Students:
Zineb Agnaou
Francesca Paola Nicoletti

Group 11

Performance Evaluation
HOMEWORK 2

Jean-Yves Le Boudec
Hossein Tabatabaee

Spring 2021

1 STATISTICS WARMUP

PROBLEM 1.

In Figure 1 are displayed the histograms of n iid Standard Normal random variables for n = 10, 20, 40, 80, 160, 320, 640, 1280 and 2560. The samples were generated using a Matlab code found in Section 5. We can see that the more samples displayed, the more we converge to a standard normal density function displayed in Figure 2.

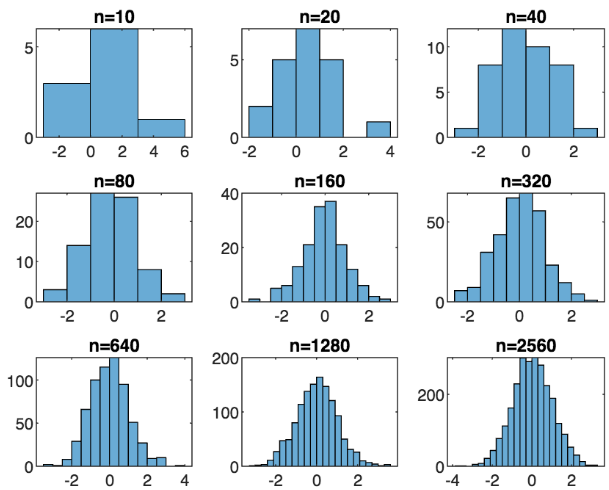


Figure 1: Histograms of n iid Standard Normal variables for n = 10, 20, 40, 80, 160, 320, 640, 1280 and 2560. The according histograms are displayed from top to bottom, left to right.

PROBLEM 2.

Plots and Distributions

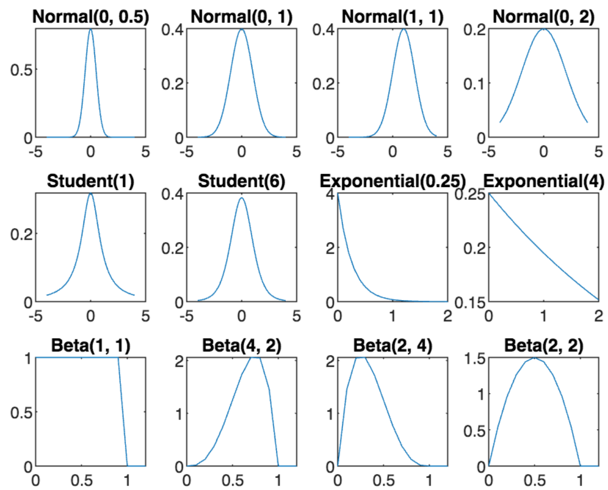


Figure 2: Densities of the following distributions from top to bottom, left to right: Normal(0, 0.5), Normal(0, 1), Normal(1, 1), Normal(0, 2), Student(1), Student(6), Exponential(0.25), Exponential(4), Beta(1, 1), Beta(4, 2), Beta(2, 4), Beta(2,2).

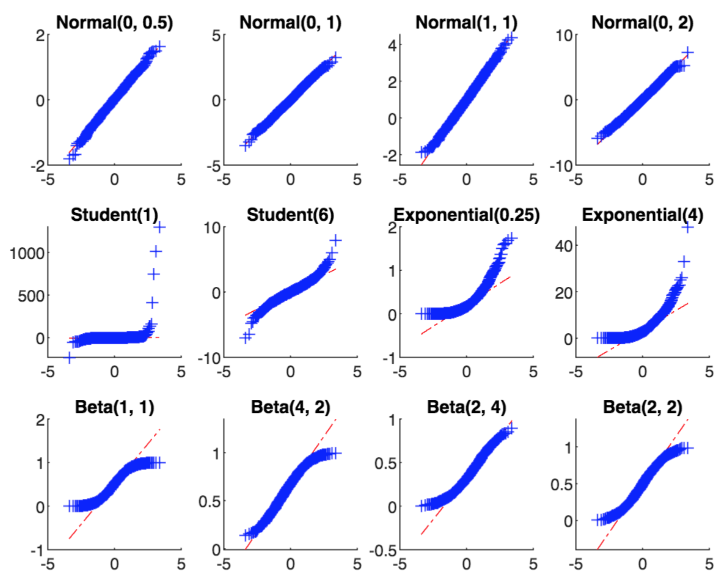


Figure 3: Standard normal QQ-plots of the following distributions from top to bottom, left to right: Normal(0, 0.5), Normal(0, 1), Normal(1, 1), Normal(0, 2), Student(1), Student(6), Exponential(0.25), Exponential(4), Beta(1, 1), Beta(4, 2), Beta(2, 4), Beta(2,2).

As expected, the QQ-plots of the normal samples fall into the diagonal axis in Figure 3. The QQ-plots for the Exponential distribution are U-shaped. It corresponds to right-skewed data relative to a normal distribution. The QQ-plots for the Beta distribution are S-shaped. It corresponds to under-dispersed data relative to a normal distribution. The QQ-plots for the Student distribution are inverse S-shaped corresponding to over-dispersed data relative to a normal distribution. Some outliers can be noticed specially in the QQ-plots corresponding to the distribution Student(1).

2 SIMULATE RANDOM WAYPOINT AND LOOK AT WHAT YOU HAVE DONE

PROBLEM 3.

Here we simulate the random waypoint model defined in the chapter “Simulation” in Le Boudec 2010 (Example 6.5 page 169). We use the following parameters:

- Number of users: N = 150,
- The area is a rectangle of dimensions $l \times L$ with $l = L = 1000$m,
- vmin = 1m/s, vmax = 2m/s.
- The simulation terminates at (simulated) time Ts = 86400 sec = 1 day.

