

Genetic Drift and Coalescence

Exercise 1: simulate binomial distribution with probability of successs p=0.1 in a sample of size 10. Find expected number of successes

```
expected <- mean(rbinom(10000, 10, 0.1))
expected

## [1] 0.9971
```

Exercise 2: write a function to simulate T generations of genetic drift for L independent SNPs. Keep track of allele freq of each SNP in each gen. All SNPs should start in the initial gen at freq q.

```
#genetic drift
pp <- 0.1
NN <- 10 # number of diploid indiv; 2*NN = number of chromosome
count <- rbinom(1, 2*NN, pp)
count

## [1] 2

count/(2*NN)

## [1] 0.1

#function
GenDriftFunc <- function(LL, NN, p0, TT){
  #matrix of allele freq of each SNP (col) in each gen (row)
  freqs <- matrix(nrow = TT, ncol = LL)
  #fill first row with p0
  freqs[1, ] <- p0
  #for loop to iterate rbinom for each subsequent row
  for (tt in 2:TT){
    for(l1 in 1:LL){
      freqs[tt, l1] <- rbinom(1, 2*NN, freqs[tt-1, l1])/(2*NN)
    }
  }
  return(freqs)
}
```

Exercise 3: use the function to simulate drift

```
NN=100
LL=1000
TT=10000
p0=0.1
sim1 <- GenDriftFunc(LL, NN, p0, TT)
```

a) How many of the 1000 SNPs are at freq 0 after 10K generations?

```
#extinct?
sum(sim1[10000, ] == 0)
```

```
## [1] 903
```

b) How many are at freq 1?

```
#fixation?
sum(sim1[10000, ] == 1)
```

```
## [1] 97
```

c) Does value agree with prediction for prob of fixation of neutral allele?

```
#agree?
sum(sim1[10000, ] == 1)/1000
```

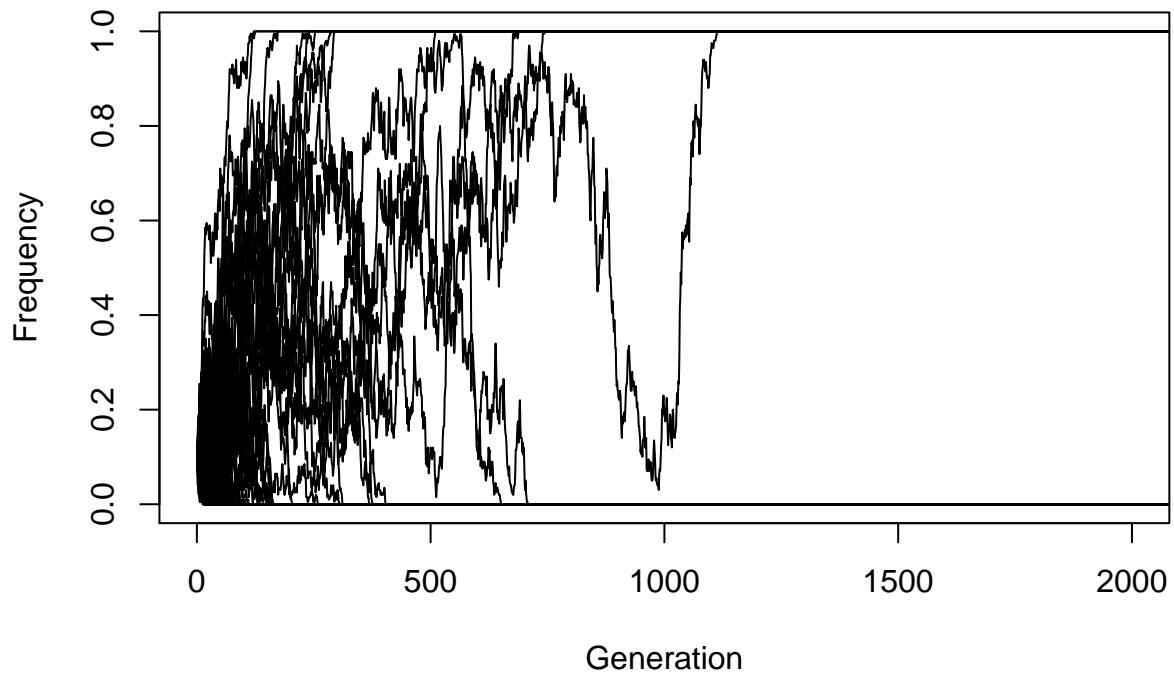
```
## [1] 0.097
```

Yes, the probability of fixation is close to 10%.

d) Plot trajectories for 100 of the SNPs

```
#plot
plot(sim1[, 1], type = "l", ylim = c(0, 1), xlim = c(0, 2000), main = "N=100", xlab = "Generation", yla
```

N=100



e) Repeat simulation with $p=0.6$. How many of the 1000 SNP are at 0?

```
#repeat
NN=100
LL=1000
TT=10000
p0=0.6
sim2 <- GenDriftFunc(LL, NN, p0, TT)
#extinct?
sum(sim2[10000, ] == 0)
```

```
## [1] 428
```

f) How many are at freq 1?

```
#fixation?
sum(sim2[10000, ] == 1)
```

```
## [1] 572
```

g) Does value agree with prediction for prob of fixation of neutral allele?

```
#agree?
sum(sim2[10000, ] == 1)/1000
```

```
## [1] 0.572
```

Yes, the probability of fixation is close to 60%.

