P8160 - Comparing Bootstrapping Methods Report

Amy Pitts, Hun Lee, Jimmy Kelliher, Tucker Morgan, Waveley Qiu

2/11/2022

Introduction

detail to be added

Simulation Planning

Aims

The primary goal of this simulation study is to assess the performance of two bootstrap methods (detailed below) in estimating the sampling variability of treatment effects obtained from a nearest-neighbor propensity-score matching (NNM). In this study, NNM will select a treated subject at random from simulated observational data. The untreated subject with the nearest propensity score is then selected to be paired with the treated subject, without replacement. Treatment effects can then be estimated by comparing outcomes (continuous or binary) between the treated and untreated subjects. The bootstrapping methods will be used to assess the variance of estimated treatment effects.

Data Generation

The data for this simulation study were generated from a parametric model. For each subject, three baseline covariates (L_1, L_2, L_3) were simulated from independent standard normal distributions, $N(0, 1)$. Two of these covariates (L_1 and L_2) affected treatment selection, while two (L_2 and L_3) affected the outcome. The probability of treatment for each subject was determined by the following model:

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \alpha_0 + \alpha_1 L_{1i} + \alpha_2 L_{2i}$$

where $\alpha_0 = \log(\frac{\pi}{1-\pi})$ as close approximation of desired treatment prevalence. **Due to the nature of the logit normal there is no closed form solution to the mean of the logit normal distibution however, calculating α_0 in this fashion gives us a very close but always biased above approximation.**

For continuous outcomes, 100 sub-populations of 100 or 1,000 subjects will be generated using the following parametric model:

$$Y_i = \beta_1 A_i + \beta_2 L_{2i} + \beta_3 L_{3i} + \epsilon_i$$

where Y_i indicates the outcome for each subject, A_i indicates the treatment status of each subject (0 or 1), L_{2i} and L_{3i} indicate observed covariate values for each subject, and ϵ_i denotes random error. Because L_{2i} affects both A_i and Y_i, it acts as a confounder in estimating the treatment effect.

Is there a β_0 value here? For binary outcomes, the same procedure will be performed using the following parametric model:

$$\log\left(\frac{\tau_i}{1-\tau_i}\right) = \beta_1 A_i + \beta_2 L_{2i} + \beta_3 L_{3i}$$

where $Y_i \sim \text{Bernoulli}(\tau_i)$. The binary outcome model does not feature an error term, as realizations of Y_i are innately subject to noise.

Here I am attempting to explain what the truth is but will need to be revised!

In the choice of parameter distributions and structure of our treatment assignment and outcome generation the true distribution can be calculated. This will allow the simulation estimates to be compared to a truth.

The treatment assignment for follows a logit normal with $\mu = \alpha_0$ and $\sigma^2 = \alpha_2^2 + \alpha_3^2$.

The outcome variable for the binary data follows at logit normal distribution with $\mu = \beta_1$ and $\sigma^2 = \beta_2^2 + \beta_3^2$.
Talk about the MCMC method that Jimmy created

The outcome variable for the continuous data data follows at normal distribution with $\mu = \beta_1$ and $\sigma^2 = \beta_2^2 + \beta_3^2$.

Thus the true treatment effect will depend on the assignment of β_1.

