UNIVERSITÀ DEGLI STUDI DI GENOVA

DIBRIS

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

RESEARCH TRACK II

# Third Assignment

## Statistical analysis

*Authors:*

Desalegn Girum Molla - 6020433
Nima Abaeian - 5967579

*Professor:*

Carmine Tommaso Recchiuto

May 31, 2024

# Contents


**1 Introduction** **2**

**2 Hypotheses** **2**

**3 Experimental Setup** **3**
3.1 Random token distribution with different number of tokens setup . . . . . . . . . . . . . . . . . . . 3
3.2 Computation of time setup . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 4

**4 Experimental Result** **4**

**5 Statistical analysis** **5**
5.1 Time completion parameter . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 5
5.2 Percentage of success rate parameter . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 6

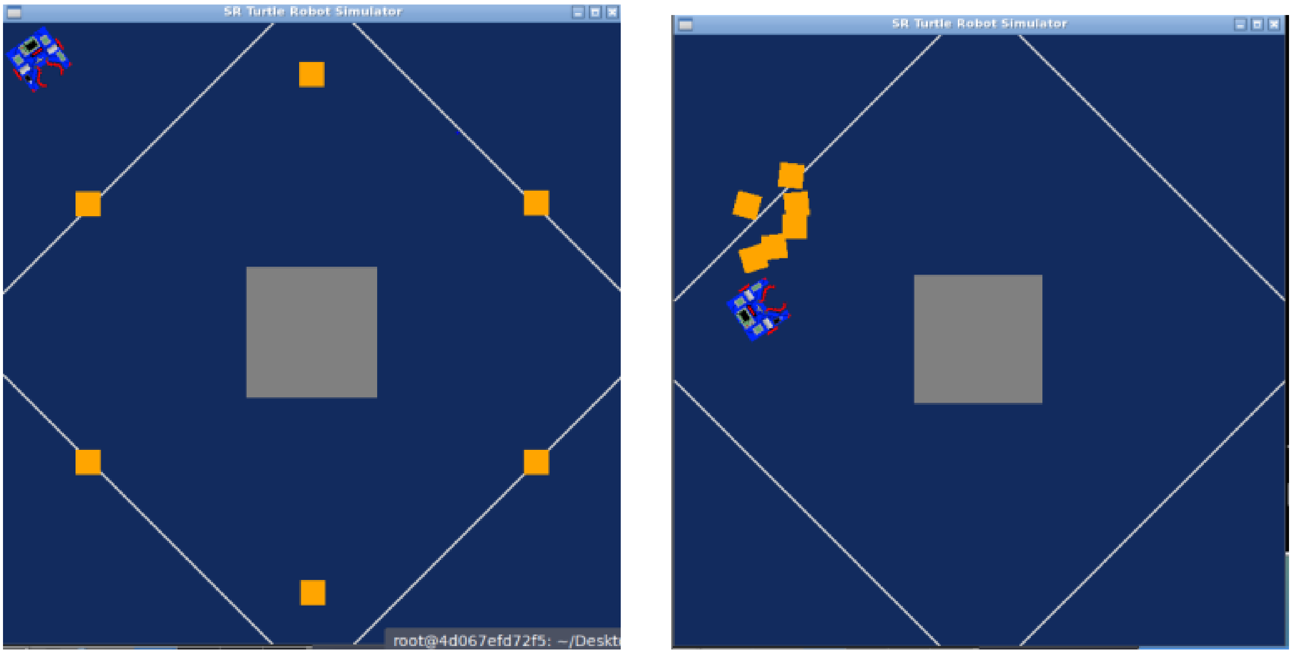
**6 Conclusion** **6**

**7 Appendix** **7**

# 1 Introduction

For the third assignment in Research Track II, our goal is to conduct an in-depth statistical analysis building on the first assignment from Research Track I. This analysis will compare two separate implementations: one by Girum Molla Desalegn (S6020433) and the other by Nima (S5967579). For simplicity, Girum's implementation will be labeled as "G-algorithm" and Nima's as "N-algorithm" throughout this report.

