Research Track 2

Part III. Final Assignment.

Statistical Analysis

Prof: Carmine Recchiuto

UNIVERSITÀ DEGLI STUDI
DI GENOVA

Carlos Ángel López de Rodas Serrano

Index

Introduction

The aim of this document is to perform a statistical analysis on the first assignment of Research Track 1. The idea behind this analysis is to compare two possible implementations, one being the solution proposed by the student[1] named *Solution 1* and one being the solution proposed by the professor[2] named *Solution 2*.

For the first assignment, a python script had to be developed in order to automatically control a mobile robot in some circuit with some obstacles, gold and silver (see *Figure 1*). Golden tokens are to be avoided while silver ones are defined as something the robot has to grab and put it on its back.

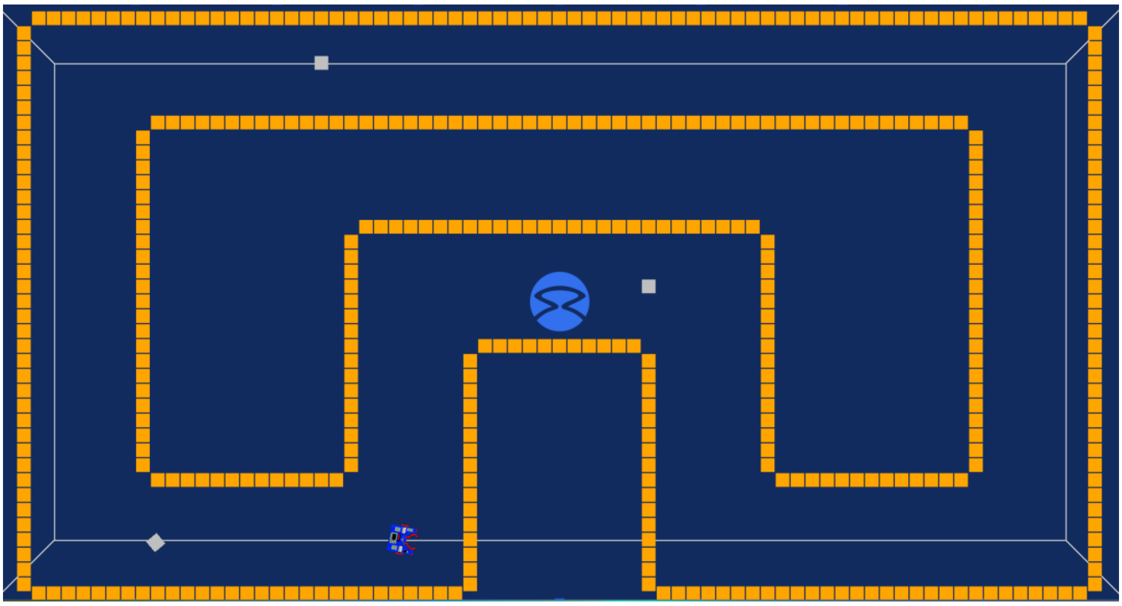

Figure 1. RT1 circuit.

Statistical Analysis

The statistical analysis is going to be performed on some random basis. That is, the two different approaches will be tested when silver tokens are randomly placed over the circuit. The number of silver tokens spawned is 3 and they can only do it on spots of the circuit which are reachable by the car.

To compare the performance of the two algorithms, a lot of evaluators can be used. In this case, the evaluator considered are the successfulness of the goal and the average time required to finish the circuit. The code will be run 50 times per each approach, so 100 in total. A run is considered successful if there are no crashes, that is, no running into golden tokens and no unexpected behaviors like not grabbing a silver token, going in another direction… Time will be measure for each run.

[1] https://github.com/crackls/RT_1_Assignment_1.git
[2] https://github.com/CarmineD8/python_simulator.git

Successfulness of the goal

After the simulation, when the *Solution 1* is run, the robot is able to successfully complete the circuit 45/50 times, while the *Solution 2* only achieves the goal in 41/50 (see *Table 1*).

Results		
	Success	Failure
Solution 1	45	5
Solution 2	41	9
Total	86	14

Table 1. Results of the successfulness of the goal.

It seems like *Solution 1* has a better performance than *Solution* 2, but in order to actually conclude whether this is true, or these results are just due to chance, a chi-squared test is performed. The null hypothesis would be that both solutions work in the same way, so it is expected for both approaches to have the same number of successful runs. That is, that out of 50 simulations, 86/2 = 43 would success.

Null Hypothesis		
	Success	Failure
Solution 1	43	7
Solution 2	43	7
Total	86	14

Table 2. Results when null hypothesis.

Now, the value of χ^2 can be computed as following:

$$\chi^2 = 2 \cdot \left(\frac{2^2}{43} + \frac{2^2}{7} \right)$$

which results in $\chi^2 = 1.3289$.

Knowing that the Dof of the *Table 2* is $Dof = (n_c - 1) \cdot (n_r - 1) = (2 - 1) \cdot (2 - 1) = 1$ and looking at the table of *Appendant 1*, it can be seen that the probability of failing when saying that the null hypothesis can be rejected is close to 0.2 (actually it is between 0.2 and 0.975 but much closer to 0.2). This value is quite large, so the null hypothesis cannot be rejected.

Average time of completing the circuit

The second evaluator to be considered in the analysis is the average time to complete the circuit. To check this, the code has been run 50 times for each solution and the time to finish the circuit has been registered. This data can be approximated into two normal distributions $\mu \sim N(\bar{x}, s)$ to be compared through a T-test.

Normal dist.	Mean	Std. Dev.
Solution 1	163.2	8.044874
Solution 2	160.72	17.86733

Table 3. Normal distribution over average time.

