Lugano Giacomo mat.5400573

UNIVERSITÀ DEGLI STUDI DI GENOVA

Robotics Engineering

AY. 2022/2023

Research Track 2

first assignment's report:

Statistical analysis of two different algorithm

0. Introduction:

In this report a brief statistical analysis will be developed to prove the thesis hereafter reported, to compare the performances of the two algorithms coded in the scripts assignment_test.py and assignment_control.py respectively called test and control algorithm.

1. Thesis:

First let's introduce the main thesis to be prove:

The assignment_test.py algorithm will hit less token during its motion respect the assignment_control.py one, but on other hand it will be less effective, meaning that in the same time window it will couple less tokens, on a normal random distribution of the tokens in the environment.

These hypotheses can be formalized by making a formal statement of both the null hypothesis and the alternative one for both:

1. The test algorithm will hit less tokens during its motion hypothesis: this thesis can be clearly stated as:

    Null hypothesis H_0^{hit}: $\alpha_{ctrl} = \alpha_{test}$
    Alternative Hypothesis H_a^{hit}: $\alpha_{ctrl} > \alpha_{test}$

    Where:

    - H_0^{hit} is the Null hypothesis, meaning that the test algorithm will hit a comparable number of tokens respect the control one.
    - H_0^{hit} is the Alternative hypothesis, meaning that the test algorithm will hit a smaller number of tokens respect the control one, which is the one that we want to prove.
    - α_{ctrl} is the mean of the number of tokens hit among all the trials by the control algorithm.
    - α_{test} is the mean of the number of tokens hit among all the trials by the test algorithm.
2. The test algorithm will couple less tokens in the same amount of time hypothesis: this thesis can be clearly stated as:

    Null hypothesis H_0^{coup}: $\beta_{ctrl} = \beta_{test}$
    Alternative Hypothesis H_a^{coup}: $\beta_{ctrl} > \beta_{test}$

    Where:

    - H_0^{coup} is the Null hypothesis, meaning that the test algorithm will couple a comparable number of tokens respect the control one.
    - H_0^{coup} is the Alternative hypothesis, meaning that the test algorithm will couple a smaller number of tokens respect the control one, which is the one that we want to prove.
    - β_{ctrl} is the mean of the number of tokens couples created among all the trials by the control algorithm.

- β_{test} is the mean of the number of tokens couples created among all the trials by the test algorithm.

For both the hypothesis the formalization suggests using a one-tailed test, since the thesis consist of a greater property of one of the algorithm respect the other.

Moreover, for both the hypothesis a significance level of 5% has been chosen, since the number of samples will not be big enough for a better level and we expect a small difference between sample means.

As stated in the thesis for the tests a normal distribution of the tokens in the environment has been chosen, further details about the testing phase can be found in the Experiment and Data acquisition part of the report.

2. Implementation:

The two algorithms, assignment_test.py and assignment_control.py, used for the experiment can be found in the relative repositories https://github.com/jek97/Robot-organizer.git and https://github.com/teolima99/first_assignment_Research-track.git .

In order to test the thesis some modification on both the algorithm and the simulation have been done:

1. Algorithms: to allow a fair comparison between the number of token's couple in the same time window in both the algorithms the *driving speed* on the **drive()** methods has been set to 20.
2. Simulation: the arena script used for the testing can be found under the path, from the root folder of the repository:

   ```
   robot-sim/sr/robot/arenas/two_colours_assignment_arena.py
   ```

   In it a new method has been created to randomly place the tokens in the environment moreover, to allow a repetition of the experiment the random distribution seed has been made explicit:

   ```python
   def place_token_random(TOKENS_PER_TYPE, type, seed):
       random.seed(seed)
       for i in range(TOKENS_PER_TYPE):
           token_type = type
           token = token_type(self, i)
           token.location = (random.uniform(-2.5, 2.5),
                             random.uniform(-2.5, 2.5))
           token.heading = random.uniform(-pi, pi)
           self.objects.append(token)
   ```

   Indeed it's possible each time to specify the number of tokens per type, the type of the token and the random seed to use for the distribution; once the seed is set the method will proceed by creating the right number of token specified, of the given type and in a random location inside the environment with a random heading.

The random seed will be specified for each trial.

3. Experiment and data acquisition:

Before explaining the data acquisition process it's essential to give some clarifications about the experiment:

We consider a collision between the robot and the token only if the token is moved, even slightly, from the contact with the robot, if the robot remain in contact with the token during its motion for a given time we still consider this as a single collision; we don't consider a collision the case in which the robot, by moving, hit a token with the grabbed token in its gripper, except if the token then hit the robot body. These choices have been made since we want to verify how much the algorithm is safety oriented for the robot and not also for the payload.

We consider a token couple a couple of two token of different type, golden token and silver token, formed by the robot, this means that any couple already present at the beginning of the simulation is discarded, moreover a couple formed by the robot, even if then disassembled by a collision or to create a new couple, will be counted as a regular couple; in the experiment may happened that the robot couple a silver token by placing it near an already formed couple, this will be considered as a valid couple; a couple formed by the collision of the robot with one or more token but not under its intention is discarded. These choices as been made since the intent is to test the ability of the robot to couple the tokens.

The time window decided for this experiment is 5 minutes.

Done these precisions the experiment will be performed by following the steps:

1. Set the random seed of the golden tokens distribution and the silver tokens one directly from the arena script specified above and annotate their values in the table.
2. launch the simulation from the robot-sim directory with the command:
```
Python run.py assignment_XXXXXXX.py
```
Where the `XXXXXXX` stand for `control` or `test` based on which we want to test.
3. Start the chronometer (manually).
4. During the execution of the simulation annotate any collision and any couple formed by the robot.
5. After 5 minutes stop the simulation by the command ctr+C.

Following this procedure it's possible to obtain the data:

TRIAL n	**ASSIGNMENT_CONTROL.PY** Token collision α_i^{ctrl} [collisions]	Token couple generated β_i^{ctrl} [couples]	**ASSIGNMENT_TEST.PY** Token collision α_i^{test} [collisions]	Token couple generated β_i^{test} [couples]	**RANDOM SILVER SEED**	**RANDOM GOLDEN SEED**
1	2	3	0	2	1	30
2	3	3	0	0	2	29
3	2	3	0	0	3	28
4	6	1	2	3	4	27
5	4	2	0	1	5	26
6	2	1	7	7	6	25

7	4	2	0	0	7	24
8	5	3	0	0	8	23
9	5	3	2	2	9	22
10	6	3	0	0	10	21
11	2	0	2	3	11	20
12	8	2	1	0	12	19
13	4	3	1	4	13	18
14	3	3	2	3	14	17
15	4	0	0	0	15	16
16	2	3	4	5	16	15
17	3	2	0	0	17	14
18	4	3	1	0	18	13
19	5	3	1	0	19	12
20	6	3	3	6	20	11
21	8	3	0	0	21	10
22	3	1	0	1	22	9
23	4	3	0	0	23	8
24	5	2	2	1	24	7
25	2	1	3	3	25	6
26	4	3	1	0	26	5
27	5	3	6	6	27	4
28	2	3	2	6	28	3
29	2	3	0	2	29	2
30	3	3	0	0	30	1

4.0 Evaluations:

Given the fact that only two samples are available and comparable for this analysis a parametric test must be preferred.

However this type of tests differ based on the condition that the population is normal distributed or not, to verify that we can perform the Lilliefors Test:

4.1 Lilliefors Test:

The Lilliefors Test will clarify if the sample belong to a normal distribution, for sake of simplicity the test has been done using MATLAB with the function lilliefortest() already available, the script can be found in the root folder of this project.

