Abstract Title Page
Not included in page count.

Title: Bayesian Propensity Score Analysis: Simulation and Case Study

Author(s): David Kaplan & Cassie J. S. Chen
Department of Educational Psychology
University of Wisconsin – Madison

Abstract Body

Limit 5 pages single spaced.

Background / Context:

Description of prior research and its intellectual context.

Rubin (1985) provides a justification for why an applied Bayesian should be interested in propensity scores, his analysis does not address the actual estimation of the propensity score equation or the subsequent causal equation from a Bayesian perspective. In a more recent paper, Hoshino (2008) argued that propensity score analysis has focused mostly on estimating the marginal treatment effect and that more complex methods are needed to handle more realistic problems. In response, Hoshino (2008) developed a quasi-Bayesian estimation method that can be used to handle more general problems { and in particular, latent variable models. More recently, McCandless, Gustafson, and Austin (2009) argued that the failure to account for uncertainty in the propensity score can result in falsely precise estimates of treatment effects. However, adopting the Bayesian perspective that data and parameters are random, uncertainty in model parameters of the propensity score equation can lead to more accurate estimates of treatment effects. Moreover, it may be possible in many circumstances to elicit priors on the covariates from previous research or expert opinion and, as such, have a means of comparing di_erent propensity score models for the same problem and resolve model choice via Bayesian model selection methods described earlier.

The paper by McCandless et al. (2009) provides an approach to Bayesian propensity score analysis for observational data. Their approach involves treating the propensity score as a latent variable and modeling the joint likelihood of propensity scores and responses simultaneously in one Bayesian analysis via an MCMC algorithm. From there, the marginal posterior probability of the treatment effect that directly incorporates uncertainty in the propensity score can be obtained. The Bayesian propensity score approach presented by McCandless et al. (2009) was examined by the authors in a simulation study and a case study with real data. In both studies, it was found that weak associations between the covariates and the treatment led to greater uncertainty in the propensity score and thus wider credibility intervals.

It should be noted that some controversy surrounds the Bayesian approach to propensity score adjustment. In particular, Gelman, Carlin, Stern, and Rubin (2003) have argued that the propensity score should provide information only regarding study design and not regarding the treatment effect, as is the case with the Bayesian procedure advocated by McCandless et al. (2009). model in the second stage.

Purpose / Objective / Research Question / Focus of Study:

Description of the focus of the research.

Propensity score analysis has been used in a variety of settings, such as education, epidemiology, and sociology. Most typically, propensity score analysis has been implemented within the conventional frequentist perspective of statistics. This

perspective, as is well known, does not account for uncertainty in either the parameters of the propensity score model or the causal model. Indeed, the conventional implementation of PSA does not allow prior information to enter into the analysis. To account for uncertainty in model parameters we must adopt a Bayesian perspective. Thus, the purpose of this paper is to provide a review and comparative investigation of frequentist and Bayesian propensity score analysis as a means of warranting causal inferences in observational settings.

Setting:
Description of the research location.

NA

Population / Participants / Subjects:
Description of the participants in the study: who, how many, key features or characteristics.
(May not be applicable for Methods submissions)

NA

Intervention / Program / Practice:
Description of the intervention, program or practice, including details of administration and duration.
(May not be applicable for Methods submissions)

NA

Significance / Novelty of study:
Description of what is missing in previous work and the contribution the study makes.

We agree with the views of Gelman eta al regarding McCandless' et al approach to Bayesian propensity score analysis and view their approach as conceptually questionable. In particular, a possible consequence of this joint modeling is that the predictive distribution of propensity scores will be affected by the outcome, which can lead to a different propensity score estimate than obtained if the outcome is not used in the analysis. Thus, to address the problem of joint modeling, this paper outlines a two-stage modeling method using the Bayesian propensity score model in the first stage, followed by the regular causal in the second stage. We show through a comprehensive simulation study that it is possible to implement a simple two-stage Bayesian propensity score model that provides good estimates of causal effects and maintains the spirit of the propensity score approach.

Statistical, Measurement, or Econometric Model:
Description of the proposed new methods or novel applications of existing methods.

