

# drift

June 5, 2024

## 1 Expected allele frequency with genetic drift

### 1.1 (1) Over one generation

We assume to have 50 diploid individuals and an allele frequency of 0.50.

```
[ ]: N <- 50  
     fA <- 0.5
```

A binomial distribution models the sampling with replacement of gene copies. The next cells illustrate how to perform a random sampling from a binomial distribution with  $2N$  samples and probability of “success” equal to  $f_A$ . Since we are interested in frequencies and not counts, we then divide by  $2N$ .

```
[ ]: rbinom(1, 2*N, fA) / (2*N)
```

```
[ ]: rbinom(1, 2*N, 0.53) / (2*N)
```

How do we calculate the expected allele frequency at the next generation? We can repeat the experiment many times and calculate the average value, then...

```
[ ]: # ...
```

```
[ ]: # ...
```

What do these results suggest? Does the expected allele frequency change or not after one generation? Do allele frequencies change or not after one generation?

Prove that your conclusions stay the same if you vary the initial allele frequency and/or sample size.

### 1.2 (2) Over many generations

Scenario similar to above, but with a starting allele frequency of 0.50, 50 diploid individuals and 100 generations.

```
[ ]: N <- 50  
     # 100 generations  
     fA <- rep(NA, 100)  
     # at t=0
```

```
fA[1] <- 0.50
```

We repeat the sampling procedure at each generation.

```
[ ]: for (t in 1:99) fA[t+1] <- rbinom(1, 2*N, fA[t]) / (2*N)
```

```
[ ]: plot(x=1:100, y=fA, type="l", ylim=c(0,1), lwd=2)
```

Wasn't the expected allele frequency supposed not to change after each generation? What's happening?

Let's assume that we repeat the whole history for 100 generations many times (i.e. 20) and record the trajectory of the allele frequency.

```
[ ]: N <- 50
gen <- 100 # generations
nrepl <- 20 # number of distinct simulations
for (j in 1:nrepl) {
  fA <- rep(NA, gen)
  fA[1] <- 0.5
  for (t in 1:(gen-1)) fA[t+1] <- rbinom(1, 2*N, fA[t]) / (2*N)
  if (j==1) plot(x=1:gen, y=fA, xlab="generations", type="l", ylim=c(0,1),
    ↪col=rainbow(nrepl)[j]) else lines(x=1:gen, y=fA, col=rainbow(nrepl)[j])
}
```

What is the chance that an allele at 50% at time 0 goes to fixation (100%) or is lost (0%)? To answer this question run the experiment above over more generations and more replicate and see what happens. Calculate the proportion of alleles that are fixed or lost at the end of your experiment. What happens if you change the initial allele frequency? Do these proportions change?

```
[ ]: # ...
```

### 1.3 (3) Effect of population size

Let's consider the effect of population size on the drift process.

```
[ ]: # small population (blue)
N <- 100
for (j in 1:10) {
  fA <- rep(NA, 100); fA[1] <- 0.5
  for (t in 1:99) fA[t+1] <- rbinom(1, 2*N, fA[t]) / (2*N)
  if (j==1) plot(x=1:100, y=fA, type="l", ylim=c(0,1), col="blue") else
    ↪lines(x=1:100, y=fA, col="blue")
}

# large population (red)
N <- 1000
for (j in 1:10) {
```

```
fA <- rep(NA, 100); fA[1] <- 0.5
for (t in 1:99) fA[t+1] <- rbinom(1, 2*N, fA[t]) / (2*N)
lines(x=1:100, y=fA, col="red")
}
```

Which population experiences more drift? Change the values of population sizes and confirm that your general statement is still true.

[ ]: