

MATH 598/782 – PROJECT 3

Please submit your solutions by 11.59 pm (EDT) on Sunday 24th March by uploading a single pdf to myCourses.

For the $p \times 1$ system of estimating equations

$$\sum_{i=1}^n \mathbf{U}_i(\theta) = \mathbf{0}$$

with $\mathbb{E}[\mathbf{U}(\theta_0)] = \mathbf{0}$, we have that

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} \text{Normal}_p(\mathbf{0}, \mathbf{V})$$

where

$$\mathbf{V} = \mathcal{J}^{-1} \mathcal{I} \mathcal{J}^{-\top}$$

with

$$\mathcal{I} = \mathbb{E}[\mathbf{U}(\theta_0) \mathbf{U}(\theta_0)^\top] \quad \mathcal{J} = \mathbb{E}[\dot{\mathbf{U}}(\theta_0)]$$

both $(p \times p)$ matrices, and

$$\dot{\mathbf{U}}(\theta_0) = - \left. \frac{\partial \mathbf{U}(\theta)}{\partial \theta^\top} \right|_{\theta=\theta_0}$$

We proceed by computing estimates, $\hat{I}_n$ and $\hat{J}_n$, of the two matrices based on the observed data, and then computing

$$\hat{\mathbf{V}} = \hat{J}_n^{-1} \hat{I}_n \hat{J}_n^{-\top}$$

to estimate the asymptotic variance; this approach is known as *sandwich* or *robust* variance estimation.

Based on the data set from Project 2

<https://www.math.mcgill.ca/dstephens/598-Causal-2024/Exercises/Project2/P2data.csv>

which comprises four predictor variables X_1, X_2, X_3, X_4 , a binary treatment variable Z , and outcome variable Y , compute an estimate of the asymptotic variance of the parameter of interest ψ based on the sandwich estimation procedure and knitr sheets for the following models and estimating procedures.

(a) An outcome regression model

$$Y_i = \beta_0 + \beta_1 X_{1i} + \psi Z_i + \epsilon_i$$

estimated using ordinary least squares.

4 Marks

(b) A propensity score regression model

$$Y_i = \beta_0 + \psi Z_i + \phi e(X_{1i}, X_{2i}; \hat{\alpha}) + \epsilon_i$$

where α is estimated using maximum likelihood in a logistic regression of Z on (X_1, X_2) , including an intercept, that is $Z|X_1 = x_1, X_2 = x_2 \sim \text{Bernoulli}(e(x_1, x_2; \alpha))$ with $\alpha = (\alpha_0, \alpha_1, \alpha_2)^\top$, and

$$e(x_1, x_2; \alpha) = \frac{\exp\{\alpha_0 + \alpha_1 x_1 + \alpha_2 x_2\}}{1 + \exp\{\alpha_0 + \alpha_1 x_1 + \alpha_2 x_2\}}.$$

6 Marks

(c) An inverse probability weighting (IPW) procedure using the propensity score model from (b).

4 Marks

(d) An augmented outcome regression model

$$Y_i = \beta_0 + \psi Z_i + \phi_0 \frac{1 - Z_i}{1 - e(X_{i1}, X_{i2}; \hat{\alpha})} + \phi_1 \frac{Z_i}{e(X_{i1}, X_{i2}; \hat{\alpha})}$$

with α estimated using the propensity score model from (b).

6 Marks