Network Analysis and Simulation - AY 2020/2021
Homework 1

Amedeo Giuliani

April 5, 2021

Exercise 1

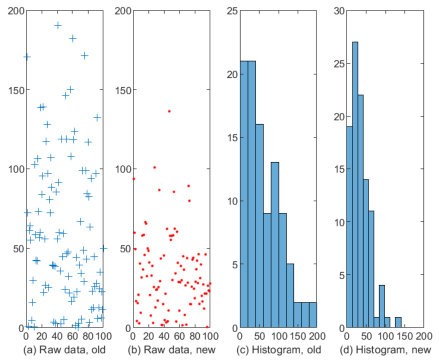


Figure 1: Figure 2.1

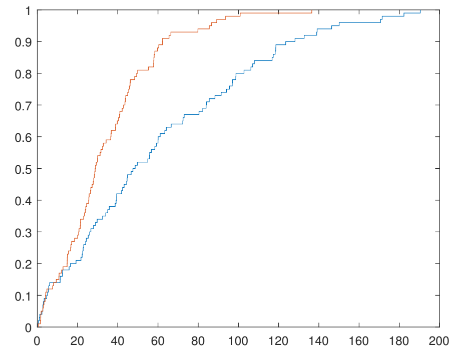


Figure 2: Figure 2.2

