UNIVERSITA DEGLI STUDI DI GENOVA

RESEARCH TRACK II

Third Assignment Report

Fundaments of Statistics

DIBRIS

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

Author

Mura Alessio

May 2023

Contents

1 Introduction

As far as the third assignment of Research Track II is concerned, we were asked to conduct a statistical analysis based on the first assignment of Research Track I. In order to achieve this analysis, we had to compare our implementation with that of another colleague. In my case, I collaborated with Pisano Davide (S4363394). The aim of the first assignment was to make a simulator for robotics. The simulator was able to simulate a robot in a 2D environment and made it move around it and interact with objects. The robot was able to sense the environment and use this information to make decisions. As far as the final result was concerned, the robot had to try to place silver tokens next to golden ones, in order to distribute them in pairs.

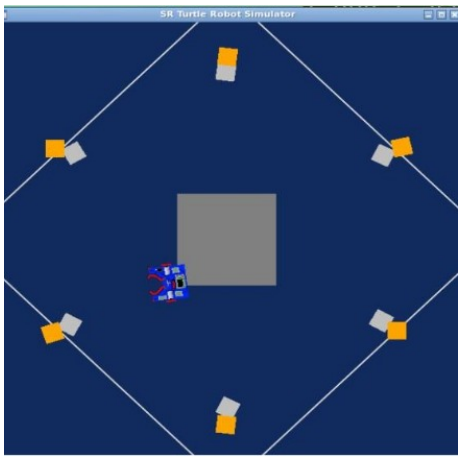

Figure 1: The arena and its setup.

The conducted experiments entailed modifying the visual aspects of the arena configuration in order to examine the effectiveness of two distinct robotic controllers. Through a comparative analysis of the results obtained from my implementation and that of my colleague, we can acquire valuable insights into the efficacy of diverse approaches when tackling the task. Utilizing statistical analysis, we can detect any notable disparities in performance between the two implementations and draw conclusive observations concerning the respective strengths and weaknesses of each approach.

2 Experiment Overview

The primary objective of the experiment was to conduct a statistical analysis focused on measuring the average time required to complete the silver-gold token pairing task. To achieve this, various simulations were performed, incorporating deliberate variations in the radius of the circle containing the golden tokens to ensure diverse scenarios and paths for the robot. Modifications were made to the "two-colours-assignment-arena" file to introduce randomness and variability in the experiments. By adjusting the random seed parameter, multiple iterations were generated, resulting in different positions for the golden tokens in each simulation. This approach allowed for randomized token configurations, adding complexity to the task. A total of 30 simulations were conducted for each token assignment to ensure a robust analysis. Each simulation utilized a distinct random seed parameter, resulting in unique token placements and environmental setups.

By conducting a sufficient number of simulations, the reliability of the findings was increased, minimizing the impact of random fluctuations on the overall results. The collected data from these simulations provided a comprehensive dataset for statistical analysis. The average time required for completion was calculated for each set of simulations with varying token assignments, enabling a comparison of the performance between the two implementations. Additionally, the variability in the results was examined to determine the statistical significance of any observed differences. By employing multiple simulations with different random seed parameters, the statistical analysis captured the overall performance trends and accounted for the influence of random factors. This approach enhanced the reliability and validity of the analysis, enabling meaningful conclusions to be drawn regarding the relative efficiency of the robotic controllers in different token assignment scenarios.

3 Testing of Hypotheses

As we clearly know, a hypothesis is a statement which makes a prediction about something which is not proven. To initiate the analysis, we have formulated hypotheses that will guide our evaluation of the two implementations. The **Null Hypothesis** assumes that there is no significant difference between the two algorithms, implying that both algorithms are equally effective in completing the task. On the other hand, the **Alternative Hypothesis** challenges the null hypothesis by proposing that one algorithm performs differently from the other. It suggests that one algorithm may demonstrate superiority or inferiority compared to the other.

Starting the analysis, we assume the null hypothesis to be true and the alternative hypothesis to be false, indicating that there is no substantial difference between the two implementations. Our goal is to verify whether the means of completion times for the task are the same for both cases, signifying equal efficiency. To assess the validity of our assumption, we will conduct a T-Test, a statistical method enabling the comparison of means between two datasets to determine if they are significantly different. By performing the T-Test on the completion times of the two implementations, we can infer whether there is a statistically significant difference between them.

Furthermore, this report aims to examine and evaluate the performance of these two distinct programs that solve the same problem. The next step involves formulating the two hypotheses that will be tested and refuted through experiments. The null hypothesis posits that there are no significant differences in performance between the two algorithms. It assumes that the average execution time is substantially similar for both algorithms, indicating no clear preference in terms of speed. On the other hand, the alternative hypothesis suggests that the first algorithm exhibits greater speed compared to the second algorithm.
It proposes that the average execution time of the first algorithm is lower than that of the second algorithm, indicating a significant speed advantage for the first algorithm. This analysis aims to provide an objective and statistical evaluation of the differences between the two algorithms, utilizing a mathematical approach and data collected during the study. By testing and analyzing these hypotheses, we aim to draw meaningful conclusions regarding the performance disparity between the two implementations.

4 Data and Analysis

As stated earlier, the experiments revolve around the average duration necessary to complete the task. For conducting the analysis, we employed Microsoft Excel. Below we can find a table presenting the time required for task completion in both my implementation and my colleague's implementation. Additionally, a comparative plot showcasing the durations in the two scenarios is available.