We propose a two-step Bayesian propensity score analysis approach, with a Bayesian propensity score model in the first step and Bayesian causal model in the second step, and compare it with the conventional propensity score analysis (PSA). Also, we fit the simple linear regression and Bayesian simple regression without any propensity score adjustment for comparative purposes. The Bayesian simple regression utilizes the Gibbs sampler within the *MCMCregress* package in R to simulate the posterior distribution of the causal model.

For both PSA and BPSA, two models are specified. The first is the propensity score model, specified as a logit model. For BPSA, we utilized the R function *MCMClogit* to simulate from the posterior distribution of a logistic regression using a random walk Metropolis algorithm. After estimating the conventional or Bayesian propensity scores, we use a causal model in the second step to estimate the causal effect via the three approaches: stratification, weighting, and optimal matching

Usefulness / Applicability of Method:
Demonstration of the usefulness of the proposed methods using hypothetical or real data.

The usefulness and applicability of the approach are demonstrated via two simulation studies and a case study. These are described in the Research Design and Findings/Results sections.

Research Design:
Description of research design (e.g., qualitative case study, quasi-experimental design, secondary analysis, analytic essay, randomized field trial).
(May not be applicable for Methods submissions)

In the two simulation studies and the case study, the frequentist based average treatment effect (ATE) and standard error are estimated via ordinary least squares regression (OLS). For the conventional PSA, propensity score stratification is conducted by forming quintiles on the propensity score, calculating the OLS treatment effect within stratum, and averaging over the strata using ``Rubin's'' rules. Propensity score weighting is performed by fitting a weighted regression with ATE weights. Propensity score matching utilizes the full optimal matching method proposed by Rosenbaum (1989).

Study I examines the effects of the Bayesian propensity score model and OLS causal model via different sample sizes, true treatment effects and priors. Study II examines a BPSA with both Bayesian propensity score model and Bayesian causal model, in which uniform priors were compared to normal priors with varying precision.
Also, the effects of different sample sizes and true values of the ATE on the causal inference are studied. Study II also contains two conditions (A and B) to examine the performance of BPSA when there is little prior information or abundant information, respectively.

The case study used for illustrating our method is from the Early Childhood Longitudinal Study Kindergarten cohort data (ECLS-K). The ECLS-K is a nationally representative longitudinal sample providing comprehensive information from children, parents, teachers and schools. The sampled children come from both public and private schools and attends both full-day and part-day kindergarten programs, having diverse socioeconomic and racial/ethnic backgrounds. We examine the treatment effect of full versus part day kindergarten attendance on IRT-based reading scores for children at the end of 1998 fall kindergarten. A sample of 600 children are randomly selected proportional to the number of children in full or part day kindergarten in the population. This resulted in 320 children in full day kindergarten and 280 children in part day kindergarten.

Thirteen covariates were chosen for the propensity score equation. These included gender, race, child's learning style, self-control, social interactions, sadness/loneliness, impulsiveness/overreactiveness, mother's employment status, whether first time kindergartner in 1998, mother's employment between birth and kindergarten, non-parental care arrangements,

social economic status and number of grandparents who live close by. We apply the BPSA approach with both Bayesian propensity score model and Bayesian causal model to obtain the treatment effects and credible interval. Noninformative uniform priors are used due to lack of strong prior information. All analyses utilized the programs available in R. Missing data were handled via the R program *MICE* (multivariate imputation by chained equations).

Data Collection and Analysis:
Description of the methods for collecting and analyzing data.
(May not be applicable for Methods submissions)

NA

Findings / Results:
Description of the main findings with specific details.
(May not be applicable for Methods submissions)

Study I reveals that greater precision in the propensity score equation yields better recovery of the frequentist-based causal effect compared to traditional PSA and compared to no adjustment. Study I also reveals a very small advantage to the Bayesian approach for N=100 versus N=250. Study II-A reveals that greater precision around the wrong causal effect can lead to seriously distorted results. Study II-B reveals that greater precision around the correct causal parameter yeilds quite good results, with slight improvement seen with greater precision in the propensity score equation. The case study reveals that the credibility intervals are wider than the confidence intervals when priors are non-informative. This was shown in McCandless et al. (2009) and is consistent with Bayesian theory.

Conclusions:
Description of conclusions, recommendations, and limitations based on findings.

