Applied Mathematical Sciences, Vol. 4, 2010, no. 71, 3503 - 3508

Coverage Probability of E-Estimator for the Effect of Dichotomous Exposure

Manoochehr Babanezhad

Department of Statistics, Faculty of Sciences
Golestan University, Gorgan, Golestan, Iran
m.babanezhad@gu.ac.ir

Abstract

One may use different estimation strategies to estimate the causal effect of a dichotomous exposure on an outcome when they control the effect of measured confounders. In this article, I investigate the E-estimator method to estimate the causal effect of a dichotomous exposure when all confounders for the association between exposure and outcome have been accurately measured. I estimate the coverage probability of E-estimator for exposure effect by fitting a semi-parametric model. I provide simulation study for supporting the method.

Keywords: E-estimator, Coverage probability, Sandwich variance estimator

1 Introduction

One may use different estimation strategies to estimate the causal effect of a dichotomous exposure on an outcome of interest. All differs in the coverage probability of effect estimate of exposure, in particular when this effect is adjusted by the measured confounder factors. This may arise this question how is the variance estimate of the effect estimated. Because the coverage probability is the frequency over many replications that the confidence interval contains the target value. It might be not possible to estimate the variance estimator through software. A common way for calculating the variance of an effect estimate is Sandwich formula [1, 2, 3]. In this formula, one often exploits the partial derivatives with respect to the considered model parameters [3, 4]. In this article, I derive the variance of the E-estimator [5, 6] through Sandwich formula when all confounders for the association between exposure and outcome have been accurately measured. I show this by extensive simulation studies. In the next Section, we introduce E-estimators for the effect of a dichotomous exposure on an outcome of interest [6, 7]. I calculate the variance of E-estimator through Sandwich formula in Section 3. Section 4 is simulation, and Section 5 is ended by conclusion.

2 E-estimators

Suppose that we are trying to estimate the effect of a dichotomous exposure X ($X = 1$ if exposed, 0 if not) on an outcome Y. Let Y_x represent the potential outcome which a subject would have had if, possibly contrary to fact, the exposure X were set to x [6]. Then for each subject either we obtain Y_0 (when $X = 0$) or Y_1 (when $X = 1$). The expected contrast $\psi = \mathrm{E}(Y_1 - Y_0)$ thus defines the average causal effect of being exposed. Suppose now that all confounders Z for the association between X and Y have been accurately measured; that is, $X \perp\!\!\!\perp (Y_0, Y_1)|Z$. (For random variables A, B and C, $A \perp\!\!\!\perp B|C$ indicates conditional independence of A and B given C). Then we may obtain an estimate of ψ by fitting the following semi-parametric model,

$$\mathrm{E}(Y_x|Z) = \psi x + g(Z) \tag{1}$$

where $\psi = \mathrm{E}(Y_1 - Y_0)$ the average causal effect of being exposed and $g(Z)$ is a unknown function of the covariates Z.

Estimates for the parameters ψ indexing model (1) may be obtained that after subtracting the causal effect ψX from the outcome, the resulting treatment-free outcome $Y - \psi X$ should be conditionally independent of X, given Z, by the no unmeasured confounders assumption [5, 6]. One may thus estimate ψ as the value for which this conditional independence holds in the observed data set. Then, we will obtain the E-estimator by solving the following estimating equation

$$\sum_{i=1}^{n} U_i(\psi) = \sum_{i=1}^{n} \{X_i - P(X_i|Z_i)\}\left[Y_i - \mathrm{E}(Y_i|Z_i) - \psi\{X_i - P(X_i|Z_i)\}\right] = 0 \tag{2}$$

for ψ. Solving (2) is tantamount to fitting the linear model (for continuous outcome)

$$\mathrm{E}(Y_i|X_i, Z_i; \gamma) = \gamma_0 + \gamma_1 X_i + \gamma_2 Z_i \tag{3}$$

when $\mathrm{E}(Y_i|Z_i) = \gamma_0^* + \gamma_1^* Z_i$ where γ_0^*, γ_1^* are unknown parameters but can be estimated. Here, $P(X_i|Z_i)$ is also unknown, but can be estimated. For a dichotomous exposure X, $P(X_i|Z_i)$ may estimate by fitting the logistic regression model,

$$P(X_i = 1|Z_i; \theta) = \frac{\exp(Z_i\theta)}{1 + \exp(Z_i\theta)}. \tag{4}$$