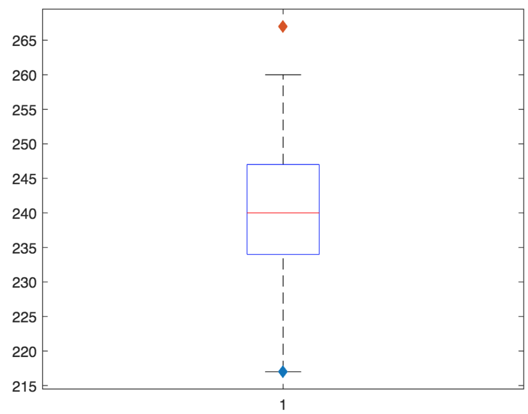


Figure 4: Box plot of the number of waypoints reached by 150 mobiles for only one simulation run. - The red line corresponds to the mean value. - The red diamond corresponds to the maximal value - The blue diamond corresponds to the minimal value.

As displayed in Figure 4, the simulation ends on average after 240 waypoints for 150 mobiles. The elapsed real time it takes to acquire simulation data of 150 mobiles is of 0.910086 seconds, this result was computed using the *tic ... toc* function in Matlab.

The waypoints and trajectories obtained are displayed in Figure 5 and 6 for one user and 5 users respectively. From these figures, we can note there is no pattern neither for the waypoints nor for the trajectories, which was expected since the walk comes from a independent uniform distribution. Furthermore, the course is very dense, especially in the center, even for one user and superposed and almost not distinguishable for the 5 users.

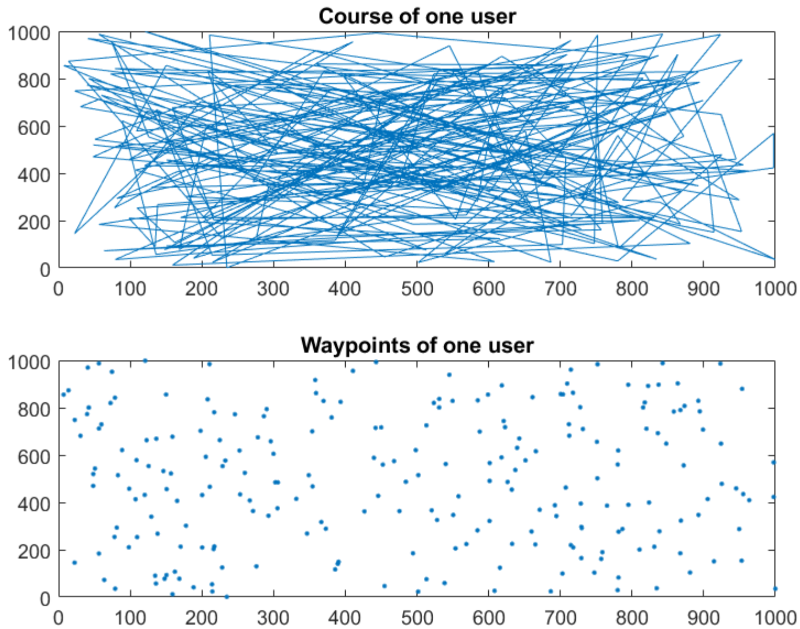


Figure 5: Course and Waypoints of one user.

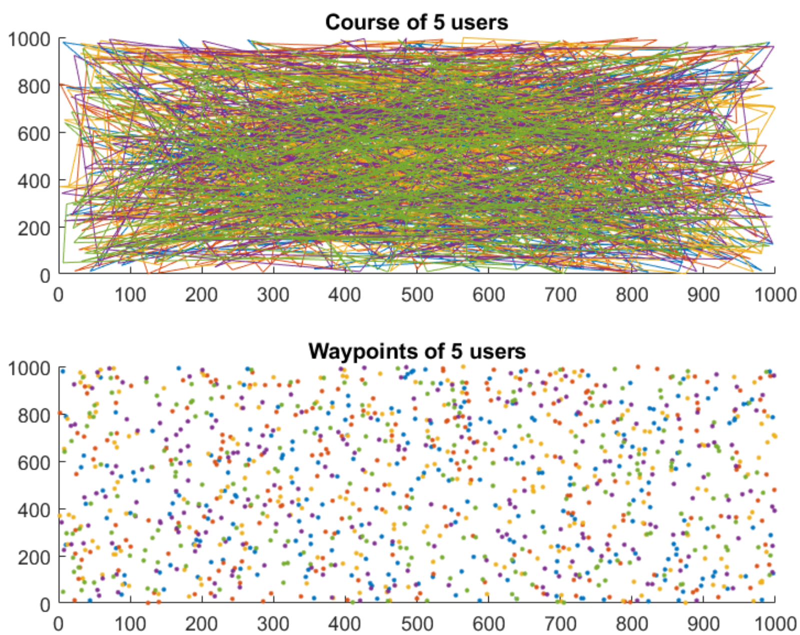


Figure 6: Course and Waypoints of 5 users.

3 DIFFERENT VIEWPOINTS

3.1 EVENT AVERAGE (PALM) VIEWPOINT

In here we display the histogram of speeds sampled at transition epochs Tn, based on the samples for one mobile and for 150 mobiles.

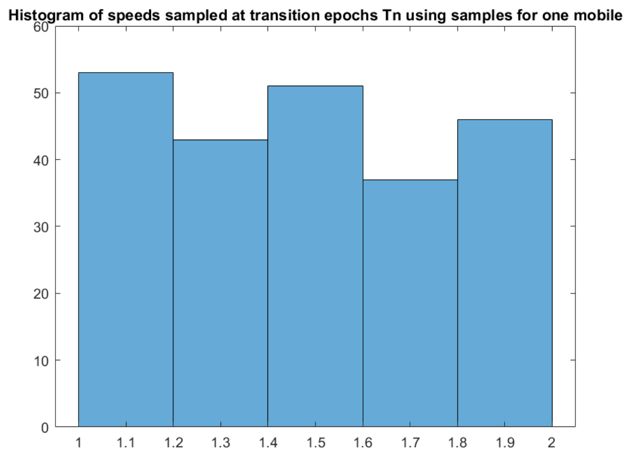


Figure 7: Histogram of speeds sampled at transition epochs Tn, based on the samples for one mobile.

