysis, we model the conditional density $P(Y, X|C)$ using a pair of logistic regression models [8]:

$$\text{logit}\left[p(Y = 1|X, C)\right] = \psi X + \xi^T g\left\{z(C, \gamma)\right\} \quad (4)$$

$$\text{logit}[p(X = 1|C)] = \gamma^T C \quad (5)$$

where $z(C, \gamma)$ is the propensity score.

Equation (5) models the probability of treatment, which depends on the measured confounding variables via the regression coefficients $\gamma = (\gamma_0, ..., \gamma_p)$. Equation (4) is a model for outcome and includes causal effect parameter ψ.

To ease modeling of regression intercept terms, we set the first component of C equal to one, so C is a $(p+1)\times 1$ vector. Accordingly, equation (4) includes the quantity $Z(C, \gamma)$ as a covariate in a regression model for the outcome via a linear predictor $g\{0\}$. We let $g\{a\} = \sum_{j=0}^{l} \xi_j g_j\{a\}$ where the quantities $g_j\{0\}$ are natural cubic spline basis functions with $l = 3$ knots (q_1, q_2, q_3) and regression coefficients $\xi = (\xi_0, \xi_1, \xi_2, \xi_3)$. The choice of $l = 3$ knots reflects a tradeoff between smoothness and complexity. To go further, the following assumptions should be satisfied;

1. No unmeasured confounding variables;

2. Propensity score model correctly specified.

We assign the prior distributions for the model parameters ψ, γ, ξ as follows;

$$\psi \sim N(0, \sigma_\psi^2)$$

$$\gamma_0, ..., \gamma_p \sim N(0, \sigma_\gamma^2)$$

$$\xi_0, ..., \xi_3 \sim N(0, \sigma_\xi^2)$$

where $\sigma_\psi^2 = \sigma_\gamma^2 = \sigma_\xi^2 = \{log(15)/2\}^2$. The value for σ_ψ^2 states that the odds ratio for the treatment effect is not overly large and lies between 1/15 and 15 with probability 0.95. The values for σ_γ^2 and σ_ξ^2 make similar modeling assumptions about the prior magnitude for the association between Y and Z given X, and also the association between C and X. The regression models in equations (4) and (5) give a likelihood function for the data. Combining the likelihood and prior distributions, we can sample from the posterior distribution of the treatment effect ψ and nuisance parameters γ, ξ; using Markov chain Monte Carlo.

Conceptually, the implementation involves two steps iterative procedures;

1. Impute the propensity score parameter,

2. Fit a complete data step to estimate the treatment effect ψ and ξ given the propensity scores.

Successive iterations average over uncertainty in the propensity scores. The approach has close connections to the EM algorithm and multiple imputations. Before applying BPSA to the data, we first select the knots used in the spline regression for the relationship between mortality and propensity score. To choose the knots, we fit the logistic regression model given in equation (5) via ML and compute the fitted values. We then apply BPSA to the data by sampling from the posterior density $P(\psi, \xi, \gamma|data)$. Note that here the posterior expected propensity score is used

$$\hat{Z}_i^{BPSA} = E\left[Z(C_i, \gamma)\right]_{\gamma|data} = \int \text{expit}(\gamma^T C_i) p(\gamma|data)$$

and compute from the posterior samples directly.

4 Simulation Design

To investigate the efficiency of two explained method we perform a hypothetical study through simulation. We consider the case where there are four measured confounding variables. We simulate matrix C, where the each columns of C are independent and identically distributed as $N(0,1)$. Note that the first component of C is equal to one. To generate X we use logistic regression model in equation (5). Also a binary outcome is generated from as follows;

$$logit\left[P(Y=1|X,C)\right] = \psi X + \xi^T C$$

where ψ is equal to 1. The results are shown in Table 1. It should be noted that $\exp(\hat{\psi})$ is reported in Table 1. By comparison the estimated value of causal effect, we can figure out that Bayesian propensity score could result in better estimation than the Stratification propensity score. To compare the

