Network analysis and simulation
Homework 1

Davide Peron

Exercise 1

In the first exercise two datasets are given and on those, a bunch of figures have been plotted, in which the data are showed and different measures of confidence are calculated on them.

The first dataset is a collection of execution times resulting from a first version of a program and a second more optimized one. In the follow the said figures are reported.

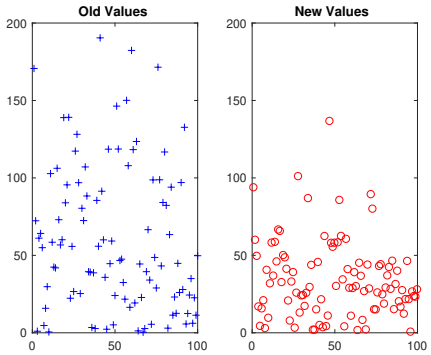

Figure 1: Plot of the data

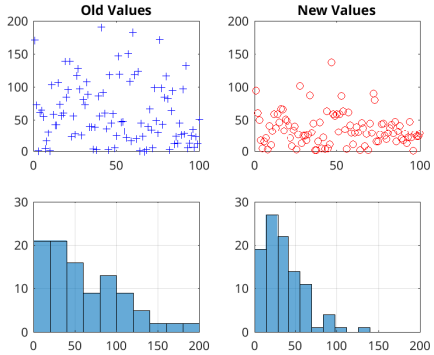

Figure 2: Data plotted also in histograms divided in 10 bins (Figure 2.1)

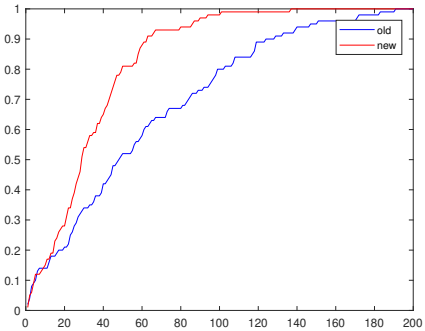

Figure 3: Empirical distribution function of the data (Figure 2.2)

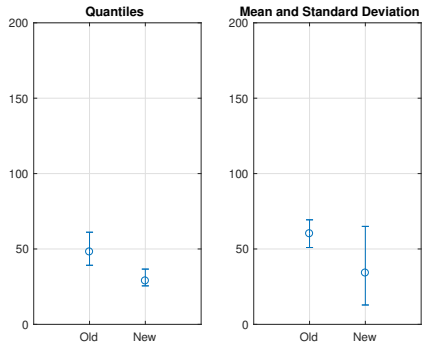

Figure 4: Box Plots of the data with Confidence Interval (CI) for median and mean (Figure 2.3)

Exercise 2

Executing the script correspondent to the second exercise, we found that 56 experiments the CI does not contain the true value of the mean.

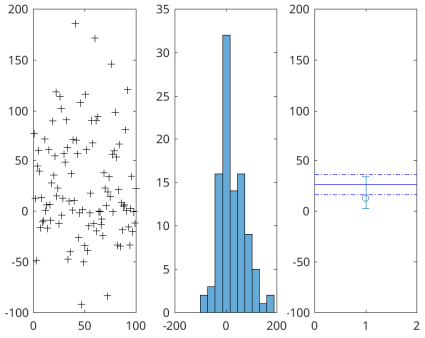

Figure 5: Difference between old and new data (Figure 2.7)

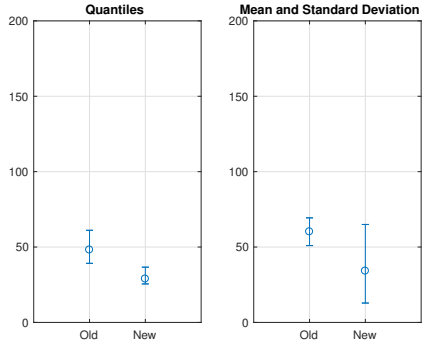

Figure 6: Box Plots of the data with CI for median and mean (Figure 2.3)

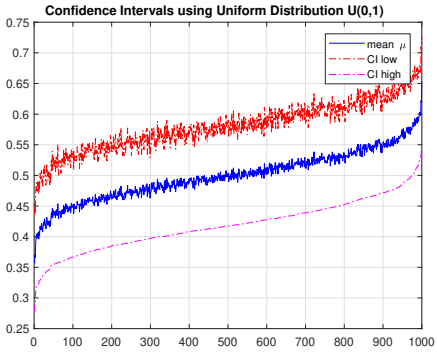

Figure 7: Results of the experiment with $n = 48$

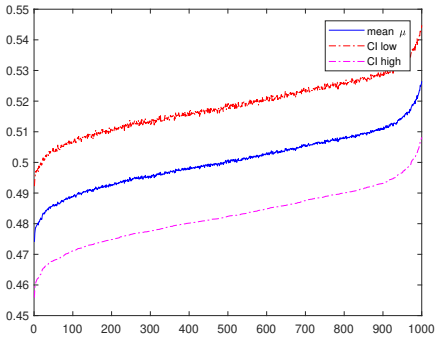

Figure 8: Results of the experiment with $n = 1000$

In Figure 7 and Figure 10 are reported the value of the sample mean and its CI (with $\gamma = 0.95$) for each experiment, sorted based on the lower extreme of the CI and using a different number of random variables in each experiment. In both cases, the sample mean is distributed around the true mean (that for this Uniform Distribution is 0.5).

