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.

In the follow the said figures are reported.

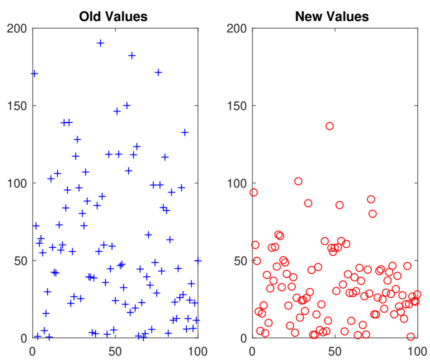

Figure 1: Plot of the data

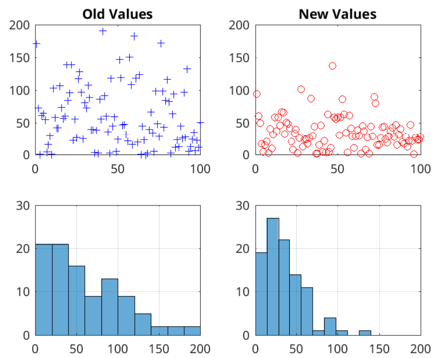

Figure 2: Data plotted also in histograms divided in 10 bins (Figure 2.1)

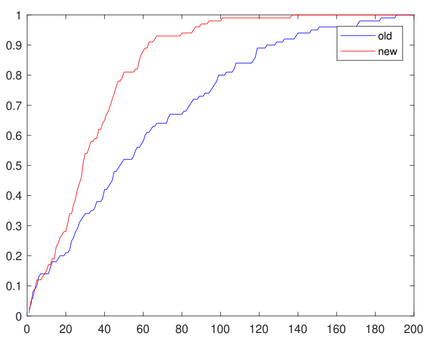

Figure 3: Empirical distribution function of the data (Figure 2.2)

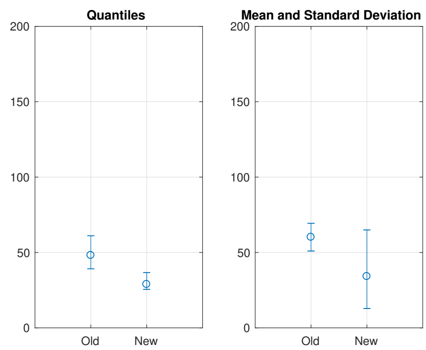

Figure 4: Box Plots of the data with Confidence Interval (CI) for median and mean (Figure 2.3)

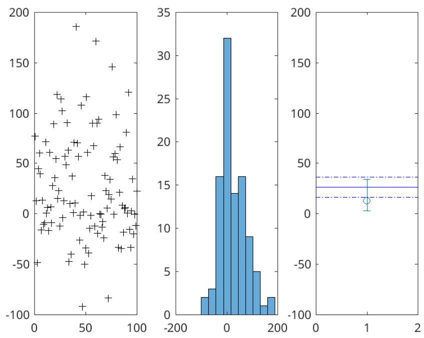

Figure 5: Difference between old and new data (Figure 2.7)

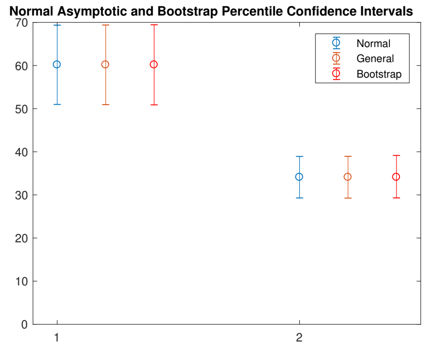

Figure 6: Normal Asymptotic and Bootstrap Percentile Confidence Intervals (Figure 2.8)

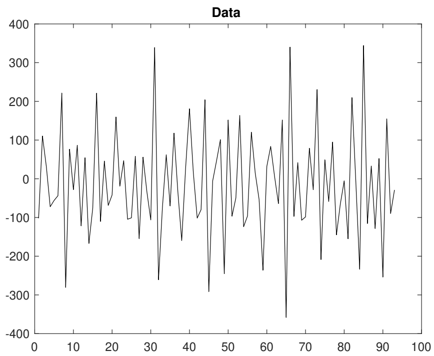

Figure 7: (Figure 2.10a)

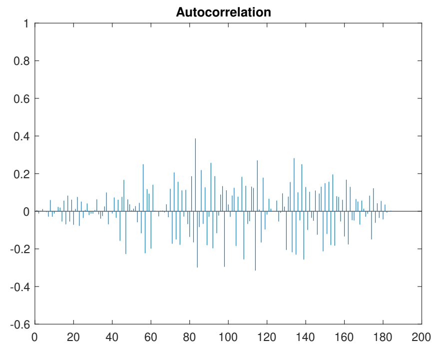

Figure 8: Autocorrelation of the plot in Figure 7(Figure 2.10c)

Exercise 2

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

In Figure 10 and Figure 11 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.

