Tools and Frameworks for Causal Inference

ECON 31720 Applied Microeconometrics

Francesco Ruggieri

The University of Chicago

September 30, 2020

1. Monte Carlo Simulations
2. Nonparametric Bootstrap
3. Frameworks for Causal Inference
4. Unobserved Determinants of the Outcome vs. Linear Regression Residual

Monte Carlo Simulations

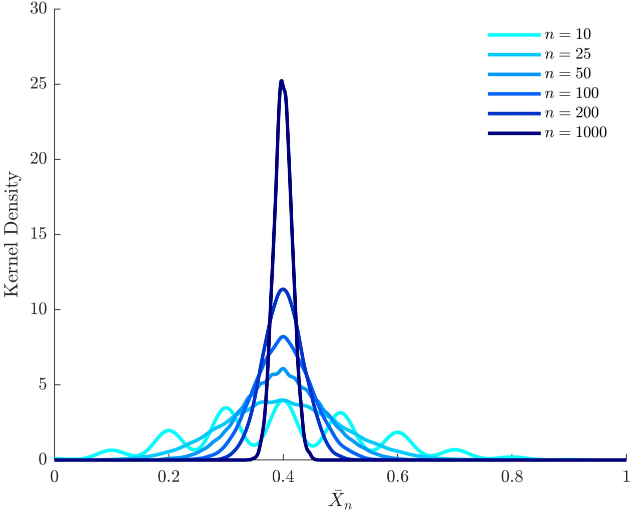
Monte Carlo experiments hinge on **repeated random sampling** to fulfill various goals:

- Provide a **numerical approximation to integrals** using empirical means
    - Applications: Method of Simulated Moments (MSM) and Maximum Simulated Likelihood (MSL)
    - The "quality" of the approximation increases with the number of simulations (via LLN)
- Investigate the **properties** of any statistical procedure via simulation

Monte Carlo Simulations: Implementation

1. Specify the number of Monte Carlo samples, $\boldsymbol{m}$, and the size of each sample, $\boldsymbol{n}$
2. Assume **knowledge of the joint distribution** of all random variables in the experiment
   - E.g., $Y = 5 + 1.6 \cdot \sin(X) \cdot \log(X) + R$, with $X \sim \mathcal{U}[0, 10]$, $R \sim \mathcal{N}(0, 9)$, and $X \perp\!\!\!\perp R$
3. For each one of the m Monte Carlo iterations:
   - **Simulate a $\boldsymbol{n}$-dimensional sample** from the joint distribution of random variables
     - E.g. $\{(Y_i, X_i, R_i)\}_{i=1}^{n}$ is an i.i.d. sample from the joint probability distribution of (Y, X, R)
   - Perform a **deterministic computation** on this sample
     - E.g. Perform local linear regression of Y on X with bandwidth $h = 0.5$, and store fitted values
4. Compute one or more **statistics** by averaging quantities across Monte Carlo samples
   - E.g. compute the Mean Squared Error of local linear regression fitted values

Monte Carlo Simulations: Example



Nonparametric Bootstrap

- A variant of Monte Carlo simulation that can be implemented with
  1. **Fewer parametric assumptions** on the Data Generating Process
  2. **Little additional code** beyond the one required to estimate the model in the first place
- Nonparametric bootstrap is typically used to compute analytically complex statistics
  - E.g. **standard errors**, when working with Continuous Mapping and Delta Method is hard
- It hinges on **sampling with replacement** from the empirical distribution of the data
  - Let $P = \{1, 2, 3, 4, 5\}$ be the population. Examples of 3-dimensional samples with replacement are $S_1 = \{4, 1, 2\}$, $S_2 = \{5, 2, 5\}$, and $S_3 = \{4, 4, 4\}$
- **Smoothness** conditions must be satisfied for nonparametric bootstrap to be valid
  - E.g. Bootstrap is invalid if applied to one-to-one propensity score matching (more later)

Nonparametric Bootstrap

- $X \in \mathbb{R}^{d_x}$ is a random vector distributed according to some **population cdf** F
- $\{X_i\}_{i=1}^n$ is a n-dimensional collection of draws from F (i.e., a **sample**)
- $\widehat{F}$ is the **empirical cdf** of X
  - This distribution places equal probability mass to each of the n draws, $\{X_i\}_{i=1}^n$
- θ is a parameter of interest that can be estimated by $\widehat{T} = T\left(\widehat{F}\left(X_1, \ldots, X_n\right)\right)$
- A **bootstrap sample** of size n, $\left\{X_i^b\right\}_{i=1}^n$, is a collection of **i.i.d.** draws from $\widehat{F}$
  - Sampling **with replacement** is necessary for $X_1^b, \ldots, X_n^b$ to be i.i.d.

