UNIVERSITÀ DEGLI STUDI DI GENOVA

DIBRIS

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

RESEARCH TRACK 2

Assignment 3

Statistical Analysis of the Robot Simulation Algorithms

Authors:	*Student ID:*
Iris Laanearu	s6350192
Mia Jane La Rocca	s6344889

Professor:

Carmine Recchiuto

May 25, 2024

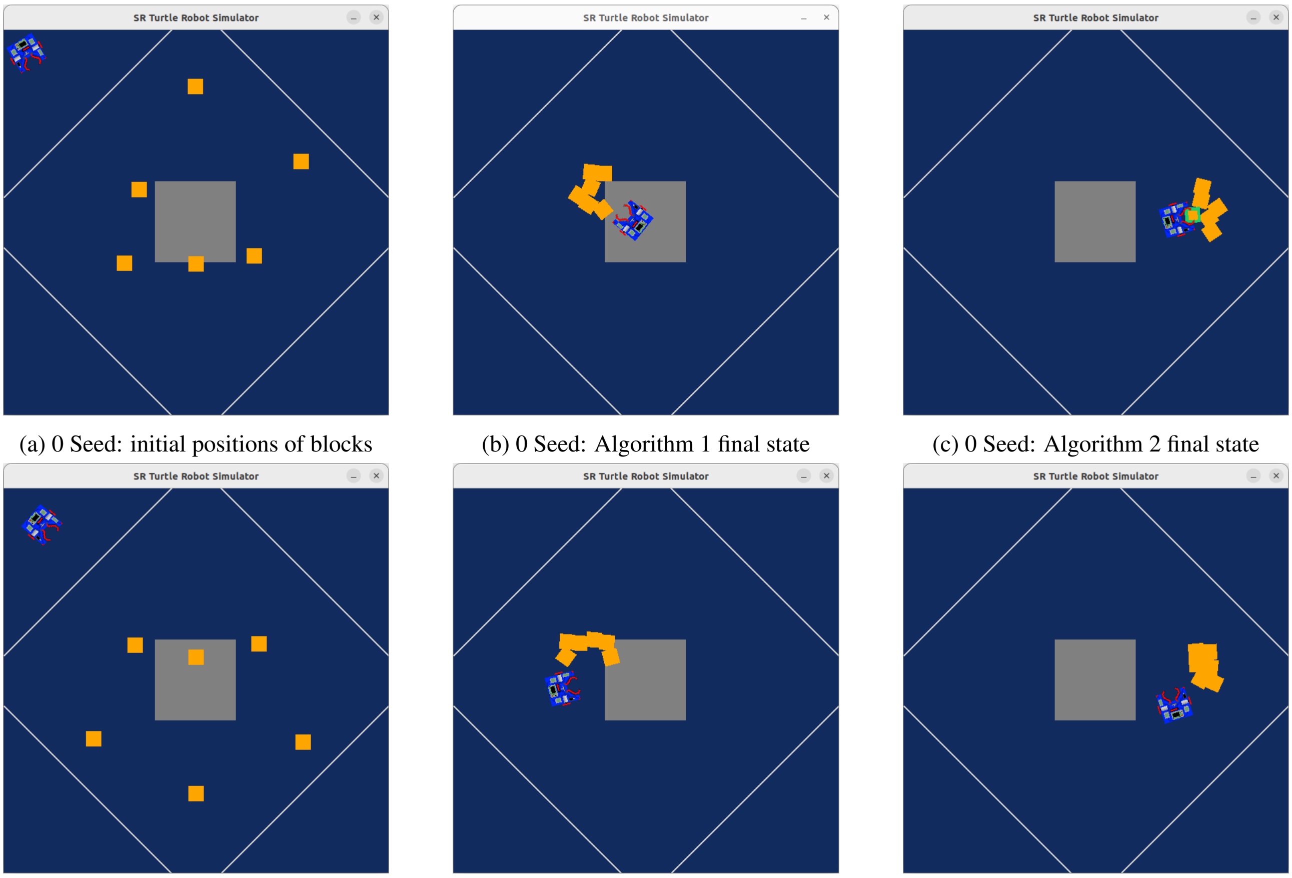
Contents

1 Hypothesis

In this report two robot motion algorithms are tested and analysed based on their performance of gathering blocks in a simulated environment. The algorithm created by Mia Jane La Rocca (student ID s6344889) will be tested in terms of speed and success/failure against the algorithm created by Iris Laanearu (student ID s6350192). These algorithms will be referred to as Algorithm 1 and Algorithm 2 respectively.
The null hypothesis is that there is no significant difference between the speed of the algorithms to a 95% level of confidence. In case the null hypothesis can be rejected the alternative hypothesis is that there is a significant difference between the speed of the algorithms with a 95% level of confidence.

