Title in Progress

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

2022-02-21

Motivation

- Identifying the effect of a treatment or intervention is one of the most fundamental tasks we encounter as biostatisticians...

Motivation

- Identifying the effect of a treatment or intervention is one of the most fundamental tasks we encounter as biostatisticians...
- ... but confounding variables can bias our estimates of treatment effects.

Motivation

- Identifying the effect of a treatment or intervention is one of the most fundamental tasks we encounter as biostatisticians...
- ... but confounding variables can bias our estimates of treatment effects.
- Propensity score matching (PSM) is a tool that can help us mitigate the effects of confounders...

Motivation

- Identifying the effect of a treatment or intervention is one of the most fundamental tasks we encounter as biostatisticians...
- ... but confounding variables can bias our estimates of treatment effects.
- Propensity score matching (PSM) is a tool that can help us mitigate the effects of confounders...
- ... but there is no consensus on the best way to estimate standard errors when using the PSM algorithm.

Motivation

- Identifying the effect of a treatment or intervention is one of the most fundamental tasks we encounter as biostatisticians...
- ... but confounding variables can bias our estimates of treatment effects.
- Propensity score matching (PSM) is a tool that can help us mitigate the effects of confounders...
- ... but there is no consensus on the best way to estimate standard errors when using the PSM algorithm.
- How can we assess which procedures reliably estimate standard errors?

Motivation

A simulation study!

Enter the Bootstrap

- Bootstrapping is one of the most common procedures for estimating standard errors.

Enter the Bootstrap

- Bootstrapping is one of the most common procedures for estimating standard errors.
- The PSM algorithm intakes an unmatched dataset and outputs a matched one.

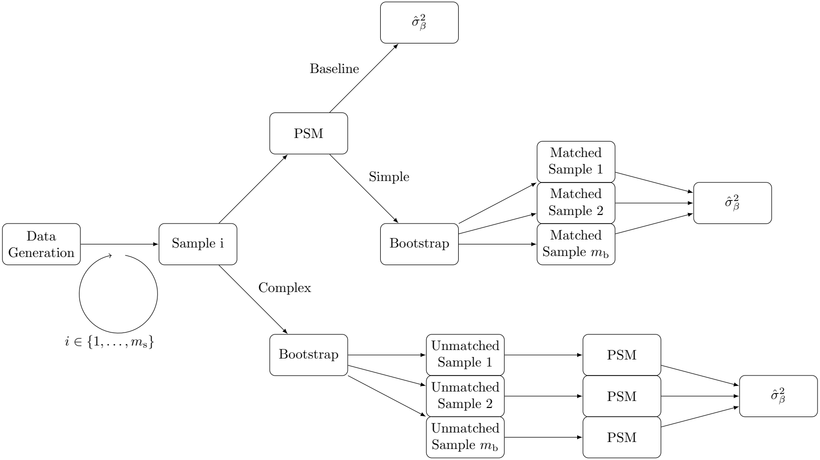
Enter the Bootstrap

- Bootstrapping is one of the most common procedures for estimating standard errors.
- The PSM algorithm intakes an unmatched dataset and outputs a matched one.
- When do we execute the bootstrap - before the match or after it?

Enter the Bootstrap

- Bootstrapping is one of the most common procedures for estimating standard errors.
- The PSM algorithm intakes an unmatched dataset and outputs a matched one.
- When do we execute the bootstrap - before the match or after it?
- Let's try both!

Roadmap of the Simulation Study



Data Generation - Continuous Outcome

For each individual $i \in \{1, \dots, n\}$, we consider covariates $L_{1i}, L_{2i}, L_{3i} \sim N(0, 1)$. Treatments are distributed according to law $A_i \sim B(\pi_i)$, where π_i - the true propensity to be treated - is subject to the data-generating process

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

Given this, we further define the data-generating process of our continuous outcome via

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

where ε_i denotes random error. Because L_{2i} effects both A_i and Y_i, it acts as a confounder in estimating the treatment effect.

Data Generation - Binary Outcome

For each individual $i \in \{1, \dots, n\}$, we consider covariates $L_{1i}, L_{2i}, L_{3i} \sim N(0,1)$. Treatments are distributed according to law $A_i \sim B(\pi_i)$, where π_i - the true propensity to be treated - is subject to the data-generating process

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

Given this, we further define the data-generating process of our binary outcome via $Y_i \sim B(\tau_i)$ where

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

Observe that we have omitted a random error term, as realizations of Y_i are innately subject to noise.