In Figure 7, we see that even the sample is relatively small it resembles already to the overall shape of the uniform distribution from which the sample was generated.

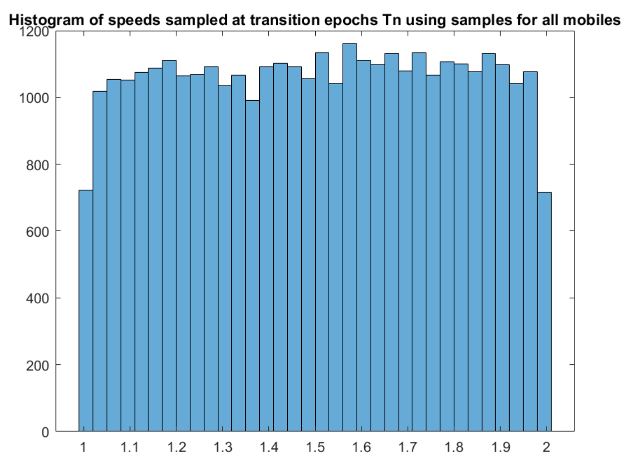


Figure 8: Histogram of speeds sampled at transition epochs Tn, based on the samples for 150 mobiles.

Figure 8 confirms the resemblance of our samples to the overall shape of the uniform distribution from which the samples were generated. However, it should be noted that the values at the edges deviate from the expected distribution.

We here plot histograms of mobile positions based on the samples for one user and for 150 users.

In order to achieve this, we first discretize the area of the waypoints into 100 x 100 square bins. We choose to represent the number of passages on each zone of this discretized area by a 100 x 100 matrix initialized at 0. Each value contained in this matrix therefore represents the number of passages on the square of the corresponding discretized area. The histogram of mobile positions is then displayed using this matrix as an argument for the Matlab $surf()$ function. This function return a three-dimensional surface plot with solid edge colors and solid face colors as shown in Figures 10 and 12 for example. The function plots the values in the matrix as heights above a grid in the x-y plane. The color of the surface varies according to the heights specified by the matrix. Using the $'gray'$ colormap and projecting the 3D plot in the 2D plane, we then have obtained a grey shade diagram where the intensity of the grey is proportional to the frequency displayed in Figures 9 and 11. The darker the bin is, the fewer it was visited. The same way, the lighter the bin is, the more it has been visited.

In the Figures 9 and 11, we see the results obtained respectively for one user and for the 150 users. On one hand, we observe as expected a lot of black and few different light shades in Figure 9 since we take the trajectory of a single mobile into account so few waypoints and thus not many passages in each zone and in the overall area. On the other hand, in Figure 11 we can distinguish a multitude of different shades of gray. Each zone is therefore not visited the same number of times, but the pattern obtained appears to be homogeneous. The weights seem to be well distributed throughout the area, which is what we expected as the waypoints follow an iid, random and uniform distribution in the given area. Furthermore, the 3D surfaces showed in Figures 10a and 12 confirm the random uniform distribution because we can see that the peaks are homogeneously distributed all over the area and that the height is quite similar for all the peaks. In Figure 10b is displayed a colored version of Figure 10a to allow better visualisation.

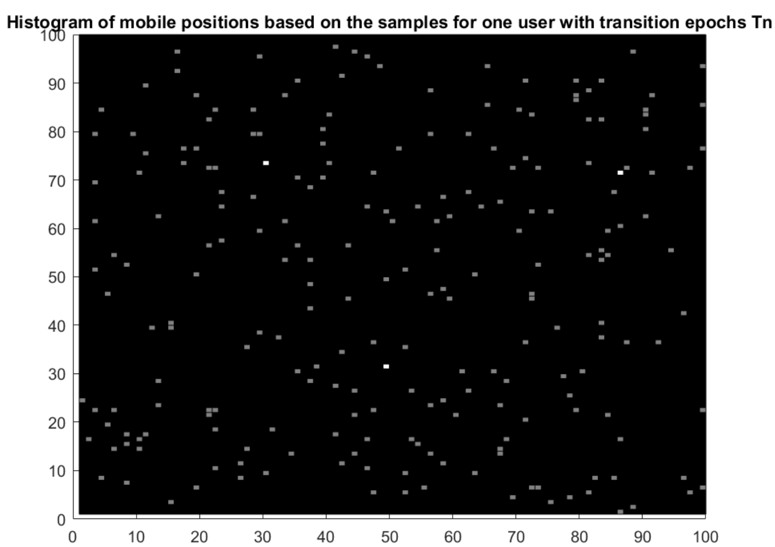
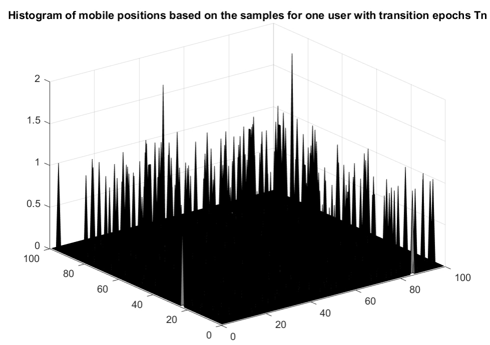
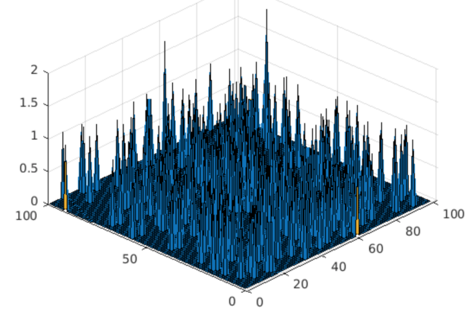


