Monte Carlo Simulations of Allele Frequencies*

Quoclinh Nguyen
Masakatsu Watanab e

This work is produced by The Connexions Project and licensed under the
Creative Commons Attribution License 2.0†

Abstract

In this assignment you will apply Monte Carlo computer simulations to study the effects of genetic drift and selection pressure on a single locus with two alleles: "A" and "a". You will be running these simulations on a computer using a simple simulation program written in Python (see appendix if this is not on your computer). Note that you do not need to know how to program in Python in order to do this lab. Each exercise will involve editing one or two lines at the top of the file and then running the simulation for a few minutes and analyzing the output

BIS 1 COMPUTER ASSIGMENT #2

1 Monte Carlo Simulations of Allele Frequencies

IMPORTANT: You must generate you own answers to all questions.

In this assignment you will apply Monte Carlo computer simulations to study the effects of genetic drift and selection pressure on a single locus with two alleles: "A" and "a". You will be running these simulations on a computer using a simple simulation program written in Python (see appendix if this is not on your computer). Note that you do not need to know how to program in Python in order to do this lab. Each exercise will involve editing one or two lines at the top of the file and then running the simulation for a few minutes and analyzing the output.

The simulation software carries out the simulation shown in the following flowchart. The initial population is 100% heterozygous (Aa). Each generation the population is exactly replaced by new individuals and all of the old individuals die off. In the simulations in Part II, in which some part of the population is eliminated by selection, the population is restored to the original population size in the next generation. This might seem artificial, but it actually is not a bad model of species that produce a huge number of offspring only a fraction of which survive to reproduce due to limitations in food supply or space.

*Version 1.4: Mar 17, 2009 11:06 am -0500
†http://creativecommons.org/licenses/by/2.0/

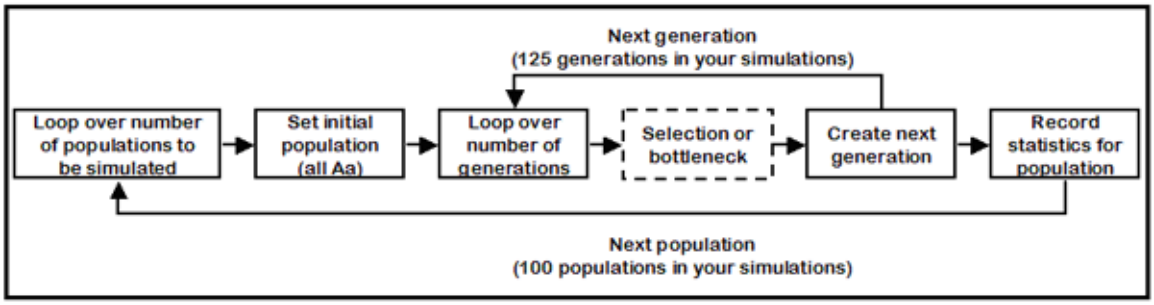

Figure 1

Figure 1 **Flowchart for simulation program. The program will simulate 100 different populations for each set of conditions you specify and simulate each population for 125 generations.**

In most cases the output will be a histogram showing the distribution of allele A frequencies across the set of different populations simulated. For example, the following histogram is the result of simulating 100 separate populations. The number ranges on x-axis give the range of allele A frequencies and the length of the bar reflects how many populations had allele A percentages in this range. So, of the 100 simulated populations, 25 ended up with allele A percentages between 50 and 60 percent, while just 1 population had an allele A percentage less than 10 percent.

