Interactive Illustration of the Sampling Properties of Estimators

Markus Mößler

July 19, 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 and animated 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 the underlying data generating process (DGP) formalized in probabilistic (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 and animated illustrations of simulation results.

One example is to understand the effect of increasing 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).

The topics, i.e. the sample distribution of estimators, is part of every statistics and econometrics course and can be found in any introductory textbook for this field. However, by using interactive and animated illustrations the results we aim to provide a deeper and more intuitive understanding of these concepts. We believe that this kind of understanding is hard to achieve by just attending a *lecture*, reading a *textbook*, or, by chatting with *ChatGPT* about this topic. Certainly, another way to achieve this is to provide the implementation of the simulation studies or to let the students to implement the simulation studies themselves. However, this is often too big a hurdle, especially for undergraduate students.

3 Subject of the learning module (What?)

3.1 General

Subject of this learning module are the sampling properties of estimators for different data generating processes (DGPs). The DGPs are based on a statistical model with a particular parameter of interest, e.g., the mean of a Bernoulli random variable. The parameter(s) of interest of the DGPs are estimated using a particular estimator, e.g., the sample average. This learning module shows the effect of changes in the DGP, e.g., increasing the sample size n, 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 Univariate random variables and sample average

Bernoulli distribution and sample average

This illustration shows the effect of changing the sample size n and the probability of success p of the Bernoulli distribution on the sampling distribution of the sample average as estimator for mean of the Bernoulli distribution.

...

Continuous uniform distribution and sample average

This illustration shows the effect of changing the sample size n and the lower bound a and the upper bound b of the continuous uniform distribution on the sampling distribution of the sample average as estimator for mean of the continuous uniform distribution.

...

3.3 Linear regression model and ordinary least squares

Illustration of the properties of the OLS estimator to estimate the slope coefficient β_1 of a linear regression model, i.e.,

$$Y_i = \beta_0 + \beta_1 X_i + u_i \tag{1}$$

Sampling distribution and sample size

This illustration shows the effect of increasing the sample size n on the sampling distribution of the OLS estimator for the slope coefficient of a simple linear regression model.

...

Sample Size and parameterization of the DGP

This illustration shows the effect of changing the parameters of the DFP, i.e.,

1. changing the sample size n,
2. changing the variance of u_i, i.e., $\sigma^2_{u_i}$, and,
3. changing the variance of X_i, i.e., σ^2_X,

on the sampling distribution of the OLS estimator for the slope coefficient of a simple linear regression model.

...

Effect of Heteroskedasticity

...

Effect of Omitted Variable Bias

...

4 Method of the learning module (How?)

4.1 General

For an interactive and immediate user experience it is useful to separate the simulation study and the presentation of the results. This two-step procedure allows also a flexible implementation, i.e., the simulation studies can be conducted with any suitable software, e.g., *R*, *python*, etc. and the results can be presented interactively using basic web development languages, i.e., *html*, *css* and *javascript*.

Appealing, flexible, interactive and animated presentation using html, css *and* javascript

While the setup and the implementation of the simulation or other studies depends on the topic of the learning module, the presentation of the results should be based on core web development tools. The content of each learning module is presented

using *html*, the most popular language to build web pages.[1] This allows a very flexible presentation of the results. The styling of the learning module is based on *css*, a popular language for styling web pages, and on *bootstrap*, a popular *css* framework. The interactive and animated feature of the learning module is based on *javascript*, a popular language for programming web pages. The *bootstrap* framework with its grid system allows the development of responsive, mobile-first websites with an appealing design. This enables an appealing experience of the learning also on mobile devices.

Structure of the html, css *and* javascript *files*

What follows in Section 4.3 to 4.3 is an explanation of the structure of the *html*, *css* and *javascript* files. The explanations are based on the illustration of the properties of the sample average as estimator for the mean of a Bernoulli random variable, i.e., based on module `ber-dis-sam-ave` which has two parameters (the sample size n and the probability of success p) and three figures (the bar plot of number of ones and zero, the histogram of sample average and the histogram of the standardized sample average). Note, the implementation is flexible in the sense that it can be used for any learning module with two parameters and three figures.

