Session 2 Simulation of random genetic drift

Dmytro Pravdyvets

Topic : Simulation of the Wright-Fisher model to describe the behavior of Genetic Drift

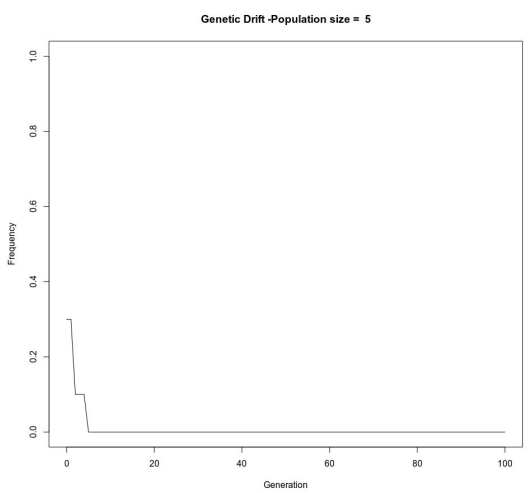
- All the scripts for this exercise can be found in the script file

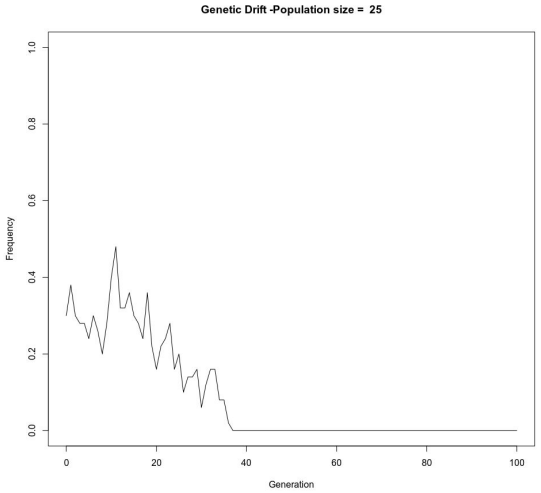
Exercises:

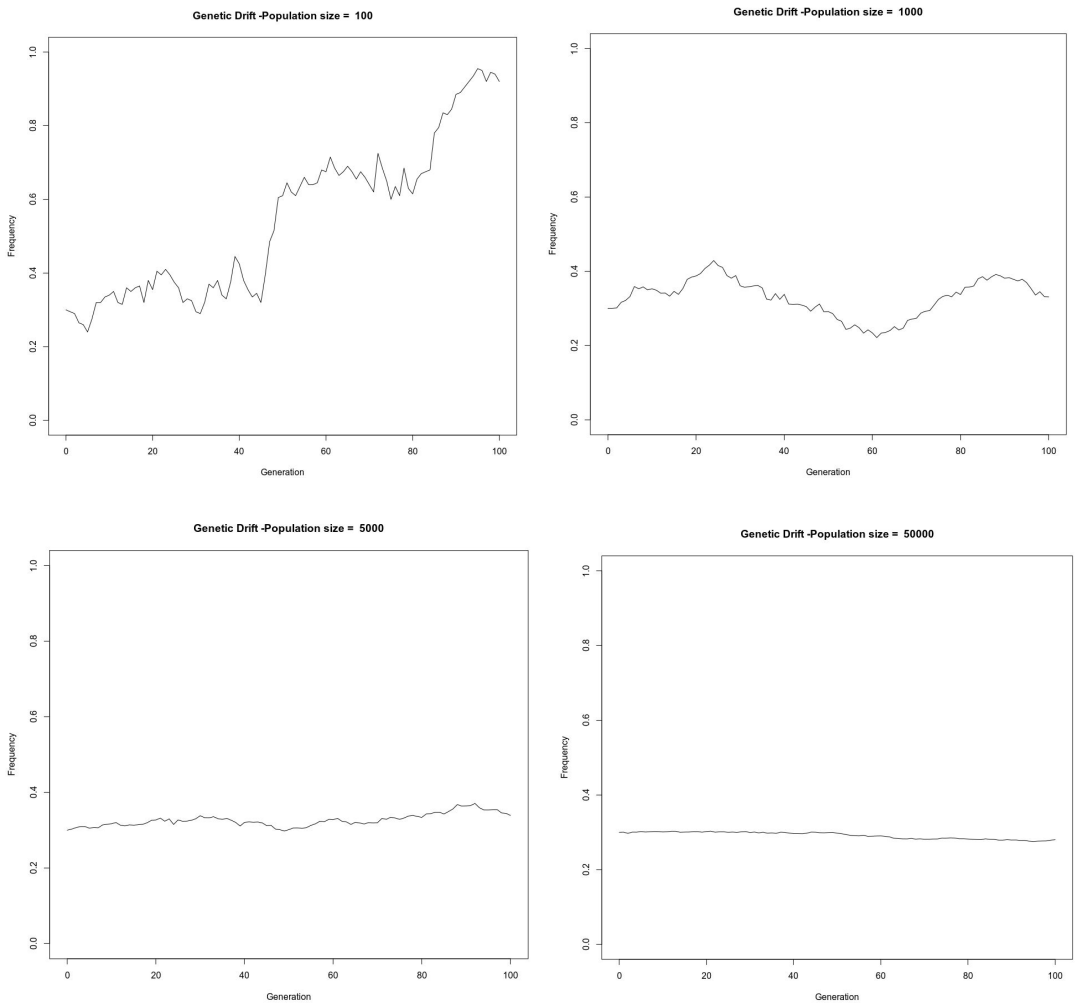
1. Explore the effects of population size on allele frequencies and describe it

