ECOM90025 Advanced Data Analysis

Tutorial 1

Zheng Fan

The University of Melbourne

Outline

Introduction

Zheng Fan

- Ph.D. student in Economics
- Email: fan.z@unimelb.edu.au

Seek help?

- Ed discussion board
- Consultations: refer to Canvas for details

Software

We will be using R for this subject.

- Feel free to use the local R studio, which has a lot of features that others don't.
- Colab is a free online platform provided by Google where you can execute your code (especially for Python) and text (LaTeXand Html) without any software installed.
- If you like the notebook style of Colab and want to do everything locally, I recommend downloading **Anaconda** - **Jupyter notebook**, which performs much faster than Colab and it has a nice-looking interface.

Sampling Distribution

This question refreshes you about repeated sampling in the frequentist framework. Suppose random variable

$$X_i \sim N(\mu, \sigma^2)$$

for $i = 1, ..., n$. The true values are known as $\mu = 1, \sigma = 3$. The sample size is $n = 100$.

1. Randomly generate a sample from the true Data Generating Process (DGP) and save the sample as 'X'. Set the pseudo-random number generator's seed[1] (E.g., your student ID number).
2. Draw a histogram of 'X' and a normal PDF from the true DGP as a reference curve.

[1] A seed is the starting point for random number generation. The same seed always ends up with the same randomly generated value. A fixed seed will guarantee that your results are replicable.

Sampling Distribution

This question refreshes you about repeated sampling in the frequentist framework. Suppose random variable

$$X_i \sim N(\mu, \sigma^2)$$

for $i = 1, ..., n$. The true values are known as $\mu = 1, \sigma = 3$. The sample size is $n = 100$.

3. Calculate the sample mean, its theoretic standard deviation, and standard error computed from data.

Sampling Distribution

This question refreshes you about repeated sampling in the frequentist framework. Suppose random variable

$$X_i \sim N(\mu, \sigma^2)$$

for $i = 1, ..., n$. The true values are known as $\mu = 1, \sigma = 3$. The sample size is $n = 100$.

4. **Monte Carlo study**. Simulate $B = 10000$ sets of samples from the same distribution. Each sample has 100 observations. - Draw a histogram of all B sets of sample means with the theoretical pdf of the sampling distribution of the sample mean. - Compare this figure to the one based on 'X' before. Discuss their difference.

Bootstrap

Use the simulated data 'X' and **Algorithm 1** from the textbook (nonparametric bootstrap) to simulate $B = 10000$ values of sample means.

- Draw a histogram based on bootstrapped sample means and the theoretical pdf of the sample mean.
- Compare this figure to the previous figure (Monte Carlo study).

Optional exercise

1. For the sampling distribution of the sample mean $\overline{X}$, repeat the Monte Carlo study
   - when μ changes, or
   - when σ changes, or
   - when the sample size n changes.

   What are the results indicate? More discussion.
2. What are the above implications for the bootstrap?

Optional exercise

Sampling Distribution of the Estimator of σ^2

- Estimate the variance σ^2.
- Bootstrap inference of the estimator.
- Bias correction of the estimator.

The end

Thanks for your attention!