Figure 9: Histogram of mobile positions based on one user samples with transition epochs Tn



(a) 3D surface corresponding to mobile positions based on one user samples with transition epochs Tn



(b) Colored 3D surface corresponding to mobile positions based on one user samples with transition epochs Tn

Figure 10

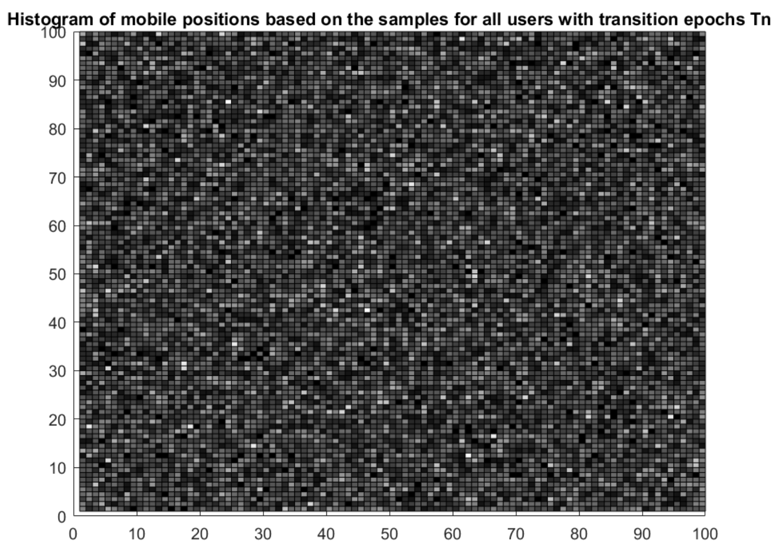


Figure 11: Histogram of mobile positions based on all users samples with transition epochs Tn

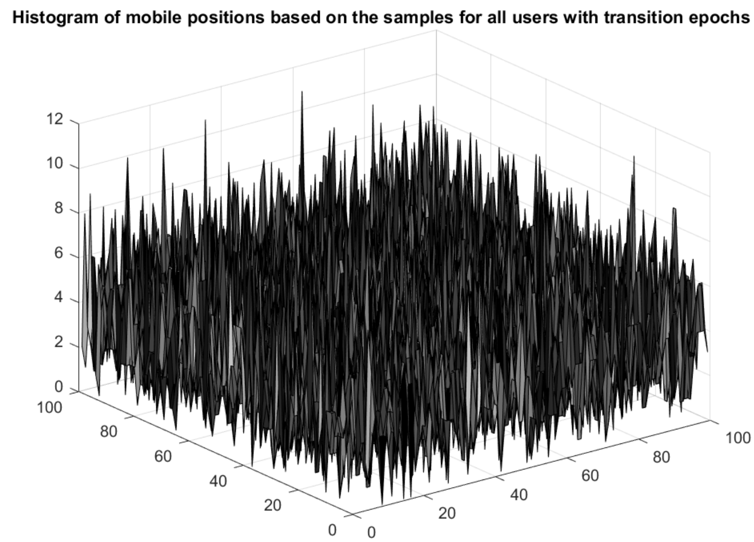


Figure 12: 3D surface corresponding to mobile positions based on all users samples with transition epochs Tn

3.2 TIME AVERAGE VIEWPOINT

We now sample the position and speed of mobiles every 10 seconds for one and 150 mobiles.

First, as we can see in Figures 13 and 14 that respectively represents the course and waypoints of one user sampled with transition epochs Tn and every 10 seconds, we notice that the trajectory is the same but much more waypoints have been recorded sampling every 10 seconds.

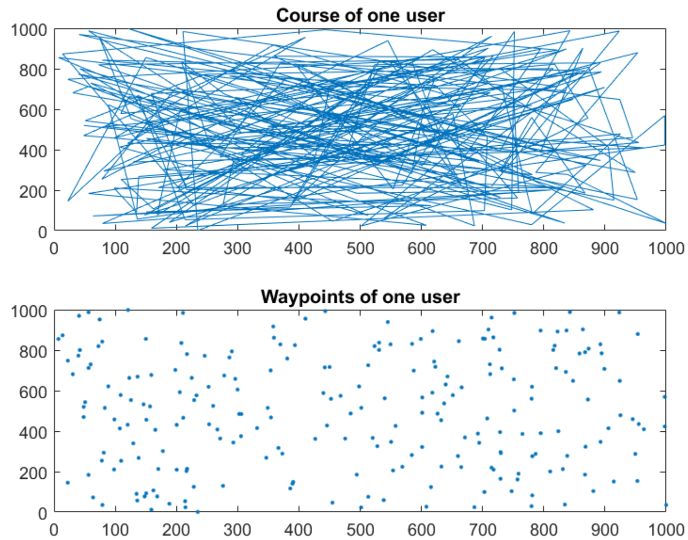


Figure 13: Course and waypoints of one user sampled with transition epochs Tn (Similar to Figure 5)

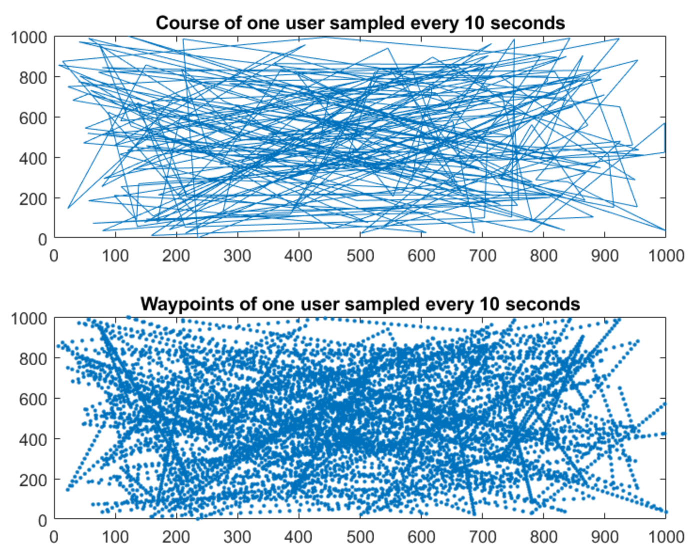


Figure 14: Course and waypoints of one user sampled every 10 seconds

Once again we display first the histogram of speeds, now sampled every 10 seconds, based on the samples for one mobile and for 150 mobiles.

