

# Genetic Drift

## 1 Introduction

All real populations are finite, and therefore, all real populations are subject to genetic drift. In this exercise, we'll be looking at the effects of population size and initial allele frequency on drift dynamics. As we went over in lecture...

1. Genetic drift will eventually lead to the loss of all alleles in the populations except one.
2. The probability that any allele will eventually become fixed in the population is equal to its current frequency.

Let's test this!

## 2 Drift Simulations

Let's do some simulations in R (in a biallelic system). Instructions:

- Open up the `drift_sim.R` file
- Install and call the `colorspace` package
- Compile the `Drift_graph()` function
- Note the arguments to this function:
  - `t` (time, aka number of generations to run)
  - `R` (number of replicates)
  - `N` (effective population size)
  - `p_init` (initial allele frequency of p)
- Run the function using the below settings
- For each: record # of times p remains stable or goes to fixation (top of graph) or extinction (bottom of graph)
- Report for class data

- Qualitatively observe when (in terms of  $t$ ) simulations go extinct compared with different values of  $N$
- Part 1:
  - For all:  $t = 1000$ ,  $R = 10$ ,  $p_{\text{init}} = 0.5$
  - Vary  $N$ : 10, 50, 100
- Part 2:
  - For all:  $t = 1000$ ,  $R = 10$ ,  $p_{\text{init}} = 0.25$
  - Vary  $N$ : 10, 50, 100
- Part 3:
  - For all:  $t = 1000$ ,  $R = 10$ ,  $p_{\text{init}} = 0.1$
  - Vary  $N$ : 10, 50, 100

Once we have class data, we can see if our simulations match our expectations!

## 2.1 Data

p	N	<i>stable</i>	<i>fixed</i>	<i>extinct</i>
0.5	10			
0.5	50			
0.5	100			
0.25	10			
0.25	50			
0.25	100			
0.1	10			
0.1	50			
0.1	100			

## 2.2 Compiled class data

1. Do all replicates eventually go to fixation or extinction?
2. Does the probability of fixation = probability of initial  $p$  frequency?

## 3 Effective Population Size

Remember:

$$Var(p_t) = p_0(1 - p_0)(1 - (1 - \frac{1}{2N_e})^t) \quad .$$

We will use the package `NB` in R to generate an example data set and look at the effective population size ( $N_e$ )

1. Install the package `NB` and call it using the `library()` function
2. Create an example data set using the `NB.example.dataset()` function
  - This will create a file in your working directory called ‘`sample_data.txt`’
  - The example has 50 loci with 4 alleles at each locus, sampled at the 0th and 8th generations
3. Use the `NB.estimator()` function to estimate the effective population size.
  - What arguments do you need?
  - use `?NB.estimator` help to see what the arguments are!
  - What is the best estimate? \_\_\_\_\_
  - What are the  $N_e$  values associated with the 95% confidence intervals? \_\_\_\_\_
4. Use the `NB.likelihood()` function to get the log-likelihood values for  $N = 10, 100, 500, 1000$ . What trends do you see?
5. Now use the `NB.plot.likelihood()` function to plot for  $N = 10 : 500$ , with a step size of 50.
  - Again, use the `?` to look at the help page to understand the arguments!
  - What does a “leveling-off” mean here?
6. Now, pretend that we’ve sampled the 0th and 2nd generations, and repeat steps 3 and 5.
  - What is the best estimate now? \_\_\_\_\_

7. Now pretend that we've sampled the 0th and 100th generations, repeat steps 3 and 5.
  - What is the best estimate now? \_\_\_\_\_
8. Here, we observed the same allele frequencies, but assumed sampling different generations. How does this affect our results? How do you expect allele frequencies to change through time under drift?
9. Aside, how many generations of data could you obtain for your study organism(s) while here at UConn?

## 4 References

- Hui, Tin-Yu J., and Austin Burt. "Estimating effective population size from temporally spaced samples with a novel, efficient maximum-likelihood algorithm." *Genetics* 200.1 (2015): 285-293.