The main objective of the first assignment in Research Track I is to develop a Python node to control the robot to collect all the golden boxes. The task involves writing code that ensures all gathered boxes are placed onto a single base box. To achieve this, the assignment includes functions and a while loop tailored to meet this goal. In the Research Track II third assignment, the tokens are now distributed randomly throughout the environment.




(a) Before (b) After

Figure 1: Research Track I First Assignment

The following sections of this report will detail the experimental setup, present the results from the data analysis, and discuss the implications of these findings. Lastly, we will conclude by summarizing the results and addressing the hypothesis.

# 2 Hypotheses

The hypotheses were designed to examine the performance differences between the "G-algorithm" and the "N-algorithm" in terms of their effectiveness in ensuring that all collected boxes are placed onto a single base box.

The null hypothesis ($H_0$) asserts that there is no significant difference in the performance of the "G-algorithm" and the "N-algorithm" in carrying out this task.

From other perspective, the alternative hypothesis ($H_a$) asserts that there is a significant difference in the performance when completing the task. This means that the performance outcomes for each algorithm differ in a measurable way.

To evaluate these hypotheses, an analysis using a T-test was performed on the completion time data for each algorithm. The T-test offers a statistical tool for comparing the means of two separate samples and determining if the observed distinctions are statistically meaningful. It's suitable for scenarios where the sample size is relatively small and when the standard deviation of the population is not known.

# 3 Experimental Setup

This experiment aimed to assess the performance of both the 'G-algorithm' and the 'N-algorithm' to finish a given specific task. The experiment involved randomly placing tokens within the environment and changing the number of tokens. The experiment was carried out over 40 iterations to gather sufficient data to evaluate the hypothesis.

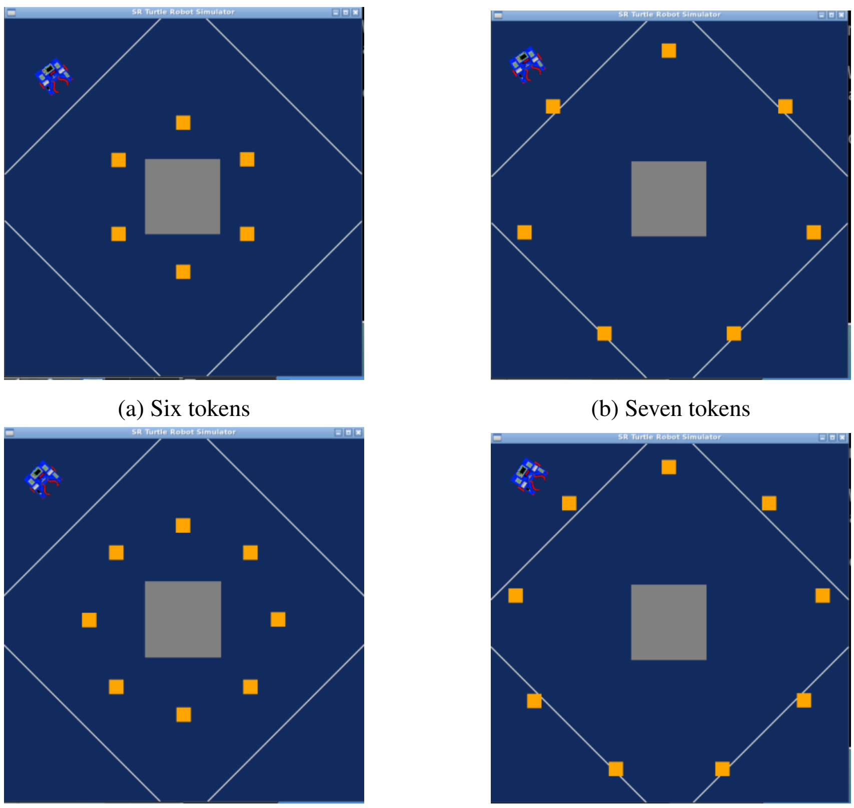
The experimental setup was designed as follows:

- The experiment was conducted in a randomly generated token environment, utilizing configurations with 6, 7, 8, and 9 tokens.
- The experiment was carried out over 40 iterations. For each iteration, the following steps were performed: 10 steps with 6 tokens, 10 steps with 7 tokens, 10 steps with 8 tokens, and 10 steps with 9 tokens.
- The experiment analyzed the success and failure rates of the two algorithms at each step. This involved evaluating and comparing the performance of each algorithm to determine their effectiveness and reliability throughout the various steps of the process.
- For each step of the experiment, the time taken was meticulously recorded. This involved measuring the duration required to complete each individual step, providing detailed insight into the efficiency and speed of the process.