Estimation thus proceeds in two stages: first we estimate θ in model (4); then, we fit model (3). Then by solving 2, the E-estimator for ψ will be obtained

$$\hat{\psi} = \frac{\sum_{i=1}^{n}\{(X_i - P(X_i|Z_i))(Y_i - \mathrm{E}(Y_i|Z_i))\}}{\sum_{i=1}^{n}\{X_i - P(X|Z_i)\}^2}.$$

3 Sandwich variance estimator

Estimator $\hat{\psi}$ is asymptotically normally distributed under standard regularity conditions with finite variance [5, 6]. We now derive the variance of $\hat{\psi}$ through Sandwich formula. For doing so, we should take into account all of the parameters in forming the associated variance matrix,

$$W(\phi^*) = \begin{pmatrix} U(\phi^*) = \{X_i - P(X_i|Z_i)\}[Y_i - \mathrm{E}(Y_i|Z_i) - \psi\,\{X_i - P(X_i|Z_i)\}] \\ \\ U_L(\theta) = \begin{pmatrix} 1 \\ Z_i \end{pmatrix} \{X_i - \mathrm{expit}(\theta_0 + \theta_1 Z_i)\} \\ \\ U_O(\gamma^*) = \begin{pmatrix} 1 \\ Z_i \end{pmatrix} \{Y_i - (\gamma_0^* + \gamma_1^* Z_i)\} \end{pmatrix}$$

where $\phi^* = (\psi, \theta, \gamma^*)$, $U(\phi^*)$ is the estimating equation of semi-parametric model (2), $U_L(\theta)$ is the estimating equation of logistic model (4), and $U_O(\gamma^*)$ is the estimating equation of ordinary linear regression model, $\mathrm{E}(Y_i|Z_i) = \gamma_0^* + \gamma_1^* Z_i$. The variance matrix for $\phi^* = (\psi, \theta, \gamma^*)$ is formed by obtaining the necessary derivatives. Therefore by Sandwich method, variance of $\phi^* = (\psi, \theta, \gamma^*)$ will be obtained,

$$V_{sand} = \frac{1}{n}\ \mathrm{E}^{-1}\{\partial W(\phi^*)/\partial\phi^*\}\ \ Var\{W(\phi^*)\}\ \ \mathrm{E}^{'-1}\{\partial W(\phi^*)/\partial\phi^*\}$$

where

$$\frac{\partial W(\phi^*)}{\partial\phi^*} = \begin{pmatrix} \partial U_G(\phi^*)/\partial\psi & \partial U_G(\phi^*)/\partial\theta & \partial U_G(\phi^*)/\partial\gamma^* \\ \\ 0 & \partial U_L(\theta)/\partial\theta & 0 \\ \\ 0 & 0 & \partial U_O(\gamma^*)/\partial\gamma^* \end{pmatrix}$$

and index ‘stand’ indicates Sandwich variance estimator. The covariance matrix V_{sand} is a 5×5 matrix. The Sandwich variance estimator of $\hat{\psi}$ then is

$$Var(\hat{\psi}) = [V_{stand}]_{11} = \frac{1}{n}\ \mathrm{E}^{-1}\{X_i - P(X_i|Z_i)\}^2\ \ [Var\{W(\phi^*)\}]_{11}\ \ \mathrm{E}^{'-1}\{X_i - P(X_i|Z_i)\}^2.$$

Table I. Coverage probability and average length of 95% confidence intervals of E-estimator for ψ in terms of the confounder-exposure association $OR_{X|Z}$ and the mean of exposure $\mathrm{E}(X)$.