2 Experimental Setup

A test environment was developed such that multiple runs of multiple algorithms could be timed. The simulation environment was modified from the original `two_colors_assignment_arena.py` such that the radius of each of the six blocks is scaled by a random number between 0 and 1. This allows for some variation in the block position while also ensuring there are minimal conflicts between blocks. The seed is selected for each set of runs so each algorithm will run on the same environment. After each set of runs, the seed is increased by 100. The block distributions as well as final states for the first two runs of each algorithm are shown in Figure 1. The runs are timed using the python `time.time()` function which tracks wall clock time. If an algorithm is running for more than 300 seconds, it is stopped and that run is counted as a failure. This is designed to avoid cases where the robot is stuck indefinitely, which can happen if a block gets in between the robot and the block it is trying to grasp. For our experiment we chose to do 50 runs of each algorithm in the hope that there would be enough failures to do a chi squared test. The runs were all done on the same laptop running Ubuntu 22.04 with no other processes running in the background.



(a) 0 Seed: initial positions of blocks

(b) 0 Seed: Algorithm 1 final state

(c) 0 Seed: Algorithm 2 final state

(d) 100 Seed: initial positions of blocks

(e) 100 Seed: Algorithm 1 final state

(f) 100 Seed: Algorithm 2 final state

Figure 1: Initial and final states of the first two runs

3 Results

Results of the experiment are in Figure 2. The time it took to complete the simulation is given in seconds and the timed out attempts are marked as FAILED.

Run	Algorithm 1	Algorithm 2
1	220.19	149.83
2	227.20	180.00
3	203.23	120.35
4	209.78	136.78
5	224.52	194.53
6	232.33	125.82
7	238.24	143.32
8	213.78	179.48
9	230.78	167.89
10	247.52	245.70
11	220.50	116.26
12	219.36	200.03
13	228.33	189.48
14	212.62	176.46
15	240.97	134.82
16	251.47	102.23
17	240.15	156.85
18	240.85	107.26
19	219.52	125.32
20	240.85	196.57
21	267.33	114.26
22	219.31	198.52
23	223.04	187.98
24	248.00	99.74
25	229.99	108.30
26	203.16	133.84
27	223.86	211.53
28	209.88	202.56
29	232.77	192.48
30	237.83	121.27
31	227.65	111.78
32	234.33	216.60
33	214.08	251.15
34	232.53	219.08
35	241.16	FAILED
36	233.57	136.30
37	254.80	136.27
38	260.40	166.46
39	239.48	FAILED
40	204.88	183.48
41	235.89	122.31
42	231.54	FAILED
43	223.48	124.28
44	238.37	141.32
45	238.69	FAILED
46	220.14	184.00
47	288.99	139.87
48	250.25	98.78
49	227.94	175.38
50	242.64	140.34

Figure 2: Results of the completion speed of algorithms

3.1 Observations

It can be seen that the Algorithm 1 did not fail any of the attempts and the Algorithm 2 failed 4 times out of 50. Furthermore, it can be noticed that the Algorithm 1 run times are more higher than of Algorithm 2. None of the run times for Algorithm 1 are below 200 seconds while for Algorithm 2 there are few attempts where the algorithm completion is below 100 seconds.

4 Statistical analysis

The paired T-test is carried out to analyse relevant results. Algorithm 1 didn't fail any attempts, but Algorithm 2 failed 4 times. The failed attempts are disregarded to have relevant statistical analysis. Therefore, the sample size is 46. Because there were fewer than ten failures for both algorithms, it is not possible to do a chi-squared test. Therefore, we decided to do a two-sided paired T-test to determine whether there was a significant difference in speed between the two algorithms.

4.1 Paired T-test

A paired T-test can be used to compare two population means where the two samples have observations that can be paired. In our experiment, each of the algorithms were run on the same set of fifty initial block conditions, so we can use the paired T-test rather than a regular T-test. The purpose of the paired T-test is to determine whether there is a statistically significant difference in the means of the paired observations.

4.2 Analysis of the algorithms

First, to test the null hypothesis, we calculate the difference between the two observations. The histograms of the times for Algorithm 1, Algorithm 2 and the difference in time between each paired run are seen in Figure 3.