Thanks to this test we've obtained that all the data recorded are belong to a normal distribution except the number of collisions of the control algorithm α_i^{ctrl} (from assignment_control.py), for that reason for the number of collisions and the number of couples will be executed a different statistical test:

4.2 Number of collisions statistical analysis:

Since in this case we've a small sample which doesn't come from a normally distributed population for that hypothesis the Wilcoxon-Mann-Whitney test (or U-test) will be performed, following the steps:

1. Ranking the data jointly, as they belong to a single sample in increasing order based on the data value, note how in case of ties the rank assigned to each of the tied observation will be the mean of the ranks which they jointly occupy:

Sample belonging	Value	rank	Sample belonging	value	rank	Sample belonging	value	rank
α_i^{test}	0	0,0714	α_i^{ctrl}	2	0,2143	α_i^{ctrl}	4	0,625
α_i^{test}	0	0,0714	α_i^{ctrl}	2	0,2143	α_i^{ctrl}	4	0,625
α_i^{test}	0	0,0714	α_i^{ctrl}	2	0,2143	α_i^{ctrl}	4	0,625
α_i^{test}	0	0,0714	α_i^{ctrl}	2	0,2143	α_i^{ctrl}	4	0,625
α_i^{test}	0	0,0714	α_i^{ctrl}	2	0,2143	α_i^{ctrl}	4	0,625
α_i^{test}	0	0,0714	α_i^{ctrl}	2	0,2143	α_i^{ctrl}	4	0,625
α_i^{test}	0	0,0714	α_i^{ctrl}	2	0,2143	α_i^{ctrl}	4	0,625
α_i^{test}	0	0,0714	α_i^{test}	2	0,2143	α_i^{test}	4	0,625
α_i^{test}	0	0,0714	α_i^{test}	2	0,2143	α_i^{ctrl}	5	1,2
α_i^{test}	0	0,0714	α_i^{test}	2	0,2143	α_i^{ctrl}	5	1,2
α_i^{test}	0	0,0714	α_i^{test}	2	0,2143	α_i^{ctrl}	5	1,2
α_i^{test}	0	0,0714	α_i^{test}	2	0,2143	α_i^{ctrl}	5	1,2
α_i^{test}	0	0,0714	α_i^{test}	2	0,2143	α_i^{ctrl}	5	1,2
α_i^{test}	0	0,0714	α_i^{ctrl}	3	0,5714	α_i^{ctrl}	6	1,75
α_i^{test}	1	0,4	α_i^{ctrl}	3	0,5714	α_i^{ctrl}	6	1,75
α_i^{test}	1	0,4	α_i^{ctrl}	3	0,5714	α_i^{ctrl}	6	1,75
α_i^{test}	1	0,4	α_i^{ctrl}	3	0,5714	α_i^{test}	6	1,75
α_i^{test}	1	0,4	α_i^{ctrl}	3	0,5714	α_i^{test}	7	8
α_i^{test}	1	0,4	α_i^{test}	3	0,5714	α_i^{ctrl}	8	4,5
α_i^{ctrl}	2	0,2143	α_i^{test}	3	0,5714	α_i^{ctrl}	8	4,5

2. Evaluate the sum of the ranks assigned to the values of the first sample, in our case the test one, under the name R_{test} and of the second one, in our case the control one, under the name R_{ctrl}:
$$R_{test} = 15{,}8032 \qquad R_{ctrl} = 29{,}1964$$

3. U-test: measure the difference between the ranked observations following the formula:
$$U = n_{test} \cdot n_{ctrl} + \frac{n_{test}(n_{ctrl}+1)}{2} - R_{test} = 1319{,}1968$$
Where:
   - n_{test} is the sample size coming from the test algorithm (30).
   - n_{ctrl} is the sample size coming from the control algorithm (30).
   - R_{test} is the sum of ranks assigned to the values of the test sample.

