12. Matching Exercise

Ian Lundberg
Cornell Info 6751: Causal Inference in Observational Settings
Fall 2022

29 Sep 2022

Learning goals for today

At the end of class, you will be able to:

1. Apply matching estimators for causal effects
2. Use simulation to evaluate the bias, variance, and mean squared error of an estimator

Simulation: Evaluating an estimator

Simulation: Evaluating an estimator

Two broad classes of statistical research

Simulation: Evaluating an estimator

Two broad classes of statistical research

- Applications: Use an estimator to study the world

Simulation: Evaluating an estimator

Two broad classes of statistical research

- Applications: Use an estimator to study the world
- Methodology: Study the performance of an estimator

Simulation: Evaluating an estimator

Two broad classes of statistical research

- Applications: Use an estimator to study the world
- Methodology: Study the performance of an estimator

To study methodology, it is often helpful to operate in a setting where the **truth is known**

Simulate a population

Simulate a population

Suppose we generate a population as follows.

$$
\begin{tikzcd}
L_1 \arrow[dr] \arrow[drr, bend left] & & \\
& A \arrow[r] & Y \\
L_2 \arrow[ur] \arrow[urr, bend right] & &
\end{tikzcd}
$$

Simulate a population

Suppose we generate a population as follows.

$$\begin{tikzcd}
L_1 \arrow[dr] \arrow[drr, bend left] & & \\
& A \arrow[r] & Y \\
L_2 \arrow[ur] \arrow[urr, bend right] & &
\end{tikzcd}$$

$$L_1 \sim \text{Normal}(\text{Mean} = 0, \text{SD} = 1)$$

Simulate a population

Suppose we generate a population as follows.

$$
\begin{tikzcd}
L_1 \arrow[dr] \arrow[drr, bend left] & & \\
& A \arrow[r] & Y \\
L_2 \arrow[ur] \arrow[urr, bend right] & &
\end{tikzcd}
$$

$$
\begin{aligned}
L_1 &\sim \text{Normal}(\text{Mean} = 0, \text{SD} = 1) \\
L_2 &\sim \text{Normal}(\text{Mean} = 0, \text{SD} = 1)
\end{aligned}
$$

Simulate a population

Suppose we generate a population as follows.

$$
\begin{tikzcd}
L_1 \arrow[dr] \arrow[drr, bend left] & & \\
& A \arrow[r] & Y \\
L_2 \arrow[ur] \arrow[urr, bend right] & &
\end{tikzcd}
$$

$$L_1 \sim \text{Normal}(\text{Mean} = 0, \text{SD} = 1)$$
$$L_2 \sim \text{Normal}(\text{Mean} = 0, \text{SD} = 1)$$
$$A \sim \text{Bernoulli}\left(\text{logit}^{-1}\left[-2 + L1 + L2\right]\right)$$

Simulate a population

Suppose we generate a population as follows.

$$
\begin{tikzcd}
L_1 \arrow[dr] \arrow[drr, bend left] & & \\
& A \arrow[r] & Y \\
L_2 \arrow[ur] \arrow[urr, bend right] & &
\end{tikzcd}
$$

$$
\begin{aligned}
L_1 &\sim \text{Normal}(\text{Mean} = 0, \text{SD} = 1) \\
L_2 &\sim \text{Normal}(\text{Mean} = 0, \text{SD} = 1) \\
A &\sim \text{Bernoulli}\left(\text{logit}^{-1}\left[-2 + L1 + L2\right]\right) \\
Y^0 &\sim \text{Normal}\left(\text{Mean} = L1 + L2, SD = 1\right)
\end{aligned}
$$

Simulate a population

Suppose we generate a population as follows.

$$
\begin{tikzcd}
L_1 \arrow[dr] \arrow[drr, bend left] & & \\
& A \arrow[r] & Y \\
L_2 \arrow[ur] \arrow[urr, bend right] & &
\end{tikzcd}
$$

$$
\begin{aligned}
L_1 &\sim \text{Normal}(\text{Mean} = 0, \text{SD} = 1) \\
L_2 &\sim \text{Normal}(\text{Mean} = 0, \text{SD} = 1) \\
A &\sim \text{Bernoulli}\left(\text{logit}^{-1}\left[-2 + L1 + L2\right]\right) \\
Y^0 &\sim \text{Normal}\left(\text{Mean} = L1 + L2, SD = 1\right) \\
Y^1 &\sim \text{Normal}\left(\text{Mean} = L1 + L2 + 1, SD = 1\right)
\end{aligned}
$$

Simulate a population

Suppose we generate a population as follows.

