Mattia Piras s4524072 31/05/2022

Robotics Engineering AY. 2021/2022

Università di Genova

Research Track 2 assignment's report

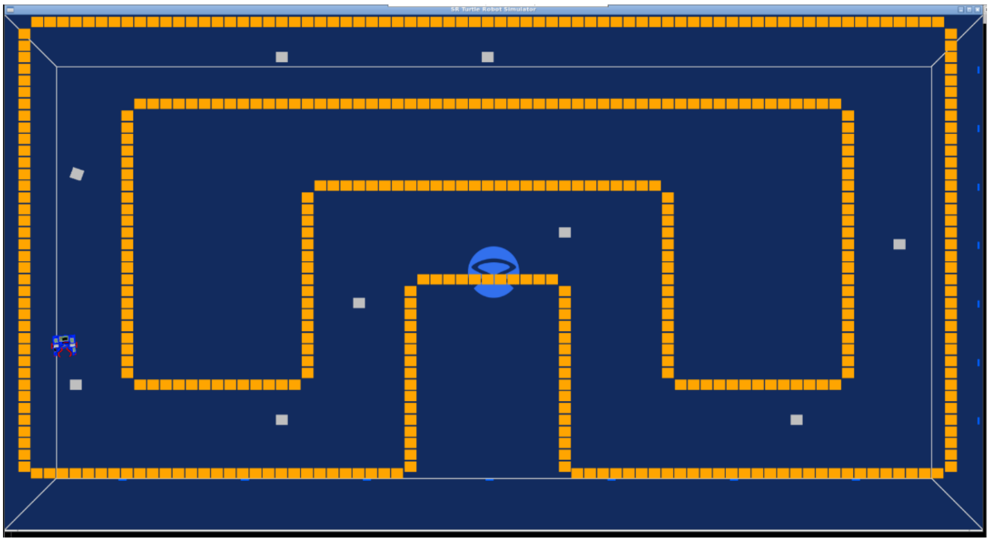
Statistical Analysis on two different approaches of the first assignment of Research Track 1

INTRODUCTION:

Firstly, let me present the problem to be analysed:

- We have a simple python code for letting a mobile robot
  - running inside a given arena, without crushing against walls
  - grabbing some boxes, randomly distributed in the middle of the path

We can see the typo of the exercise in the following picture:



Then after some tests I compered my code with the one written by the professor of this course.

What I have compered:

1) Average distances from walls
2) Average completion lap time
3) Crushes or number of direction inversion

If you are interested in the code, I leave here both GitHub repositories:

- https://github.com/MattiP99/rt-assignment1

- https://github.com/CarmineD8/python_simulator (**on rt2 branch**)

Important Questions:

- How to perform a quite suitable analysis without so many values?
- Are these values related to chances or to real information?

Hypothesis Testing

It is a key procedure in inferential statistics. It is based on the idea that we can tell things about the population from a sample taken from it. It can be explained in five steps

- Hypothesis
- Significance
- Sample
- P-value
- Decide

1) Hypothesis

Decide on your hypothesis: You need a null hypothesis: H0 and an alternative hypothesis H1 or Ha . Inferential statistics is based on the premise that you cannot prove something to be true, but you can disprove something by finding an exception.
You decide what you're trying to provide evidence for… which is the alternative hypothesis, then you set up the opposite as the null hypothesis and find evidence to disprove that. It can be said that the alternative hypothesis is usually a thing we're trying to prove or find out about, while the null hypothesis is the opposite or status quo.

NOTE: The hypothesis always about the population parameters, not the sample values or statistics
NOTE2: The null hypothesis usually refers to the status quo – the thing we're trying to find evidence against. It generally represents no effect.
NOTE3: The null hypothesis should include a statement of equality and the alternative should not.

2) Significance

Decide on the level of significance unless there is a good reason not to, people generally use 0.05 as the significance level also known as the alpha value. The significance value is the probability that you will say that the null hypothesis is wrong when really it is correct.

3) Sample

Take a sample from a population to provide the statistics you need.