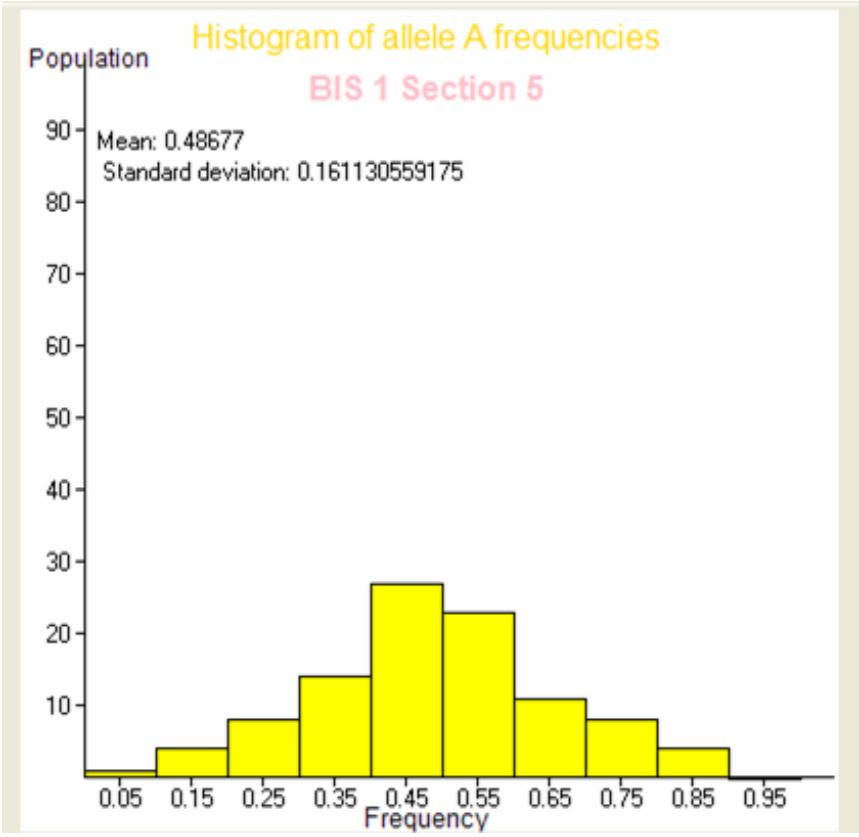

Figure 2

Figure 2 **sample output**

General instructions for using the simulation software:

Step 1: First, Go to http://bioinformatics.ucmerced.edu/resources/biological_sciences_1[1] . Then, download "Population Allele Simulation file" of Assignment 2 on your Desktop.

Step 2:Then, start the Python interpreter: Start::Programs::Python2.4::IDLE(Python GUI). This will start a Python Shell window. (This location might vary slightly for different versions of Python.) If you don't have Python in your computer, please refer to Appendix 1 of this assignment to install Python.

Step 3:In the File menu on the Python Shell, select Open and then navigate to Desktop and open the file called Population Allele Simulation (which you downloaded in step 2). This will open a new window showing the simulation program

Note: You may see Python icon with your saved files, so just double-click the icon to run the program.

[1] http://bioinformatics.ucmerced.edu/resources/biological_sciences_1

Figure 3

Figure 3 **python icon**
Step 4:Then, two windows will open: console and graphic user interface (GUI) windows.

