Research Track 2 – Assessment 3

Statistical Analysis of Algorithms

Student Name:

Nafiseh Monavari

Student ID:

5174160

May 2023

1 Introduction

In Assignment 3, the aim is to assess the effectiveness of algorithms presented for problem-solving through statistical analysis. For instance, in the token arrangement problem, two algorithms can be compared based on their average execution time. Statistical measures such as mean and variance are commonly employed to calculate features like task execution time. These measures are then utilized to conduct statistical analyses such as t-test or chi-square test, which helps to determine the validity of the hypothesis that one algorithm is equally or more efficient compared to the others.

2 The Reacher Hypothesis (H0)

To evaluate the efficiency of algorithms through statistical analysis, two hypotheses are proposed:

- **Null Hypothesis (H0)**, which states that the efficiency of two algorithms is similar,
- **Alternative Hypothesis (H1)**, which suggests that one algorithm is more efficient than the other.

To verify these hypotheses, data must be collected through multiple experiments, and statistical parameters like average or variance are calculated based on features such as task duration or success rate. A statistical test is then conducted using these parameters, with the choice of the test depending on the shape of the statistical density function and the number of tests repeated. The t-test is preferred for situations where population variance is uncertain and at least 30 tests are conducted. Hypotheses are accepted or rejected based on a significance level, which is typically 1% or 5% in engineering applications.

3 Design the Experiments

To perform statistical analysis, it is necessary to carry out experiments and gather data. For the purposes of our discussion, we will be conducting a test involving the random arrangement of tokens in a playground, so that the radius of the arrangement of silver tokens will vary in each trial.

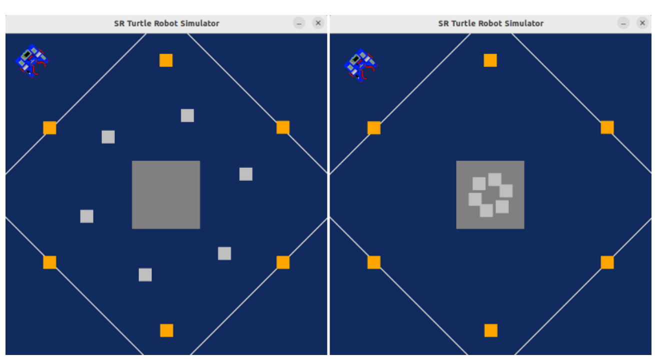

Figure 1 Random arrangement of silver tokens

The algorithms will be executed in the same playing field for each test, and the duration of task execution will be recorded for each experiment.

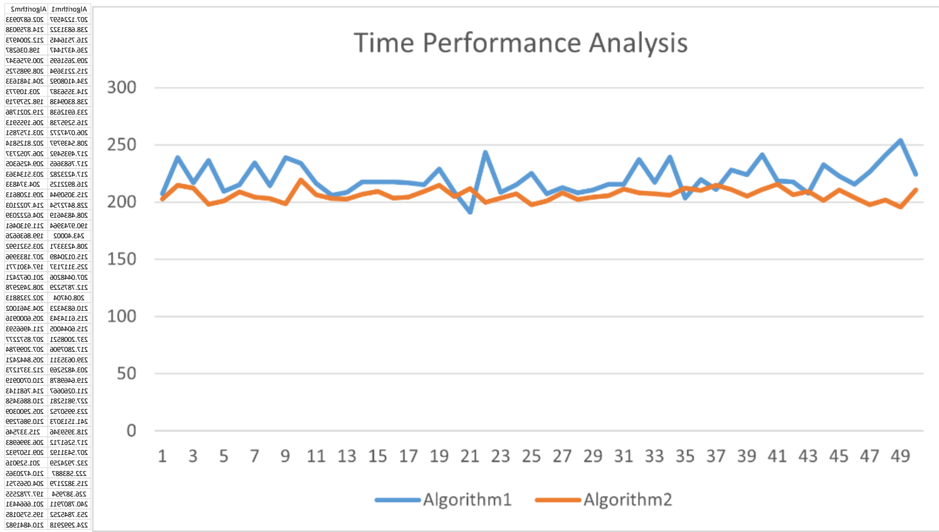

Figure 2 The Time performance analysis of the algorithms

According to the test results, both algorithms exhibit similar levels of time efficiency. However, to validate this hypothesis, it is imperative to conduct a statistical analysis using the t-test methodology:

$$mean_1 = 220.6267,\ \mathrm{var}_1 = 161.0993,\ n_1 = 50$$
$$mean_2 = 206.3551,\ \mathrm{var}_2 = 27.88595,\ n_2 = 50$$

t 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	[illegible]	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	[illegible]	1.962	2.330	2.581	3.098	3.300

Figure 3 t-table values

The table above displays the t-values for various degrees of freedom and tiles. To validate the null hypothesis, we use a t-test. With a confidence level of 5%, the critical value is 1.6600. The next step is calculating the t-value for the designed experiments :

$$t-value = \frac{|mean_1 - mean_2|}{\sqrt{\frac{\text{var}_1}{n_1} + \frac{\text{var}_2}{n_2}}} = 7.340811$$

Since the obtained t-value 7.340811 is much larger than the critical reference value 1.6600, we can confirm that the two algorithms are statistically significantly different. Therefore, we can reject the null hypothesis that states there are no significant differences between the performance of the two algorithms.

This conclusion leads us to support the alternative hypothesis, which suggests that one of the algorithms performs better than the other.