

Topic #3 Assignment

ECON 5753 — University of Arkansas

Prof. Kyle Butts

These assignments should be completed in groups of 1 or 2 but submitted individually.

Overview of Monte Carlo Simulation

The goal of this assignment is for you to write a Monte Carlo simulation to better understand the sampling distribution of the ordinary least squares estimator.

Our simulation will involve generating data for $n = 100$ individuals. For each individual, we will generate two covariates, X_{i1} and X_{i2} and an error term ε_i . How we generate these will vary across different versions of the simulation. With those, we will generate these according to the linear model $y_i = 10 + X_{i1} * 1 + X_{i2} * 1 + \varepsilon_i$, i.e. the true value of $\beta = [\alpha \ \beta_1 \ \beta_2]'$ is $\beta_0 = [10 \ 1 \ 1]'$.

We will generate $B = 1000$ datasets and for each dataset we will estimate the ordinary least squares model using `mod <- lm(y ~ 1 + X1 + X2, data = df_b)`. With the model, we can grab the three estimates using `coef(mod)`.

We want to store each iteration's estimates. At the top of the simulation, we will create a matrix using `ests <- matrix(nrow = B, ncol = 3)`. Then, for each iteration, we will store the coefficients into the b -th row using `ests[b,] <- coef(mod)`.

Here is an outline of what the code will look like, with only the data-generation part left for you:

```
B <- 1000
n <- 100
ests <- matrix(nrow = B, ncol = 3)
for (b in 1:B) {
  # Generate data

  # Estimate OLS
  mod <- lm(y ~ 1 + X1 + X2, data = df_b)
```

```
ests[b, ] <- coef(mod)
}
```

Then, we can visualize the sample distribution as follows. I will use ‘ggplot2’ because it makes nicer looking graphs.

```
library(ggplot2)
# Marginal sample distribution
ggplot() +
  geom_histogram(aes(x = ests[, 2]))
ggplot() +
  geom_histogram(aes(x = ests[, 3]))

# Joint sampling distribution of  $\hat{\beta}_1$  and  $\hat{\beta}_2$ 
ggplot() +
  geom_bin_2d(aes(x = ests[, 2], y = ests[, 3]), bins = 50)
```

This starter template is available in the script 03-assignment_template.R

We are going to do a bunch of simulations. I encourage you to try out different values and explore; this will help build your intuition. I don’t need all of your code for this, but do want to see at least one example. In other words, I would recommend having code for the "baseline" case and then a second code that you change around. That way you can make comparing to the baseline case easier.

For the error term, we will draw from the normal distribution with `rnorm`. Unless otherwise stated, draw `rnorm(n, mean = 0, sd = 1)`. For the two covariates, we want to be able to change the correlation between them (zero and non-zero). To do so, we will use the `rmvnorm` function from the `mvtnorm` package. It requires you to pass a vector of means (in our case $[1 \ 1]'$) and the variance-covariance matrix of X_1 and X_2 . For example, here we are drawing X_1 and X_2 with a positive covariance:

```
Sigma_X = matrix(c(1, 0.5, 0.5, 1), nrow = 2, ncol = 2)
X <- rmvnorm(n = n, mean = c(1, 1), sigma = Sigma_X)
```

Simulations

1. First, generate the two covariates independently (i.e. sigma is the identity matrix). Think of this as the "baseline" simulation. We will adjust many of the models and want to compare to this. Take note of (1) the standard deviation of the sample distribution of each coefficient and (2) the shape of the joint distribution of coefficients
2. Second, let's experiment with changing the sample size. Try out a few values $n = 10, 100, 1000, 10000$ and comment on what happens to the standard deviation of the sample distribution and the joint distribution.
3. Third, let's try changing the variance of the error term. Try $\sigma_\varepsilon^2 = 1, 2$, and 3 . Comment on what happens to the standard deviation of the sample distribution and the joint distribution.
4. Fourth, let's experiment with the variance-covariance matrix of X . Try simulations with the following:

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0.25 \\ 0.25 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0.99 \\ 0.99 & 1 \end{bmatrix}, \text{ and } \begin{bmatrix} 1 & -0.25 \\ -0.25 & 1 \end{bmatrix}$$

Comment on what happens to the standard deviation of the sample distribution and the joint distribution.

5. For the case of an extreme positive correlation (0.99), I want you to think about why the results look the way they do. Try to explain why in your own words. Does this change as we increase n ?