The histogram peaks of Figures 15 and 16 differ in shape compared to the ones in Figures 7 and 8, the distribution looks considerably less uniform distribution. The peaks corresponding to low speeds are considerably higher than those corresponding to high speeds. Moreover, the decrease in peak height appears to have a linear trend. This can be explained as such: when the speed of an trajectory is low, it takes more time to go from one waypoint to another. Suppose that a trajectory A takes 15 secs with speed v_A when another trajectory B takes 40 secs with speed v_B, so $v_B < v_A$ since the next waypoints are picked independently from the distance. When sampling every 10 secs, v_A will be collected once versus v_B that will collected 4 times. Thus, slow speeds are more likely to be more represented in the Histogram. This can not be remarked when samples are taken according to the transition epochs Tn.

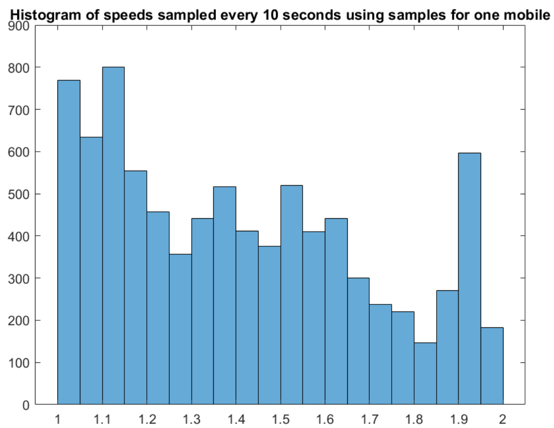


Figure 15: Histogram of speeds sampled every 10 seconds using samples for one mobile

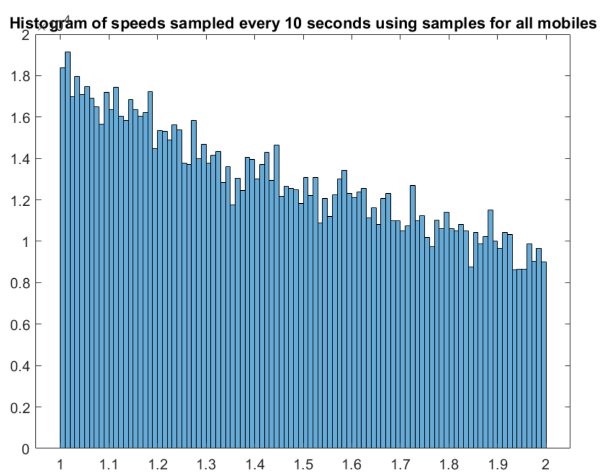


Figure 16: Histogram of speeds sampled every 10 seconds using samples for all mobiles

Now, similarly as done to get Figures 9 to 12, we display in Figures 17 to 22 histograms of mobile positions sampled every 10 seconds based on samples for one mobile and for 150 mobiles.

For histograms for only one mobile, results can be quite different : Figures 17 to 20 show two different examples of the same case study. As we can see in our Figures 17 and 19, very few lighter squares stand out clearly. These are surrounded mostly by black and a few lighter shades forming lines rather in the center of the image. This large difference in color can be explained by the very large values of number of passages reached in certain areas. Indeed, as one can observe it with the 3D surfaces of the Figures 18 and 20, the difference between the few very high peaks and the other peaks is very large : some areas are considerably more visited than others. Then, since the scale of the shades of gray is proportional to this difference, the few very high values will appear very light and the rest of the values rather dark. This might not have a specific explanation rather than just the randomness in each run. This did not happen in samples taken at transition epochs Tn, so the histograms of Figure 9 and the ones in Figures 17 and 19 are quite different. However, they keep some similarities : both show a lot of black and few different shades since we take a single race into account so few waypoints, and not many passages on the overall area. We can also notice that some areas which are not black in Figure 17 are the same as those in Figure 9. This is normal since these histograms correspond to the same path, just sampled at different times.

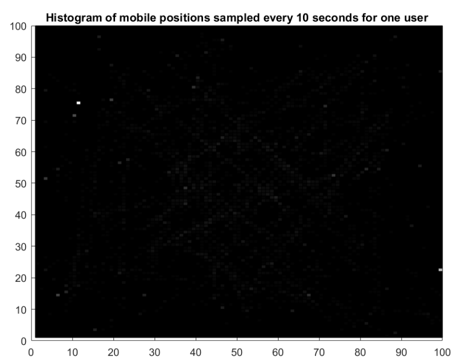


Figure 17: Histogram of mobile positions sampled every 10 seconds for one user

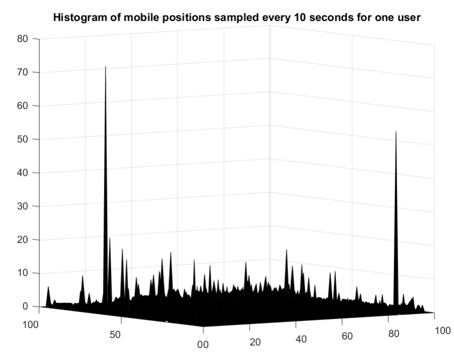


Figure 18: 3D surface corresponding to the mobile positions sampled every 10 seconds for one user