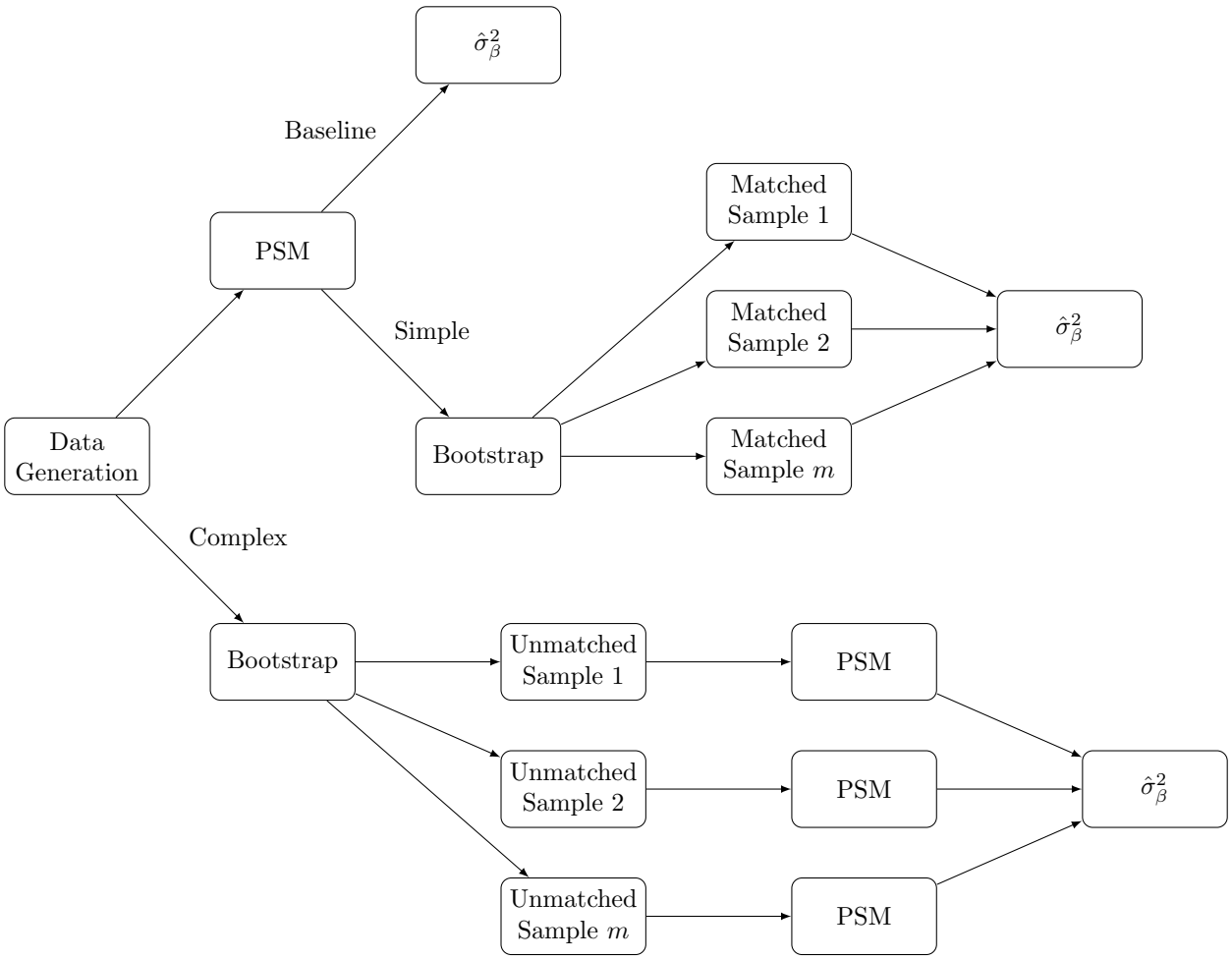
Methods for Evaluation

Two bootstrap methods will be assessed in this simulation: the simple bootstrap and the complex bootstrap.

In the simple bootstrap, one draws repeated samples from an original sample with replacement to imitate the process of drawing samples from a population. Here, 500 repeated samples (m_{boot}) of matched pairs ($n_{boot} = n_{sample} \cdot P(A = 1)$) will be drawn from the matched pairs of observations for each of the 100 initial samples (m_{sample}). The distribution of the estimated treatment effect ($\hat{\beta}_1$) across the 500 bootstraps is assessed for each of the 100 initial samples.

The complex bootstrap considers two additional sources of variability compared to the simple bootstrap. In this approach, a sample is drawn with replacement from the original, unmatched observational data. The propensity-score model is estimated using this bootstrap sample, and NNM proceeds as before. The treatment effect is estimated from the newly matched sample. This process is repeated 500 times (m_{boot}) for each of the 100 samples (m_{sample}).

The resulting standard error estimates, $\hat{\sigma}_\beta$, will be the primary targets of this analysis.



Parameters of Interest

In simulations, three parameters will vary to aid in the comparison of the two bootstrap techniques. These three parameters are dataset sample size (n_{sample}), population proportion of treated individuals (π) and the true average treatment effect (β_1). Varying the sample size and proportion of treated individuals will help make suggestions for studies that have varying amounts of participants and amount of treatment. The range the number of the sample size tried is $n_{\text{sample}} \in \{100, 1000\}$, one a low sample and one a larger sample. The range of population proportion of treated individuals $\pi \in \{0.113, 0.216, 0.313\}$. This is to approximate low, medium, and high treatment levels in a study. The true average treatment effect will take on $\beta_1 \in \{0.15, 0.30\}$ in a binary study and $\beta_1 \in \{-1, 1\}$ in a continuous study.

To compare these three parameters there are number of other parameters that will be held constant from simulation to simulations. Such as the number of datasets ($m_{\text{sample}} = 100$), the number of bootstrap re-sample ($m_{\text{boot}} = 500$), the strength of covariate correlation on treatment status (α_1, α_2), and strength of covariate correlation on outcome variable (β_2, β_3). Without loss of generality for the continuous data ($\alpha_1 = 1, \alpha_2 = 2, \beta_2 = 2, \beta_3 = 1$), and for the binary data ($\alpha_1 = \log(1.25), \alpha_2 = \log(1.75), \beta_2 = \log(1.75), \beta_3 = \log(1.25)$).

Performance Measures

The standard error estimates from each bootstrap method will be assessed in two ways. First, coverage rates of confidence intervals will be analyzed to assess how frequently the true average treatment effect (β_1) is

included in confidence intervals using the bootstrap-estimated treatment effect ($\bar{\hat{\beta}}_1$) and estimated standard errors ($\hat{\sigma}_\beta$). Second, standard error estimates from each bootstrap method will be compared to the sample standard deviation of treatment effects of the initial samples to determine how bootstrapping aligns with a simpler approach.

Bias is also calculated using the true treatment effect. This measure helps confirm that each method is able to accurately identify the treatment effect. A 95% percent confidence interval is also constructed around the bias using the standard error.

Simulation Execution

discussion of coding will go here, perhaps even excerpts of code

Results

detail to be added

Discussion

conclusions to be added