Exercise 3

To calculate $\mathbf{E}(U_{(j)})$ we have firstly to compute the probability that at least an order statistic $U_{(j)}$ falls in $[u, u + du]$, that is $P_{U_{(j)}}(u)$. Actually the probability that more than one element falls in this interval is negligigble since is an $O(du^2)$, so we can take in account only the probability that exactly one element falls in $[u, u + du]$.

This probability is given by the following expression:

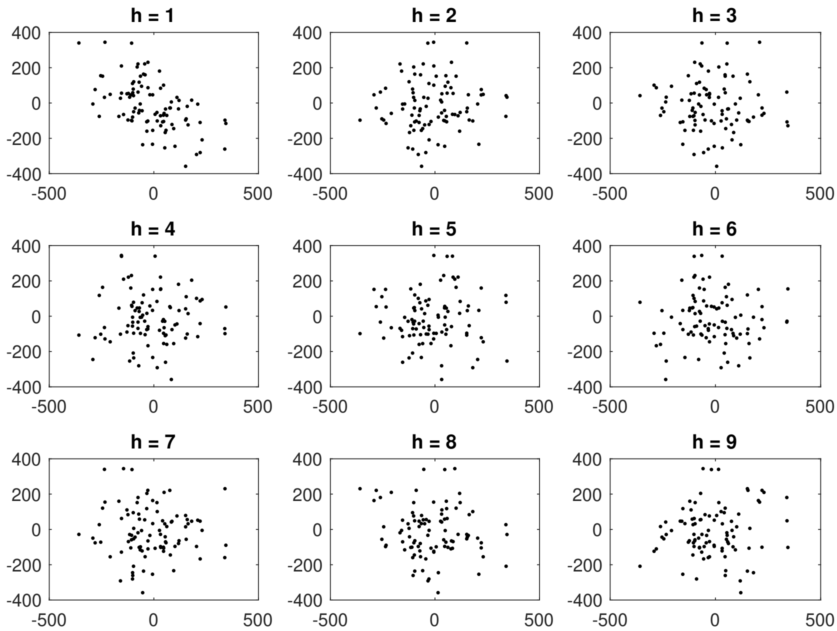

Figure 9: Lag plots (Figure 2.10d)

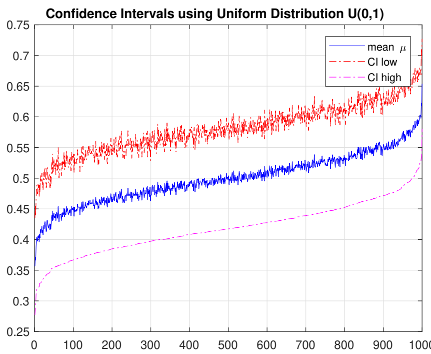

Figure 10: Results of the experiment with $n = 48$

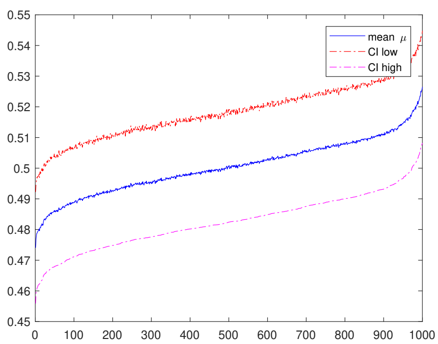

Figure 11: Results of the experiment with $n = 1000$

$$P_{U_{(j)}}(u) = P\Big[j-1 \text{ order statistics are in } [0,u]\Big] \cdot P\Big[1 \text{ order statistic is in } [u, u+du]\Big] \cdot P\Big[n-j \text{ order statistics are in } [u+du, 1]\Big] \tag{1}$$

The probability that a realization of a random variable $\mathbf{U}(0,1)$ is in $[0,u]$ is u, we want $j-1$ realization over n to be in this interval, so the resulting probability is:

$$P\Big[j-1 \text{ order statistics are in } [0,u]\Big] = \binom{n}{j-1} u^{j-1}$$

Now we want the next order statistic to be in $[u, u+du]$, so we can choose from the remaining $n-j+1$ realization an element with probability du, that is:

$$P\Big[1 \text{ order statistic is in } [u, u+du]\Big] = (n-j+1)du$$

Finally, the last $n-j$ elements have to be in $[u+du,1]$, so:

$$P\Big[n-j \text{ order statistics are in } [u+du,1]\Big] = (1-u-du)^{n-j}$$

but since du is very small by definition, we can ignore it rewriting this last equation as $(1-u)^{n-j}$.

