

# FINAL PROJECT

## The Moran Model - Neutral Evolution

*Computational Methods*

Prof. Andrea Sánchez-Tapia

Prof. Sara Mortara

Amanda Araujo Silva

2022

## 1 Introduction

This project aims to investigate concepts in evolutionary dynamics, especially biological effects that arise from stochasticity, such as neutral drift. The goal is to study the Moran model and get an intuition of its behavior and biological implications.

### 1.1 Theoretical basis

Evolution dynamics depend on two processes: genetic variability by mutations and changes in the frequency of alleles within the population over time. In a simplistic way, a mutation can be advantageous, disadvantageous, or neutral, in terms of altering the fitness of the individual. The first two's fate is subjected to natural selection, while in the last one alterations in allele frequencies are just due to random processes, called genetic drift. Both cause changes in the frequency of alleles, but while natural selection is directional, genetic drift is driven by stochasticity [1].

The Moran model was proposed in 1958, by the Australian statistician Patrick Moran, and it is the simplest stochastic formulation of finite population evolution. It can describe the probabilistic dynamics of two alleles A and B competing for dominance. The model allows us to ask questions such as how/whether/when a gene is lost in a population. [2][3].

### 1.2 The model

Assumptions of the Moran model:

- ✓ Total size of the population is constant,  $N$ ;
- ✓ Composition is fixed in terms of allele variants A and B, no mutations;
- ✓ Asexual reproduction;
- ✓ Generations overlap.

Frequencies:

- $i \equiv$  absolute frequency of A
- $N - i \equiv$  absolute frequency of B

At each iteration randomly:

- Choose 1 individual to reproduce
- Choose 1 individual to replace

### 1.3 Implementation

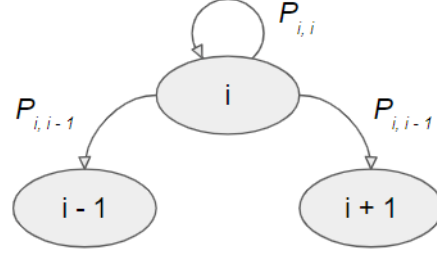
In order to implement the model, the state of the population is tracked by the frequency of allele A, given by  $i$ , the only stochastic variable of the system.

At each iteration, the allele A can gain or lose one unit or stay the same, with the probabilities of changing state given by:

$$P_{i,i+1} = \frac{i}{N} \frac{(N-i)}{N}$$

$$P_{i,i-1} = \frac{(N-i)}{N} \frac{i}{N}$$

$$P_{i,i} = 1 - P_{i,i+1} - P_{i,i-1}$$



In the neutral evolution case, it is assumed that alleles A and B present the same fitness, *i.e.*  $P_{i,i+1} = P_{i,i-1}$ . The decision of which event will take place is made computationally through the use of a uniform number generator.

## 2 Results and discussion

### 2.1 Moran model neutral evolution

The Moran model was implemented in two programming languages, *R* and *Python*.

