UNIVERSITÀ DEGLI STUDI DI GENOVA

DIBRIS

DEPARTMENT OF COMPUTER SCIENCE AND TECHNOLOGY,
BIOENGINEERING, ROBOTICS AND SYSTEM ENGINEERING

RESEARCH TRACK 2

First Assignment

Statistics

Author:

Luca Petruzzello - 5673449

Professors:

Carmine Recchiuto

May 2, 2023

Contents

1 Introduction

In this report, I compared my algorithm from the first assignment of Research Track 1 and that of a colleague called Carmine. I modified the environment so that for the first experiment I had four tokens per circle, for the second five, for the third six and for the last seven and it is essential to remember that the position of the tokens in the circles changes randomly after each run. For each environment obtained, I ran both algorithms 10 times and measured the time required to complete the assignment task, then, again for each environment, I ran both algorithms 25 times to compare how often they succeeded or failed. I then plotted from the data, performed the Lilliefors test to ensure that the acquired time data belonged to a normal distribution, and finally performed the t test to compare the times in each environment. As for the data concerning successes and failures, I ran chi square text with contingency tables and, since the latter are non-parametric, I did not need to run the Lilliefors test on these data.
All calculations were verified with MatLab, while plots were performed on Jupyter Notebook.

This is the link for my assignment: `https://github.com/luk1897/Research_Track_1-Assignment-1`
This is the link for my colleague's assignment: `https://github.com/Carmine00/ResearchTrack-Assignment-1`

2 Environments

Below I will graphically show all the environments used.