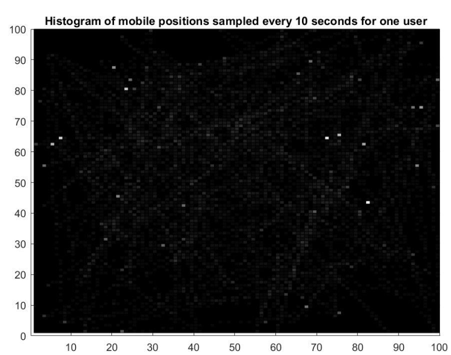


Figure 19: Histogram of mobile positions sampled every 10 seconds for one user

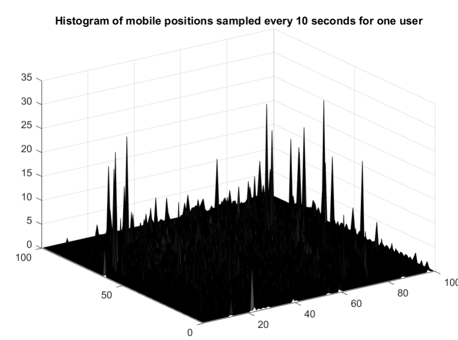


Figure 20: 3D surface corresponding to the mobile positions sampled every 10 seconds for one user

On the other hand, as noticed for Figure 11 corresponding to all users, in Figure 21 we can distinguish a multitude of different shades of gray. Each zone is therefore not visited the same number of times. However, unlike the Figure 11 where the pattern seemed homogeneous, we clearly observe in Figure 21 that the central area is much lighter than the edges of the image. This means that there are many more passages in the center of the image than at the edge. This can also be seen clearly in Figure 22. Knowing that when we sample every 10 seconds, we consider almost the entire trajectory of each interval, and that almost all the intervals have parts passing towards the center of the image because to go from one extreme to another on the plane $l \times L$, we always pass by the centre points, thus the center has more visits, so it seems correct to observe this result. But again, this seems to show us that the distribution that we obtain is no longer uniform.

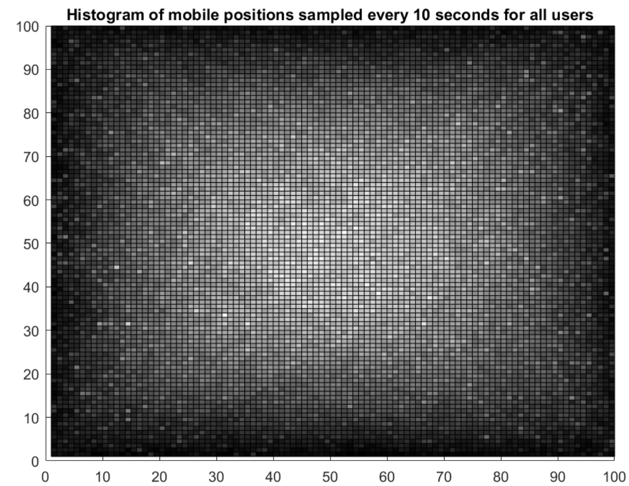


Figure 21: Histogram of mobile positions sampled every 10 seconds for all users

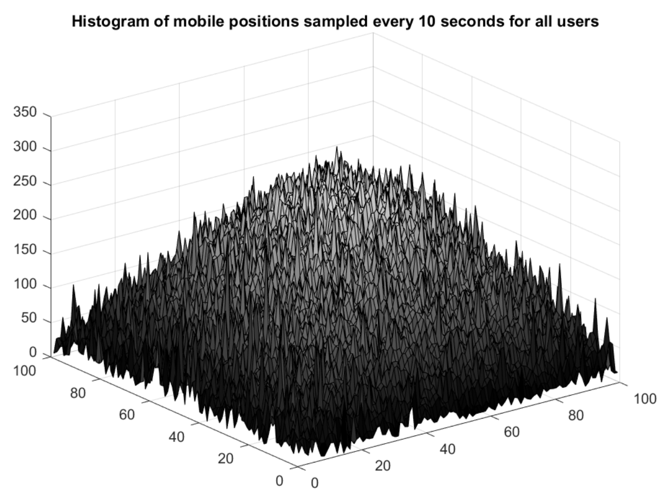


Figure 22: 3D surface corresponding to the mobile positions sampled every 10 seconds for all users

4 CONFIDENCE INTERVALS

4.1 CONFIDENCE INTERVALS FOR MEDIANS AND MEANS

For each of the N mobiles, we compute the average of the speed values sampled at transition epochs Tn (event average speed for each user). This gives us a data sequence $X_1, X_2, \ldots, X_N$. Similarly, for each of the N mobiles we compute the average of the instant speed values sampled at arbitrary instants of time (time average speed for each user). This gives us another data sequence Y_1, Y_2,..., Y_N.

First we consider N = 150. For the sequences X and Y :

For the mean confidence interval, we use the Theorem 2.2 in Le Boudec 2010 (page 34). To apply this theorem :
1) the observations have to be independent.
2) the observations come either from a large sample size or follow a Gaussian distribution.
3) the mean:

$$\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

and variance:

$$s_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \hat{\mu}_n)^2$$

are well defined.

The speed variables from which we sampled the sequence $(X_n)_n$ and $(Y_n)_n$ are iid and come from the same uniform distribution. In addiction, the data sequence (X_n) and (Y_n) appears normal as seen the QQ-plots in Figure 23. Both the mean and variance are well defined since we are working with finite sequence taking finite values.

The confidence interval for the mean at level $\gamma = 0.95$ is:

$$\hat{\mu}_n \pm \eta \frac{s_n}{\sqrt{n}}$$

where η is the $\frac{1+\gamma}{2}$ quantile of the standard normal distribution. In our case $\eta = 1.96$.