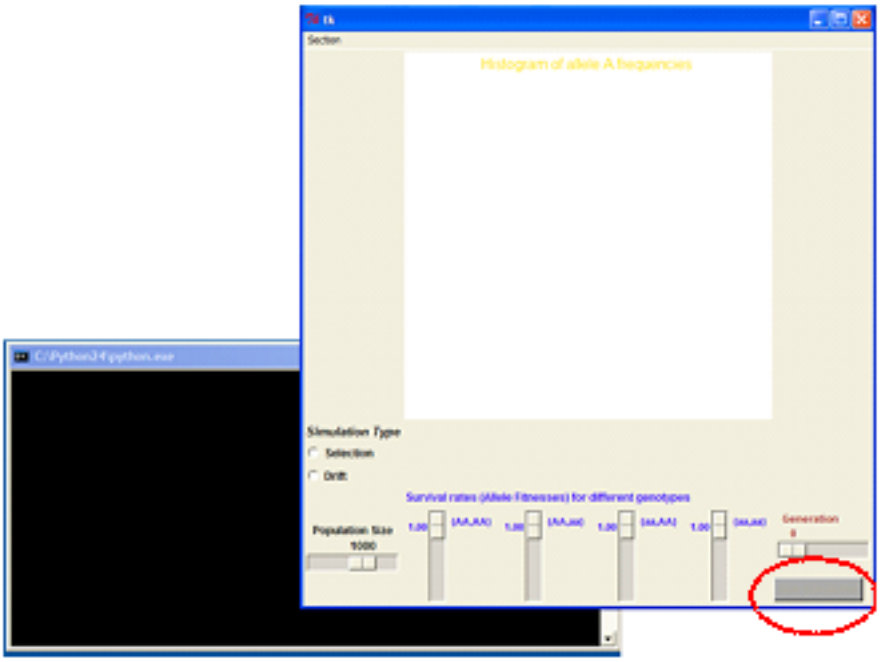

Figure 4

Figure 4. **This Python program opens console and graphic user interface windows.**
Step 5: Now select your BIS 1 section from the menu bar. **This is very important. Otherwise, the "Run Simulation" button in the dotted circle of the above figure cannot be turned on.**
Step 6:You can run a simulation by clicking Run Simulation button from the GUI window. The simulation will now start running. In the Python console window, you will see a series of numbers printing out on the first line—this is just an indicator of the number of populations completed—since we're running 100 populations, you'll be done shortly after this number reaches 90.

2 Part I: Modeling Genetic Drift

In this set of experiments, you will be testing the effect of population size of the genetic drift. This will involve running a number of simulations with different population sizes without selection or bottlenecks. Since genetic drift arises from random fluctuations in allele frequencies, it's not surprising that the size of the population is a key parameter in the rate of genetic drift.

Step 1:Open the Population Allele Simulation as described above. Check to make sure that the simulation type is set to be "Drift".

Step 2: Set the population size to 1000: Next, run the simulation, recording in the table below, the mean (Ave) and standard deviation (SD) for each generation are printed on the Python GUI, for example:

```
Number of generation:  125
Mean:  0.51
Standard deviation:  0.24
```

Be sure to record the data for every generation including the final case (generation 125) – **you can check histograms, Means, and Standard deviations of the different generations by sliding the "Generation" scale on the GUI.**

Step 3:Run three more simulations with different population sizes (500, 200, 100, 50, and 25) and then fill in the table below with means and standard deviations of the allele distributions at each of the generations printed. If you feel like it, you can run some additional simulations with other populations, but note that populations over 1000 will start to take a lot of computer time. Note: After you are done the simulation of each population size, be sure to save a histogram of each generation by selecting Print:Save as a Postscript in the menu. Then, you will have a record of your results and histograms. To print these postscripts files, please read Appendix 2 or 3.

Pop	Generation											
	0		25		50		75		100		125	
	Ave	SD	Ave	SD	Ave	SD	Ave	SD	Ave	SD	Ave	SD
1000												
500												
250												
100												
50												
25												

Table 1

3 Part II. Selection Pressure

In the genetic drift simulations above, all genotypes (AA, Aa, aA, and aa) had the same probability of surviving and reproducing. In this set of simulations you will adjust the "fitness" of different genotypes and observe the effects on the allele distributions. Note that in this study the population is set to 500.

Step 1:Start the simulation program using the same steps you did above.

Step 2: Now set the simulation type to "Selection"

Step 3:First run a "control" simulation with all fitnesses set to 1.0 (Default). Record the results in the table below.

Step 4: Set the survival rates to slightly disfavor the homozygous recessive genotype aa and record the results in the table below

```
# Set survival rates for different genotypes
 selection={('A','A'):1.,('A','a'):1.,('a','A'):1.,('a','a'):0.98}
```

Also run this simulation with the aa fitness set to 0.95.

Step 5:Set the survival rates to slightly disfavor both homozygous genotypes and record the results in the table below

```
# Set survival rates for different genotypes
 selection={('A','A'):.95,('A','a'):1.,('a','A'):1.,('a','a'):0.95}
```

Step 6:Choose four other allele fitness combinations and run simulations on them and record the results in the table.

Results table for Selection Simulations (Part II)

Allele Fitnesses	Allele A frequencyMean	Allele A frequencyStandard Deviation
AA=1; Aa=aA=1; aa=1 (control)		
AA=1; Aa=aA=1; aa=0.98		
AA=1; Aa=aA=1; aa=0.95		
AA=0.95; Aa=aA=1; aa=0.95		
AA=___; Aa=aA=___; aa=___		
AA=___; Aa=aA=___; aa=___		
AA=___; Aa=aA=___; aa=___		
AA=___; Aa=aA=___; aa=___		

Table 2

3.1 Appendix 1 – Python

To do this assignment, you need to have Python in your computer. Python is an interpreted, interactive, object-oriented programming language. If Python is not installed in your computer yet, please follow the instructions below to install Python. The installation of Python is very simple and easy:

1 If you don't have Python in your computer, go to http://www.python.org/download/[2]

2 Select appropriate Python software for your computer.

For Windows users, select "Python 2.4.2.