4.2 Structure of the file with the simulation study

Depending on the subject of the leaning module the number of interactive parameters and the number of figures and/or tables for the illustration of the results are choosen. For the learning module `ber-dis-sam-ave` there are two interactive parameters, i,e, sample size n and probability of success p and three figures, i.e., the bar plot of number of ones and zero, the histogram of sample average and the histogram of the standardized sample average.

In this setup the variable and fixed inputs of the simulation study can be defined as in *R* Code Snippet 1. The simulation study is based on a loop which runs over the

[1] An introduction to the most important web page development languages can be found here: https://www.w3schools.com/

variable inputs (here `n.vec` and `p.vec`). For each variable input combination `RR=10000` realizations of the DGP are simulated and `RR=10000` estimates are calculated. Based on the simulation results for `RR=10000` realizations the illustration of interest is stored as `figure_01_y_z.svg`, where `0x` indicates the `0x`th figure, `y` indicates the `y`th index value of the first parameter and `z` indicates the `z`th index value of the second parameter. E.g. for `ii = 1` and `jj = 1` the results for the parameters $n = 5$ and $p = 0.2$ are stored in `figure_01_1_1.svg` (see also *R* Code Snippet 2 below). Note, for the integration into the *html* file and the animation using *javascirpit* file later the naming of the figure and/or tables is important (see Section 4.3).

4.3 Structure of the *html* file

Integration of the figures

The figure are integrated using an *img* tag with an *src* attribute with the link to the figure following a certain naming convention (see Section 4.2 above) and with an *id* attribute also following a certain naming convention (see below). The *html* Code Snippet 3 below shows the integration of figure 1, i.e., the bar plot of zeros and ones of the module `ber-dis-sam-ave` for the parameterization $n = 25$ and $p = 0.4$. The *id* attribute `imageL1N1Id` will be used in the *javascript* file to change the source link of the *src* attribute. The other figures are integrated accordingly, e.g. see *html* Code Snippet 4 below for figure 2.

Integration of parameter sliders

Slider are integrated using an *input* tag

A *bootstrap-slider* input[2] for each interactive parameter is added to change the *src* attribute of each *img* tag, i.e., the figure shown in the *html* file. The *id* attribute `imageL1N1Id` will be used in the *javascript* file to change the source link of the *src* attribute.

[2]See: https://github.com/seiyria/bootstrap-slider

Code Snippet 1: *R* code snippet simulation setup

```
# inputs (variable)
n.vec <- c(5, 10, 25, 50, 100)
p.vec <- c(0.2, 0.4, 0.6, 0.8)

# inputs (fixed)
RR <- 10000
```

Code Snippet 2: *R* code snippet simulation and illustration implementation

```
# simulation/illustration

for (ii in 1:length(n.vec)) {

        for (jj in 1:length(p.vec)) {

                NN <- n.vec[ii]
                p <- p.vec[jj]

                # simulation
                tmp.sim <- Y_bar_ber_sim_fun(RR = RR, NN = NN, p = p)

                # plot no 01
                plt.nam <- paste(fig.dir, "figure_01_", ii, "_", jj, ".svg", sep = "")
                svg(plt.nam)

                ...

                dev.off()

                ...

        }

}
```

Code Snippet 3: *Html* code snippet for integration of figure 1

```
<img src="./figures/figure_01_3_2.svg" alt="" id="imageL1N1Id" style="text-align: center; max-width: 100%;">
```

Code Snippet 4: *Html* code snippet for integration of figure 2

```
<img src="./figures/figure_02_3_2.svg" alt="" id="imageL1N1Id" style="text-align: center; max-width: 100%;">
```