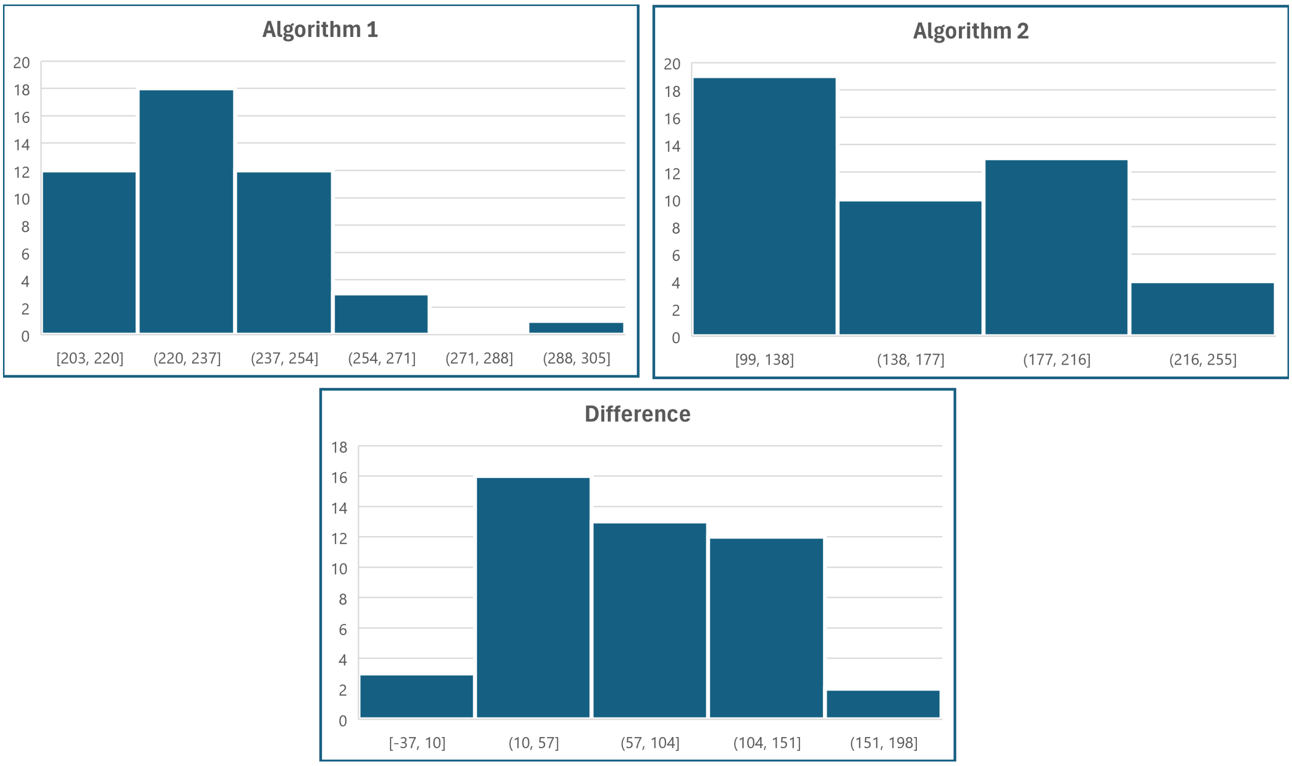


Figure 3: Data graphs of the speed of algorithms and their difference

Then we take the mean $\bar{d} = 73.49$ and the standard deviation $s_d = 47.75$ of the differences. The full table can be seen in Figure 4.

	Algorithm 1	Algorithm 2	Difference
Mean	231.46	157.98	73.49
Standard deviation	17.14	40.17	47.75

Figure 4: Mean $\bar{d}$ and standard deviation s_d

Then we calculate the standard error using the equation 1, where the number of generated scenarios are $n = 46$.

$$SE(\bar{d}) = \frac{s_d}{\sqrt{n}} \tag{1}$$

Which gives the standard error $SE(\bar{d}) = 7.04$.
In order to calculate the t-value, the following equation is used:

$$t = \frac{\bar{d}}{SE(\bar{d})} \tag{2}$$

This gives the t-value as $t = 10.44$. Now the result is compared with 95% of confidence to the values of the two tailed in the t-table shown in Figure 5. The degrees of freedom is around 40 because $n - 1 = 46 - 1 = 45$.

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	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 5: T-values table
Source: https://www.sjsu.edu/faculty/gerstman/StatPrimer/t-table.pdf

It can be seen that the t-value with 95% confidence and DoF of 45 is $t = 2.02$. This is far from our calculated value $t = 10.44$, which means that we can reject with 95% confidence the previously stated null hypothesis: there is no significant difference of speed between the algorithms. Therefore, we can say with 5% of error probability that there is a significant difference of speed between two algorithms.

5 Conclusion

Given our set of test initial conditions and a time out of 300 seconds, Algorithm 1 did not fail any of the attempts, and Algorithm 2 failed 4 times. Therefore, it became clear that there was not enough data on failed attempts to do a chi-squared test and the sample size 50 was reduced to 46. Paired T-test was performed to statistically analyse the results. The outcome of the paired T-test is $t = 10.44$, which is higher than the two tailed t-value of 95% confidence, and 40 DoF in the table. As a result, the originally set null hypothesis ”There is no significant difference between the speed of the algorithms”, was rejected. The alternative hypothesis ”There is a significant difference between the speed of the algorithms”, was accepted.

5.1 Improvements

A further improvement would be to compare the failure/success of the algorithms. For this case, there should be more experiments to get additional failure cases. Then chi-squared test can be performed to statistically analyse the results.