## 3.1 Random token distribution with different number of tokens setup

To randomize the placement of the tokens and to adjust the number of tokens modifications were made to the $robot - sim/sr/robot/arenas/two_{c}olours_{a}ssignment_{a}rena.py$ file.

```
import random
INNER_CIRCLE_RADIUS = round (random.uniform(0.2,1),1)
OUTER_CIRCLE_RADIUS = round (random.uniform(1,2.5),1)
TOKENS_PER_CIRCLE = 6  # if we want to increase the token, set 7/8/9
```



(a) Six tokens (b) Seven tokens

(c) Eight tokens (d) Nine tokens

Figure 2: Randomly distributed token with different number of tokens

### 3.2 Computation of time setup

In order to accurately measure the time needed to complete a specific task for all steps, we have implemented the following code for both the ”G-algorithm” and the ”N-algorithm.”

```
# when while loop begin
start_time = time.time()  #time start
# After the end of task is completed
end_time = time.time()
time_taken = round((end_time - start_time),2)
print('time taken:', time_taken,'seconds')
```

## 4 Experimental Result

The experimental results derived from the performance analysis of the ”G-algorithm” and ”N-algorithms” in terms of their task completion capabilities are presented in this section. These results are illustrated in the accompanying graph, which provides a visual comparison of the algorithms’ efficiencies and effectiveness. The analysis covers the time required to complete each iteration step of the tasks and the percentage of success rates for both algorithms.

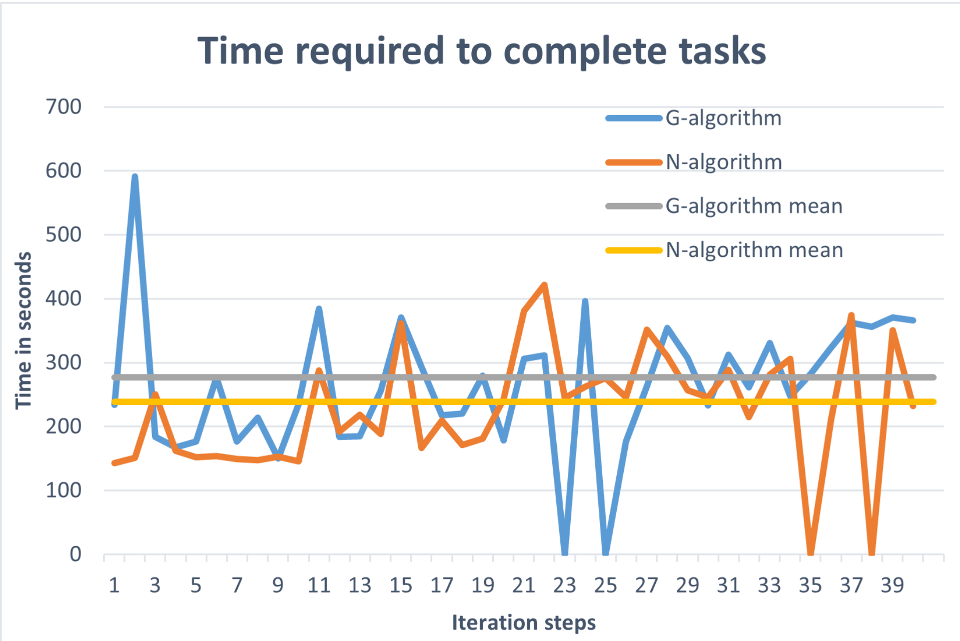



Figure 3: Time required for every step of iteration

In Figure 3, the x-axis represents the repetitions, while the y-axis denotes the completion time for each algorithm. The graph provides a clear visualization of the trends and patterns observed in the completion time data.

