Università di Genova

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

RESEARCH TRACK 2

First Assignment Statistical Analysis on the First Assignment (RT1)

Author : Roumaissa BENKREDDA

Professor: Carmine Tommaso Recchiuto

Student ID: s5434673

June 2nd,2023

Contents :

- # 1. Introduction :

The goal of this assignment is to make a statistical evaluation. In particular the goal is to compare two different algorithms developed for the first Assignment of Research Track 1. The two algorithms taken into account are:

1. Robot 1: https://github.com/benkredda/assignment.py

2. Robot 2: https://github.com/AmirRad1998/Research-Track-1---Amir-Rad

In the following paragraphs the steps made in order to make a statistical evaluation are shown. In particular an Hypothesis is made and to prove it, with a certain significance level, a test is designed. In order to find the data to perform the test some experiment are done.

- # 2. Hypotheses Formulation :

The Hypothesis made is that: ”the algorithm developed by me (1) has a higher rate of success than the one developed by my colleague (2) when the position and number of tokens in the map change” The rate of success is the number of silver-golden tokens paired over the number of tokens in the environment. A couple of silver-gold tokens is considered paired if during its execution the robot takes the silver token and releases it to the gold, once a couple is paired what happens to it doesn’t matter,and for the sake of experiments will be considered paired in the end. This is lead to the following definition of null and alternative Hypothesis:

• H0 : $\mu 1 = \mu 2$ is the null Hypothesis;

• H1 : $\mu 1 > \mu 2$ is the alternative Hypothesis.

This means that the test that should be used is a one-tailed test, in particular a right-tailed test because the goal is to prove that the rate of success of algorithm 1 is greater than the rate of success of algorithm 2. To prove H1 the null Hypothesis H0 must be rejected with a significant level of 5%. An alternative could be to use a 1% level of significance but, since the values of rate of success go between [0,1] and it’s not expect to vary too much, it’s better to choose 5%. The distribution we aim to use is a t-distribution and in order to apply the central limit theorem, so that the shape of the sampling distribution is more similar that the one of a t-distribution, we need a number of sampling greater than 30. This must be kept in mind designing the number of experiments to make. Lastly a paired T-test is chosen to test the hypothesis since we are comparing two different approaches applied to the same scenario.

- # 3. Experimental Setup :

During the experiments, both algorithms are executed with varying arrangements of tokens in the environment and different numbers of tokens. For each execution, the success rate is calculated as described earlier, by dividing the number of paired tokens

by the total number of tokens. Now, let's examine the different configurations considered in the experiments.

To evaluate how the position of tokens affects the performance of the algorithms, four different maps have been selected. In these four maps, the radius of the inner circle, where the silver tokens are placed, is gradually increased until both tokens are almost on the same circle. Figure 1 provides a visual representation of the various token positions considered in the experiment.