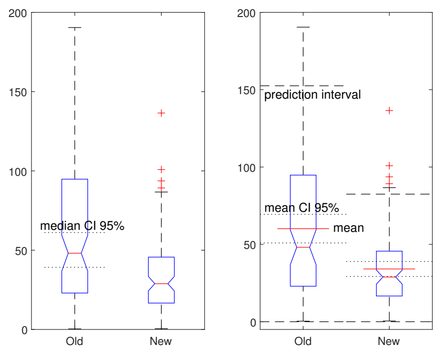


Figure 3: Figure 2.3

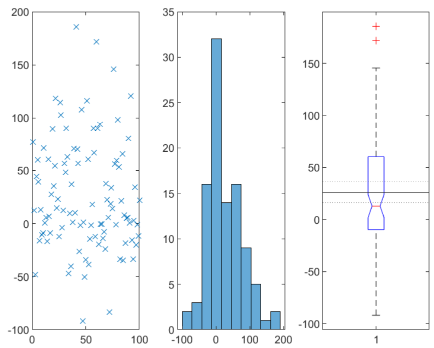


Figure 4: Figure 2.7

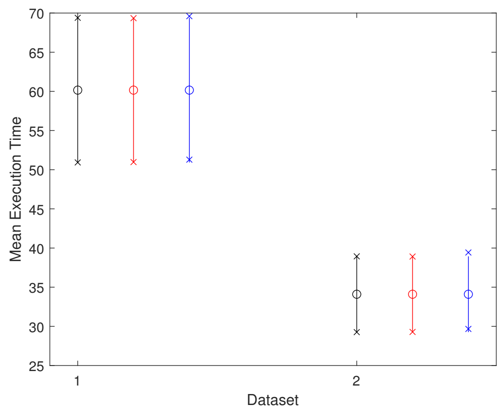
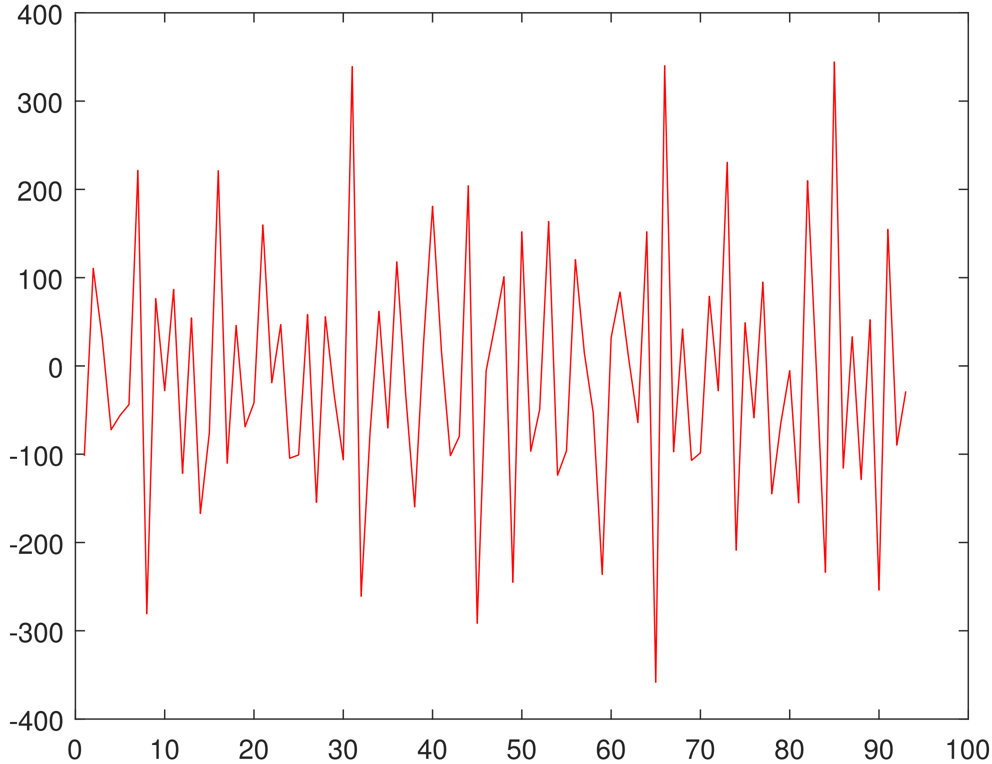
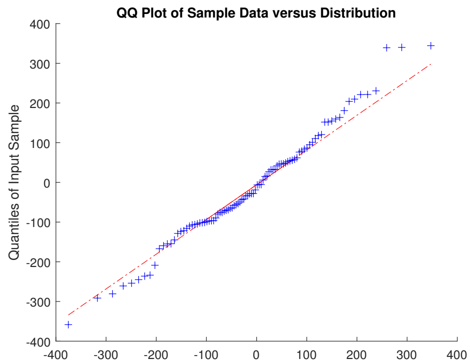