This graph demonstrates the N-algorithm consistently outperforms the G-algorithm in terms of time required to complete tasks across all iteration steps. This is evident from the N-algorithm’s mean line being constantly lower than the G-algorithm’s. The N-algorithm accomplishes the same tasks as the G-algorithm relatively in a shorter amount of time. The N-algorithm might be using a more efficient approach to solve the problem and having lower computational complexity.

To give a thorough overview of the results, a detailed Excel table has been included in the Appendix. This table contains the data recorded for each algorithm.

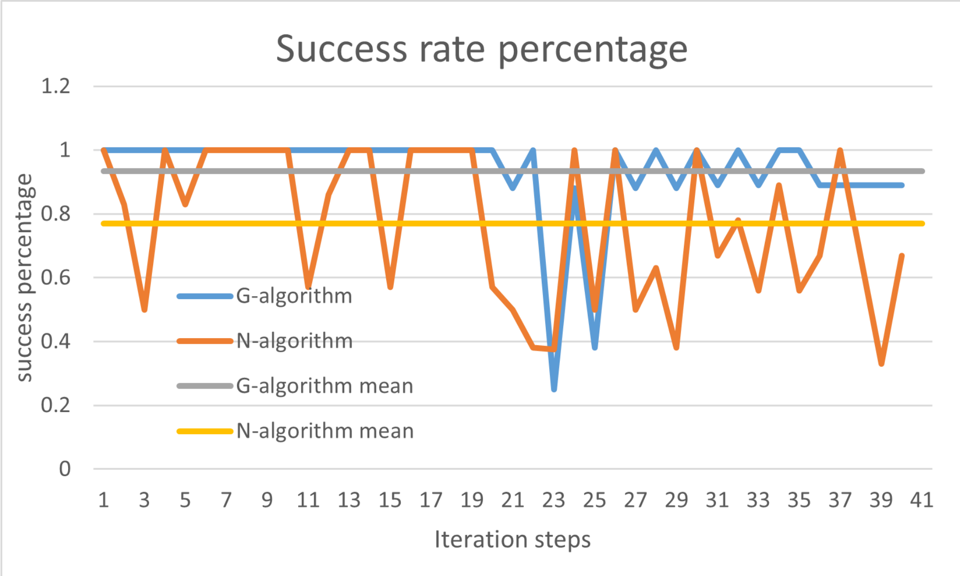



Figure 4: Percentage of success rate

In figure 4, a graph showing the success rate of different iterations for two algorithms, G-algorithm and N-algorithm. The y-axis is labeled "Success rate percentage", and the x-axis is labeled "Iteration steps". There are two lines plotted on the graph, one for each algorithm, and a horizontal line representing the average success rate across all iterations for each algorithm.

The G-algorithm has a higher success rate than the N-algorithm across all iterations. This is evident from the G-algorithm's line being consistently higher than the N-algorithm's line.

# 5 Statistical analysis

A t-test is a crucial tool for comparing the performance of "The G-algorithm" and "The N-algorithm" by assessing whether their mean performance metrics are significantly different. It involves a null hypothesis($H_0$) that assumes no difference between the means of the "G-algorithm" and the "N-algorithm" and an alternative hypothesis($H_a$) suggesting a significant difference.

For this dataset, a two-sample t-test is appropriate because the G-algorithm and N-algorithm are independent, the population parameters are unknown, and we need to compare their performance metrics. Specifically, the two-sample t-test is designed to determine whether there is a significant difference between the means of two independent groups.