Multiplying these three terms, as specified in Equation 1, the final probability is:

$$P_{U_{(j)}}(u) = \frac{n!}{(j-1)!(n-j)!}u^{j-1} \cdot du \cdot (1-u)^{n-j} \tag{2}$$

The keypoint of the proof is to note that Equation 2 is the PDF of a Beta distribution. Indeed a general Beta distribution is defined as

$$f(x) = \frac{1}{\mathbf{B}(\alpha,\beta)}x^{\alpha-1}(1-x)^{\beta-1} \tag{3}$$

where $\mathbf{B}(\alpha,\beta)$ is called *Beta function* and is defined as

$$\mathbf{B}(\alpha,\beta) = \frac{(\alpha-1)!(\beta-1)!}{(\alpha+\beta-1)!} \tag{4}$$

Given this definition, we can derive that $P_{U_{(j)}}(u)$ is distributed as a $Beta(j, n+1-j)$ and since the expected value of a *Beta* distribution is defined as $\mathbf{E}[x] = \frac{\alpha}{\alpha+\beta}$, we conclude that

$$\mathbf{E}\Big[U_{(j)}\Big] = \frac{j}{n+1} \tag{5}$$

Exercise 4

In Figure 12 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:

$$A_i = |\overline{x_i} - \overline{x}| \tag{6}$$

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 error from the true value 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.

In Figure 13 is reported the variance of each experiment and the relative confidence interval computed using bootstrap method. Finally, in Figure 14 and Figure 15 are reported the prediction intervals at level $\gamma = 0.95$ using theory and using bootstrap.

Exercise 5

Redoing Exercise 2, the plots in Figure 16 and Figure 17 have been obtained. Note that the sample mean now is distributed around the new true value, that is 0.

For the Exercise 4 the results are reported in Figure 18 and Figure 19. As in exercise 4, is visible how, 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.

For what regards the Prediction Interval (PI), we can see that using the relative theory, the PI is (as expected) between [-2,2], that is the γ quantile of the Normal Distribution $N(0,1)$. In the bootstrap case, the result is very similar even if not so accurate as the theoretical result.

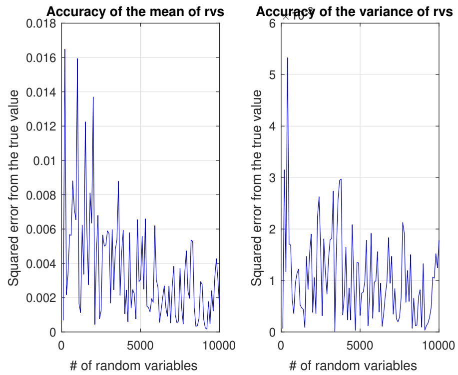

Figure 12: Accuracy of the estimation versus n

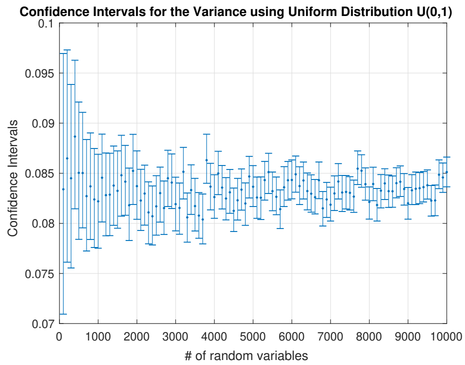

Figure 13: Confidence intervals for the variance using rvs $U(0,1)$

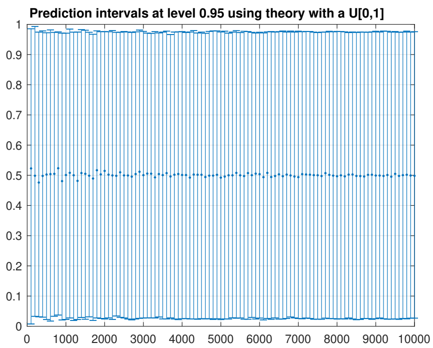

Figure 14: Prediction interval for a $U(0,1)$ at level $\gamma = 0.95$ using theory

