Homework 1
Simulation and Performance Evaluation – University of Trento

You can solve the following assignments using any programming language. In doing so, make sure to implement the formulas explained in class, and try not to use functions made available by the languages to achieve the required tasks.

Exercise 1

Load the iid data samples from the CSV file `data_ex1.csv` and provide:

1. the median and a 95% confidence interval for the median

2. the mean, a 95% and a 99% confidence interval for the mean

Explain which formulas you used and why.

Exercise 2

Load the data from the CSV file `data_ex2.csv`. This should be a matrix with 6001 rows and 200 columns. All entries are iid.

1. Find a 95% confidence interval for the data of the first row.

2. Find the mean of all the remaining 6000 rows, and check how many of these means fall inside the confidence interval you computed for the first row. Is this what you expected? Explain.

Exercise 3

Load the data from the CSV file `data_ex3.csv`.

1. Compute the CoV for the data, Jain's fairness index and the Lorenz curve gap

2. Plot the Lorenz curve; discuss the correspondence with the values of the previous point, and the difference between the values of the Lorenz curve gap and of Jain's fairness index.

3. Implement the bootstrap algorithm seen in class and use it to compute 95% and 99% confidence intervals for:

   (a) the Lorenz curve gap;
   (b) Jain's fairness index;
   (c) the mean of the data;
   (d) the standard deviation of the data,

4. Compare the obtained confidence interval for the mean with the one obtained via the asymptotic formulas. Discuss.

Exercise 4

Load the data from the CSV file `data_ex4.csv`.

1. Give a 95% and a 99% confidence interval for the mean using the bootstrap method.

2. Compare the results against the value obtained from the asymptotic formulas. Explain.

3. Transform the data using a log transformation and give the a confidence interval for the transformed mean.

Exercise 5

Load the data from the CSV file `data_ex5.csv`. Note that this data represents successes (ones) and insuccesses (zeros) of some experiment.

1. Find the probability of success and confidence intervals for it at both a 95% and at a 99% level.

2. Consider now only the first 15 data points. Find a confidence interval for the probability of success. Explain.