Exercise 2: Sergio López Padilla

1. Introduction

The previous exercise proposed to study how the time of fixation and the probability of fixation evolved for different alleles in the face of a process of genetic drift, a simple case where other evolutionary forces such as mutation and selection are not considered. This exercise poses a more complex challenge in which the 3 evolutionary forces studied during the course are considered: mutation, selection and drift. More specifically, this exercise proposes the simulation of the Muller's Ratchet phenomenon, common in small populations. The main objective of this study is, using the scripts of lesson 6, to evaluate how the average fitness varies according to different values in population size (N) and fitness advantage (f_ben).

The process followed to develop this study has been, first of all, to generate a script that iterates on 2 lists with the different values in order to simulate all the combinations of the recommended parameters:

- Days: 25000
- replicas: 100
- N: 10, 30,100,300 and 1000
- f_ben: 0.005,0.01,0.02,0.04,0.08

The result has been a table showing how the average fitness varies according to the combination of the parameters to be studied (N and f_ben). In order to facilitate the study of these results, some graphs have been created that provide representativeness to the results and correlation tests have been carried out between the variables to be studied and the target variable (avg_fitness), in order to check the significance in the relationship between variables that the graphs seem to show.

Finally, some general conclusions of the study have been raised, as well as considerations to be taken into account to improve this methodology in future simulations.

2. Script

The script related to this exercise can be found on the following GitHub link:

https://github.com/shevelp/Comp-Evo

This script has been developed using R, and RStudio as IDE and it is divided in the same parts as the exercise 1, previously delivered:

- Parameters definition
- Modeling
- Data transformation
- Relationship Analysis
- Plotting