Code Snippet 5: *Html* code snippet for slider 1

```
<p style='font-size: 12pt; text-align: center;'>Sample size \(n\)</p><p style='font-size: 12pt; text-align: center' id="sliderValue01Id"></p>
<input id="slider01Id" type="text" data-slider-min="0" data-slider-max="4" data-slider-step="1" data-slider-value="2"/>
```

For each figure and parameter a *audio* file with an explanation of the figure and the simulation results can be added

Other content such as explanations as plain text or formula or as audio files can be added and can be manipulated, e.g. made visible/invisible inside tabs or collapsibles using *javascript*.

4.4 Structure of the *javascript* file

The values of the parameters can be changed using *bootstrap-slider* inputs[3] and the *javascript* `par02Fig03Interactivation.js`. For example for the input of the sample size the following *html* code snippet is used.

Code Snippet 6: *Html* code snippet for *bootstrap-slider* input

```
<p style='font-size: 12pt; text-align: center;'>Sample size \(n\)</p><p style='font-size: 12pt; text-align: center' id="sliderValue01Id"></p>
<input id="slider01Id" type="text" data-slider-min="0" data-slider-max="4" data-slider-step="1" data-slider-value="2"/>
```

Changing the slider input values invokes *javascript* function.

Code Snippet 7: *Javascript* code snippet for interactive sliders (no 1): Initialize slider values

```
var nVec = [5, 10, 25, 50, 100]

// set initial value for slider 1
document.getElementById("sliderValue01Id").innerHTML = "\\(n = " + nVe
var element = document.getElementById("sliderValue01Id");
MathJax.typeset([element]);
```

Code Snippet 8: *Javascript* code snippet for interactive sliders (no 2): Update slider values

```
var slider01 = new Slider("#slider01Id", {
        tooltip: "never",
        formatter: function(value) {
                return nVec[value]
        }
});
```

[3] See: https://github.com/seiyria/bootstrap-slider

The interactive experience of the illustrations is based on three *javascript* function which are linked to the sliders of the parameters. There is a

4.5 Explained and animated user experience using *javascript*

The following explanations are based on the illustration of the properties of the sample average as estimator for the mean of a Bernoulli random variable, i.e., based on module `ber-dis-sam-ave`. The explanation of the figure is based on a simple *javascript* function `explainButto01Click`.

The explained animation of the parameters is based on the *javascipt* function `sliderLoop()` and the functions `animateButto01Click()` and `animateButto02Click()`. Note, for each parameter a function `animateButtoXXClick()` is used, i.e., for the module `ber-dis-sam-ave` there are two functions, `animateButto01Click` and `animateButto02Click`. Furthermore, the body of both functions, `sliderLoop()` and `animateButtoXXClick()` depends on the number of figures (for `ber-dis-sam-ave` equals three) and the number of parameters (for `ber-dis-sam-ave` equals two) used in the module.

4.6 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 *GitHub* 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.

Code Snippet 9: *Javascript* code snippet for interactive sliders (no 3): Change illustration

```
slider01.on("slide", function() {

        var sliderValue01 = slider01.getValue() + 1;
        var sliderValue02 = slider02.getValue() + 1;

        document.getElementById("imageL1N1Id").setAttribute("src", "./figures/figure_01_" + sliderValue01 + "_" + sliderValue02 + ".svg");
        document.getElementById("imageL1N2Id").setAttribute("src", "./figures/figure_02_" + sliderValue01 + "_" + sliderValue02 + ".svg");
        document.getElementById("imageL1N3Id").setAttribute("src", "./figures/figure_03_" + sliderValue01 + "_" + sliderValue02 + ".svg");

        document.getElementById("sliderValue01Id").innerHTML = "\\(n = " + nVec[slider01.getValue()] + "\\)";
        var element = document.getElementById("sliderValue01Id");
        MathJax.typeset([element]);

})
```

