UNIVERSITÀ DEGLI STUDI DI GENOVA

DIBRIS

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

RESEARCH TRACK II

Assignment 3

Perform a statistical analysis

Author:

Parisa Farvardin (S6101141)

May 21, 2024

1 Introduction

The third assignment of the Research Track II is to carry out a statistical analysis of the first assignment of the Research Track I class. This analysis aims to compare two distinct approaches: one is my code and the other one is developed by my classmate. The goal is to ascertain which approach exhibits superior performance (which means the time it consumes) under varying quantities of the tokens in the given scenario. To recall the situation, there was a Robot who had to move and collect all the Tokens which was spread in the environment and place them in the center (or any other place but together).

1.1 Hypothesis

Hypothesis testing is a statistical method used to make inferences about a population based on sample data. It involves formulating two competing hypotheses: the **null hypothesis (H0)** and the **alternative hypothesis (Ha)**. The two hypotheses are mutually exclusive, meaning that if one is true, the other one must be false.
The process of hypothesis testing involves collecting sample data and using statistical techniques to assess the likelihood of observing such data if the null hypothesis is true. This is done by calculating a test statistic, which measures the difference between the observed sample data and what would be expected under the null hypothesis.
Once the test statistic (e.g. a t-statistic, z-statistic, chi-square statistic, etc.) is calculated, its probability of occurrence is determined assuming the null hypothesis is true. This probability, known as the p-value, indicates the likelihood of obtaining the observed data or more extreme results if the null hypothesis is true. A small **p-value** (probability of getting a test statistic) suggests that the observed data is unlikely under the null hypothesis, providing evidence against it.
Based on the calculated p-value and a predetermined significance level (usually denoted as α), a decision is made whether to reject the null hypothesis in favor of the alternative hypothesis or to fail to reject the null hypothesis due to insufficient evidence. If the p-value is less than or equal to the significance level (α), the null hypothesis is rejected, indicating that there is sufficient evidence to support the alternative hypothesis. Conversely, if the p-value is greater than the significance level, the null hypothesis is not rejected.
Hypothesis testing is a powerful statistical tool that allows researchers to make decisions based on data. By specifying the null and alternative hypotheses, calculating a test statistic, and calculating a p-value, we can determine whether or not the evidence supports the claim made by the alternative hypothesis.

$$H0 : \alpha_p = \alpha_c$$

$$Ha : \alpha_p < \alpha_c$$

Where,

- H0 is the Null hypothesis, meaning that there is no relevant distinction between the two algorithms. The averages of the time required for completion are similar.
- Ha is the Alternative hypothesis, stating that there is a significant difference between the two implementations. One algorithm is faster than the other in reaching the goal.
- α_p is the average time taken to reach the goal by my algorithm.
- α_c is the average time taken to reach the goal by my colleague's algorithm.

2 Data Collection

The experiment was conducted using various arena configurations, with the number of tokens ranging from 2 to 7. To modify the number of tokens in the arena, it is required to change the value of **TOKENS PER CIRCLE** in the **two colors assignment arena.py** file inside the **arenas** folder. Data collection comprised two projects: SimulationA, my project, and SimulationB, developed by a classmate. Each project underwent five executions for every token configuration, resulting in a total of 30 trials per project. The analysis focused on measuring the total time required by the robot to fulfill its task. To record this data, simple commands were added to the main function of the code:

```
import time

start_time = time.time()
# Robot completes its task
end_time = time.time()

total_time = end_time - start_time
print ("Time:", total_time )
```

Figure 1 indicates the result of the comparison between two codes in order to find which one behaves better in the way that the time of doing the task for the Robot is shorter. However, for certain configurations, one algorithm appears to be more efficient than the other, which is probably due to how the algorithms for moving the robot in the environment have been implemented.