Table 1: Odds ratios for association between Y and X given C to compare Bayesian and Stratification propensity score methods when $\psi = 1$

$\exp(\hat{\psi}_B) = 2.49$		$\exp(\hat{\psi}_S) = 1.05$	
$\exp(\xi_1)$	1.80	$\exp(\beta_1)$	1.16
$\exp(\xi_2)$	2.68	$\exp(\beta_2)$	1.16
$\exp(\xi_3)$	2.94	$\exp(\beta_3)$	1.07
$\exp(\xi_4)$	5.75	$\exp(\beta_4)$	1.89

Bayesian propensity score and the Stratification propensity score, we choose; $\exp(\psi) = 1, 1.5, 2, 2.5, 3, 3.5$. We plot of ψ versus mean square error (MSE) of both estimator [10, 11, 12]. In fact the mean squared error (MSE) of two estimators is way to quantify the difference between values implied by an estimator and the true values of the quantity being estimated. Whether an estimator of a causal effect is consistent estimator of the causal effect of the treatment on the outcome of interest [12, 13, 14].

The MSE of Bayesian and stratification propensity score method have the same fluctuation when the $\exp(\hat{\psi})$ is over 5. Whereas the MSE of stratification propensity score method tended to increase until $\exp(\hat{\psi})$ gain nearly 5.

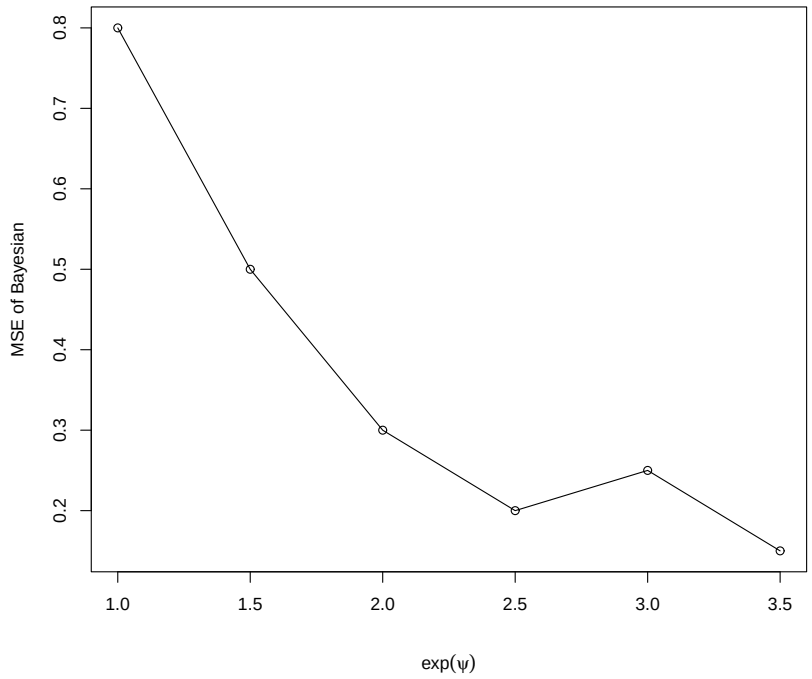

Figure 1: exp(ψ) versus the MSE of Bayesian propensity score method

5 Conclusion

The control of confounding is essential in many statistical problems, especially in those that attempt to estimate exposure effects. The objective of this study was to compare two estimations based on Propensity score method. To challenge of this issue, several methods of adjustment are available whether confounders are measured or unmeasured. One of which is to adjusting effect estimates for unmeasured confounding with validation data using propensity score calibration [14, 15]. One case is to use secondary sample in addition to the primary sample in which though having no direct information about the treatment effect contains information about the confounding factors [16, 17]. Our case-study reveals that treatment is associated with all components covariate C, and this association persists even after adjustment for four considered measured confounders.

