Computer Exercise 1: Monte Carlo Study

In this computer session, we will investigate the estimators and the t-test statistic for the linear regression model under different sample sizes (different Ns):

$$y_i = \beta_1 + \beta_2 x_i + \epsilon_i, \qquad i = 1, 2, \ldots, N, \quad \epsilon_i \sim i.i.d.N(0, \sigma^2), \tag{1}$$

where $\beta_1 = 1$, $\beta_2 = 2$ and $\sigma^2 = 100$.

1. Set the working directory which contains the program file. For sample size $n = 5$, set the number of simulations $s = 100$ and run the program by `source('mc2.R')`. Check the sample averages of $b1hat$, $b2hat$, $sigmahat$ and $sigmasqhat$. Are they close to the true values? You can plot the histograms of these estimates and carry out Jarque-Bera tests.

2. What is the theoretical value of $t1err$ (probability or frequency of making Type I error in testing H_0 : $slope = slope_{true}$) given the level of significance $a = 5\%$? What is the actual value you obtain?

3. How will you interpret the seemingly inverted bell shape curve of the power function for H_0 : $slope = slope_{wrong}$? What is the minimum value of bp? What value in $btest$ does this minimum correspond to? Is this value far away from the true value? How is the minimum compared to the level of significance?

4. Run the program again without changing anything. Are the new results very different from the results obtained before?

5. Change the number of simulations (s) to $1e4$ and re-run the program. Recalculate the estimates above. Are the sample averages closer to the true values with the increase of s? What about $t1err$? How does the curve of power function look like? For large $!s$, can you observe any big changes in the results if we run the program again? For large $!s$, can you observe any big changes in the results if we run the program again? Are there any biased estimators?

6. If you think the estimates are not accurate enough and the power curve is not smooth enough, you can continue to increase $!s$ bearing in mind

that it will take longer for the computer to finish running the program. Now decide the number of simulations (s) for the rest of the questions which you think the computer can manage and is sensible to obtain Monte Carlo results. For $n = 5$, plot the histograms for $b1hat$, $b2hat$, $sigmahat$ and $sigmasqhat$. Do you think any of them follow normal distribution?

7. Increase the sample size (n) to 20. Run the program and investigate the estimators: $b1hat$, $b2hat$, $sigmahat$ and $sigmasqhat$ (their sample averages and their histograms). What can you observe about the power function?

8. Repeat 7 with sample size (n) equal to 50, 100,.... Are all the estimators consistent? Can you say that the t-test is a consistent test?

9. Do you think if we change the true values of the intercept, slope and the variance of the error term, our results obtained above will change? Verify your claim by experiments. What about the change of significance level (a)? Will it affect anything?