Research Track II: Statistical Analysis

Mobina Alinaghian

May 25, 2024

1 Introduction

This report presents a static analysis of the first assignment completed in the previous semester for the course *Research Track I*. The analysis compares the performance of two algorithms: one implemented by myself and the other by Hoda Mostafanezhad. Both algorithms aim to navigate a robot in a virtual environment to collect and gather tokens. The primary objective of this experiment is to evaluate the effectiveness of the two scripts under varying conditions, specifically the number of tokens and the radius at which they are generated. By comparing these variations, we aim to understand how the algorithms adapt to different spatial configurations and to assess their overall success in achieving the intended task.

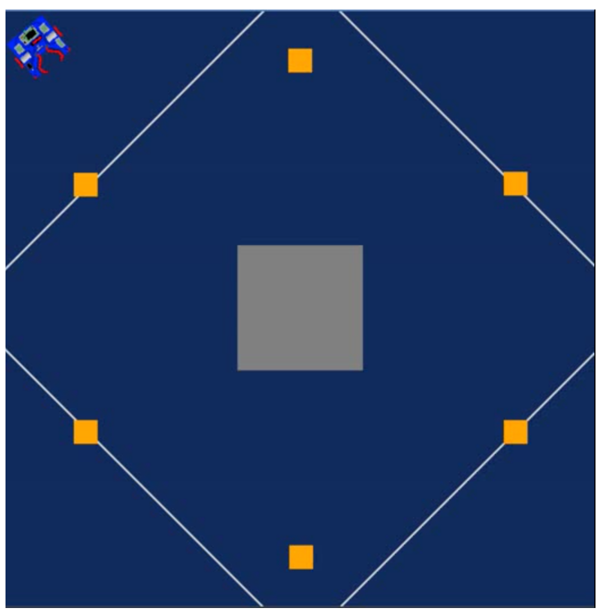


Figure 1: The environment of the project

2 Hypotesis

2.1 Explaination

Hypothesis testing is a fundamental statistical tool in scientific research that enables researchers to make informed decisions based on data. This process involves evaluating a claim or prediction about a population parameter, such as the mean or variance, using collected data. Typically, this evaluation is structured around two competing hypotheses: the null hypothesis and the alternative hypothesis.

The null hypothesis (H0) assumes no effect or no difference, serving as the default claim. **The alternative hypothesis (Ha)** proposes a specific effect or difference and is considered if the null hypothesis is rejected. These hypotheses are **mutually exclusive**; if one is true, the other must be false.

By conducting hypothesis testing, researchers can make probabilistic statements about population parameters, thereby aiding in the validation or rejection of hypotheses.

In the current analysis, the following hypotheses are used to assess the performance of the algorithms:

- **Null Hypothesis (H0):** The outcomes from both algorithms show high similarity, making it difficult to ascertain the superior performer overall. This hypothesis assumes that there is no significant difference between the performance of the two algorithms.
- **Alternative Hypothesis (Ha):** The two algorithms demonstrate notable differences in performance. This hypothesis indicates that there is a statistically significant distinction between how well the two algorithms perform.

2.2 Hypothesis

In this section, we compare the efficiency of two codes.

Thesis:

Hoda's code is significantly more efficient in completing its tasks compared to my code.

Hypothesis:

Null Hypothesis (H0):

There is no significant difference between the two algorithms. The average performance metrics are similar.

Alternative Hypothesis (Ha):

There is a significant difference between the two implementations. Hoda's algorithm is more efficient than mine in reaching the goal.

3 Testing

3.1 Method

Testing the hypotheses was conducted using a paired T-test. This statistical test is used to compare the means of two sets of paired observations. In this case, the performances of both programs were evaluated by running them on the same token configurations with varying radius and numbers of tokens.

We utilized a T-distribution for the analysis. To apply the central limit theorem to the sampling distribution and ensure it approximates a T-distribution, we required more than 30 samples.

To assess performance, we measured the time taken to complete each task. Additionally, we implemented a function to record the time consumed for task execution.

3.2 Collecting Data

We systematically measured the performance of both programs by varying the token radius and number. Each measurement from one program was paired with a corresponding measurement from the other. This pairing allowed us to analyze the performance of the programs under different experimental conditions.

To ensure accurate measurements of success, we conducted 5 runs of the algorithms for each configuration. This yielded a total of 45 overall states, exceeding the minimum threshold of 30. Consequently, we could apply the central limit theorem, leading to the conclusion that the sample distribution follows a T-distribution shape.

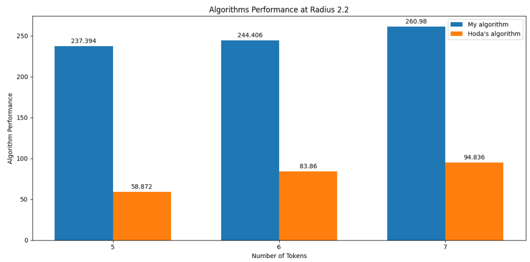
The table below presents the obtained results.