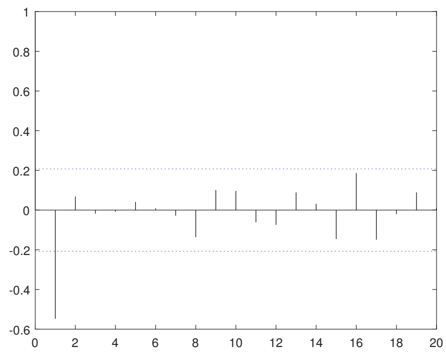
Figure 5: Figure 2.8



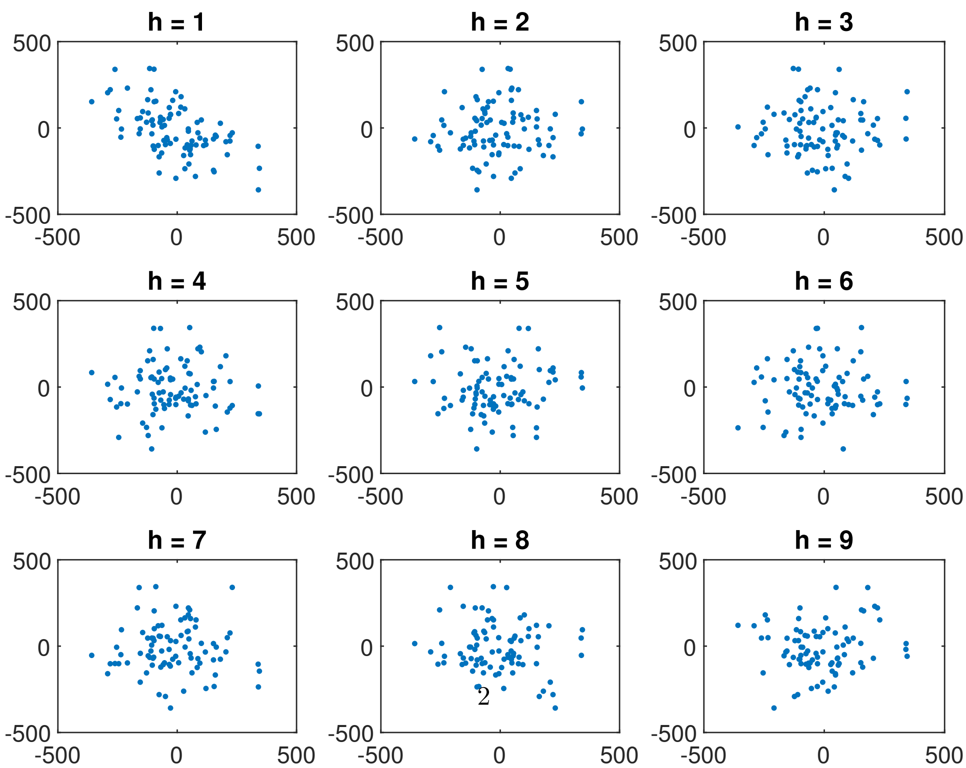
(a) Data



(b) Data



(c) Autocorrelation



(d) Lag plots

Exercise 2

We have $n = 48$ samples $\mathrm{U}_1, ..., \mathrm{U}_\mathrm{n} \sim \mathcal{U}(0, 1)$. To make the Central Limit Theorem hold we need at least 30 samples, so in this case we can assume the theorem works. Thus, for the Theorem 2.2.2 from [1] we can state that the confidence interval for the true mean is:

$$\mathrm{CI}_{\mathrm{mean}} = \hat{\mu} \pm \eta \frac{s_n}{\sqrt{n}} \tag{1}$$

where $\hat{\mu}$ is the sample mean, s_n is the sample standard deviation and $\eta = 1.96$ as we are looking for a 95% confidence level. We repeat this experiment independently for 1000 times and for every iteration we test if the true mean falls in the computed CI or not: if not, then we count this failure. We do this in order to test if the number of times that the true mean falls outside the CI, let's call it $\mathrm{N}_{\mathrm{out}}$, is the one we expect, that is 5% of the times.
However, since we are computing the CI by relying ourselves on an asymptotic result and our number of samples is not quite large, we have as result that $\mathrm{N}_{\mathrm{out}}$ oscillates a lot around 5% and in one experiment it has even reached the value of 6.8%. Ultimately in Fig. 7 and in Fig. 8 are depicted the CI lower and upper bound points, ordered in ascending order according to the lower bound, in each of the experiments for the uniform distribution and the normal distribution, respectively.

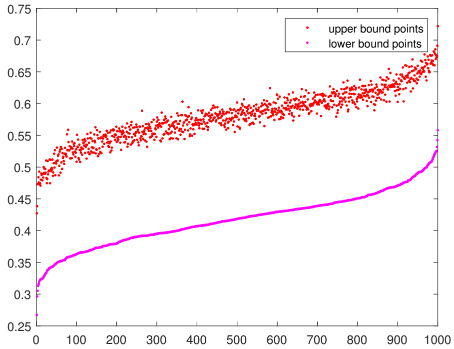


Figure 7: Graph of upper and lower bound points of CI of mean for uniform distribution in each experiment.