Windows installer (Windows Binary version)".

For Mac users, select "Python 2.3 OS X 10.2 installer".

3 Save the installer file on your local machine, and then double-click python-2.4.2.msi

for Windows users or Macpython-OSX-2.3-1.dmg for Mac users.

[2] http://www.python.org/download/

For Windows users, double-clicking of MSI file doesn't work. Go to http://www.python.org/2.4.2/[3] . Then, follow the instructions in the section of "Download the release".

4 Done! You are ready to go.

As a general computer language, Python combines remarkable programming power with very clear syntax. If you want to learn more about Python programming, there is a book called, "How to Think Like a Computer Scientist: Learning with Python", available for free on the web. You can go to http://greenteapress.com/thinkpython/html/[4] and read this web-book. It is also available as a pdf file for download at: http://greenteapress.com/thinkpython/thinkCSpy.pdf[5] .

3.2 Appendix 2 – Printing Postscript files @ UC Merced Library computer labs

In this assignment, you have saved all histograms in a postscript format. Postscript is a programming language optimized for printing graphics and text. To print these files at the UCM library computer lab, please follow these steps:

1. Double-click the postscript file.
2. Then you may have the following "Open With" window. Now select "ArcPress extension for Arc/Info" and press OK. (You may not see this window. If this is a case, go to Step 3)

[3] http://www.python.org/2.4.2/
[4] http://greenteapress.com/thinkpython/html/
[5] http://greenteapress.com/thinkpython/thinkCSpy.pdf