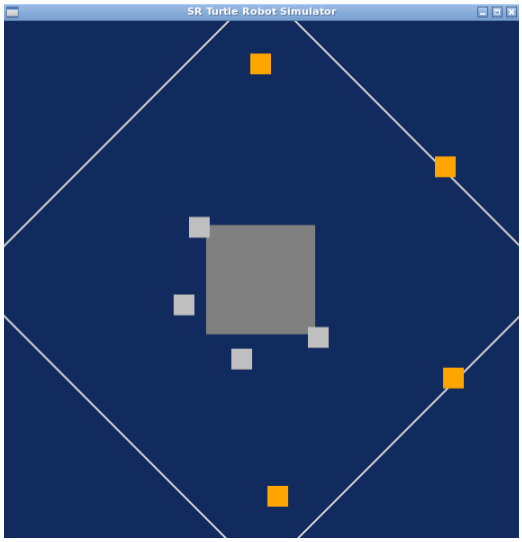

Figure 1: One of 4 tokens environments used

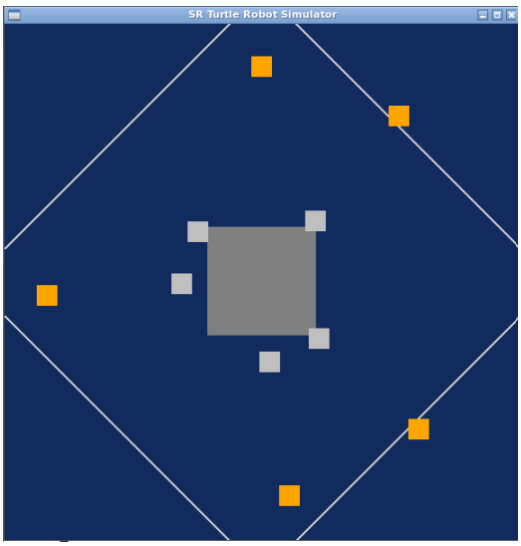

Figure 2: One of 5 tokens environments used

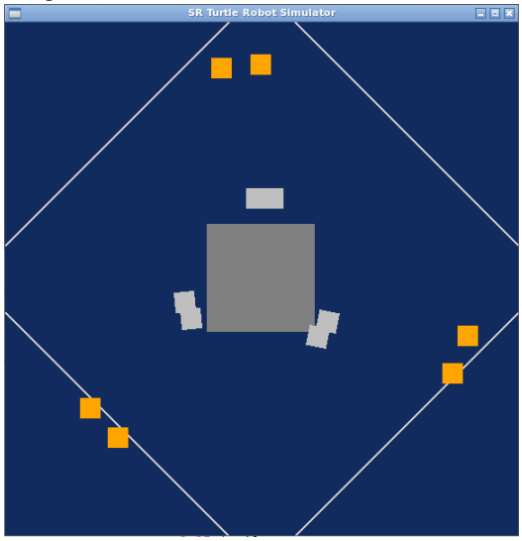

Figure 3: One of 6 tokens environments used

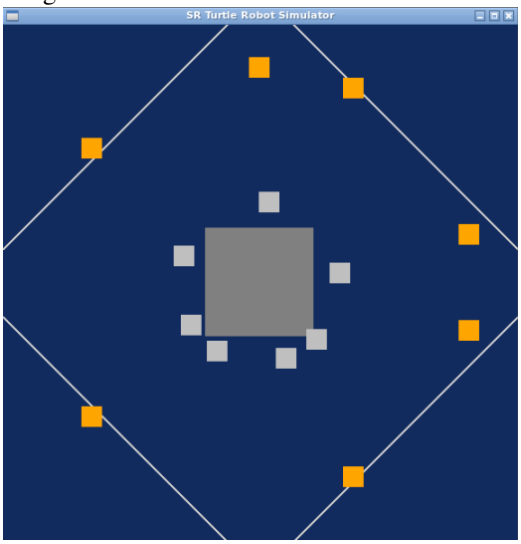

Figure 4: One of 7 tokens environments used

3 Data and hypothesis

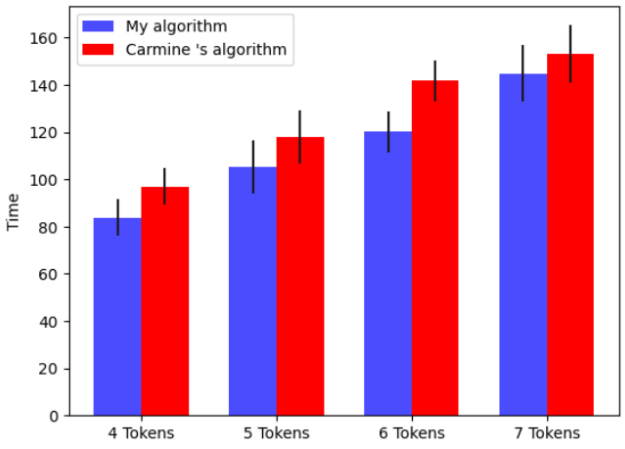

Figure 5: Average time of both algorithms for each environment

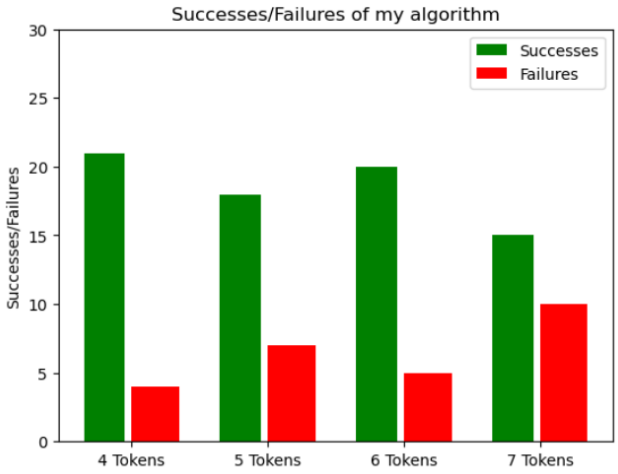

Figure 6: Successes/Failures of my algorithm

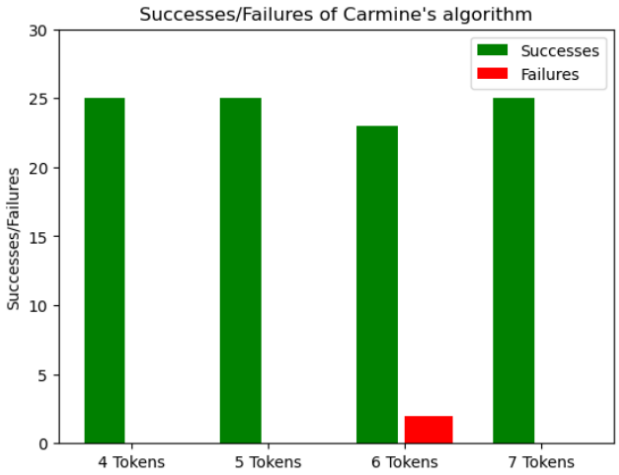

Figure 7: Successes/Failures of Carmine's algorithm

In Figure 5, we can observe the average time required to complete the task for each algorithm and in each environment. We can see that the times of my algorithm are shorter than those of Carmine's algorithm with a difference that tends to increase in the environment with 6 tokens. The standard deviation of each environment shows that randomly placing tokens at each run results in quite different values.
In figure 6 it is possible to see that my algorithm had problems when varying the tokens, because as the number of tokens varied, so did their position, and it happened a few times that the robot could not move forward due to the proximity of two or more tokens.
Looking at figure 7, on the other hand, it can be said that my colleague's algorithm had no problem finishing tasks in any environment, except in the environment with 6 tokens where the algorithm failed once.