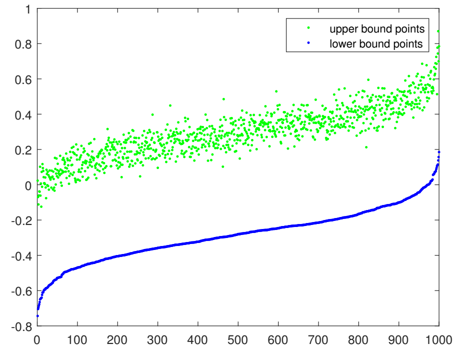


Figure 8: Graph of upper and lower bound points of CI of mean for normal distribution in each experiment.

Exercise 3

Let $\mathrm{U}_1, ..., \mathrm{U}_\mathrm{n}$ be independent uniform random variables $\mathcal{U}(0, 1)$. Let $\mathrm{U}_{(1)}, ..., \mathrm{U}_{(\mathrm{n})}$ be the order statistics of $\mathrm{U}_1, ..., \mathrm{U}_\mathrm{n}$. The expected value of the k-th order statistics $\mathrm{U}_{(\mathrm{k})}$ is:

$$\mathbb{E}[\mathrm{U}_{(\mathrm{k})}] = \frac{k}{n+1} \tag{2}$$

Proof. The PDF of the order statistics for n independent uniform random variables is:

$$f_k(u) = \begin{cases} n\binom{n-1}{k-1} u^{k-1}(1-u)^{n-k} & u \in [0,1] \\ 0 & \text{elsewhere} \end{cases} \tag{3}$$

and if we apply the definition of the expected value, then:

$$\mathbb{E}[\mathrm{U}_{(\mathrm{k})}] = \int_{-\infty}^{\infty} u f_k(u)\, du = n\binom{n-1}{k-1} \int_0^1 u^k (1-u)^{n-k}\, du \tag{4}$$

We recall that the Beta function is defined as:

$$B(,b) = \int_0^1 t^{-1}(1-t)^{b-1}\, dt = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}\ ,\ \Gamma(l) = (l-1)!\ \forall l \in \mathbb{N} \tag{5}$$

and that a beta random variable $\mathrm{X} \sim \beta(,b)$ has density[2, p.226]:

$$f_{\mathrm{X}}(x) = \begin{cases} \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1} & x \in [0,1] \\ 0 & \text{elsewhere} \end{cases} \tag{6}$$

whose expected value is:

$$\mathbb{E}[\mathrm{X}] = \frac{a}{a+b} \tag{7}$$

We notice that (3) and (6) are quite similar. Therefore, if we further develop the first:

$$\begin{aligned} f_k(u) &= n\frac{(n-1)!}{(k-1)!(n-k)!}u^{k-1}(1-u)^{n-k} = \frac{n!}{(k-1)!(n-k)!}u^{k-1}(1-u)^{n-k} = \\ &= \frac{\Gamma(n+1)}{\Gamma(k)\Gamma(n-k+1)}u^{k-1}(1-u)^{n-k} \end{aligned} \tag{8}$$

and if we set $a = k$ and $b = n - k + 1$, which also yields $a + b = n + 1$, we obtain:

$$f_k(u) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}u^{a-1}(1-u)^{b-1} \Rightarrow \mathrm{U}_{(\mathrm{k})} \sim \beta(k, n-k+1) \tag{9}$$

Therefore, (4) can be expressed as:

$$\mathbb{E}[\mathrm{U}_{(\mathrm{k})}] = \frac{k}{k+n-k+1} = \frac{k}{n+1} \tag{10}$$

□

Exercise 4

We wan to study the accuracy of the estimation of the mean and variance of a uniform distribution between 0 and 1. To do so, we generate $X_1, ..., X_n \sim \mathcal{U}(0, 1)$, we compute sample mean $\bar{X}$ and sample variance S^2 and we iterate this steps with a growing number of samples. As expected, as n grows, the sample mean converges to the true mean $\mu$and the sample variance converges to the true variance σ^2. This behaviour is shown in Fig. 9 and Fig. 10 and it is thanks to the fact that the two estimators we are using are unbiased, that is $\mathbb{E}[\bar{X}] = \mathbb{E}[X] = \mu$ and $\mathbb{E}[S^2] = \sigma^2$.