4) P-value

Calculate the p-value. This is almost always done by a computer package

5) Decide

Use the p-value to decide whether to reject the null hypothesis. If the p-value is less than significance level you chose earlier, you will reject the null hypothesis. The sample has given you evidence that the null hypothesis is wrong

T-TEST (2 samples)

T-Test, also known as Student's Test, is based on t-distribution, and is considered an appropriate test for judging the significance of a sample mean or for judging the significance of difference between the means of two samples in case of small sample(s) when population variance is not known (in which case we use variance of the sample as an estimate of the population variance). The relevant test statistic, t, is calculated from the sample data and then compared with its probable value based on t-distribution at a specified level of significance for concerning degrees of freedom for accepting or rejecting the null hypothesis.

How does it work?

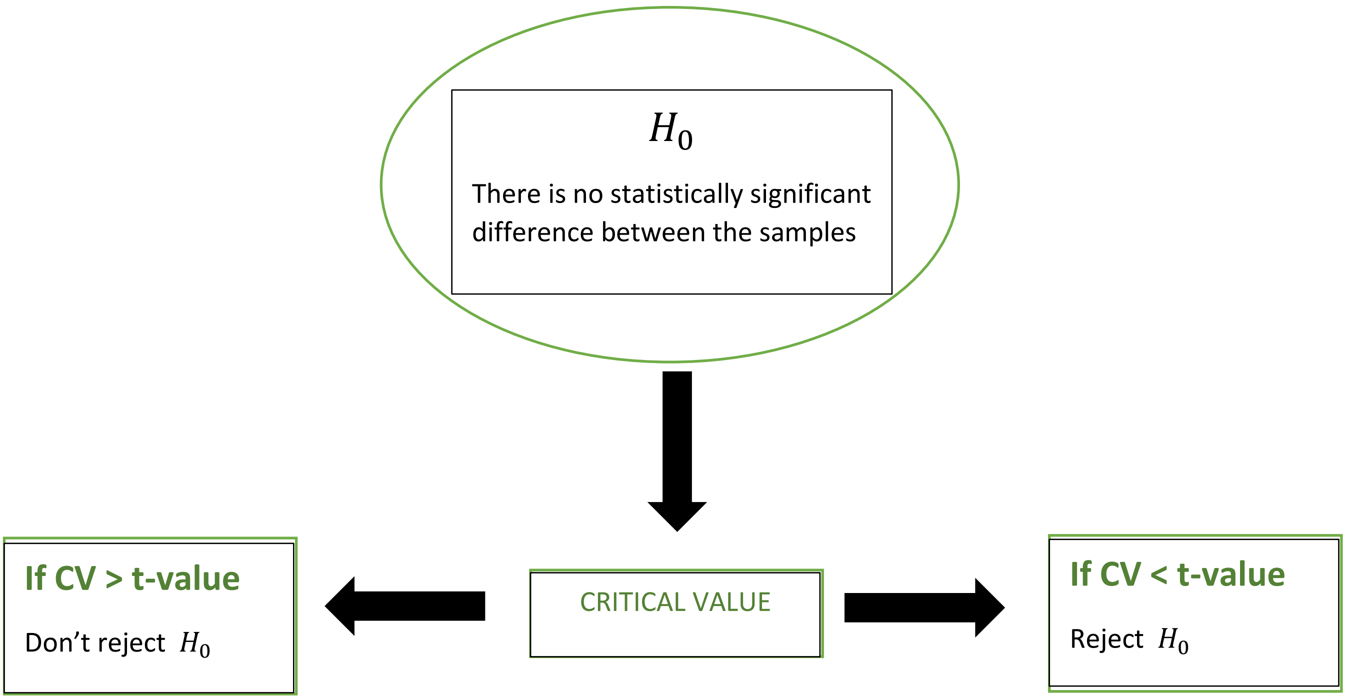
It is based on the t-value: it is like checking how many data are important with respect to the one that are in the sample by chance.

$$t - value = \frac{signal}{noise} = \frac{difference\ between\ 2\ group\ mean}{variability\ of\ groups} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{{s_1}^2}{n_1} + \frac{{s_2}^2}{n_2}}}$$