## 5.1 Time completion parameter

To perform two sample t-test for time completion parameter:

$$H_0 : \mu_G = \mu_N$$
$$H_a : \mu_G \neq \mu_N$$

**Mean value:** $\mu_G = \frac{\sum_{i=1}^{N_G}}{N_G} = 277.249$ and $\mu_N = \frac{\sum_{i=1}^{N_G}}{N_N} = 239.017$

**Standard deviation:** $\sigma_G = \sqrt{\frac{\sum_{i=1}^{N_G}(X_i - \mu_G)^2}{N_G}} = 88.652$ and $\sigma_N = \sqrt{\frac{\sum_{i=1}^{N_N}(X_i - \mu_G)^2}{N_N}} = 77.415$

**Degree of freedom:** $N_G + N_N - 2 = 38 + 38 - 2 = 74$

**Mean value difference:** $\mu_G - \mu_N = 277.249 - 239.017 = 38.232$

Throughout the experiment, 40 iterations were carried out for each algorithm. Nevertheless, it is crucial to highlight that both the "G" and "N" algorithms encountered 2 failed iterations each. These setbacks led to a reduction in the effective sample size for both algorithms.
The pooled standard deviation calculation allows for the creation of a unified estimate of the overall standard deviation. This adjusted estimate takes into account variations in group sizes.

**Pooled variance:** $\hat{\sigma}^2_{pooled} = \frac{(N_G-1)\sigma_G^2+(N_N-1)\sigma_N^2}{N_G+N_N-2} = 6926.130$

**Standard Error of Difference:** $\hat{\sigma}_{\mu_G-\mu_N} = \sqrt{\hat{\sigma}^2_{pooled}(\frac{1}{N_G}+\frac{1}{N_N})} = 19.093$

Finally, the t-value of the two sample t-test may be computed as:

**t value:** $t_{\mu_G-\mu_N} = \frac{\mu_G-\mu_N}{\sigma_{\mu_G-\mu_N}} = \frac{277.249-239.017}{19.093} = 2.00$

With a significance level ( $\alpha$) of 0.05 and 74 degrees of freedom, we compared the calculated t-value of 2.00 to the critical t-value from the t-distribution table. The critical t-value for a two-tailed test at this significance level is approximately $\pm$1.9925.

$$|t| > 1.9925$$

Since the absolute value of the calculated t-value is greater than the critical t-value, there is a substantial difference between the performance of the "G-algorithm" and "N-algorithm" in terms of this parameter.

## 5.2 Percentage of success rate parameter

To perform two sample t-test for percentage of success rate parameter:

$$H_0 : \mu_G = \mu_N$$
$$H_a : \mu_G \neq \mu_N$$

**Mean value:** $\mu_G = \frac{\sum_{i=1}^{N_G}}{N_G} = 0.9345$ and $\mu_N = \frac{\sum_{i=1}^{N_G}}{N_N} = 0.7699$

**Standard deviation:** $\sigma_G = \sqrt{\frac{\sum_{i=1}^{N_G}(X_i-\mu_G)^2}{N_G}} = 0.1534$ and $\sigma_N = \sqrt{\frac{\sum_{i=1}^{N_N}(X_i-\mu_G)^2}{N_N}} = 0.2344$

**Degree of freedom:** $N_G + N_N - 2 = 40 + 40 - 2 = 78$

**Mean value difference:** $\mu_G - \mu_N = 0.9345 - 0.7699 = 0.1646$

Throughout the experiment, 40 iterations were carried out for each algorithm.
The pooled standard deviation calculation allows for the creation of a unified estimate of the overall standard deviation. This adjusted estimate takes into account variations in group sizes.

**Pooled variance:** $\hat{\sigma}^2_{pooled} = \frac{(N_G-1)\sigma_G^2+(N_N-1)\sigma_N^2}{N_G+N_N-2} = 0.0392$

**Standard Error of Difference:** $\hat{\sigma}_{\mu_G-\mu_N} = \sqrt{\hat{\sigma}^2_{pooled}(\frac{1}{N_G}+\frac{1}{N_N})} = 0.0443$

Finally, the t-value of the two sample t-test may be computed as:

**t value:** $t_{\mu_G-\mu_N} = \frac{\mu_G-\mu_N}{\sigma_{\mu_G-\mu_N}} = \frac{0.9345-0.7699}{0.0443} = 3.7156$

Likewise a significance level ( $\alpha$) of 0.05 and 78 degrees of freedom, we compared the calculated t-value of 3.7156 to the critical t-value from the t-distribution table. The critical t-value for a two-tailed test at this significance level is approximately $\pm$1.9908.

$$|t| > 1.9925$$

Since the absolute value of the calculated t-value is greater than the critical t-value, there is a substantial difference between the performance of the "G-algorithm" and "N-algorithm" in terms of this parameter.