$OR_{X\|Z}$	E(X)	Coverage	Average length CI
	0.75	0.90	13.50
1.12	0.50	0.85	16.60
	0.25	0.75	24.50
	0.75	1.00	5.50
2.25	0.50	0.93	6.50
	0.25	0.85	12.50
	0.75	0.99	10.20
5.50	0.50	0.97	11.50
	0.25	0.98	10.50

4 Simulation

To investigate the estimated variance of the considered estimator in finite samples, we conducted a simulation experiment. Data were simulated to mimic the body mass index (BMI) study reported in Babanezhad et al. [8] with Y representing birth weight, dichotomous exposure X representing BMI of maternal at last menstrual period, and Z confounder scores based on maternal age, child sex, history of maternal hypertension, and thalassemia minor of father in the north area of Iran. The exposure X was generated to be dichotomous adjusted with covariates Z by $P(X = 1|Z) = \text{expit}(\theta_0 + \theta_1 Z)$, where $\theta_0 = 1.26$ and $\theta_1 = 0.84$. The outcome Y was chosen to be normally distributed from X adjusted covariates Z. Table I summarizes the results for different choices of the confounder-exposure association, $OR_{X|Z} = \exp(\theta_1)$. Table I summarizes the results with known causal effect value $\psi_0 = 1$. Table I shows the coverage probability and the average length of 95 % confidence interval of E-estimator $\hat{\psi}$. The simulation has also carried out with severe confounding $OR_{X|Z} = 0.50$ and 0.75. We show that when the quasilikelihood model $P(X|Z)$ is correct, the Sandwich covariance matrix estimate is approximately less variable than the usual parametric variance estimate, and its coverage probabilities seem to be close to 0.95.

5 Conclusion

The estimated variance through Sandwich formula, often known as inefficient variance compared to the estimator computed directly from the software. But it might not generally be true. For instance for case where the estimator itself is robust this may not be true and justifiable. The E-estimator is robust for the causal effect of an exposure [5, 6, 7] compared to the ordinary least square estimator because working model $P(X|Z)$ or $E(Y|X, Z)$ are correctly specified not necessarily both under semi-parametric model (1). Therefore, when there is no direct way to find the variance estimators (this is the case for E-estimator) I show that the Sandwich variance estimator is useful. This is confirmed by the simulation results. Table I shows that the Sandwich variance estimator is approximately inefficient just for case where the confounder-exposure association is 1.12 and E(X)=0.25. For other cases this estimator has high converge probability and approximately small average length of 95% confidence intervals. With small samples and/or non-robust estimation strategy the Sandwich estimator often leads to coverage probabilities lower than the usual estimator [1, 2].

References

[1] X. Guo, W. Pan, J.E. Connett, P.J. Hannan and S.A. French, Small-sample performance of the robust score test and its modifications in generalized estimating equations.*Statist. Med.* 24 (2005), 34793495.

[2] R.J. Carroll and G. Kauermann, A note on the efficiency of sandwich covariance matrix estimation. *Journal of the American Statistical Association*, 96 (2001), 1387–1396.

[3] J.W. Hardin, The robust variance estimator for two-stage models. *Stata Journal* 2 (2002), 253–265.

[4] R.J. Carroll, D. Ruppert, L.A. Stefanski and C.M. Crainiceanu, *Measurement Error in Nonlinear Models, Second Edition*. CRC Press, 2006.

[5] J.M. Robins, S.D. Mark and W.K. Newey, Estimating exposure effects by modeling the expectation of exposure conditional on confounders. *Biometrics*, 48 (1992), 479–495.

[6] M. Babanezhad, S. Vansteelandt and E. Goetghebeur, Comparison of causal effect estimators under exposure misclassification, *Journal of Statistical Planning and Inference*, 140 (2010), 1306–1319.

[7] J.M. Robins and R. Rotnitzky, (2004) Estimation of treatment effects in randomised trials with non-compliance and a dichotomous outcome using structural mean models. *Biometrika* 91 (2004),763–783.

[8] M. Babanezhad, Z. Atarod and Z. Nazari, (2010), The effect of body mass index on pregnancy outcomes in the city of north of Iran, *Technical Report*.

Received: July, 2010