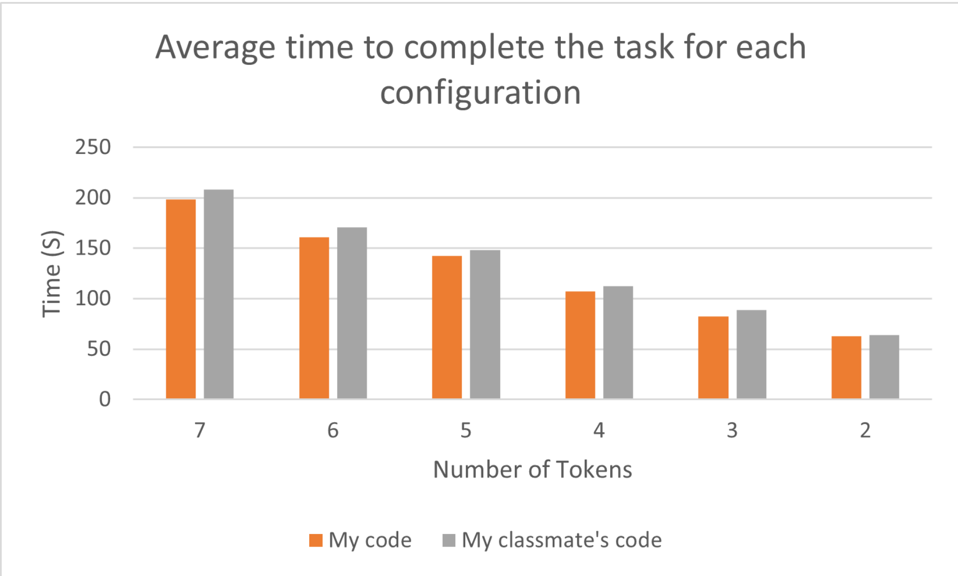
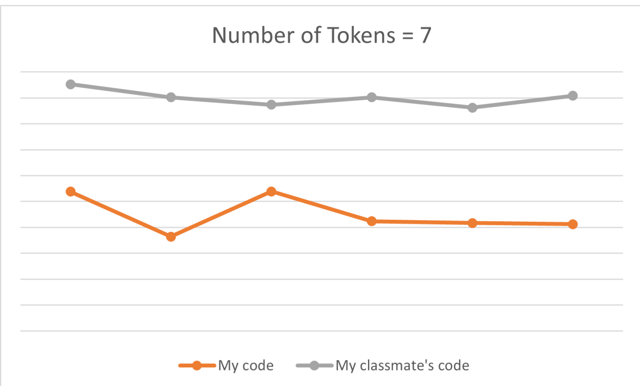
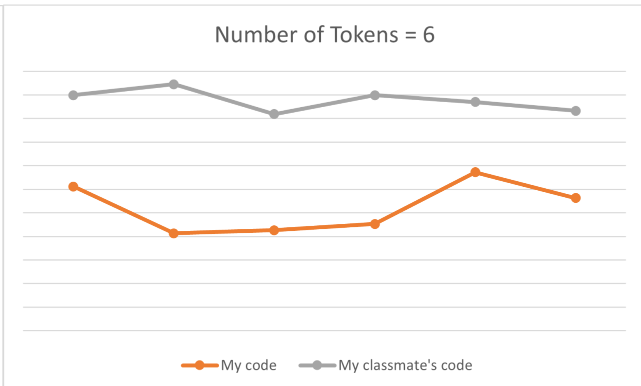


Figure 1: The graph represents the average time taken to complete the task for each configuration in the arena.

These line graphs indicate the detailed difference between the two algorithms for each token configuration in every 5 intervals:





3 T-Test

Since the analysis deals with the comparison of two different approaches applied to the same scenario, the method it is used is the Paired T-Test. A Paired T-Test is used to compare two population means where you have two samples in which observations in one sample can be paired with observations in the other sample. The t-test calculates the t-statistic, which measures the difference between the means of the two groups relative to the variation within each group. By comparing the t-statistic to a critical value derived from the t-distribution, we can determine whether the observed difference in means is statistically significant or simply due to random chance. To test the null hypothesis that the true mean difference is zero, the procedure is as follows:

- Calculate the difference ($di = yi - xi$) between the two observations on each pair, distinguishing between positive and negative differences.
- Calculate the mean difference, $\bar{d}$.
- Calculate the standard deviation of the differences, s_d, and use this to calculate the standard error of the mean difference, $SE(\bar{d}) = \sqrt{s_d}/\sqrt{n}$
- Calculate the t-statistic, which is given by $T = \bar{d}/SE(\bar{d})$. Under the null hypothesis, this statistic follows a t-distribution with n-1 degrees of freedom.
- Use tables of the t-distribution to compare the value for T to the $t_{(n-1)}$ distribution. This will give the p-value for the paired t-test.

The code was executed 30 times each and data was collected on the execution time and now all ingredients are ready to start the calculations:

- First calculate the mean of the differences:

$$\bar{d} = \frac{\sum_{i=1}^{n}(x_i)}{n} = -6.31377434$$

- Then, the standard deviation of the differences:

$$s_d = \frac{\sum_{i=1}^{n}(x_i - \mu)^2}{n} = 13.6399802$$

- And then, the standard error of the mean difference:

$$SE(\bar{d}) = \frac{s_d}{\sqrt{n}} = 2.49030828$$

- Finally, the t-value:

$$t - value = \frac{\bar{d}}{SE(\bar{d})} = -2.535338452$$

4 Conclusion

As can be seen from the graphs and results, my algorithm is faster than the implementation done by my colleague. In the scenarios of this test, the number of Tokens has changed during the test and the run time recorded, and as a result the algorithm still works for each Token but, obviously, with a different amount of time consumption. The t-value obtained must be compared with the values available in the t-table. The row with Degrees of Freedom given by $DoF = N - 1 = 29$ (the sample population is $N = 30$) is taken, and the value closest to that obtained from the calculations is looked for. In this case, $t = -2.53$. The paired test is used since the thesis asserts that my algorithm has shorter run times than my colleague's. Thus, we can reject the Null Hypothesis with a confidence of 98%.

t-test table

cum. prob	$t_{.50}$	$t_{.75}$	$t_{.80}$	$t_{.85}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$	$t_{.999}$	$t_{.9995}$
one-tail	0.50	0.25	0.20	0.15	0.10	0.05	0.025	0.01	0.005	0.001	0.0005
two-tails	1.00	0.50	0.40	0.30	0.20	0.10	0.05	0.02	0.01	0.002	0.001
df											
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.31	636.62
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.000	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.000	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.000	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.000	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.000	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.000	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.000	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.000	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.000	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.000	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.000	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.000	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.000	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.000	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.000	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.000	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.000	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.000	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.000	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.000	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.000	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.000	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.000	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.000	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.000	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.000	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	0.000	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	0.000	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.232	3.460
80	0.000	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.195	3.416
100	0.000	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174	3.390
1000	0.000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	3.098	3.300
z	0.000	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.090	3.291
	0%	50%	60%	70%	80%	90%	95%	98%	99%	99.8%	99.9%
	Confidence Level										

Figure 2: T-test table