For the median confidence intervals $[X_{(j)},\ X_{(k)}]$ for N=150, we use the approximation of j and k in Theorem 2.1 in Le Boudec 2010 (page 33) considering that 150 is a quite large sample. Here $X_{(1)} < X_{(2)} < \ldots < X_{(n)}$ is the order statistics. In this theorem, only the independence is to be assumed. The Independence assumptions hold as explained for the mean confidence interval.

The confidence interval for the median at level $\gamma = 0.95$ is therefore:

$$[X_{(j)},\ X_{(k)}]$$
$$\text{s.t. } j = \left\lfloor np - \eta\sqrt{np(1-p}\right\rfloor,\ k = \left\lceil np + \eta\sqrt{np(1-p}\right\rceil + 1$$

where η is the $\frac{1+\gamma}{2}$ quantile of the standard normal distribution. In our case $\eta = 1.96$.

When N=30, the number of samples is quite small, using the approximation in Theorem 2.1 in Le Boudec 2010 will lead to absurd intervals. We then take the values in corresponding to N in the table page 313 in Le Boudec 2010 to get the median confidence interval which gives us j=10 and k=21.

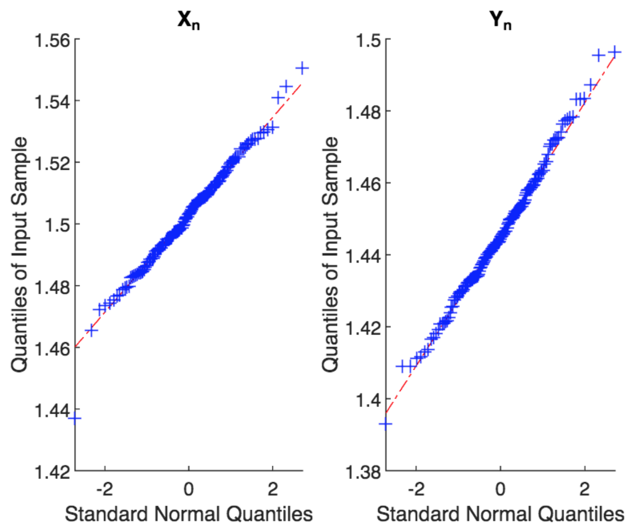


Figure 23: QQ-plots of the sequence $(X_n)_n$ on the left and $(Y_n)_n$ on the right - for 150 mobiles

For N=150 in Figure 24, we can see median, the confidence interval for the median at level 0.95, the mean and confidence interval for the mean at level 0.95. A first thing to notice is that the values of $(Y_n)_n$ are smaller than $(X_n)_n$ for both the mean and the median (for mean: ≈ 1.5 m per s for $(X_n)_n$ vs ≈ 1.45 m per s for $(Y_n)_n$, same for median. This can be explained again with the fact that lower speeds are more likely to be observed when sampling at an arbitrary interval and thus these values draw the mean and median down. The mean CI are also wider than the median CI. All but the median CI for $(Y_n)_n$ seem quite symmetric apart. For N=30, the features. The values of $(Y_n)_n$ are lower and the CI of the median is not symmetrical. When comparing the CI of $(X_n)_n$ and $(Y_n)_n$, we notice that CI for $(Y_n)_n$ are significantly wider see Table 1 and Table 2. This is explained by the fact that the more samples we have, the more information we gather and thus less uncertainty leading to thinner and precise intervals.

	N=30	N=150
$(X_n)_n$	[1.4979, 1.5075]	[1.5003,1.5056]
$(Y_n)_n$	[1.4390,1.4520]	[1.4429,1.4489]

Table 1: Mean CI for $(X_n)_n$ and $(Y_n)_n$ at level 0.95 based on N Samples.

	N=30	N=150
$(X_n)_n$	[1.4943,1.5078]	[1.4983,1.5078]
$(Y_n)_n$	[1.4344,1.4514]	[1.4411,1.4500]

Table 2: Median CI for $(X_n)_n$ and $(Y_n)_n$ at level 0.95 based on N Samples.