Angle-Offset	Execution time 1 (Davide)	Execution time 2 (Alessio)
value 1	124.152205944	123.747475147
value 2	159.195235014	116.804877996
value 3	164.201638937	123.395332098
value 4	221.377575874	233.639571905
value 5	162.199706078	123.160826921
value 6	149.176033974	120.351647854
value 7	132.163321972	118.255588055
value 8	161.22991395	120.699299097
value 9	121.148519039	122.128663063
value 10	147.165704012	117.783178091
value 11	179.233612061	116.354319096
value 12	169.184362888	125.264173031
value 13	181.251331806	125.269979
value 14	117.128409147	119.554562092
value 15	174.235455036	121.574173927
value 16	169.245977879	122.968338013
value 17	168.253568172	172.331265926
value 18	143.16601181	127.462956905
value 19	199.246370792	117.754606962
value 20	150.157207012	121.04955101
value 21	114.144917011	124.151800156
value 22	179.222509146	122.35408783
value 23	125.145918131	124.468497038
value 24	197.22224021	121.451179028
value 25	147.151827812	125.957015038
value 26	139.158722878	120.75873518
value 27	134.208396196	125.157833099
value 28	162.207538128	124.248339891
value 29	201.222958088	124.968111992
value 30	164.163652897	119.761034012

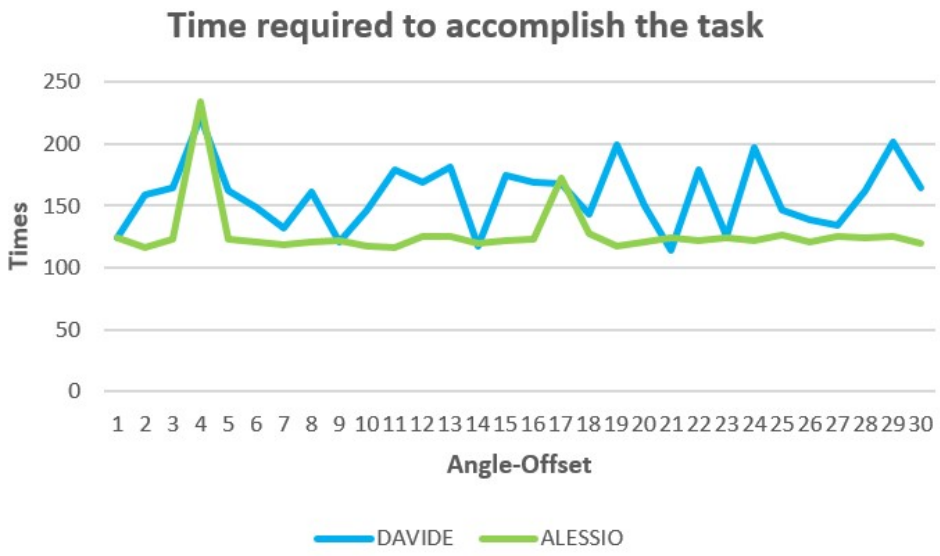

Figure 2: Task completion comparison

After gathering the time data, I proceeded to compute the arithmetic mean (x) by dividing the sum of the observed values by the total number of observations (n). The formula employed in both instances (Davide: 1, Alessio: 2) is as follows:

$$x_1 = \frac{\sum_{n=1}^{N_1} X_i}{N_1} = 158{,}5653613965 \qquad x_2 = \frac{\sum_{n=1}^{N_1} X_i}{N_2} = 127{,}4275673151$$

Subsequently, the standard deviation was calculated to provide an indication of the proximity of the entire dataset to the mean value. The formula utilized in both scenarios is as follows:

$$\sigma_1 = \sqrt{\frac{\sum_{n=1}^{N_1} (X_i - \mu_1)^2}{N_1}} = 26{,}6915511 \qquad \sigma_2 = \sqrt{\frac{\sum_{n=1}^{N_1} (X_i - \mu_2)^2}{N_2}} = 22{,}24373644$$

where N1, N2 = 30 and Xi are the values of the times.
We will now proceed to conduct a statistical analysis, specifically utilizing the two-sample T-test, in order to ascertain the statistical significance of the observed difference. Performing this statistical analysis is crucial in obtaining a more precise conclusion regarding the equality or disparity between the two implementations. Furthermore, I have computed the pooled variance, represented by the following value:

$$\sigma^2_{pooled} = \frac{(N_1 - 1) \cdot s_1 + (N_2 - 1) \cdot s_2}{N_1 + N_2 - 2} = 645{,}2397249$$

The computation of the pooled variance contributes to the determination of the pooled and estimated Standard Error of the sampling distribution of the difference of means. This estimation can be obtained using the following formula:

$$\sigma_{\bar{x_1} - \bar{x_2}} = \sqrt{\frac{\sigma^2_{pooled}}{N_1} + \frac{\sigma^2_{pooled}}{N_2}} = 6{,}558657001$$

Given our focus on examining the differences, the t-statistic transforms into the following form:

$$t^2_{\bar{x_1} - \bar{x_2}} = \frac{x_1 - x_2}{\sigma_{x_1 - x_2}} = 4{,}747586903$$

5 Final Inference

Given the knowledge that there are 30 degrees of freedom, we can determine the critical value of the t-score and ascertain the statistical significance of the observed difference. With a sufficiently large sample size ($N = 30$) that allows for a safe t-test and an approximate normal distribution of the data, we can employ a two-tailed t-test to evaluate statistical significance.

Considering the 30 degrees of freedom, the critical value of the t-score for a significance level of 5% ($\alpha = 0.05$) is approximately ± 2.042 (obtained from the T-test table).

In your specific case, the calculated t-score is 4.747586903. Since the absolute value of the t-score surpasses 2.042, we can conclude that the observed difference between the two implementations is statistically significant. Based on the data provided in the report, we can reject the null hypothesis and favor the alternative hypothesis, which suggests that the first algorithm is significantly faster than the second algorithm. Based on the results depicted in the aforementioned graph, it is evident that the algorithm implemented by Alessio is significantly more performant and efficient than that of his colleague.