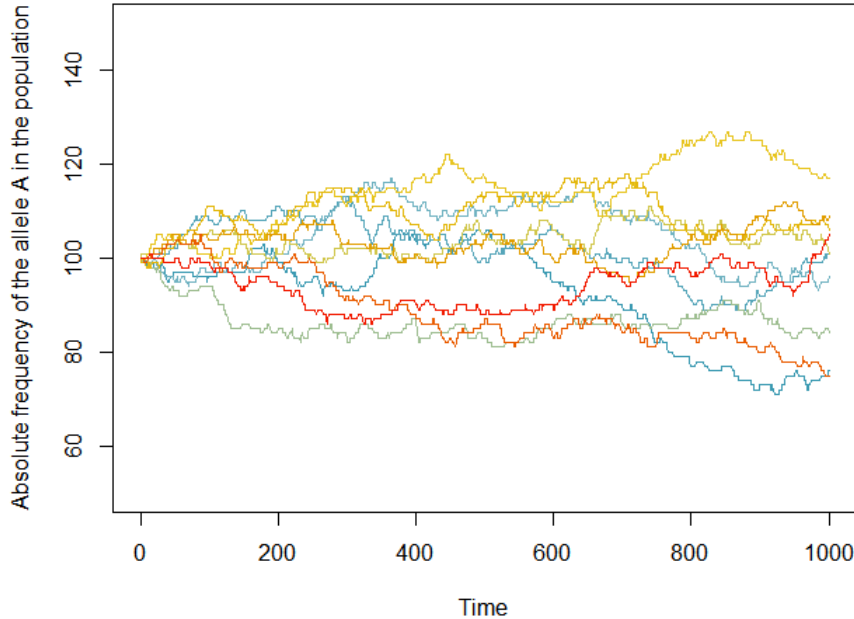


Figure 1: Moran process of neutral evolution for a population size  $N = 1000$ . Initial value of the number of allele A in the polutation  $i = 100$ . 1000 iterations. 10 independent *R* simulations, each one represent by a different line color.

In Fig. (1) we can see the dynamics of the alleles' proportion within a fixed population size. The system behaves as a “random walk”: starting with the same  $i$  value leads to different dynamics of allele frequencies within the population. Due to neutral evolution, the chance of allele A or B being chosen at each iteration is the same, reflected in no observation of a preferential direction in the 10 independent runs.