According to the data collected, I hypothesise that my algorithm has better time performance than Carmine's, while I hypothesise that Carmine's algorithm is more reliable, i.e. can finish the task under any condition, than mine.

4 Lilliefors Test

As mentioned above, the Lilliefors test is used to prove that samples belong to a normal distribution since only when this condition occurs I can use the parametric T-test. In this case, I used this test for the time data. On the other hand, for the success and failure data, I used contingency tables and therefore did not use this test. If h is equal to zero, then the null hypothesis is accepted and the sample in question belongs to normal distribution, if, on the other hand, h is equal to one, then the sample does not belong there. I used a 5% significance level.

Below are the results of this test.

Lilliefors Test		
Algorithm	Number of token	h
My algorithm	4 tokens	0
Carmine's algorithm	4 tokens	0
My algorithm	5 tokens	0
Carmine's algorithm	5 tokens	0
My algorithm	6 tokens	0
Carmine's algorithm	6 tokens	0
My algorithm	7 tokens	0
Carmine's algorithm	7 tokens	0

Table 1: Lilliefors Test for both algorithms

As shown in table 1, all results demonstrate that, for each sample, the null hypothesis of the Lilliefors test is accepted, therefore all samples belong to a normal distribution.

5 T-Test

The T-test is an instrument for evaluating the averages of one or two populations by means of hypothesis testing and can be used to determine whether a single group differs from a known value (one-sample T-test), whether two groups differ from each other (two-sample independent T-test) or whether there is a significant difference in paired measurements (paired T-test). In order to compare the time averages of two algorithms in each environment, I used the two-sample one-tail T-test to test which of the two algorithms takes the shortest time to complete the task.
First of all, I defined the hypothesis to be tested:

$$\begin{cases} h0 : \mu_1 \neq \mu_2 \\ h1 : \mu_1 < \mu_2 \end{cases}$$

where h0 is the null hypothesis, h1 is the alternative hypothesis, $\mu 1$ is the time average of my algorithm and $\mu 2$ is the time average of Carmine's algorithm. These hypothesis were used for each environment. I chose a 5% risk of drawing an incorrect conclusion. Furthermore, there must be homogeneity in the variance, which is why I calculated the pooled variance and as mentioned earlier, I collected 10 time data for each algorithm in each environment, so the degrees of freedom turn out to be 18.
Below there are the test results obtained.

T-Test			
Token Map	h	p-value	dof
4 tokens	1	0.0079	18
5 tokens	1	0.0286	18
6 tokens	1	0.0258	18
7 tokens	0	0.2368	18

Table 2: Results of T-Test for each environment

With h=1, I accept the alternative hypothesis that my algorithm takes less time to complete the task than Carmine's algorithm, and this is true for the environments with 4, 5 and 6 tokens, therefore my hypothesis is accepted for this three environments,
Instead h=0 indicates that the null hypothesis is accepted and that therefore my algorithm and Carmine's algorithm have the same time performance and this applies to the environment with 7 tokens. Therefore in the latter case my hypothesis is rejected.
P-value represents the value below which the data from a statistical test lead to the rejection of the null hypothesis.

6 Chi Square Test

With regard to the successes and failures of the two algorithms applied to each environment, I used the Chi-test for the contingency tables to understand whether two nominal qualitative variables are associated or independent of each other. I performed the test by initially accepting the null hypothesis, then treating the data obtained as if it were part of a single algorithm and applied this reasoning for each environment. With these data, I calculated the chi-square estimator for each environment at a 5% significance level.

Chi-Test		
Token Map	h	dof
4 tokens	1	1
5 tokens	1	1
6 tokens	0	1
7 tokens	1	1

Table 3: Results of Chi-Test for each environment

With h=1, the null hypothesis is rejected and therefore the two algorithms do not have the same level of reliability and, considering the data, this means that Carmine's algorithm is more reliable than mine in the environments with 4, 5 and 7 tokens and my hypothesis is verified. With h=0, the null hypothesis is accepted and therefore the two algorithms have the same level of reliability in the environment with 6 tokens and my hypothesis is rejected.

7 Conclusions

Thanks to the T-test, I proved that my algorithm is faster than Carmine's algorithm in all environments except the one with 7 tokens, while thanks to the Chi-test, I proved that Carmine's algorithm is more reliable in all environments except the one with 6 tokens. So my assumptions proved true except in the environments mentioned. I can say with certainty that in principle my algorithm is faster than Carmine's one and that Carmine's one is more reliable than mine.

8 Appendix

[Comment] Add here additional material (if needed)

8.1 Appendix A

8.2 Appendix B