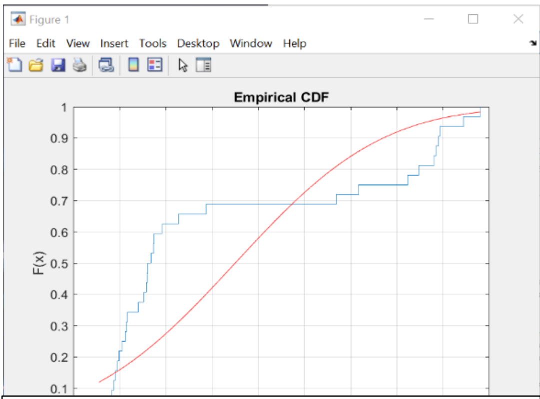
Where:

- $\bar{x}_1 and\ \bar{x}_2$ are the two mean values of are population
- s_1 and s_2 are the two standard deviations
- ${s_1}^2$ and ${s_2}^2$ are the two variances: Increasing them means decreasing the t-value

- n_1 and n_2 are the number of samples that have been considered. Increasing them means increasing the t-value up to a certain point

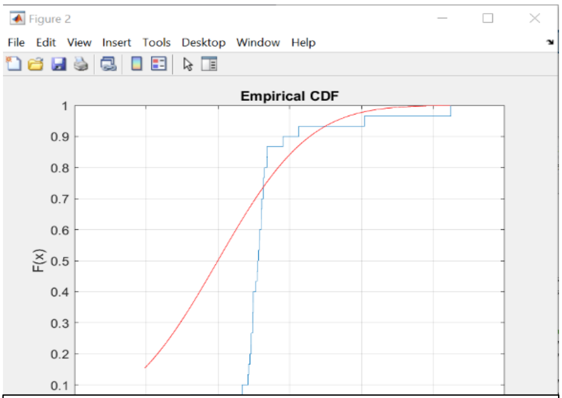


Firstly, I had to assure that I was working with normal distribution data since t-distribution is based on the normal one.



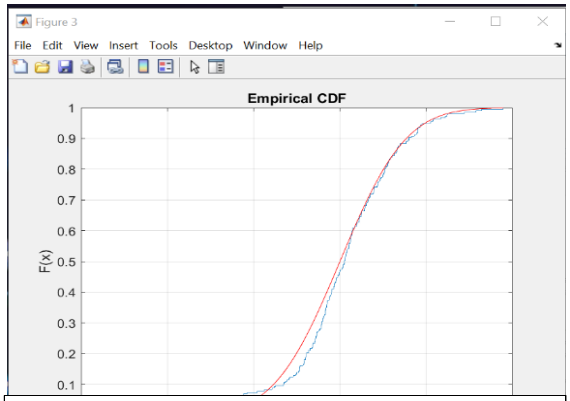
Gaussian: `normcdf(mT7_values,75,5)`

Sample: `mT7_values`



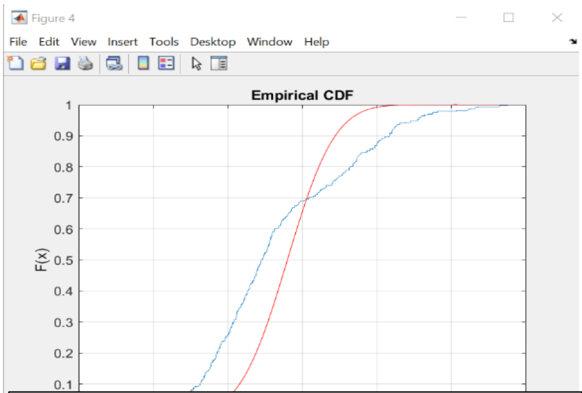
Gaussian: `normcdf(mT9_values,70,10)`

Sample: `mT9_values`



Gaussian: `normcdf(mD7_values,1.5,0.3)`

Sample: `mD7_values`



Gaussian: `normcdf(mD9_values,1.4,0.25)`

Sample: `mD9_values`

For sake of space and simplicity I reported only values about my code and not the professor ones too.

They refer to time taken for a lap with 7 and 9 boxes and to average distances from walls (here with 7 or 9 boxes on the path too). They both concern 30 values corresponding to 30 laps.