## 2.2 Size matters: exploring different population sizes $N$ 's

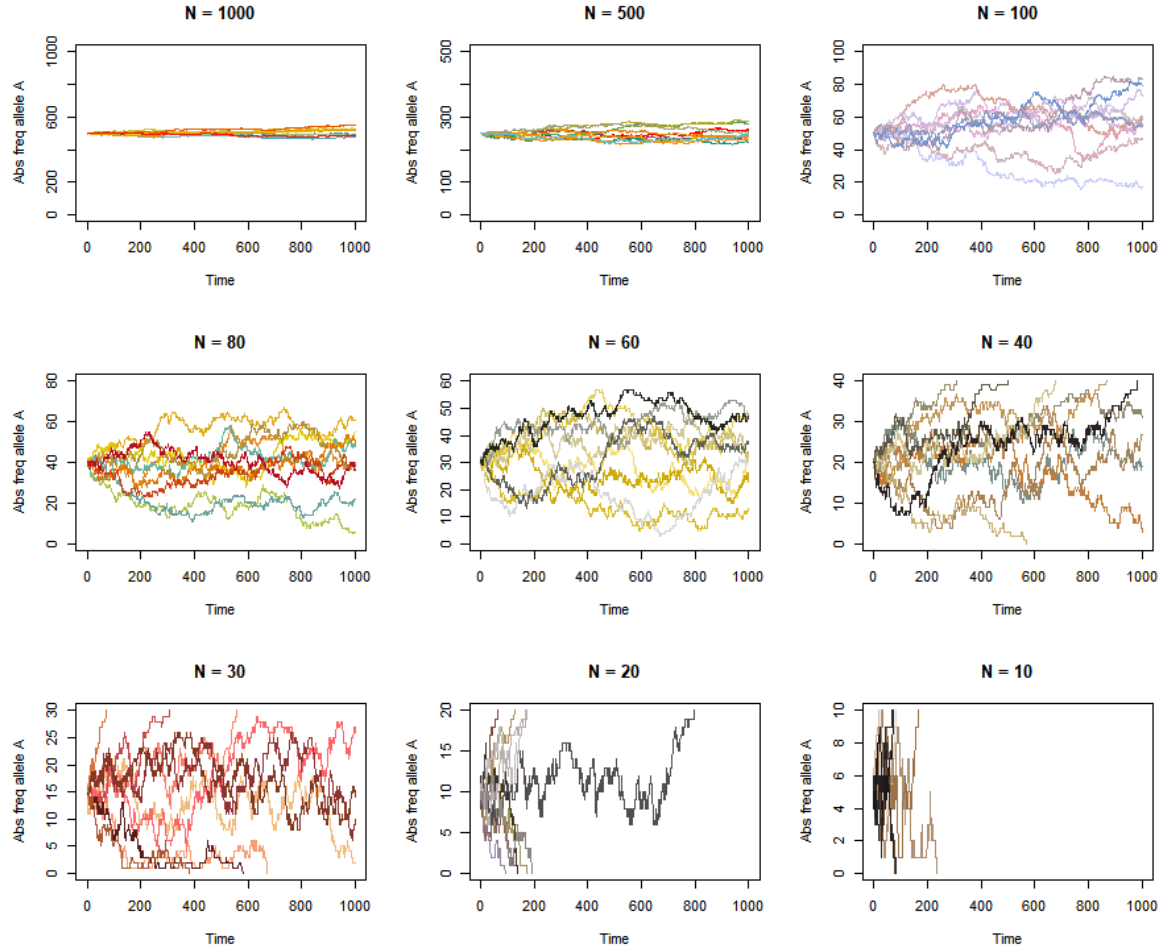


Figure 2: Moran process neutral evolution simulated in  $R$  for different population sizes  $N = 1000, 500, 100, 80, 60, 40, 30, 20, 10$ . Initial value of the number of allele A in the polutation  $i$  corresponding to the proportion of 50% for all  $N$ 's. 1000 iterations.

Simulating the model for different population sizes Fig. (2), when we can see that the fluctuations affect the systems at different levels. The system presents two absorbing states, 0 and 1, *i.e.* states from which the system cannot leave after reaching it. Biologically, it means the elimination or fixation of the allele, respectively, and that

there's no coexistence of alleles. All other states are transient.

For larger populations, the transient time is longer; we can see that even after 1000 iterations, in populations with sizes such as  $N = 1000, 500, 100$ , none of the runs reached the final stage, the effect of  $\pm 1$  is less felt by the system. On the other hand, we observe that the smaller the population size, the more notable the effect of stochasticity. For  $N \leq 40$  we observe most of the simulations reaching one of the absorbing states, with  $\downarrow N$  resulting in shorter mean times to reach the culmination of the evolutionary process.

## 2.3 Probability of fixation

Knowing that the genetic drift leads to one allele dominating over the other, it is worth questioning what would it be the allele's A fixation probability given its initial frequency in the population  $i$ .

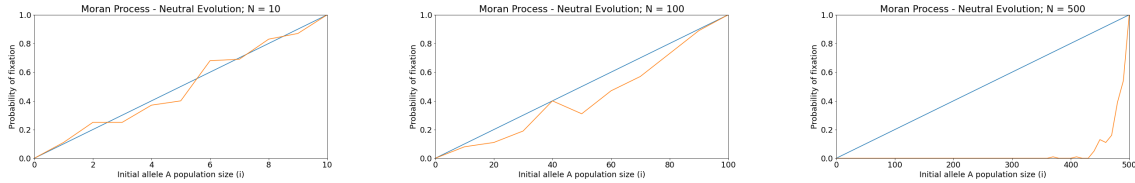


Figure 3: Probability of fixation for different population sizes  $N = 10, 100, 500$  given the initial frequency of allele  $i$ . Blue line indicates theoretical probability, orange line the probability obtained after simulating the model in *Python* with 10000 iterations, 100 runs for each  $i$ .

It is deeply related to the initial proximity to the absorbing states, with a fixation probability of  $i/N$  (theoretical), but simulating it Fig. (3) we can observe that the result depends on the size of the population, with fluctuations for more or less the theoretical value for small  $N$ , and the number of iterations: for large  $N$  if we don't give it enough time we won't see allele A fixating unless for values of  $i$  starting close to 100% of total population size.

## 3 Conclusion

After all the simulations, we can have as a take-home message that stochastic processes appear in multi aspects of biology, but in evolution, stochasticity is particularly important, yielding to unique mechanisms that can shape evolutionary dynamics.

This short project was a way of exercising the tools and knowledges seen in the Computational Methods course, in the Serrapilhiera/ICTP-SAIFR Quantitative Biology and Ecology Program. It is an Open Science initiative, all the project's codes can be found in: <https://github.com/amanda-araujo/stochastic-mathematical-modeling>.

## References

- [1] Laurent Duret. Neutral theory: the null hypothesis of molecular evolution. *Nature education*, 1:803–806, 2008.
- [2] Patrick Alfred Pierce Moran. Random processes in genetics. In *Mathematical proceedings of the cambridge philosophical society*, volume 54, pages 60–71. Cambridge University Press, 1958.
- [3] Martin A Nowak. *Evolutionary dynamics: exploring the equations of life*. Harvard university press, 2006.