Nonparametric Bootstrap

1. Specify the number of bootstrap samples $\overline{b}$
2. Let $b \in \left\{1, 2, \dots, \overline{b}\right\}$. For each bootstrap iteration:
   - Extract a **bootstrap sample** of size n, $\left\{X_i^b\right\}_{i=1}^n$
   - Perform a **deterministic computation** on this bootstrap sample
     - E.g. Regress X_1 on $X_2, \dots, X_k$ and store the coefficient associated with X_2, i.e., $\widehat{\beta_2^b}$
3. Compute one or more **statistics** by averaging quantities across bootstrap samples
   - E.g. compute the standard error associated with $\widehat{\beta}_2$ as $\sqrt{\frac{1}{\overline{b}} \sum_{b=1}^{\overline{b}} \left(\widehat{\beta_2^b} - \frac{1}{\overline{b}} \sum_{j=1}^{\overline{b}} \widehat{\beta_2^j}\right)^2}$
   - As above, approximation quality increases with the number of bootstrap samples $\overline{b}$

Nonparametric Bootstrap

Object	Data	Bootstrap
Population Distribution	F	$\widehat{F}$
Sample	$\{X_i\}_{i=1}^n$	$\{X_i^b\}_{i=1}^n$ i.i.d.
Empirical Distribution	$\widehat{F}$	$\widehat{F}^b$
Parameter	$\theta = T(F)$	$\widehat{T} = T\left(\widehat{F}\right)$
Estimator	$\widehat{T} = T\left(\widehat{F}\right)$	$\widehat{T}^b = T\left(\widehat{F}^b\right)$

Source: Lecture notes by Charles J. Geyer

Frameworks for Causal Inference

Two main frameworks for causal inference:

1. **Latent variables**, or **all-causes**, model:

$$Y = g\left(D, U\right)$$

where D and U denote the **observed** and **unobserved** determinants of Y, respectively. Together, D and U **exhaustively** cause the outcome.

2. **Potential outcomes** model:

$$Y = \sum_{d \in \mathcal{D}} Y(d) \mathbb{I}\left[D = d\right]$$

where $Y(d)$ is the counterfactual Y associated with treatment state d.

Example: from Potential Outcomes to All-Causes

- Consider a binary treatment $D \in \{0, 1\}$
- Assume that D and Y are linked by **potential outcomes** $Y(0)$ and $Y(1)$
- Let us derive a **linear all-causes model** from the potential outcomes model:

$$\begin{aligned} Y &= DY(1) + (1 - D)Y(0) \\ &= \underbrace{\mathbb{E}[Y(0)]}_{\equiv \alpha} + \underbrace{(Y(1) - Y(0))}_{\equiv \beta} D + \underbrace{Y(0) - \mathbb{E}[Y(0)]}_{\equiv U} \\ &= \alpha + \beta D + U \equiv g(D, U) \end{aligned}$$

- β could be assumed to be a **deterministic constant** (homogeneous treatment effects) or a **nondegenerate random variable** (heterogeneous treatment effects)

Example: from All-Causes to Potential Outcomes

- Maintain the assumption that the treatment is binary

- Assume a **linear all-causes model** of the observed and unobserved determinants of Y,

$$Y = \alpha + \beta D + U$$

- To derive a **potential outcomes model**, it is sufficient to define

$$\begin{aligned} Y(0) &\equiv g(0, U) = \alpha + U \\ Y(1) &\equiv g(1, U) = \alpha + \beta + U \end{aligned}$$

- Both $Y(0)$ and $Y(1)$ are random variables because U is a random variable

- In addition, β may be a random variable if the effect of D is heterogeneous

Unobserved Determinants of the Outcome vs. Linear Regression Residual

- In general, the conditional expectation function is **nonlinear**:

$$\mathbb{E}[Y|D] = h(D, \theta)$$

where $h(\cdot)$ is some function of a random vector D and a parameter vector θ

- $h(D, \theta)$ is typically unknown but can be approximated:
  - A convenient approximation choice is the “best” **linear approximation**

$$\beta^* \in \arg\min_{\beta \in \mathbb{R}^{d_d}} \mathbb{E}\left[\left(\mathbb{E}[Y|D] - D'\beta\right)^2\right] \tag{BLA}$$

  - This minimization problem is equivalent to the **linear prediction** problem

$$\beta^* \in \arg\min_{\beta \in \mathbb{R}^{d_d}} \mathbb{E}\left[\left(Y - D'\beta\right)^2\right] \tag{BLP}$$

Unobserved Determinants of the Outcome vs. Linear Regression Residual

- The first-order necessary conditions associated with both minimization problems are

$$\mathbb{E}[D\left(Y - D'\beta^*\right)] = \mathbb{E}[DU] = 0$$

where $U \equiv Y - D'\beta^*$ is a **statistical residual**

- U captures the **"quality"** of the linear approximation to $\mathbb{E}[Y|D] = h\left(D, \theta\right)$
- Being a statistical residual, U has **no causal interpretation**
- Analogously, β^* is the solution to a Mean Squared Error minimization problem

Unobserved Determinants of the Outcome vs. Linear Regression Residual

- Consider the linear causal model

$$Y = D'\beta + U$$

- If U were interpreted as encompassing the **unobserved determinants** of Y, then
  - $\mathbb{E}[DU] = 0$ would imply that observed and unobserved determinants of Y are **linearly unrelated**
  - This is not a statistical property, but a causal one, and its credibility is **assessed subjectively**

- The first part of this course will be devoted to studying cases in which $\boldsymbol{\beta^* = \beta}$
  - Under selection on observables, β^* identifies only some components of β