$$
\begin{tikzcd}
L_1 \arrow[dr] \arrow[drr, bend left] & & \\
& A \arrow[r] & Y \\
L_2 \arrow[ur] \arrow[urr, bend right] & &
\end{tikzcd}
$$

$$
\begin{aligned}
L_1 &\sim \text{Normal}(\text{Mean} = 0, \text{SD} = 1) \\
L_2 &\sim \text{Normal}(\text{Mean} = 0, \text{SD} = 1) \\
A &\sim \text{Bernoulli}\left(\text{logit}^{-1}\left[-2 + L1 + L2\right]\right) \\
Y^0 &\sim \text{Normal}\left(\text{Mean} = L1 + L2, SD = 1\right) \\
Y^1 &\sim \text{Normal}\left(\text{Mean} = L1 + L2 + 1, SD = 1\right) \\
Y &= \begin{cases} Y^0 & \text{if } A = 0 \\ Y^1 & \text{if } A = 1 \end{cases}
\end{aligned}
$$

Simulate a population

Suppose we generate a population as follows.

$$
\begin{tikzcd}
L_1 \arrow[dr] \arrow[drr, bend left] & & \\
& A \arrow[r] & Y \\
L_2 \arrow[ur] \arrow[urr, bend right] & &
\end{tikzcd}
$$

$$
\begin{aligned}
L_1 &\sim \text{Normal}(\text{Mean} = 0, \text{SD} = 1) \\
L_2 &\sim \text{Normal}(\text{Mean} = 0, \text{SD} = 1) \\
A &\sim \text{Bernoulli}\left(\text{logit}^{-1}\left[-2 + L1 + L2\right]\right) \\
Y^0 &\sim \text{Normal}\left(\text{Mean} = L1 + L2, SD = 1\right) \\
Y^1 &\sim \text{Normal}\left(\text{Mean} = L1 + L2 + 1, SD = 1\right) \\
Y &= \begin{cases} Y^0 & \text{if } A = 0 \\ Y^1 & \text{if } A = 1 \end{cases}
\end{aligned}
$$

Generate a population of $N = 100,000$ cases

Simulate samples

1. Sample $n = 100$ cases from the population

Simulate samples

1. Sample $n = 100$ cases from the population
2. Apply an estimator

Simulate samples

1. Sample $n = 100$ cases from the population
2. Apply an estimator
3. Produce a sample-based estimate $\hat{\tau}$

Simulate samples

For replication $1, \ldots, R$,

1. Sample $n = 100$ cases from the population
2. Apply an estimator
3. Produce a sample-based estimate $\hat{\tau}$

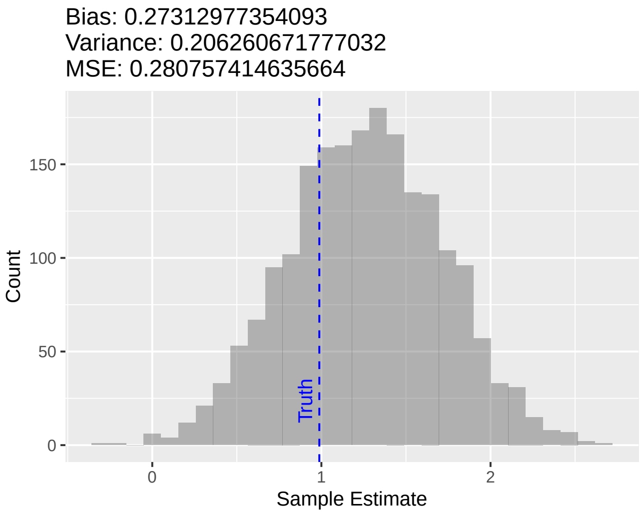
Simulate samples

For replication $1, \ldots, R$,

1. Sample $n = 100$ cases from the population
2. Apply an estimator
3. Produce a sample-based estimate $\hat{\tau}$

Compare the distribution of our sample-based estimates to the known population truth τ

Performance of the estimator: Visual summary



Performance of the estimator: Numeric summaries

Bias	$\frac{1}{R}\sum_{r=1}^{R}(\hat{\tau}_r - \tau)$ Average error
Variance	$\frac{1}{R}\sum_{r=1}^{R}\left(\hat{\tau}_r - \bar{\hat{\tau}}\right)^2$ Sampling variation
Mean Squared Error	$\frac{1}{R}\sum_{r=1}^{R}(\hat{\tau}_r - \tau)^2$ How far off, on average

Exercise: Simulate performance of matching estimators

In small groups, we will

- apply a matching estimator in a simulated setting
- and assess its performance across repeated samples

Exercise: tinyurl.com/MatchingSim

Learning goals for today

At the end of class, you will be able to:

1. Apply matching estimators for causal effects
2. Use simulation to evaluate the bias, variance, and mean squared error of an estimator

Let me know what you are thinking

tinyurl.com/CausalQuestions

Office hours TTh 11am-12pm and at
calendly.com/ianlundberg/office-hours
Come say hi!