Interactive Illustration of the Sampling Properties of Estimators

Markus Mößler

July 8, 2023

Abstract

This is the documentation of the implementation of a learning module for an interactive illustration of the sampling properties of estimators

1 Introduction

This repository contains the implementation of a learning module with interactive illustrations of fundamental statistical concepts and properties.

2 Goal of the learning module (Why?)

Understanding the concept and properties of sampling distributions of estimators is one of the most important concepts of statistical inference, i.e, of learning from data about a (population) model of interest, and thus, a fundamental part of empirical studies in social science in general and econometrics in particular. Practice, however, has shown that students often have difficulty understanding the concept of sampling distributions and their properties. One potential reason for this is that the sampling properties of estimators are often formulated and analyzed in an abstract way only. The goal of this learning module is to give a more intuitive understanding of the sampling properties of estimators using interactive illustrations of simulation results.

An example is to understand the effect of the sample size N on the sampling properties of the sample average $\overline{X}$ as an estimator for the mean μ of a random variable of interest as stated in the law of large numbers (LLN) and the central limit theorem (CLT). Using interactive illustrations of simulation results the students can increase the sample size and observe how the sample average gets closer to the mean (LLN) and how the sampling distribution of the standardized statistic of the sample average gets closer to the standard normal distribution (CLT).

3 Subject of the learning module (What?)

3.1 General

Subject of this learning module is the sampling properties of different estimators for different specifications of the data generating process (DGP). The DGP is based on a statistical model with a particular parameter of interest, e.g., the mean of a Bernoulli random variable. Based on the DGP the parameter of interest is estimated using a particular estimator, e.g., the sample average. This learning module shows the effect of changes in the DGP on the sampling distribution of an estimator.

Note, the interactive illustration of the sampling properties of the sample average for increasing the sample size N, i.e., the large sample properties of the sample mean, is only one subject of interest. Other subjects are to understand the effect of omitted variable bias (OVB) and heteroskedasticity on the sampling distribution of the OLS estimator in a simple linear regression model.

3.2 Bernoulli Distribution and Sample Average

3.3 Continuous Uniform Distribution and Sample Average

3.4 Cross Section Linear Regression Model and Ordinary Least Squares

3.4.1 Parameterization and the Effect of Sample Size

3.4.2 Effect of Heteroskedasticity

3.4.3 Effect of Omitted Variable Bias

4 Method/implementation of the learning module (How?)

For an interactive and immediate user experience it is useful to separate the simulation study and the interactive presentation of the simulation results.

This two-step procedure allows also a flexibel implementation. I.e., the simulation studies can be conducted with any software such as *R*, *python*, etc. and the interactive presentation of the results can be achieved using basic web development languages such as `.html`, `.css` and `.js`.

Simmulation study and illustration/reporting of the simulation results:

- The DGP based on a given statistical model is simulated and the values of a given estimator is calculated.

- The simulation results, i.e., the sampling distribution of the estimator, are illustrated using barplots, scatterplots and/or histograms.

- Other simulation results can be reported using tables.

- The simulation study is performed for different specifications of the statistical model or for different estimators.

- For each specification or estimator the results, e.g., figures and/or tables are save in different `.svg` and/or `html` files.

- The simulation study is performed using the programming language for statistical computing and graphics *R*.

Interactive presentation of the simulation results:

- The results for different specifications, e.g., figures and/or tables, are interactively embedded in a `.html` file and linked to a slider input tag.

- The effect of different specifications can be studied by changing the slider, i.e., the specification, and thus, the embedded results.

- The interactive integration of the illustrations into the `.html` file is based on javascript.

Additional material, e.g., for explanation purposes:

- Verbal text or mathematical formula.

- Audio or video explanations

- Any additional can be added into the `html` file interactively or non-interactively.

Integration in the lecture:

- The material of this learning module can be provided on a gradual basis using links to the specific illustrations/sub modules or as a complete course/module with a starting page and links to the sub modules.

- The material can be hosted on *GutHub* or on a learning platform such as *ILIAS*. The easiest way to host the material on a learning platform such as *ILIAS* is using a import interface for `.html` structures. In the case of the learning platform *ILIAS* this procedure is quite easy and flexible.

Structure of the learning module:

- The learning module contains different sub modules where each sub module has a specific learning goal, e.g., "understand the effect of the sample size on the sample properties of the sample average to estimate the mean of a Bernouli distribution".

- The material of a sub module is collected in a sub folder, e.g., `lln-clt-ber`.

- The sub folder contains:

    - `.R` file, e.g., `lln_clt_ber_ber_rscript.R`, with the simulation study and the results stored in the `figures` and/or `tables` subdirectory
    - `figures` subdirectory with the illustrations of the simulation results
    - `tables` subdirectory with the reports of the simulation results (optional)
    - `.html` file, e.g., `lln_clt_ber_html.html`, with the interactive representation of the illustrations
    - `myScript.js` with the javascript for the interactive illustrations
    - `myStyle.css` with the css styles for the interactive illustrations
    - Additional assets, e.g.,:
        * `.png` file with a logo for the header of the `.html` file
        * ...

References

[SW19] James H. Stock and Mark W. Watson. *Introduction to Econometrics.* Pearson, 2019.