Sample Num.	Radius	Num. of Tokens	My algorithm	Hoda's algorithm	Success rate (mine)	Success rate (Hoda's)	Difference
1	2.4	6	262.59	90.91	1	1	0
2	2.4	6	262.54	90.93	1	1	0
3	2.4	6	262.59	90.92	1	1	0
4	2.4	6	262.57	90.89	1	1	0
5	2.4	6	262.58	90.92	1	1	0
6	2.4	5	250.01	60.27	0.6	0.8	0.2
7	2.4	5	249.99	60.27	0.6	0.8	0.2
8	2.4	5	252.08	62.24	0.6	0.6	0
9	2.4	5	249.99	60.21	0.6	0.8	0.2
10	2.4	5	251.90	63.02	0.6	0.6	0
11	2.4	7	270.65	93.81	0.428	0.857	0.429
12	2.4	7	270.65	93.82	0.428	0.857	0.429
13	2.4	7	270.63	93.85	0.428	0.857	0.429
14	2.4	7	271.05	95.88	0.428	0.714	0.286
15	2.4	7	270.71	96.82	0.428	0.714	0.286
16	2.7	6	268.59	92.83	1	1	0
17	2.7	6	269.49	93.83	1	1	0
18	2.7	6	262.59	93.79	1	1	0
19	2.7	6	262.61	93.81	1	1	0
20	2.7	6	262.78	93.81	1	1	0
21	2.7	5	268.01	71.77	0.6	0.8	0.2
22	2.7	5	267.79	71.82	0.6	0.8	0.2
23	2.7	5	267.78	71.70	0.6	0.6	0
24	2.7	5	268.03	71.82	0.6	0.8	0.2
25	2.7	5	268.03	71.83	0.6	0.6	0
26	2.7	7	298.29	98.10	0.428	0.857	0.429
27	2.7	7	293.45	97.84	0.428	0.857	0.429
28	2.7	7	299.51	97.81	0.428	0.857	0.429
29	2.7	7	295.88	102.82	0.428	0.714	0.286
30	2.7	7	298.67	104.02	0.428	0.714	0.286
31	2.2	5	237.01	58.77	0.6	1	0.4
32	2.2	5	236.79	58.81	0.6	1	0.4
33	2.2	5	238.08	58.73	0.8	1	0.2
34	2.2	5	237.03	59.02	0.6	0.8	0.2
35	2.2	5	238.06	59.03	0.8	0.8	0
36	2.2	6	244.69	83.36	0.6	1	0.4
37	2.2	6	243.09	83.34	0.8	1	0.2
38	2.2	6	245.59	84.31	0.6	0.8	0.2
39	2.2	6	246.07	84.34	0.6	0.8	0.2
40	2.2	6	242.59	83.95	0.8	0.8	0
41	2.2	7	260.66	93.81	0.428	1	0.572
42	2.2	7	261.65	93.82	0.428	1	0.572
43	2.2	7	260.63	93.85	0.428	1	0.572
44	2.2	7	261.25	95.88	0.428	0.857	0.429
45	2.2	7	260.71	96.82	0.428	0.857	0.429

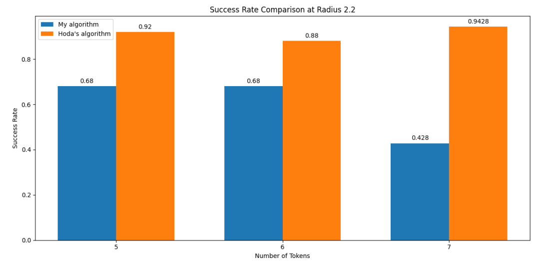
Table 1: Collected Data

3.3 Comparison

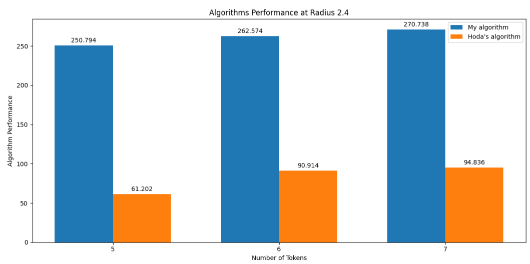
After collecting all the data, the next step involved comparing the two simulations through statistical analysis. Utilizing Python, I generated a bar graph to visually represent the comparison between the time taken to complete the task for each configuration in the area. This graph offers a clear visualization of the results, as depicted in Figure 2.



