Simulation Study Setup

We use the following simulation setup

$$\begin{bmatrix} x \\ e_1 \\ e_2 \end{bmatrix} \overset{ind}{\sim} N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & \rho \\ 0 & \rho & 1 \end{bmatrix} \right)$$
$$y_1 = x + e_1$$
$$y_2 = \mu + x + e_2$$

This yields outcome variables Y_1 and Y_2 that are correlated both with X and additionally with each other. To generate the missingness pattern, we draw from a categorical distribution with selection probabilities $(\pi_{00}, \pi_{10}, \pi_{01}, \pi_{11})$. Each selection probability, π_{j_1,j_2} indicates the probability that an observation is selected into a segment A_{j_1,j_2} where A_{00} indicates that only X is observed, A_{10} means that X and Y_1 are observed, A_{01} has only X and Y_2 observed, and A_{11} observes X, Y_1, and Y_2.

In addition to varying the values of the parameters μ and ρ, there are two main factors of the simulation that change: the distribution of the categories and the parameter of interest, θ. In a "balanced" distribution, we have the following selection probabilities: $\pi_{00} = 0.2$, $\pi_{10} = 0.2$, $\pi_{01} = 0.2$, and $\pi_{11} = 0.4$. In the "unbalanced" distribution we have $\pi_{00} = 0.3$, $\pi_{10} = 0.4$, $\pi_{01} = 0.1$, and $\pi_{11} = 0.2$. The two types of θ values that we consider are a linear estimate, $\theta = E[g(Z)] = E[Y_2]$ and a non-linear estimate, $\theta = E[g(Z)] = E[Y_1^2 Y_2]$. For notational purposes, we define $Z = (X, Y, Z)$ and occasionally we will use the notation $G_r(Z)$ to indicate the general form of the observed data in rth segment. We also use n to denote to total number of observations in the sample. While n is known, due to our setup, the number of elements in each segment A_{j_1,j_2} is random.

There are several algorithms for comparison which are defined as the following:

$$Oracle = n^{-1} \sum_{i=1}^{n} g(Z_i)$$
$$CC = \frac{\sum_{i=1}^{n} \delta_{11} g(Z_i)}{\sum_{i=1}^{n} \delta_{11}}$$
$$IPW = \sum_{i=1}^{n} \frac{\delta_{11}}{\pi_{11}} g(Z_i)$$