Note how, increasing the number of random variables per experiment, the mean width of the CI get lower. Furthermore, the width of the CI is not constant.

Exercise 3

Exercise 4

In Figure 9 is plotted the accuracy of the sample mean versus the number of random variables in each experiment. This measure have been made based on the following formula:

$$A_i = |\overline{x_i} - \overline{x}| \qquad \text{(Equation 1)}$$

where A_i is the accuracy of the experiment i and x_i is the sample mean of the same experiment. As can be seen in the said figure, the accuracy get lower as the number of random variables in each experiment increase, this happens since the higher is the number of random variables, the higher is the precision of the experiment.

Exercise 5

Redoing Exercise 2 the plot in Figure 11 and Figure 12 have obtained. Note that the sample mean now is distributed around the new true value, that is 0.

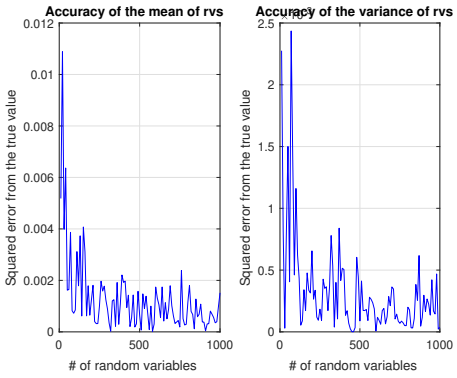

Figure 9: Accuracy of the estimation versus n

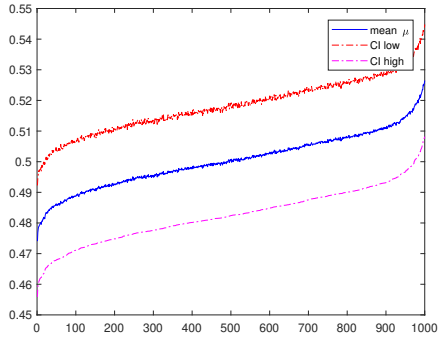

Figure 10: Results of the experiment with $n = 1000$

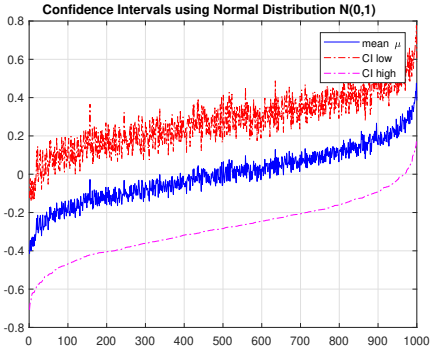

Figure 11: Results of the experiment with $n = 48$ rvs $N(0, 1)$

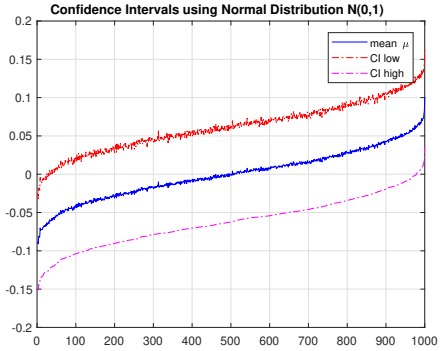

Figure 12: Results of the experiment with $n = 1000$ rvs $N(0, 1)$

For the Exercise 4 the results are reported in Figure 13 and Figure 14.

Is visible how, also in this case, the accuracy get higher (that is, the distance between the sample mean and the true value get lower) and the precision of the variance get higher (the CI becomes smaller) as n increases. Now the sample variance oscillates around 1 that is in fact the new true value.

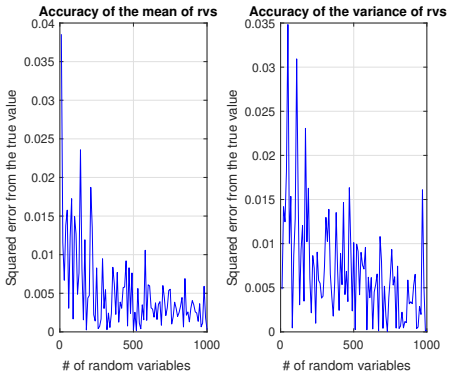

Figure 13: Accuracy of the estimation versus n using rvs $N(0, 1)$

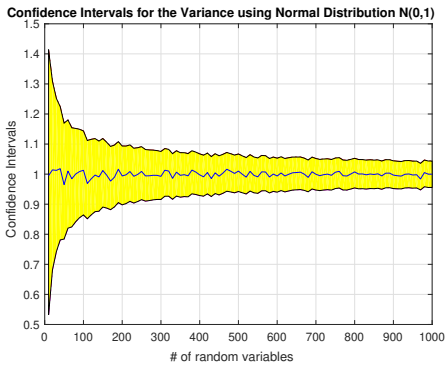

Figure 14: Confidence intervals for the variance using rvs $N(0, 1)$