Data Simulation

Waveley Qiu (wq46)

2023-11-27

Treatment and Propensity Score

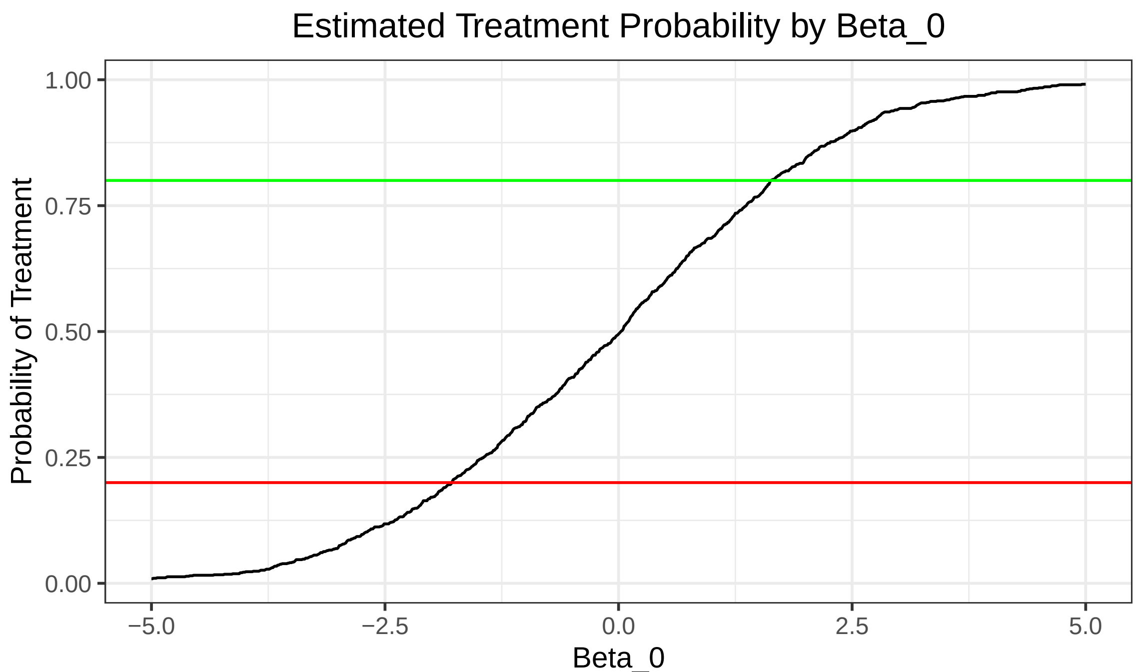
The propensity score is defined as follows:

$$
\begin{aligned}
e(\mathbf{X}) &= P(Z=1|\mathbf{X}) = E[Z=1|X] \\
\beta X^T &= \text{logit}(E[Z=1|X]) \\
&\implies E[Z=1|X] = \frac{\exp(\boldsymbol{\beta X}^T)}{1+\exp(\boldsymbol{\beta X}^T)} \\
E\Big[P(Z=1)\Big] &= E_X\Big[E_Z\Big(Z=1|X\Big)\Big] \\
&= E_X\left[\frac{\exp(\boldsymbol{\beta X}^T)}{1+\exp(\boldsymbol{\beta X}^T)}\right] = p
\end{aligned}
$$

We see that though it is straightforward to created a propensity score model, depending on the distribution of the covariates, it may be very difficult (or impossible) to achieve a closed-form solution determine the β coefficients needed to explicitly specify the proportion of individuals treated, p.

Then, we will select β coefficients in an empirical fashion after fixing the distributional forms of the covariates we are interested in.

Let $X_1 \sim N(0,1)$, $X_2 \sim \text{Bernoulli}(0.6)$, representing one continuous and one binary covariate upon which treatment is determined for each subject. Then, we see that we will need to select three β coefficients to satisfy the form $g^{-1}(E[Z=1|X]) = \beta_0 + \beta_1 X_1 + \beta_2 X_2$, where $g(\mathbf{X}) = \frac{\exp(\boldsymbol{\beta X}^T)}{1+\exp(\boldsymbol{\beta X}^T)} = E[Z=1|X]$. Fixing β_1 at 2 and β_2 at 3 for simplicity of calculation, we see the following:



Then, we will define our propensity score models as follows:

Low treatment (~20%): $g^{-1}(E[Z=1|X]) = \beta_{0,20} + \beta_1 X_1 + \beta_2 X_2 = -1.66 + 2X_1 + 3X_2$

High treatment (~80%): $g^{-1}(E[Z=1|X]) = \beta_{0,80} + \beta_1 X_1 + \beta_2 X_2 = 1.74 + 2X_1 + 3X_2$