Exercise 4: Effect of pop size on gen drift. Repeat simulation, set N=10, 500, and 1000. Keep other parameters the same. (N is number of diploids)

```
# N=10
NN=10
LL=1000
TT=10000
p0=0.1
sim3 <- GenDriftFunc(LL, NN, p0, TT)

# N=500
NN=500
LL=1000
TT=10000
p0=0.1
sim4 <- GenDriftFunc(LL, NN, p0, TT)

# N=1000
NN=500
LL=1000
TT=10000
p0=0.1
sim5 <- GenDriftFunc(LL, NN, p0, TT)
```

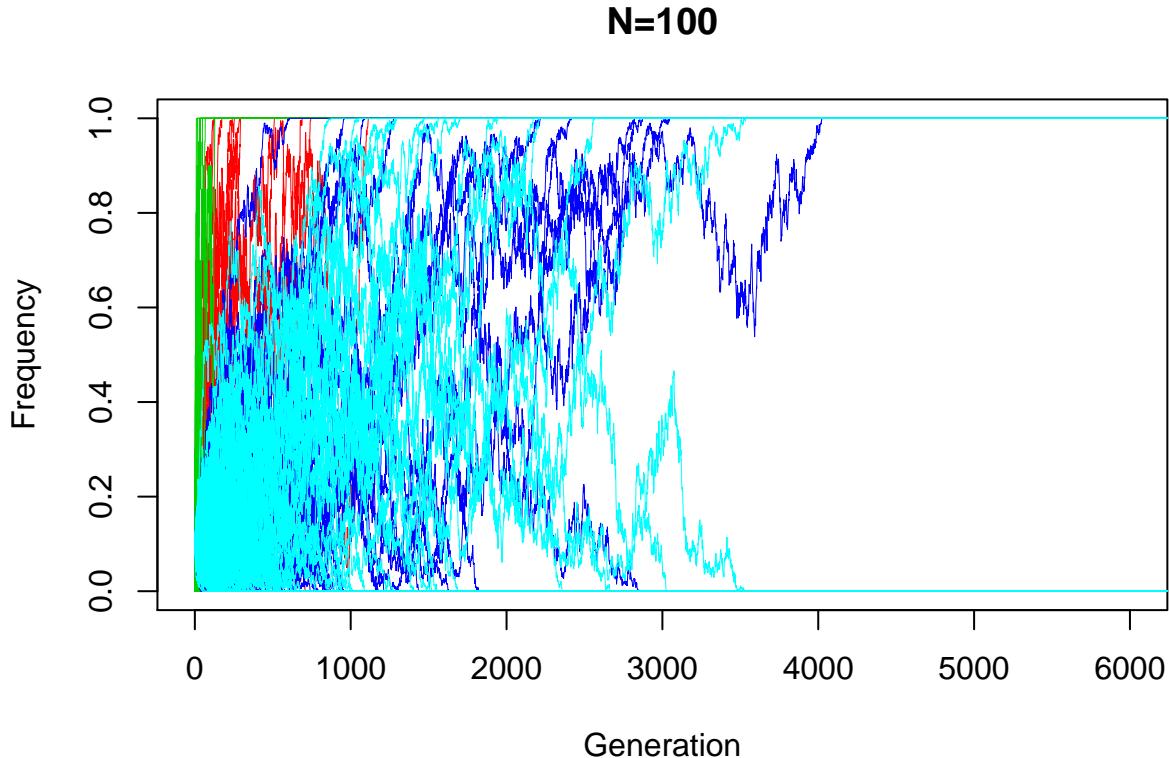
a) Plot all 4 pop sizes on the same page.

```
#plot sim1, sim3, sim4, sim5
#100=red; 10=green; 500=blue; 1000=cyan
plot(sim1[, 1], type = "l", ylim = c(0, 1), xlim = c(0, 6000), main = "N=100", xlab = "Generation", yla
for(ii in 2:100){
  lines(sim1[, ii], col = 2, lwd = 0.01)
  adjustcolor(col = 2, alpha.f = 0.8)
}
for(ii in 1:100){
  lines(sim3[, ii], col = 3, lwd = 0.01)
  adjustcolor(col = 3, alpha.f = 1)
}
for(ii in 1:100){
  lines(sim4[, ii], col = 4, lwd = 0.01)
  adjustcolor(col = 4, alpha.f = 0.6)
}
for(ii in 1:100){
  lines(sim5[, ii], col = 5, lwd = 0.01)
```

```

    adjustcolor(col = 5, alpha.f = 0.6)
}

```



- b) How does pop size affect allele freq change?

The larger the population size, the longer it takes for alleles to either reach fixation or go extinct.

- c) For each pop size, in what proportion of simulation replicates did the derived allele become fixed by the end?

```

#fixation?
fix1 <- sum(sim1[10000, ] == 1)/1000
fix3 <- sum(sim3[10000, ] == 1)/1000
fix4 <- sum(sim4[10000, ] == 1)/1000
fix5 <- sum(sim5[10000, ] == 1)/1000
popsize <- c(10, 100, 500, 1000)
fixedproportion <- c(fix3, fix1, fix4, fix5)
table <- rbind(popsize, fixedproportion)
table

##          [,1]      [,2]      [,3]      [,4]
## popsize    10.000 100.000 500.000 1.0e+03
## fixedproportion 0.096   0.097   0.101 9.5e-02

```

d) How is this probability affected by the pop size?

It isn't affect by the population size.

e) How does this probability of fixation estimated from the simulations match with theoretical prediction?

They are all close to 10%, matching the theoretical prediction.

Coalescent simulations