# 6 Conclusion

Based on this experiment, we concluded that there is a significant difference between the "G-algorithm" and the "N-algorithm" in terms of task completion time and success rate. We reject the null hypothesis and accept the alternative hypothesis. Therefore, the "N-algorithm" exhibits a significantly different completion time performance compared to the "G-algorithm." However, in terms of the success rate, the "G-algorithm" outperformed the "N-algorithm," demonstrating a higher percentage of successful completions.

# 7 Appendix

| Experiments | | G-algorithm | | N-algorithm | |
|---|---|---|---|---|---|
| Number of steps | Number of tokens | Time takes | Success Percentage | Time takes | Success Percentage |
| 1 | 6 | 233.92 | 1 | 142.5 | 1 |
| 2 | 6 | 591.19 | 1 | 150.98 | 0.83 |
| 3 | 6 | 183.75 | 1 | 250.53 | 0.5 |
| 4 | 6 | 167.3 | 1 | 162.41 | 1 |
| 5 | 6 | 176.45 | 1 | 152.2 | 0.83 |
| 6 | 6 | 276.93 | 1 | 154.24 | 1 |
| 7 | 6 | 176.6 | 1 | 149.52 | 1 |
| 8 | 6 | 214.37 | 1 | 147.41 | 1 |
| 9 | 6 | 149.89 | 1 | 152.92 | 1 |
| 10 | 6 | 234.68 | 1 | 145.46 | 1 |
| 11 | 7 | 384 | 1 | 287.31 | 0.57 |
| 12 | 7 | 183.9 | 1 | 191.7 | 0.86 |
| 13 | 7 | 185.09 | 1 | 218.9 | 1 |
| 14 | 7 | 255.51 | 1 | 188.62 | 1 |
| 15 | 7 | 370.55 | 1 | 361.88 | 0.57 |
| 16 | 7 | 293.36 | 1 | 166.55 | 1 |
| 17 | 7 | 217.94 | 1 | 209.81 | 1 |
| 18 | 7 | 220.43 | 1 | 171.08 | 1 |
| 19 | 7 | 279.18 | 1 | 180.75 | 1 |
| 20 | 7 | 178.55 | 1 | 243.07 | 0.57 |
| 21 | 8 | 305.97 | 0.88 | 380.86 | 0.5 |
| 22 | 8 | 311.09 | 1 | 422.05 | 0.38 |
| 23 | 8 | failed | 0.25 | 245.1 | 0.375 |
| 24 | 8 | 396.5 | 0.88 | 262.61 | 1 |
| 25 | 8 | failed | 0.38 | 275.88 | 0.5 |
| 26 | 8 | 176.65 | 1 | 245.71 | 1 |
| 27 | 8 | 263.04 | 0.88 | 351.75 | 0.5 |
| 28 | 8 | 354.01 | 1 | 309.59 | 0.63 |
| 29 | 8 | 307.24 | 0.88 | 256.74 | 0.38 |
| 30 | 8 | 233.43 | 1 | 245.56 | 1 |
| 31 | 9 | 312.62 | 0.89 | 289.12 | 0.67 |
| 32 | 9 | 261.36 | 1 | 215.05 | 0.78 |
| 33 | 9 | 330.19 | 0.89 | 281.09 | 0.56 |
| 34 | 9 | 248.05 | 1 | 305.76 | 0.89 |
| 35 | 9 | 282.64 | 1 | failed | 0.56 |
| 36 | 9 | 323.31 | 0.89 | 210.32 | 0.67 |
| 37 | 9 | 362.26 | 0.89 | 374.67 | 1 |
| 38 | 9 | 356.07 | 0.89 | failed | 0.67 |
| 39 | 9 | 370.92 | 0.89 | 350.39 | 0.33 |
| 40 | 9 | 366.51 | 0.89 | 232.57 | 0.67 |

Table 1: Comparison of G-algorithm and N-algorithm