We propose that a simple and reasonable strategy for Bayesian propensity score analysis is a two-step approach. This approach preserves the basic idea that the propensity score should provide information only regarding study design and not regarding the treatment effect. Bayesian PSA is easy to implement and addresses the issue of uncertainty in the propensity score equation and causal model equation. Elicitation of priors is essential to demonstrate the value of the Bayesian approach (O'Hagan et al., 2006).

Appendices
Not included in page count.

Appendix A. References
References are to be in APA version 6 format.

Cochran, W. G. (1968). The effectiveness of adjustment by subclassification in removing bias in observational studies. Biometrics, 24 , 295{313.
Dawid, A. P. (1982). The well-calibrated Bayesian. Journal of the American

Statistical Association, 77 , 605{610.
Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2003). Bayesian data analysis, 2nd edition. London: Chapman and Hall.
Hansen, B. B. (2004). Full matching in an observational study of coaching for the SAT. Journal of the American Statistical Association, 99 , 609{618.
Hansen, B. B., & Klopfer, S. O. (2006). Optimal full matching and related designs via network flow. Journal of Computational and Graphical Statistics, 15 , 609{627.
Heckman, J. J. (2005). The scientific model of causality. In R. M. Stolzenberg (Ed.), Sociological methodology (Vol. 35, p. 1-97). Boston: Blackwell Publishing.
Hirano, K., & Imbens, G. W. (2001). Estimation of causal effects using propensity score weighting: An application to data on right heart catheterization. Health Services & Outcomes Research Methodology, 2 , 259{278.
Hirano, K., Imbens, G. W., & Ridder, G. (2003). Efficient estimation of average treatment effects using the estimated propensity score. Econometrica, 71 , 1169{1189.
Holland, P. W. (1986). Statistics and causal inference. Journal of the American Statistical Association, 81 , 945-960.
Horvitz, D. G., & Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. Journal of the American Statistical Association, 47 , 663-685.
Hoshino, T. (2008). A Bayesian propensity score adjustment for latent variable modeling and MCMC algorithm. Computational Statistics & Data Analysis, 52 , 1413{1429.
Kaplan, D., & Chen, C. J. S. (2010, March). A Bayesian perspective on methodologies for drawing causal inferences in experimental and non-experimental settings. Paper presented at the 2010 annual research conference of the Society for Research on Educational Effectiveness, Washington, DC.
Lewis, D. (1973). Counterfactuals. Oxford: Blackwell.
Little, R. J. A. (2004). To model or not to model? competing modes of inference for finite population sampling. Journal of the American Statistical Association, 99 , 546{556.
McCandless, L. C., Gustafson, P., & Austin, P. C. (2009). Bayesian propensity score analysis for observational data. Statistics in Medicine, 28 , 94{112.
Morgan, S. L., & Winship, C. (2007). Counterfactuals and causal inference: Methods and principles for social research. Cambridge: Cambridge University Press.
NCES. (2001). Early childhood longitudinal study: Kindergarten class of 1998-99: Base year public-use data _les user's manual (Tech. Rep. No. NCES 2001-029). U.S. Government Printing Office.
Neyman, J. S. (1923). Statistical problems in agriculture experiments. Journal of the Royal Statistical Society, Series B, 2 , 107{180.
R Development Core Team. (2008). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Available from http://www.R-project.org (ISBN 3-900051-07-0)
Rassler, S. (2002). Statistical matching: A frequentist theory, practical applications,

and alternative Bayesian approaches. New York: Springer.
Rosenbaum, P. R. (1987). Model-based direct adjustment. Journal of the American Statistical Association, 82 , 387{394.
Rosenbaum, P. R. (1989). Optimal matching for observational studies. Journal of the American Statistical Association, 84 , 1024{1032.
Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. Biometrika, 70 , 41-55.
Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. Journal of Educational Psychology, 66 , 688-701.
Rubin, D. B. (1985). The use of propensity scores in applied Bayesian inference. Bayesian Statistics, 2 , 463{472.
Rubin, D. B. (2006). Matched sampling for causal effects. Cambridge: Cambridge University Press.
Woodward, J. (2003). Making things happen: A theory of causal explanation. Oxford: Oxford University Press.

Appendix B. Tables and Figures

Not included in page count.