As the simulation shows, the Bayesian estimation of the causal effect can present better estimate, whereas the Stratification propensity score method

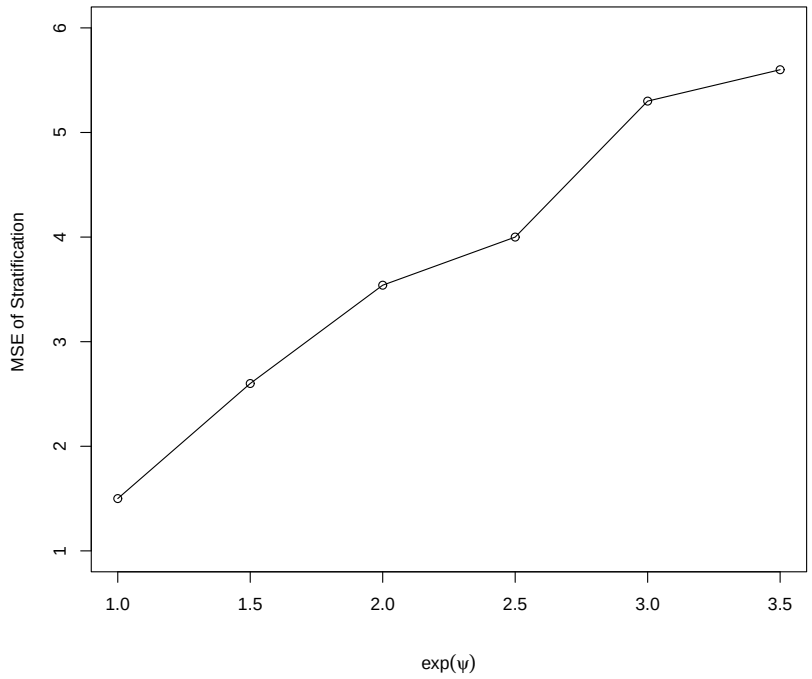

Figure 2: $\exp(\psi)$ versus the MSE of Stratification propensity score method

has large bias in comparison with the Bayesian propensity score method. Since both estimations have bias we prefer to consider the MSE of both estimators. As plots shown, the MSE of each estimator increases as the $\exp(\psi)$ grows. But the fluctuations in first values of $\exp(\psi)$ are related to definition of g function in equation 5. After estimating Propensity score through equation 4 and using these values in g function, we are expected to have increase sinusoidal wave. The similar trend exists for the Stratification propensity score method. If the bias estimators are large in magnitude, meaning that C contains powerful confounders, then the variability of $P(Y|X, C)$ and $P(X|C)$ will increase. To capture this variability, one extension would be to model uncertainty in the quantity in equation using a Bayesian bootstrap [13, 17].

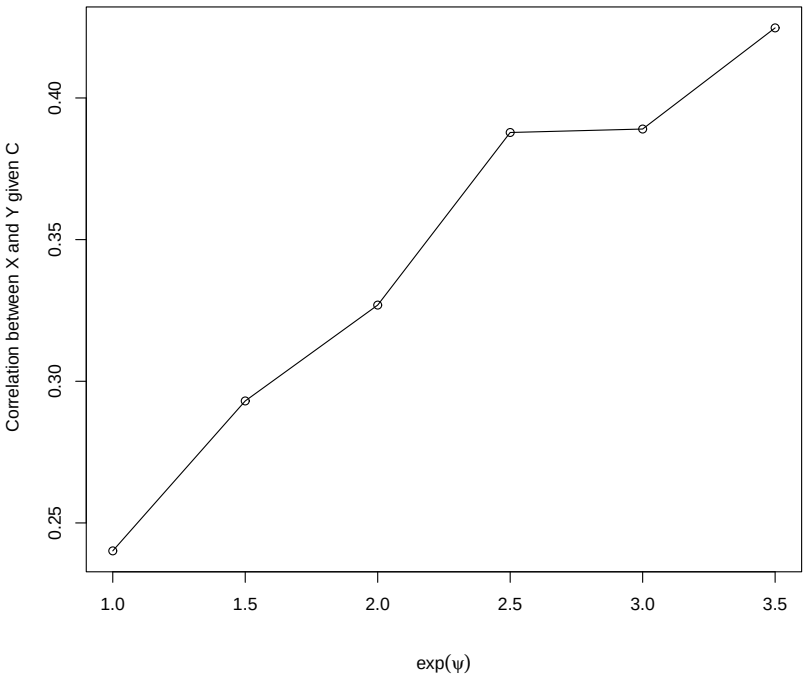

Figure 3: $\exp(\psi)$ versus the correlation between X and Y given C

References

[1] P.W. Holland, *Causal Inference, Path Analysis, and Recursive Structural Equations Models*, (with discussion) in Clogg CC ed., Sociological Methodology, 1998.