Code Snippet 10: *Javascript* snippet for explanation

```
explainButto01Click = function(audioId) {
        var audio = document.getElementById(audioId);
        audio.play();
}
```

Code Snippet 11: *Javascript* snippet for explanation and animation (part I)

```
function sliderLoop(loopSliderValue01, loopSliderValue02) {

        var sliderValue01 = loopSliderValue01 + 1;
        var sliderValue02 = loopSliderValue02 + 1;

        document.getElementById("imageL1N1Id").setAttribute("src", "./figures/figure_0
        document.getElementById("imageL1N2Id").setAttribute("src", "./figures/figure_0
        document.getElementById("imageL1N3Id").setAttribute("src", "./figures/figure_0

        document.getElementById("sliderValue01Id").innerHTML = "\\(n = " + nVec[slider
        var element = document.getElementById("sliderValue01Id");
        MathJax.typeset([element]);

        document.getElementById("sliderValue02Id").innerHTML = "\\(p = " + pVec[slider
        var element = document.getElementById("sliderValue02Id");
        MathJax.typeset([element]);

        var slider = document.getElementById("slider01Id");
        slider.value = sliderValue01;

        var slider = document.getElementById("slider02Id");
        slider.value = sliderValue02;

};
```

Code Snippet 12: *Javascript* snippet for explanation and animation (part II)

```
animateButto01Click = function(org, start, stop, audioId) {
        var audio = document.getElementById(audioId);
        audio.play();
        var ind = start;
        var loopSliderValue02 = slider02.getValue();
        var outerInterval = setInterval(function() {
                var loopSliderValue01 = ind;
                slider01.setValue(ind);
                sliderLoop(loopSliderValue01, loopSliderValue02);
                ind++;
                if (ind > stop) {
                        var innerInterval = setInterval(function() {
                                slider01.setValue(ind);
                                slider01.setValue(org);
                                sliderLoop(org, loopSliderValue02);
                                clearInterval(innerInterval);
                        }, 1000);
                        clearInterval(outerInterval);
                }
        }, 1000);
}
```

Code Snippet 13: *Javascript* snippet for explanation and animation (part III)

```
animateButto02Click = function(org, start, stop, audioId) {
        var audio = document.getElementById(audioId);
        audio.play();
        var ind = start;
        var loopSliderValue01 = slider01.getValue();
        var outerInterval = setInterval(function() {
                var loopSliderValue02 = ind;
                slider02.setValue(ind);
                sliderLoop(loopSliderValue01, loopSliderValue02);
                ind++;
                if (ind > stop) {
                        var innerInterval = setInterval(function() {
                                slider02.setValue(ind);
                                slider02.setValue(org);
                                sliderLoop(loopSliderValue01, org);
                                clearInterval(innerInterval);
                        }, 1000);
                        clearInterval(outerInterval);
                }
        }, 1000);
}
```

4.7 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 Bernoulli distribution".
- The material of a sub module is collected in a sub folder, e.g., `ber-dis-sam-ave`.
- The sub folder contains:
  - `.R` file, e.g., `ber_dis_sam_ave.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
  - `audios` subdirectory with the audio text and files for the explanations using interactive animations
  - `tables` subdirectory with the reports of the simulation results (optional)
  - `.html` file, e.g., `ber_dis_sam_ave.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
    - ...

4.8 Structure of the `.html` file

The content is divided across four bootstrap containers:

- 1st Container: Header of the module
- 2nd Container with three sections

  - Topic of the module
  - Data generating process (DGP)
  - Estimator and parameter of interest

- 3rd Container:
  - Parameter panel with the slider to change the parameter of interest
  - Illustration panel with four tabs for the illustration of
    * Sample draw: Particular outcome
    * Residual: Particular diagnostic
    * Estimates: Histogram of parameter estimates
    * Std. Estimates: Histogram of standardized parameter etimates

- 4th Container: More Details
  - Simulation Exercise:

References

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