The purpose of this session is to introduce you to the large sample behaviour of OLS estimator and provide an illustration for the Central Limit Theorem.

Weak Law of Large Numbers and Central Limit Theorem

Consider four different distributions

1. Normal distribution $u_1 \sim \mathcal{N}(3, 1)$
2. Student-t distribution $u_2 \sim t(5)$
3. Chi-squared distribution $u_3 \sim \chi^2(2)$
4. Gamma distribution $u_4 \sim \Gamma(1, 0.5)$

a) Plot theoretical densities of u_j, $j = 1, .., 4$. Note, Gamma and Chi-squared distributions are defined only on $\mathbb{R}^+$.

b) In order to examine whether the WLLN holds, simulate $n = 10000$ observations from each of the four distributions. Define a function which computes a sample moving average of an $n \times 1$ vector. Plot the moving average evolution over n for each distribution.

c) Simulate a distribution of the sample mean for each u_j, $j = 1, .., 4$. For each mean, compute the stabilizing transformation and plot its histogram together with $\mathcal{N}(0, 1)$ density.

d) Does the CLT hold for $u_5 \sim t(1)$? What about WLLN?

Exercises to solve at home

Illustrate conclusions of Theorem 3.2 and Theorem 3.3 with the help of a simulation study. For this exercise set the number of simulation draws to $mc = 10000$. Consider a bivariate linear regression model:

$$y_i = 1 - 3x_i + u_i, \tag{1}$$

where $u \sim \mathcal{N}(0, 10)$, $x \sim \mathcal{N}(5, 16)$.

a) Simulate the distribution of the estimated OLS coefficients for slope and intercept (over mc simulation draws) for $n = 10$. For each simulation step save the theoretical variances of the OLS coefficient, i.e. $V\left[\hat{\beta}|X\right] = \sigma_u^2(X'X)^{-1}$.

b) Standardize the coefficients (re-center and scale with the $\sqrt{V\left[\hat{\beta}|X\right]}$). Plot the histograms of the stabilizing transformation of estimated slope and intercept next to each other. Use `breaks = 50` option for histograms. On top of each histogram draw a theoretical density of the standard normal distribution and Student-t distribution with $n-2$ degrees of freedom. Compute the densities over an equidistant grid {-5, -4.8, ..., 4.8, 5}. Which distribution best describes the distribution of the estimated OLS coefficients?

c) Repeat b) for a different standardization: $\sqrt{\hat{V}\left[\hat{\beta}|X\right]}$, i.e. re-run the simulation and replace σ_u^2 with its estimate $\hat{\sigma}_N^2 = \frac{\sum_{i=1}^n \hat{u}_i^2}{n-2}$. Which distribution is a better fit for this case?

d) Increase the sample size to $n = 5000$. Explain this result using Theorems 3.2 and 3.3.

Compare your results with the homework answers which will be available on Blackboard on Friday.