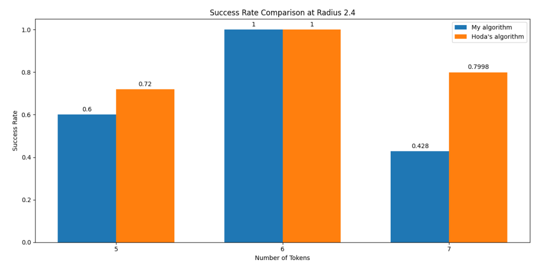
(a) Time Comparison_Radius 2.2



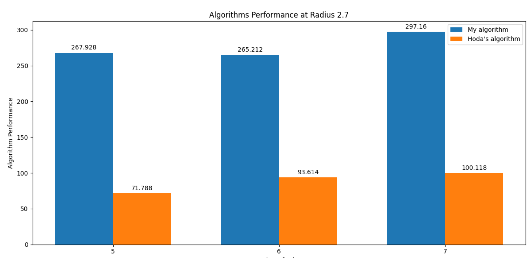
(b) Success Rate Comparison_Radius 2.2



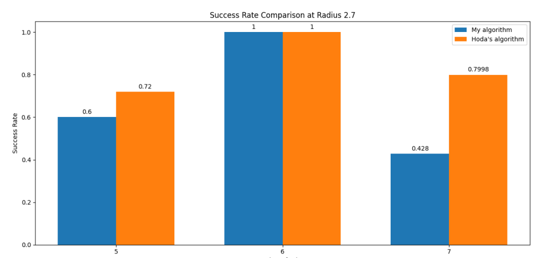
(c) Time Comparison_Radius 2.4



(d) Success Rate Comparison_Radius 2.4



(e) Time Comparison_Radius 2.7



(f) Success Rate Comparison_Radius 2.7

Figure 2: Comparison Between Two Algorithms

The data presented in the graph strongly supports the hypothesis. Ms. Hoda's code consistently demonstrates significantly more efficient task completion compared to mine. This is evident from the time of completing the task and the success rate of total task completion, where Hoda's code consistently outperforms mine by a significant margin.

4 T_Test

4.1 Explaination

The t-test assesses whether there's a significant difference between the means of two groups or populations, often used for small sample sizes or when the population standard deviation is unknown. It calculates the t-statistic, which measures the difference between group means relative to within-group variation.

By comparing this t-statistic to a critical value from the t-distribution, it determines whether the observed difference in means is statistically significant or due to random chance. Assumptions include data following an approximately normal distribution, independent and random sample selection, and approximately equal variances within each group.

Different Types:

- If the groups come from a single population, perform a *paired t-test*;
- If the groups come from two different populations, perform a *two-sample t-test* (or independent t-test);

- If there is one group being compared against a standard value, perform a *one-sample t-test*.

4.2 Computing

The mean difference ($\bar{d}$) and standard deviation (SE($\bar{d}$)) of the paired differences were calculated. Using these values, the T-value was computed, measuring the difference between the observed mean difference and the hypothesized mean difference under the null hypothesis.

The calculated T-value was compared to the critical T-value at a chosen significance level (5%). If the calculated T-value exceeded the critical T-value, the null hypothesis was rejected in favor of the alternative hypothesis, indicating a significant difference in performance.

We have computed the following values as part of our analysis:

1. Calculation of the difference between the two observations on each pair is done in the last column of the table.

2. Calculation of the mean of the difference:
   $\bar{d} = \frac{\sum_{i=1}^{n} d_i}{n}$
   $\bar{d} = \frac{9.692}{45} = 0.21538$

3. Calculation of the standard deviation of the difference:
   $S_d = \sqrt{\frac{\sum_{i=1}^{n}(d_i - \bar{d})^2}{n-1}}$
   Where:
   d_i : Individual difference
   $\bar{d}$: Mean difference
   n : Number of differences
   $S_d = 0.347$

4. Calculation of the standard error of the difference:
   $SE(\bar{d}) = \frac{S_d}{\sqrt{n}}$
   SE($\bar{d}$) $= \frac{0.347}{\sqrt{45}} = 0.0517$

5. Calculation of the T-value: $t = \frac{\bar{d}}{SE(\bar{d})} = 4.1659574468 (on 44 DOF)$

4.3 Results

The paired T-test was performed on the collected data, yielding a calculated Tvalue of 4.1659574468. With a significance level of $\alpha = 0.05$ and degrees of freedom (dof) = N − 1= 45 -1 = 44, where N represents the number of paired observations (45), the critical T-value at a two-sided test was 2.015.

5 Conclusion

Upon analysis, the calculated T-value (4.1659574468) surpassed the critical T-value (2.015). Consequently, we reject the null hypothesis, indicating a significant difference in the performance of the two programs across varied token counts and radius settings.

These findings suggest notable discrepancies in the programs' performance as the token count and radius vary. It implies that the programs' ability to match silver and golden tokens is notably affected by both the number of tokens and their spatial distribution within the arena.