3. Results

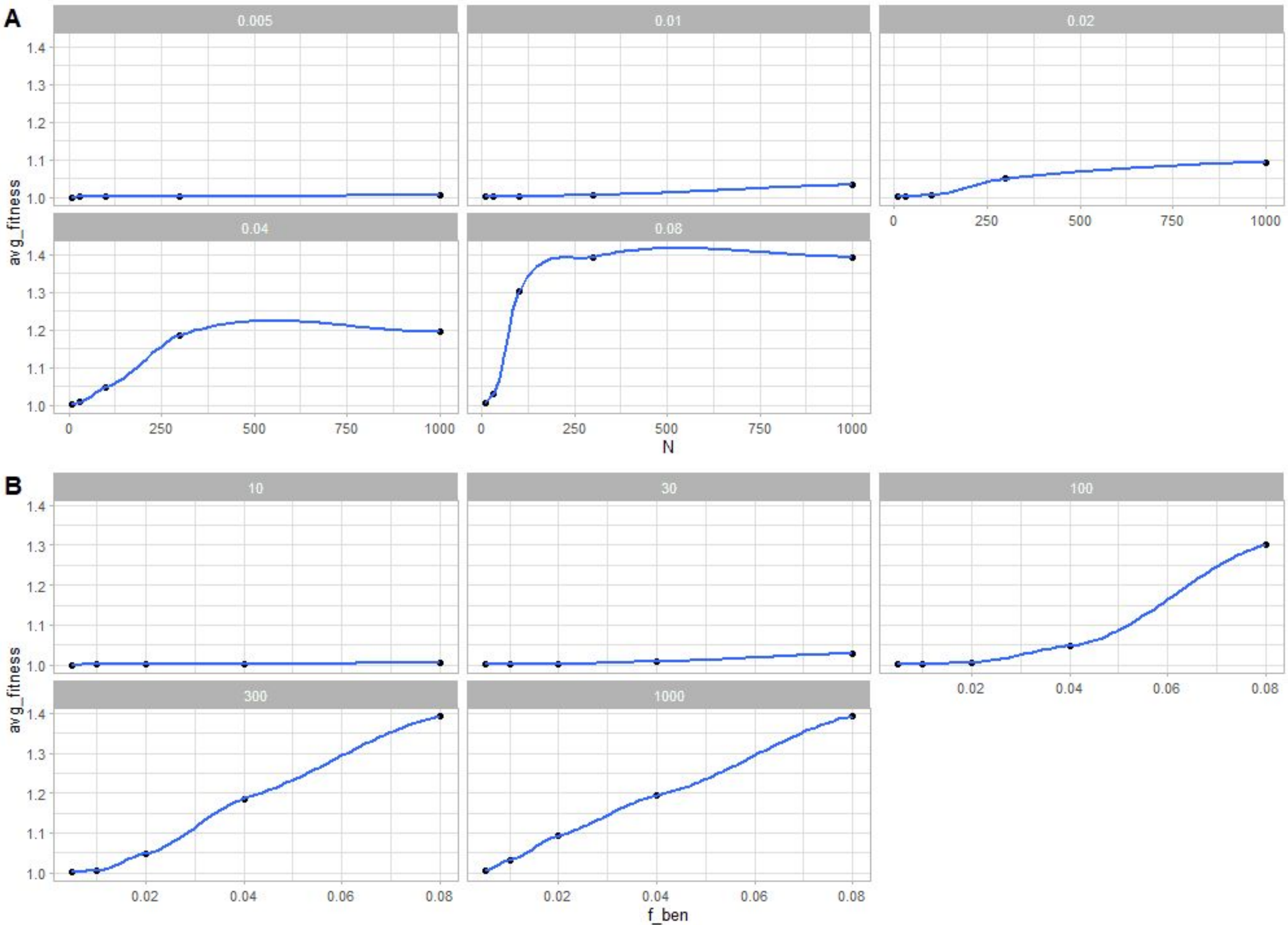
a. Estimated values

The estimated values related to this exercise are represented in the following table (Table 1). It contains the values related to the average fitness (avg_fitness) related to different values in the key parameters: population size (N) and fitness advantage (f_ben).

N	f_ben	avg_fitness
10	0.005	1.000150
10	0.010	1.000600
10	0.020	1.002000
10	0.040	1.002000
10	0.080	1.004800
30	0.005	1.000400
30	0.010	1.000460
30	0.020	1.001453
30	0.040	1.008280
30	0.080	1.028187
100	0.005	1.000321
100	0.010	1.000850
100	0.020	1.006074
100	0.040	1.047188
100	0.080	1.300952
300	0.005	1.000671
300	0.010	1.005844
300	0.020	1.048562
300	0.040	1.185797
300	0.080	1.392461
1000	0.005	1.005462
1000	0.010	1.032222
1000	0.020	1.093127
1000	0.040	1.193931
1000	0.080	1.393231

Table 1. Estimated values.

In order to interpret these results more easily, various line charts have been developed, which allow to evaluate the relationship between the parameters: N - avg_fitness (Figure 1.A) and f_ben - avg_fitness(Figure 1.B).



To make the representation clearer, the charts have been grouped around groups that consider the two variables. A shows how the average fitness evolves with respect to N but considering different values of f_ben, of the same but in reverse, Figure 1. B. Both representations agree that for small values of N and f_ben the avg fitness is very low and hardly varies. On the other hand with intermediate values of f_ben (0.02) and from a population size of 300 the avg fitness increases considerably. Finally, the maximum average fitness is around 1.4. This is reached with population values of 300 and 1000 but is only reached when the f_ben is 0.08. This indicates that from a certain population size the determining value and which has the most influence on the average fitness is f_ben.

b. Relationship analysis

To corroborate the type of relationship existing between the variables, the R correlation test (cor.test) has been carried out. The results of these correlation tests (Figure 2) are in line with the conclusions obtained from the graphical representation of the results. The regulatory parameter f_ben is related with the average fitness because the p-value for this test is 0.00017 while the population size is in the limit of the test and the p-value is above 0.05, concretely is 0.05171.

```
> #-----------relationship -------------------#
> cor.test(x = df$f_ben, y = df$avg_fitness, method = "pearson")

        Pearson's product-moment correlation

data:  df$f_ben and df$avg_fitness
t = 4.4778, df = 23, p-value = 0.0001711
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.3934135 0.8487215
sample estimates:
      cor
0.6824566

> cor.test(x = df$N, y = df$avg_fitness, method = "pearson")

        Pearson's product-moment correlation

data:  df$N and df$avg_fitness
t = 2.0521, df = 23, p-value = 0.05171
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.002055683  0.682444784
sample estimates:
      cor
0.3933948
```

Figure 2. Results obtained from correlation-test between different variables.

4. Considerations

This experiment fulfils its objective and makes it possible to estimate the correlation between the fitness advantage and the average fitness. However, it does not allow us to determine if there is a relationship between population size and average fitness as the Pearson correlation test performed is not significant. In order to study more precisely this latter relationship, possible extensions to the code could include more N values, focusing the study only on this parameter.