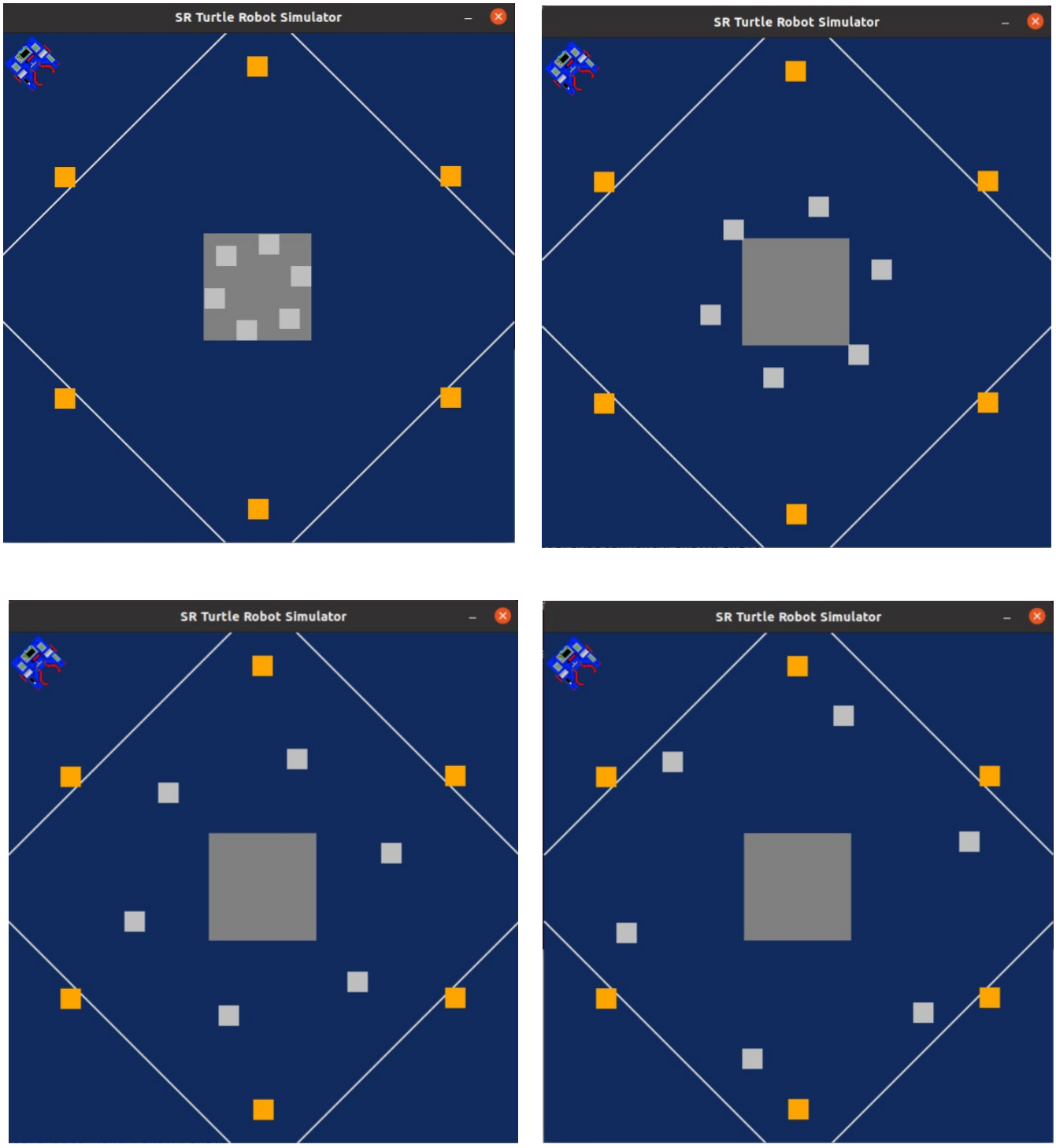

Figure 1: Position of tokens in the environment

In order to examine how the success rate varies based on the number of tokens on the map, the four different environments mentioned earlier were tested with three different numbers of tokens: the original number, two additional tokens, and two fewer tokens. The experiment was conducted with 6 tokens, 4 tokens, and 8 tokens. To ensure a more reliable measurement of the success rate for each configuration (position and number of tokens), the algorithms were executed three times. Overall, a total of 36 states were evaluated (4 environments x 3 numbers of tokens x 3 repetitions), surpassing the minimum requirement of 30. This allows us to apply the central limit theorem, indicating that the sample distribution follows a t-distribution. The outcomes of the experiments are presented in Table 1.

Number of tokens	position	Robot 1	Robot 2	Difference
		Rate success	Rate success	X1- X2
6	original map	1 1 1	1 1 1	0 0 0
	inner 1.2	1 1 1	1 1 1	0 0 0
	inner 2	1 1 1	0 0 0	1 1 1
	inner = outer	0,5 1 0.5	0 0 0	0,5 1 0.5
4	original map	1 1 1	1 1 1	0 0 0
	inner 1.2	1 1 1	1 1 1	0 0 0
	inner 2	1 1 1	1 1 1	0 0 0
	inner = outer	1 1 1	1 1 1	0 0 0
8	original map	0,75 0,75 0,75	0 0,625 0	0.75 0.125 0.75
	inner 1.2	0,75 0,75 0,75	0,875 1 1	-0.125 -0.25 -0.25
	inner 2	0,25 0,25 0,25	1 1 0,875	-0.75 -0.75 -0.625
	inner = outer	0,625 0,75 0,75	0,125 0,125 0,125	0.5 0.625 0.625

Table 1: Experiment Result

- # 4. Paired T-test :

A paired t-test is conducted to test the hypothesis using the data obtained from the experiments . The objective is to prove H1: $\mu 1 > \mu 2$ and reject the null hypothesis H0: $\mu 1 = \mu 2$ with a significance level of 5%.

To accomplish this, the following values are calculated:
• the difference d = x1 − x2 between the two observation of each pair, last column of Table 1;
• the mean of the difference it $\bar{d} = 0, 154$

• the standard deviation of the difference sd = 0, 451
• the standard error of the difference SE(d) = sd/ √N = 0, 075
• the t-value t = d^- /SE(d) = 2,05

The t-value is then compared with the value from the t-table, considering 35 degrees of freedom and a 5% significance level, which is 1.697.

Note :

However, since there is no specific entry for 35 degrees of freedom in a standard t-table, we can make an approximation by using the closest available value. In this case, we can use the critical value for 30 degrees of freedom or 40 degrees of freedom, depending on which one is closer.

Let's consider the critical value for 30 degrees of freedom:

Looking at the t-table for a one-tailed test with a 5% significance level and 30 degrees of freedom, the critical value is approximately 1.697

Since the computed t-value is smaller than the value from the t-table, the null hypothesis H0 cannot be rejected with a 95% confidence level.

In particular, the p-value associated with the t-value is 0.0455, indicating that the null hypothesis can be rejected with an 85% confidence level. This corresponds to a 15% level of significance, which is three times the requested significance level.

• 5. Conclusion :

In conclusion, based on the conducted experiments, the hypothesis has not been proven. To ensure that the null hypothesis cannot be rejected, increasing the sample size and conducting more experiments would enhance the power of the test in detecting false null hypotheses. As the sample size increases, the mean and standard error become more stable, leading to more accurate t-test results.

There are several ways to extend the experiment:
- Designing another map with different token positions, resulting in a sample size of 54 (6 environments x 3 numbers of tokens x 3 repetitions).
- Taking additional measurements for each map configuration, which would increase the sample size to 60 (5 environments x 4 numbers of tokens x 3 repetitions).
- Considering different numbers of tokens on the map, such as adding or removing one token (5 tokens and 7 tokens), resulting in a sample size of 75 (4 environments x 5 repetitions x 2 numbers of tokens).

To determine if it is worth taking additional samples or if the results obtained from the paired t-test remain the same, an optimistic approach is adopted. In this approach, all new measurements favor Robot 1, meaning that the highest success rate is attributed to Robot 1 and the lowest to Robot 2. Under this optimistic assumption and with an ideal dataset of 60 samples, the resulting t-value is calculated to be 1.73, which is greater than the value in the t-table. The corresponding p-value is found to be 0.044, suggesting that the null hypothesis could be rejected with a significance level of 5%.

However, it is important to note that this optimistic scenario is highly unlikely to occur in practice. Even if only a few experiments favor Robot 2, the results would be similar to those obtained previously, and the null hypothesis still cannot be rejected with a 95% confidence level.