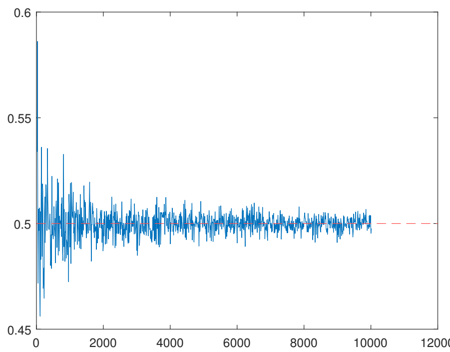


Figure 9: Sample mean vs true mean as n grows.

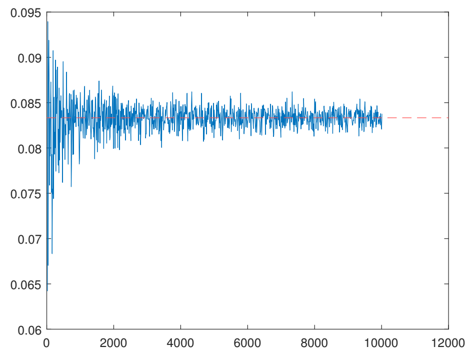


Figure 10: Sample variance vs true variance as n grows.

A similar result has been obtained by deriving the confidence interval for the sample variance S^2[3], which shrinks with a growing number of samples, as shown in Fig. 11, except for $n = 11, 21$ for which they are already narrow. In fact, a "small note": for $n = 11, 21$ the Central Limit Theorem does not hold and we cannot handle X as it is sampled from a standard normal distribution, which means that $\frac{(n-1)S^2}{\sigma^2} \nsim \chi^2_n$ and so the confidence interval for the variance for a standard normal distribution is not a good approximation. Hence, the bootstrap method has been used to compute the confidence intervals for $n = 11, 21$ and of course here we can be as precise as we want since all we have to do is to increase the number of sets of samples, obtained through sampling with replacement from the dataset, calculate the statistic on each of these sets and average to obtain a single precise result.

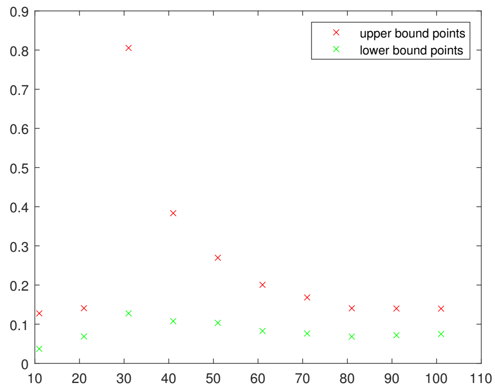


Figure 11: Confidence interval for the variance at level 95% vs n

Ultimately, to find the prediction interval at level 95% we can use the result of Theorem 2.4.1 from [1] or the bootstrap method. In the first case, it is that, for $\alpha \geq \frac{2}{n+1}$, $[X^n_{(\lfloor (n+1)\frac{\alpha}{2} \rfloor)}, X^n_{(\lceil (n+1)(1-\frac{\alpha}{2}) \rceil)}]$ is a prediction interval of at least level $\gamma = 1 - \alpha$. If we let $\alpha = 0.05$ this yields $\gamma = 0.95$, i.e. a confidence of 95% which it is the one we are pursuing. Considering $n = 999$ we obtain $X^n_{(25)}$ and $X^n_{(975)}$, which are the lower bound and the upper bound of the prediction interval, respectively. In this specific case, the theory yields the interval $\mathrm{PI}_{\mathrm{theo}} = [0.206, 0.9748]$. In the second case the following empirical method has been implemented:

Algorithm 1: Bootstrap algorithm

```
sort the vector of uniform rvs in ascending order;
set i=1;
set c=0;
set k = 10000;
while i ≤ n do
    chose lb as the i-th item from the order statistics;
    choose ub as the (999-i)-th item from the order statistics;
    for j ≤ k do
        sample with replacement from the dataset;
        if sample is not in [lb,ub] then
            c = c + 1;
        end
    end
    set c = c / k;
    if c ≥ 0.05 then
        break;
    end
end
```

which yields the interval $\mathrm{PI}_{\mathrm{boot}} = [0.203, 0.9748]$ that is almost the same as the prediction interval obtained in theory.

Now we repeat the experiment with $X_1, ..., X_n \sim \mathcal{N}(0, 1)$. In Fig. 12 and Fig. 13 is shown the accuracy of the mean and variance estimations with a growing number of samples, respectively, which have the same behaviour as before. For what concerns the confidence interval for the variance, we are now implementing the result of [3] in the correct way, that is with a dataset from a standard normal distribution and thus the output will be exact. In fact, as we can see from Fig. 14, the CIs are stll shrinking as n grows, but now for $n = 31, 41, 51$ they are more narrow with respect to the case with a dataset from a uniform distribution while for $n = 11, 21$ they are wider than before because here the dataset is too small and we cannot be as precise as bootstrap.

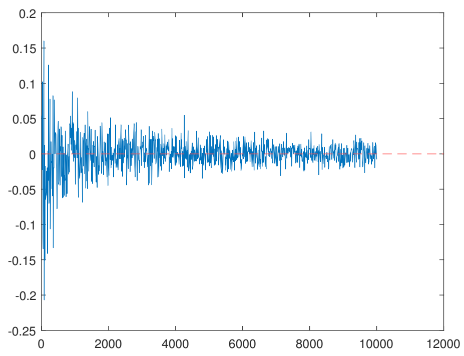


Figure 12: Sample mean vs true mean as n grows.

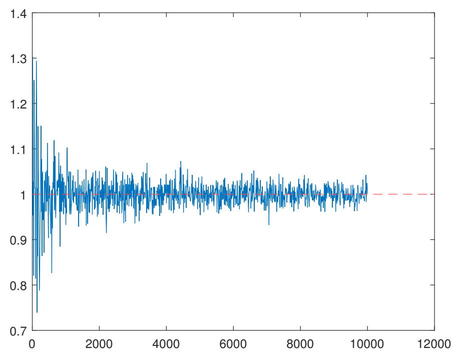


Figure 13: Sample variance vs true variance as n grows.

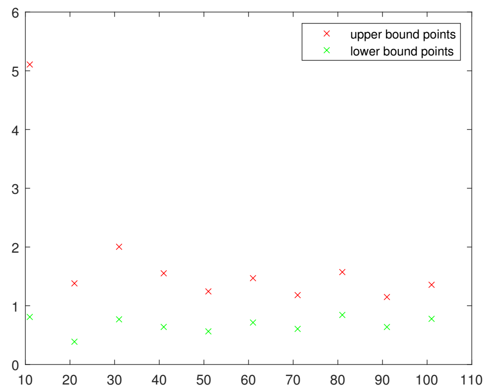


Figure 14: Confidence interval for the variance at level 95% vs n

Ultimately, to compute the prediction interval we can still use the previous algorithm by replacing the uniform random variables with the standard normal ones. Or we can use the results of Theorem 2.4.2 from [1] and in particular the approximate result as we are considering $n = 999$ which is a fairly large number.

References

[1] Jean-Yves Le Boudec, *Performance evaluation of computer and communication systems*, Epfl Press, 2010.

[2] Sheldon Ross, *A first course in probability*, Prentice Hall, 1998.

[3] Loughborough University, *Interval estimation for the variance*, Helping Engineers Learn Mathematics Workbooks, 2008.