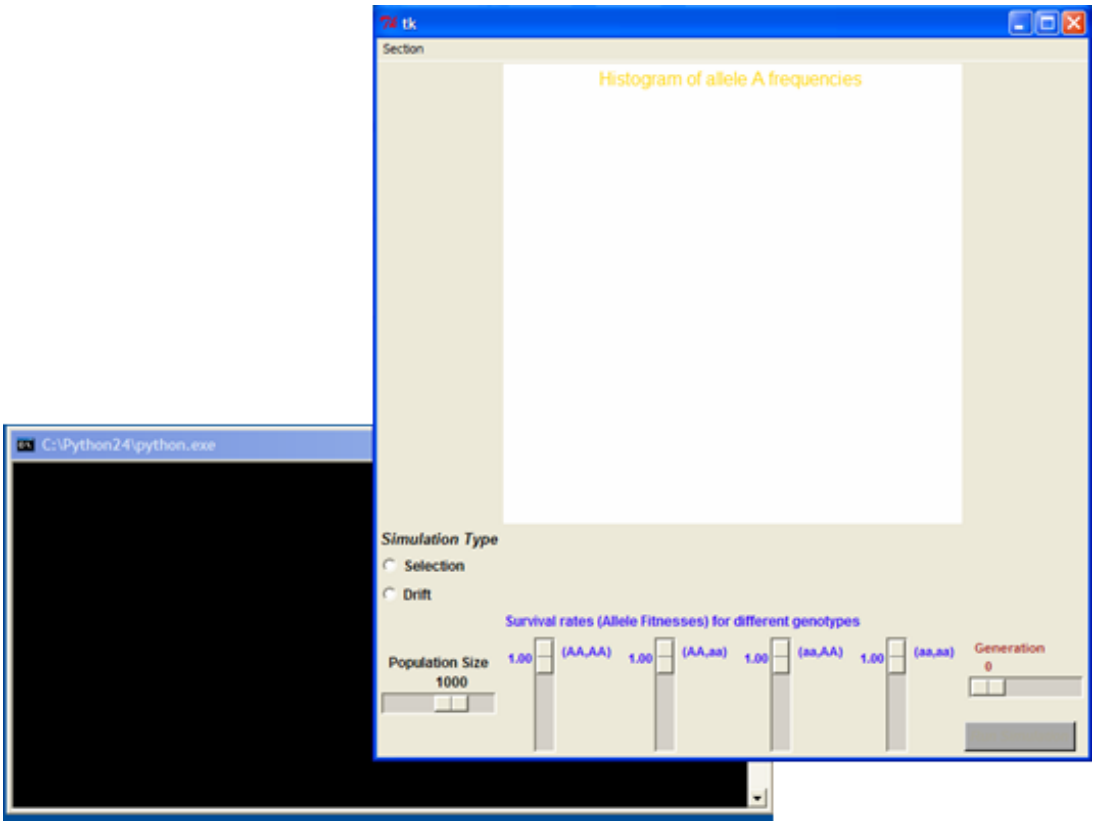

Figure 5

Figure 5 **how to postscript**

1. Then, you will see "Print" windows. Now select "Postscript" from the Driver menu. Then, press "Print.." to print the file.

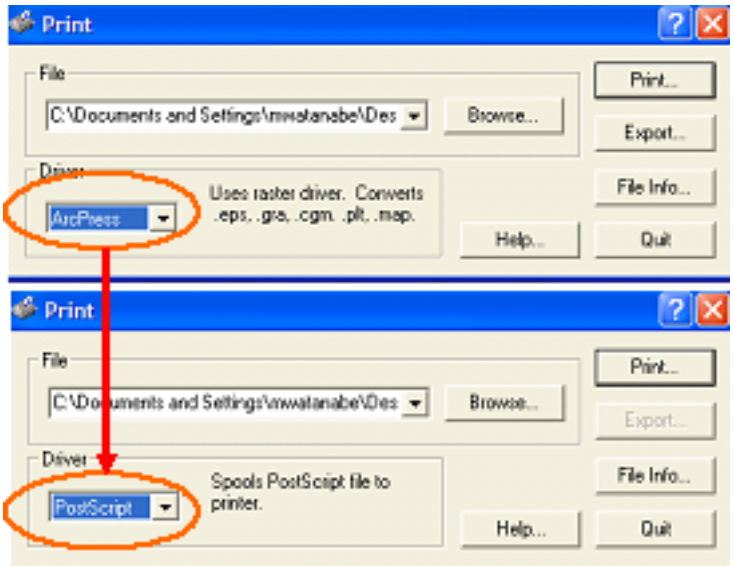

Figure 6

Figure 6 **how to postscript**

3.3 Appendix 3 – Printing and Viewing Postscript files from your computer

You may need to have Postscript interpreter in your computer to view or print postscript files at your computer. The installation of Postscript interpreter is very simple and easy:

For Windows users,

1. Go to http://www.rops.org/download/rops65c.exe[6] and save the executable file on your desktop.
2. Double-click Rops executable to install the program.
3. Done! You are ready to go.

Exercise 1 ***(Solution on p. 11.)***

In Part I, what do your results qualitatively (i.e. non-mathematically) indicate about the rate of genetic drift in different size population?

Exercise 2 ***(Solution on p. 11.)***

Look at the histograms you saved out in Part I. Do they seem to follow a Gaussian distribution (i.e. a bell-shaped curve)?

Exercise 3 ***(Solution on p. 11.)***

On a piece of graph paper (or using Excel on the computer). Graph the mean and standard deviations in allele frequencies versus the population size. Does the mean change allele frequency change with the size of the population? Does the standard deviation? Is there a simple mathematical relationship you can see? (Feel free to run more simulations to test this.)

Exercise 4 ***(Solution on p. 11.)***

In Part II, do your simulation results make intuitive sense? (explain in a few words why or why not)

[6] http://www.rops.org/download/rops65c.exe

Exercise 5 ***(Solution on p. 11.)***

Insert Problem Text Here

Solutions to Exercises in this Module

Solution to Exercise (p. 9)
Insert Solution Text Here
Solution to Exercise (p. 9)
Insert Solution Text Here
Solution to Exercise (p. 9)
Insert Solution Text Here
Solution to Exercise (p. 9)
Insert Solution Text Here
Solution to Exercise (p. 10)
How big an effect was caused by changing the selection pressure against genotype "aa" from 0.99 to 0.95? How would you expect this to change if you had a much larger population (your simulation was of a population of 500 individuals).