Advanced Topics in Statistics

Assignment 2

Submission: on Brightspace by 2 March 2022, 13:45

Consider the following model. For simplicity, assume that we have a single standard normal covariate $L \sim \mathcal{N}(0,1)$. The binary treatment participation A and the potential outcomes Y^0, Y^1 are given by

$$
\begin{aligned}
A &= \mathbb{1}_{\eta_0+\eta_1 L+\epsilon>0}, \quad (\epsilon \mid L) \sim \text{Logistic}(0,1), \\
Y^0 &= \alpha_0 + \beta_0 L + \nu_0, \quad (\nu_0 \mid L, \epsilon) \sim \mathcal{N}(0,1), \\
Y^1 &= \alpha_1 + \beta_1 L + \nu_1, \quad (\nu_1 \mid L, \epsilon, \nu_0) \sim \mathcal{N}(0,1),
\end{aligned}
$$

where the indicator $\mathbb{1}_B$ is one if B is true, zero otherwise, and $\eta_0, \eta_1, \alpha_0, \alpha_1, \beta_0, \beta_1 \in \mathbb{R}$ are fixed parameters. Assume consistency $Y = (1-A)Y^0 + AY^1$.

1. Derive the propensity score $f_\eta(1 \mid L) = \mathbb{P}(A = 1 \mid L)$ where $\eta = (\eta_0, \eta_1)^\intercal$.

2. Show that $ATE = \mathbb{E}(Y^1 - Y^0) = \alpha_1 - \alpha_0$.

3. Using a programming language of your preference[1], perform a Monte Carlo simulation to illustrate that the variance of the Horvitz-Thompson estimator of ATE is lower the estimated propensity score is used instead of the true one. You may use your favourite programming language. Use the following steps.

   Step 1: set $n = 100$ and fix $\eta_0 = 0.01$, $\eta_1 = 1$, $\alpha_0 = 4$, $\alpha_1 = 10$ and β_0, β_1 to your favourite non-zero value such that $|\beta_0| \neq |\beta_1|$.

   Step 2: generate data (L_i, A_i, Y_i), $i = 1, \ldots, n$, according to the above model and estimate η with $\hat{\eta}$, the maximum likelihood estimator[2]. Compute the two estimators

$$
T_{\mathrm{HT},\eta} = \frac{1}{n} \sum_i \left(\frac{Y_i \mathbb{1}_{A_i=1}}{f_\eta(1 \mid L_i)} - \frac{Y_i \mathbb{1}_{A_i=0}}{1 - f_\eta(1 \mid L_i)} \right) \tag{1}
$$

$$
T_{\mathrm{HT},\hat{\eta}} = \frac{1}{n} \sum_i \left(\frac{Y_i \mathbb{1}_{A_i=1}}{f_{\hat{\eta}}(1 \mid L_i)} - \frac{Y_i \mathbb{1}_{A_i=0}}{1 - f_{\hat{\eta}}(1 \mid L_i)} \right). \tag{2}
$$

   Step 3: repeat Step 2 $R = 5000$ times (or more, as many as feasible on your hardware), so that you have R values of both estimators: $T^{(1)}_{\mathrm{HT},\eta}, \ldots, T^{(R)}_{\mathrm{HT},\eta}, T^{(1)}_{\mathrm{HT},\hat{\eta}}, \ldots, T^{(R)}_{\mathrm{HT},\hat{\eta}}$. Give

[1] E.g. Julia, R, Python, Matlab, C.

[2] Use the log-likelihood. For large n, the likelihood is too small to work with. Use any standard maximizer (or minimizer) to compute the MLE

(a) the biases $B_\eta := R^{-1}\sum_{r=1}^R T_{\mathrm{HT},\eta}^{(r)} - ATE$ and $B_{\hat\eta} := R^{-1}\sum_{r=1}^R T_{\mathrm{HT},\hat\eta}^{(r)} - ATE$;

(b) the variances $V_\eta := R^{-1}\sum_{r=1}^R (T_{\mathrm{HT},\eta}^{(r)} - \bar{T}_{\mathrm{HT},\eta})^2$ and $V_{\hat\eta} := R^{-1}\sum_{r=1}^R (T_{\mathrm{HT},\hat\eta}^{(r)} - \bar{T}_{\mathrm{HT},\hat\eta})^2$, where $\bar{T}_{\mathrm{HT},\cdot} = R^{-1}\sum_{r=1}^R T_{\mathrm{HT},\cdot}^{(r)}$.

(c) the mean squared errors $M_\eta := B_\eta^2 + V_\eta$ and $M_{\hat\eta} := B_{\hat\eta}^2 + V_{\hat\eta}$.

Explain the observed differences.

4. What changes in the model would increase (further) the difference in asymptotic variances of the two estimators?

5. Bonus: show that Y^a is independent of A given L for any a.

Submit one pdf with your answers and the code you used to Brigthspace.