Another test that has been computed is the kstest(X) that confirms if or not your distribution is equal (or at least is similar) to a normal distribution.

We can notice from diagrams that whenever I considered more values, my distribution is more like a normal one.

So, the two samples are equal or is there any statistical difference between them?

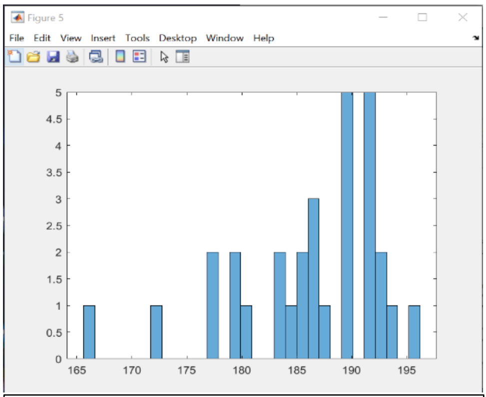
Then, instead of passing through each computation I simply run the ttest2 command in MATLAB with an alpha value of 0.05.

Here my results:

	h = rejection of H_0 or not (1 or 0)	p-value
Mine_dist7/Professor_dist7	1	1.3574e-06
Mine_dist9/Professor_dist9	1	9.7459e-04
Mine_time7/Professor_time7	1	4.7706e-64
Mine_time9/Professor_time9	1	4.2016e-55

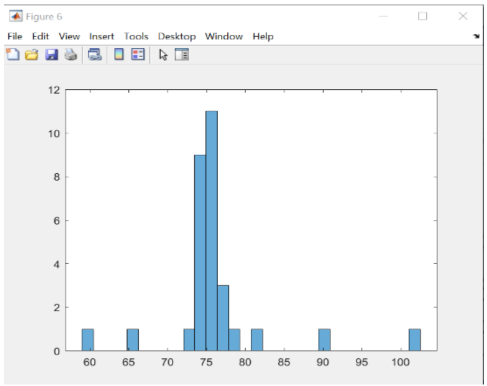
So, as we can see there are differences between the two algorithms.

What I've noticed about time:



