MATH 598 - PROJECT 3

Please upload a pdf with your solutions to myCourses by 10.00pm on Monday 22nd March.

Consider a data generating model for iid data $\{(x_i, y_i, z_i), i = 1, ..., n\}$, where

- $X \sim Normal(0.5, 0.5^2)$.
- Given X = x, Z is Bernoulli(e(x)) with

$$e(x) = \frac{\exp\{-1.5 + 2x + x^2\}}{1 + \exp\{-1.5 + 2x + x^2\}}$$

• Given X = x and Z = z, Y is conditionally Normally distributed with mean

$$\mathbb{E}_{Y|X,Z}[Y|X=x,Z=z] = 1 + x + z + xz$$

and variance $\sigma^2 = 1$.

For this model, carry out a simulation study to demonstrate the performance of G-estimation of parameters $\psi = (\psi_0, \psi_1)^{\top}$ based on the (mis-specified) mean model

$$\beta_0 + Z(\psi_0 + \psi_1 X)$$

and assuming correct specification of the propensity score model. Include in your analysis a computation of the estimated asymptotic variance of the estimator $\hat{\psi}$ assuming

- (a) parameters in the propensity score model are known, and take their data generating (true) values;
- (b) parameters in the propensity score model are **unknown** and estimated from the data.

You may adapt R code from the knitr sheets from the course.

25 Marks