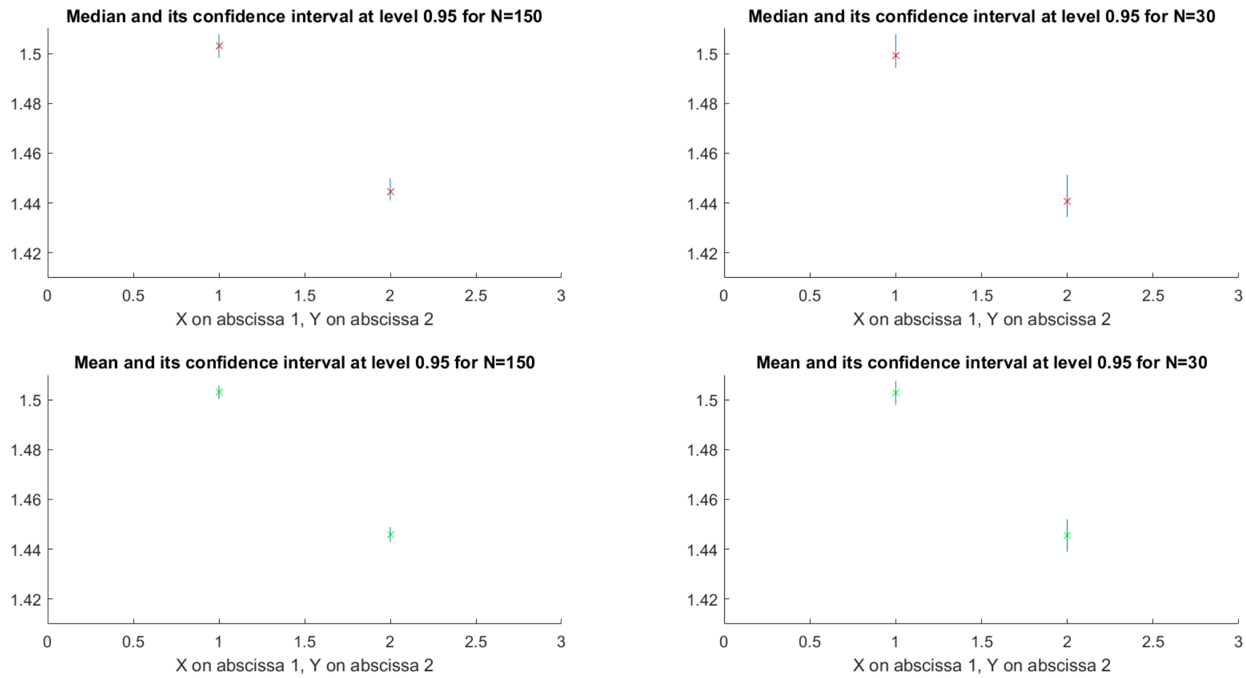


Figure 24: Median and Mean of $(X_n)_n$ and $(Y_n)_n$ and their confidence intervals for 150 and 30 users.

4.2 PREDICTION INTERVALS FOR SAMPLES

For each of the N mobiles we compute the average of the instant speed values sampled at arbitrary instants of time (time average speed for each user). This gives you a data sequence Y_1, Y_2,..., Y_N.

In Figure 25 are displayed the prediction intervals at level 0.95 of Y_n for N=150, 60 and 30 assuming the data is normal and using the order statistics.

To compute these intervals, the following formulas have been used :

- Prediction interval for a sample at level 95% assuming the data are iid and normal : $[\mu \pm 1.96\sigma]$, where μ is the mean and σ the standard deviation of the sequence, from Theorem 2.6 in Le Boudec 2010 (page 52)
- Prediction interval for a sample at level 95% computed using order statistic assuming the sequence is iid, has common distribution with density and is sorted in ascending order : $[Y_{(j)}, Y_{(k)}]$ where $\frac{k-j}{n+1} \geq 0.95$, from Theorem 2.5 in Le Boudec 2010 (page 51)

For N=30, as explained in slide 48 of our course about confidence intervals, there's no prediction interval defined using the order statistics.

As a first observation, prediction intervals are wider than confidence interval as displayed in Table 3. For both N=150 and N=60, the interval computed using the order statistic are wider than intervals computed assuming the normal distribution.

The definition of the PI assuming the normal distribution is independent of N and thus we don't see much width difference between samples of size N=150, 60, 30. The PI using the order statistics is more narrow for N=60 compared to N=150, this can be explained by the dependence of the definition on N.

Assuming normal	order Statistic
$[1.4092, 1.4826]$	$[1.3931, 1.4963]$

Table 3: PI for $(Y_n)_n$ at level 0.95 based on N=150 Samples.

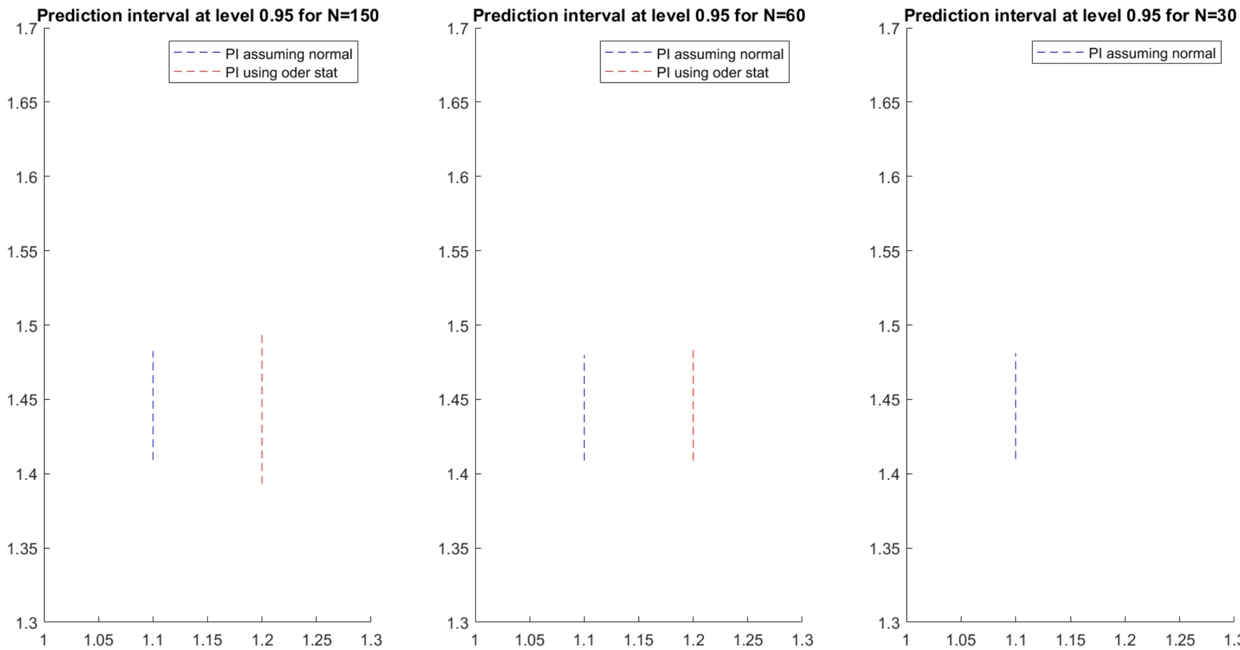


Figure 25: Prediction interval for a sample Y_1, Y_2, . . . , Y_N at level 0.95 computed using order statistic for N = 30, 60, 150.

5 Code

Code to generate data and display figures : https://github.com/ZinebAg/Homework2

References

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