   (note: the same formula can be used by replacing n_{test} with n_{ctrl} and R_{test} with R_{ctrl})

4. We can also evaluate the mean of the ranks of the two samples obtaining:
$$\bar{R}_{test} = 0{,}5268 \qquad \bar{R}_{ctrl} = 0{,}9732$$

These results will be explained in the conclusions.

4.3 Number of couples statistical analysis:

Since we can pair the observations of the two samples, and both of them belong to a normal distribution we can test the alternative hypothesis by the paired T-test following the steps:

1. To test, and then invalidate, the null hypothesis that the true mean difference is zero we first evaluate the difference $d_i = \beta_i^{ctrl} - \beta_i^{test}$ between the two observations on each pair:

Token couple generated β_i^{ctrl} [couples]	Token couple generated β_i^{test} [couples]	Difference d_i [couples]	Token couple generated β_i^{ctrl} [couples]	Token couple generated β_i^{test} [couples]	Difference d_i [couples]
3	2	1	3	5	-2
3	0	3	2	0	2
3	0	3	3	0	3
1	3	-2	3	0	3
2	1	1	3	6	-3
1	7	-6	3	0	3
2	0	2	1	1	0
3	0	3	3	0	3
3	2	1	2	1	1
3	0	3	1	3	-2
0	3	-3	3	0	3
2	0	2	3	6	-3
3	4	-1	3	6	-3
3	3	0	3	2	1
0	0	0	3	0	3

2. Evaluate the difference mean $\bar{d}_i$ with its standard deviation:
$$\bar{d} = \frac{\sum_{i=1}^{i=n} d_i}{n} = \frac{\sum_{i=1}^{i=30} d_i}{30} = 0{,}5\overline{33}$$
$$\sigma_{\bar{d}} = \sqrt{\frac{\sum_{i=1}^{i=n} (d_i - \bar{d})^2}{n}} = \sqrt{\frac{\sum_{i=1}^{i=30} (d_i - \bar{d})^2}{30}} = 2{,}4730$$
3. Evaluate the standard the standard error of the mean difference:
$$SE(\bar{d}) = \frac{\sigma_{\bar{d}}}{\sqrt{n}} = 0{,}4515$$
Where:
   - n is the size of each sample
4. Evaluate the t-statistic T and its degree of freedom:
$$T = \frac{\bar{d}}{SE(\bar{d})} = 1{,}1812 \qquad dof = n - 1 = 29$$
5. Use the t-distribution table (that you may find at the link https://en.wikipedia.org/wiki/Student%27s_t-distribution) to compare the T value

obtained above with the T_{n-1} distribution, obtaining the p-value for the paired t-test, in detail we can set the confidence level at 95% and obtain a T_{n-1} value of $T_{n-1} = 1{,}699$
Which is bigger than the value $T = 1{,}1812$ obtained from our data.
Moreover we can check that to find a value smaller then $T = 1{,}1812$ under the same degrees of freedom $dof = 29$ we obtain a confidence level of 85%.
These results will be explained in the conclusions.

5.0 Conclusions:

Before giving any conclusions it's important to state hoe the behavior of the two algorithms frequently result in a block of the robot, introducing a small bias in the data acquisition, a possibility for avoid this problem would be to take into account this phenomena by normalizing the samples based on the time of activity, or by further develop them to avoid these stuck situations.

5.1 Number of collisions statistical results:

From the comparison of the ranks mean of the two sample we can reject the null-hypothesis H_0^{hit} in favor of the alternative one H_a^{hit}, allowing to say that the test algorithm (assignment_test.py) hit less tokens while performing the coupling task.

(unfortunately I was not able to find a U-test table with a big enough size of the samples to verify this hypothesis with the actual U-test)

5.2 Number of couples statistical results:

By observing the T values obtained it's not possible to reject the null-hypothesis H_0^{coup} with the desired level of confidence of 95%, while it's possible to reject it in favor of the alternative-hypothesis with a different level of confidence equal to 85%.

A further investigation of this thesis could be done to better evaluate the thesis by increasing the sizes of the samples.