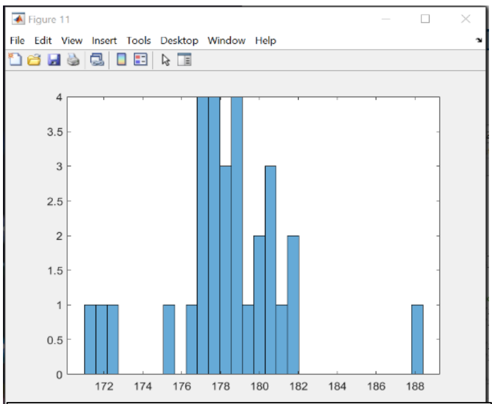
Professor average time with 9 boxes

avgP9 = 186.0155



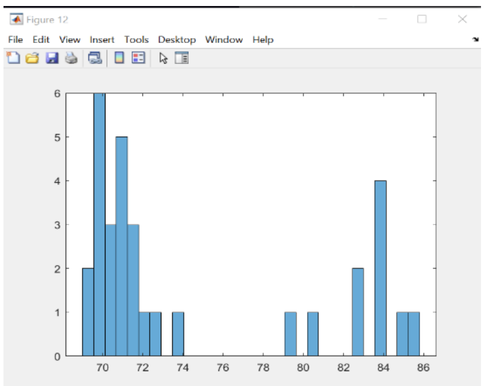
Mattia average time with 9 boxes

avgM9 = 76.3056



Professor average time with 9 boxes

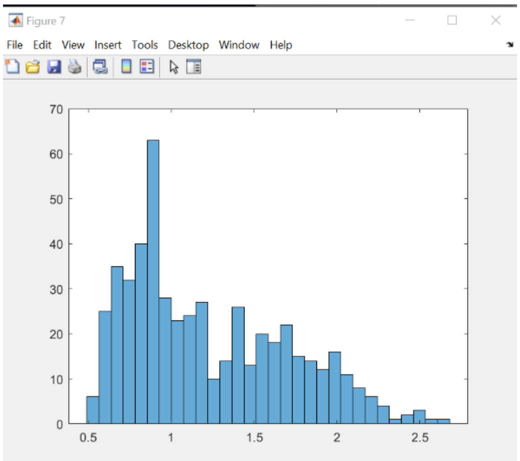
avgP7 = 178.3136



Mattia average time with 9 boxes

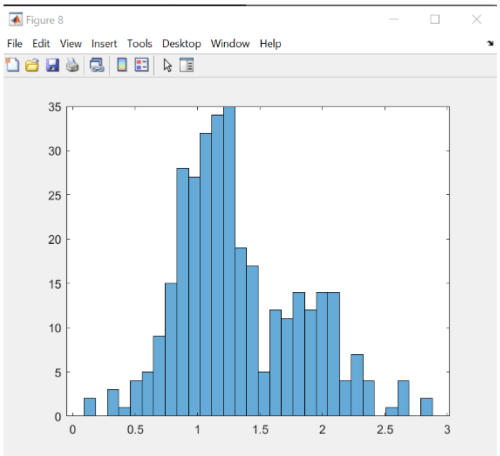
avgM7 = 73.5933

What I've noticed about distances:



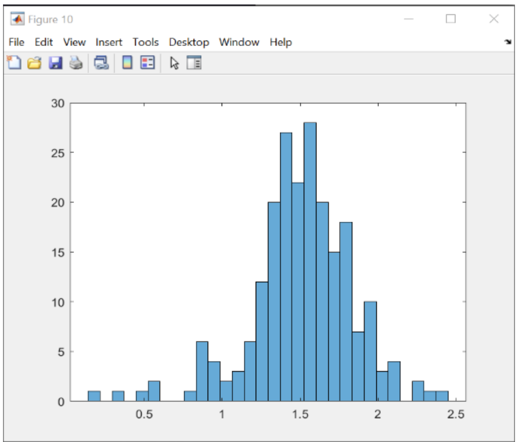
Prof. average distance with 7 boxes

avgP7 = 1.6506



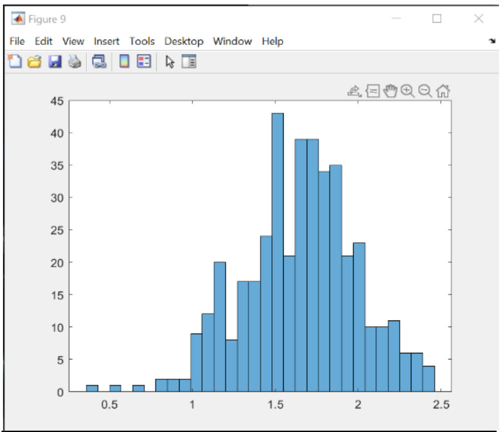
Mattia average distance with 7 boxes

avgM7 = 1.5126



Prof. average distance with 9 boxes

avgP7 = 1.2238



Mattia average distance with 9 boxes

avgM7 = 1.3372

CONCLUSIONS:

The tests performed reject in all 4 cases the null hypothesis with a pretty low probability of making some mistakes in doing that.

About the differences : I'm able to state that my algorithm is faster than the Professor 'one and the trajectory performed by the mobile robot is more accurate: the car, in my case, does turn with a more accurate logic and performs an adjustment of its angle with respect to the wall in order to stay as much parallel to the wall directions as possible while the script of my professor let the robot turning until the obstacle is no more in front of its self. This last approach, unfortunately, leads to a non-optimal path (neither mine is optimal, but it is slightly better) since the car moves in a more zigzag way.

In both cases there are some issues after some lap: In my case the robot tends to drive with a too high velocity letting it to crush (this happened 5 times in 60 laps). Instead, regarding professor algorithm, its robot tends to stop or to change direction without a specific reason (this happened 15 in 60 laps).

IMPROVEMENTS:

Some improvements can be done. For example:

- it would be good considering in each section of the arena what happens.
- Various changings of the velocity level can be done to notice differences of both algorithms a different speed.
- More samples could be considered to have a distribution more similar to the normal one.
- I'd have liked to see both algorithms performing without any boxes inside the path. To evaluate only the different approaches to the trajectories.