- First step was to see how changing the Ne will affect the way random drift effects on the population we are simulating. After checking a range of sizes between 5 and 50000,it is pretty easy to see, that by increasing the Ne, the effect of genetic drift becomes less and less noticeable if increasing Ne . As a summary we can say that a range 2-25 is the most sensible to genetic drift and everything beyond 500 is very little sensible to it.

  Here you can see some of the graphs for different Ne:







2. Consider the whole set of replicates of genetic drift for a given initial conditions. How do average allele frequencies and allele frequency variance among replicates change over generations? Add on the previous script the code to graph the average (mean) allele frequency and variance over generations. When is the maximum variance in allele frequencies among replicates reached? Which is its value?

- After modifying the script (which can be found in the script file by the name Script 2) and running it for different Ne, we can see, that by increasing Ne the variation of the frequency gets smaller and smaller, for example a population of size 5 has a variation of "var: 0.0317059028915135", while a population of size 50000 is at "var: 7.21241618953687e-05", meaning that by increasing Ne, we get more stable frequency, which eventually makes sense

3. Simulate repeatedly a bottleneck at a given t generation and compares its impact with the same population without bottleneck.

- After modifying the script and checking for different Ne, amount of generations and frequencies it is possible to say, that depending on how big is the impact of the bottleneck, it is has more or less effect on the population. Smaller decrease in population at a bottleneck point makes little to no difference in the frequencies, while decreases higher than 10 in my example can completely kill homozygosity in a population. (Everything done in script number 3, be careful with numbers, it can crash a computer)

4. Simulate variation in population size over generation. Is the harmonic mean a good predictor of the behaviour of variation in sample size?

- By modifying the original script into Script 4, we can generate our random Ne and than taking the harmonic mean of this and use it in the original script, we can say, that it can be a hit or miss, with smaller sample sizes the probability of the harmonic mean not following the same results as the original series of numbers is higher, while for big sample sizes, it is more than often very similar or equal

5. Consider a new mutation in a previous monomorphic population. Which is its probability of fixation? How much time in generations would take it?

- After doing some research on my own and checking the scripts i've come to the conclusion that depending on the effect of the mutation, the percentage of the population it is affecting in the beginning and more variables (that make things more complicated, so i'll stick to the previous two), the time and probability of fixation change. For larger effect mutations with high probability the time of fixation is lower, because of the higher probability of it, while for small ones it goes vice versa.