[2] M.A. Hernán and J.M. Robins, Estimating Causal Effect from Epidemiological Data, *J epidemiology Community Health*, **60**, (2006), 578–586.

[3] C. McCandless, Lawrence, et al, Propensity Score Adjustment for Unmeasured Confounding in Observational Studies, *ESRC National Center for Research Methods NCRM Working Paper Series*, **2**, (2008).

[4] D.Y. Lin, B.M. Psaty and R.A. Kronmal, Assessing the Sensitivity of Regression Results to Unmeasured Confounders in Observational Studies, *Biometrics*, **54**, (1998), 948-963.

[5] J.K. Lunceford and M. Davidian, Stratification and Weighting via the Propensity Score in Estimation of Causal Treatment Effects: A Comparative Study, *Stat. Med.*, **23**, (2004), 2937–2960.

[6] C. McCandless, P. Gustafson and P.C. Austin, Bayesian Propensity Score Analysis for Observational Data, (Under review).

[7] C. McCandless, P. Gustafson and A.R. Levy, Bayesian Sensitivity Analysis for Unmeasured Confounding in Observational Studies for unbalanced repeated measures, *Stat. Med.*, **26**, (2007), 2331–2347.

[8] C. McCandless, P. Gustafson and P.C. Austin, Covariate Balance in a Bayesian Propensity Score Analysis of Beta Blocker Therapy in Heart Failure Patients, *Epidemiologic Perspectives and Innovations*, **6**(5, (2009).

[9] P.R. Rosenbaum and D.B. Rubin, The Central Role of the Propensity Score in Observational Studies for Causal Effects, *Biometrika*, **70**, (1983), 41–57.

[10] P.R. Rosenbaum and D.B. Rubin, Assessing Sensitivity to an Unobserved Binary Covariates an Observational Study with Binary Outcome, *J. R. Stat. Soc. Ser. B*, **45**, (1983), 212–218.

[11] P.R. Rosenbaum, Propensity Score. In Encyclopedia of Biostatistics, *Armitage P., Colton T. (eds)*, **5**, Wiley, New York, (1998), 3551–3555.

[12] S. Greenland, Multiple bias modeling for analysis of observational data (with discussion), *J. R. Stat. Soc. Ser. A*, **168**, (2005), 267–306.

[13] S. Schneeweiss, R.J. Glynn, E.H. Tsai, J. Avorn and D.H. Solomon, Adjusting for Unmeasured Confounders in Pharmacoepidemiologic Claims Data Using External Information: The Example of Cox2 Inhibitors and Myocardial Infarction, *Epidemiol*, **16**, (2005), 17–24.

[14] S. Schneeweiss, Sensitivity Analysis and External Adjustment for Unmeasured Confounders in Epidemiologic Database Studies of Therapeutics, *Pharmacoepidemiol Drug Saf*, **15**, (2006), 291–303.

[15] T. Sturmer, S. Schneeweiss, J. Avorn and R.J. Glynn, Adjusting Effect Estimates for Unmeasured Confounding with Validation Data using Propensity Score Calibration, *Am. J. Epidemiol*, **162**, (2005), 279–289.

[16] L. Yin, R. Sundberg, X. Wang and D.B. Rubin, Control of Confounding through Secondary Samples, *Stat. Med.*, **25**, (2006), 3814–3825.

[17] T.J. Vanderweele, The Sign of the Bias of Unmeasured Confounding, *Biometrics*, **64**(3), (2008), 702–706.