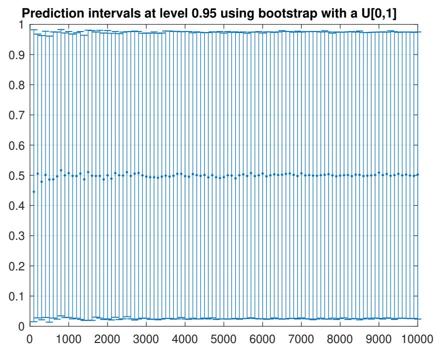

Figure 15: Prediction interval for a $U(0,1)$ at level $\gamma = 0.95$ using bootstrap

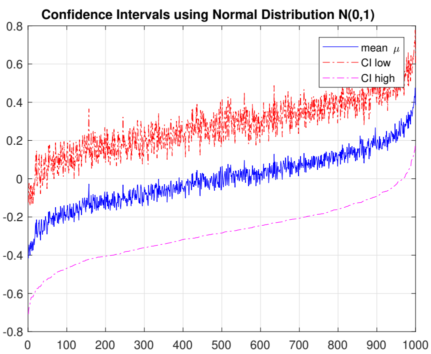

Figure 16: Results of the experiment with $n = 48$ rvs $N(0,1)$

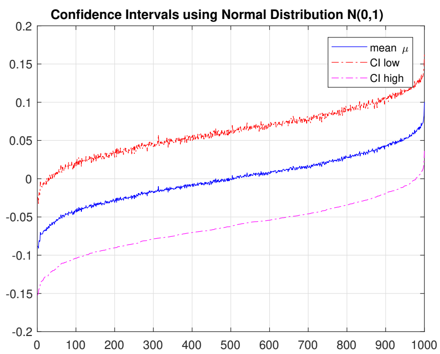

Figure 17: Results of the experiment with $n = 1000$ rvs $N(0,1)$

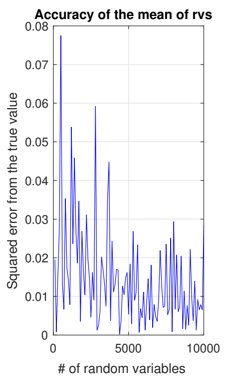

Figure 18: Accuracy of the estimation versus n using rvs $N(0,1)$

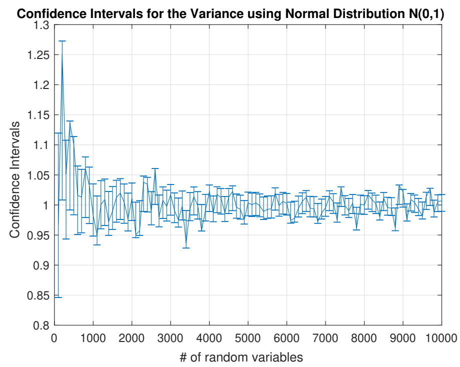

Figure 19: Confidence intervals for the variance using rvs $N(0,1)$

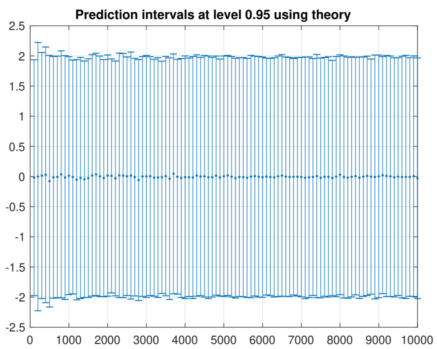

Figure 20: Prediction interval for a $N(0,1)$ at level $\gamma = 0.95$ using theory

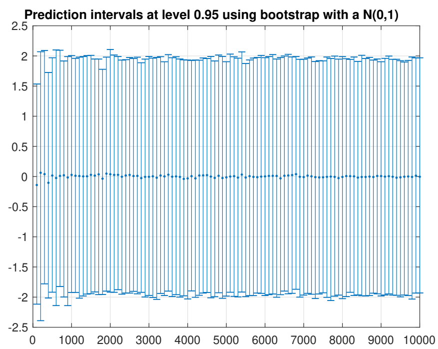

Figure 21: Prediction interval for a $N(0,1)$ at level $\gamma = 0.95$ using bootstrap