Note that the time is measured in seconds.

$$\mu_1 \sim N(163.2, 8.044) \qquad \mu_2 \sim N(160.72, 17.87)$$

Hence, the null hypothesis is that both distributions are the same, that is $H_0: \mu_1 = \mu_2$, which is the same as $H_0: \mu_1 - \mu_2 = 0$. The pooled variance is going to be computed to prove whether the hypothesis can be rejected or not, that is $H_1: \mu_1 - \mu_2 \neq 0$.

$$\hat{\sigma}^2_{pooled} = \frac{(N_1 - 1) \cdot s_1^2 + (N_2 - 1) \cdot s_2^2}{N_1 + N_2 - 2}$$

And knowing that $N_1 = N_2 = 50$, the pooled variation is $\hat{\sigma}^2_{pooled} = 191.9807395$.

This value is used to find the estimated standard deviation of the distribution of the difference of the of the means ($H_0: \mu_1 - \mu_2 = 0$):

$$\hat{\sigma}_{\bar{x}_1 - \bar{x}_2} = \sqrt{\hat{\sigma}^2_{pooled}\left(\frac{1}{N_1} + \frac{1}{N_2}\right)}$$

Which results in $\hat{\sigma}_{\bar{x}_1 - \bar{x}_2} = 38.3961479$.

All the conditions to perform a T-test to compare both distributions are met so the t-statistic can be computed:

$$t_{\bar{x}_1 - \bar{x}_2} = \frac{\bar{x}_1 - \bar{x}_2}{\hat{\sigma}_{\bar{x}_1 - \bar{x}_2}}$$

That leads to a value of $t_{\bar{x}_1 - \bar{x}_2} = 0.06459$.

Knowing that the Dof of the experiment is $Dof = N_1 + N_2 - 2 = 98$ and looking at the values of the table of *Appendant 2*, with a level of confidence of 95%, $t > 1.66$, which is not the case for this analysis. Hence, the null hypothesis cannot be rejected.

Conclusion

Having the analysis done, in both cases the probability of failing in rejecting the null hypothesis is quite large, so there is not enough confidence to say that the performance of *Solution 1* is better than *Solution 2* or vice versa. This means that the conclusion is that both algorithms have similar performance when completing the circuit.

Appendants

Appendant 1. Chi-squared test table.

	P										
DF	0.995	0.975	0.2	0.1	0.05	0.025	0.02	0.01	0.005	0.002	0.001
1	.0004	.00016	1.642	2.706	3.841	5.024	5.412	6.635	7.879	9.55	10.828
2	0.01	0.0506	3.219	4.605	5.991	7.378	7.824	9.21	10.597	12.429	13.816
3	0.0717	0.216	4.642	6.251	7.815	9.348	9.837	11.345	12.838	14.796	16.266
4	0.207	0.484	5.989	7.779	9.488	11.143	11.668	13.277	14.86	16.924	18.467
5	0.412	0.831	7.289	9.236	11.07	12.833	13.388	15.086	16.75	18.907	20.515
6	0.676	1.237	8.558	10.645	12.592	14.449	15.033	16.812	18.548	20.791	22.458
7	0.989	1.69	9.803	12.017	14.067	16.013	16.622	18.475	20.278	22.601	24.322
8	1.344	2.18	11.03	13.362	15.507	17.535	18.168	20.09	21.955	24.352	26.124
9	1.735	2.7	12.242	14.684	16.919	19.023	19.679	21.666	23.589	26.056	27.877
10	2.156	3.247	13.442	15.987	18.307	20.483	21.161	23.209	25.188	27.722	29.588
11	2.603	3.816	14.631	17.275	19.675	21.92	22.618	24.725	26.757	29.354	31.264
12	3.074	4.404	15.812	18.549	21.026	23.337	24.054	26.217	28.3	30.957	32.909
13	3.565	5.009	16.985	19.812	22.362	24.736	25.472	27.688	29.819	32.535	34.528
14	4.075	5.629	18.151	21.064	23.685	26.119	26.873	29.141	31.319	34.091	36.123
15	4.601	6.262	19.311	22.307	24.996	27.488	28.259	30.578	32.801	35.628	37.697
16	5.142	6.908	20.465	23.542	26.296	28.845	29.633	32	34.267	37.146	39.252
17	5.697	7.564	21.615	24.769	27.587	30.191	30.995	33.409	35.718	38.648	40.79
18	6.265	8.231	22.76	25.989	28.869	31.526	32.346	34.805	37.156	40.136	42.312
19	6.844	8.907	23.9	27.204	30.144	32.852	33.687	36.191	38.582	41.61	43.82
20	7.434	9.591	25.038	28.412	31.41	34.17	35.02	37.566	39.997	43.072	45.315

Appendant 2. T-test table.

25	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725
26	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.690
28	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.659
30	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646
40	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551
50	0.679	0.849	1.047	1.299	1.676	2.009	2.403	2.678	2.937	3.261	3.496
60	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	2.915	3.232	3.460
80	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	2.887	3.195	3.416
100	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	2.871	3.174	3.390
120	0.677	0.845	1.041	1.289	1.658	1.980	2.358	2.617	2.860	3.160	3.373
∞	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.291
One-sided	**75%**	**80%**	**85%**	**90%**	**95%**	**97.5%**	**99%**	**99.5%**	**99.75%**	**99.9%**	**99.95%**
Two-sided	**50%**	**60%**	**70%**	**80%**	**90%**	**95%**	**98%**	**99%**	**99.5%**	**99.8%**	**99.9%**