

Exercise

Population Genetics (Sim)

Simulation of genetic drift

Perform a simulation study on genetic drift. In particular, assume a SNP with two alleles of frequency 50% at generation 0, population sizes of 10, 100, 1000 and 10,000 chromosomes, respectively, and 50 generations. For each population size, perform 100 replications. Graph the allele frequencies by generation for all 100 replications of size 1000 within a single graph. Summarize replication results by generating a histogram and calculate summary statistics for the allele frequencies after 50 generations for each of the three population sizes.

The required R script is placed in the file **popgen_drift.q**. Open this file in a text editor and copy-paste the commands in the R console (you can also use the `source` command in R). Inspect the resulting graphs.

Frequency changes caused by selection

Explore the impact of the strength of selection on the change in allele frequency. To this end, assume selection coefficients s of 0.001, 0.01 and 0.1, respectively, a heterozygote effect of $h=0.5$ and a deleterious allele frequency of 50% at generation 0. Calculate the allele frequencies within the next 100 generations and graph them by generation. Give the allele frequencies after 100 generations for each of the three coefficients.

$$f_{s=0.001}(100) = \underline{\hspace{2cm}} \quad f_{s=0.01}(100) = \underline{\hspace{2cm}} \quad f_{s=0.1}(100) = \underline{\hspace{2cm}}$$

The required R script is placed in the file **popgen_selection.q**. Open this file in a text editor and copy-paste the commands in the R console (you can also use the `source` command in R). Inspect the resulting graph.

popgen drift.q

```
N.s = c(10, 100, 1000, 10000)
n.gen = 50
n.rep = 100

# === Simulate genetic drift === #
for (n in N.s) {
  freqs = matrix(NA, ncol=n.gen+1, nrow=n.rep)
  for (i in 1:n.rep) {
    alleles = c(rep(0,n/2), rep(1,n/2))
    freqs[i,1] = sum(alleles==1) / length(alleles)
    for (j in 1:n.gen) {
      alleles = sample (alleles, length(alleles), replace=T)
      freqs[i,j+1] = sum(alleles==1) / length(alleles)
    }
  }
  assign (paste("freqs.n", n, sep=""), freqs)
}

summary(freqs.n10 [,n.gen+1])
summary(freqs.n100 [,n.gen+1])
summary(freqs.n1000 [,n.gen+1])
summary(freqs.n10000[,n.gen+1])

# === Graph allele frequency changes === #
pdf("drift_plot.pdf", paper="special", height=4*2, width=4*2, onefile=F)
split.screen(c(2,2))
  screen(1)
    plot(x=0, y=0, type="n", xlim=c(0,n.gen), ylim=c(0,1),
         xlab="Generation", ylab="Allele frequency", main="N=10")
    lines (c(0,n.gen), rep(0.5, 2), lty=3, col="#AAAAAA")
    freqs = get(paste("freqs.n", 10 , sep=""))
    for (i in 1:n.rep) { lines(0:n.gen,freqs[i,], col="#44AAAA") }
  screen(2)
    plot(x=0, y=0, type="n", xlim=c(0,n.gen), ylim=c(0,1),
         xlab="Generation", ylab="Allele frequency", main="N=100")
    lines (c(0,n.gen), rep(0.5, 2), lty=3, col="#AAAAAA")
    freqs = get(paste("freqs.n", 100 , sep=""))
    for (i in 1:n.rep) { lines(0:n.gen,freqs[i,], col="#44AAAA") }
  screen(3)
    plot(x=0, y=0, type="n", xlim=c(0,n.gen), ylim=c(0,1),
         xlab="Generation", ylab="Allele frequency", main="N=1000")
    lines (c(0,n.gen), rep(0.5, 2), lty=3, col="#AAAAAA")
    freqs = get(paste("freqs.n", 1000 , sep=""))
    for (i in 1:n.rep) { lines(0:n.gen,freqs[i,], col="#44AAAA") }
  screen(4)
    plot(x=0, y=0, type="n", xlim=c(0,n.gen), ylim=c(0,1),
         xlab="Generation", ylab="Allele frequency", main="N=10000")
    lines (c(0,n.gen), rep(0.5, 2), lty=3, col="#AAAAAA")
    freqs = get(paste("freqs.n", 10000, sep=""))
    for (i in 1:n.rep) { lines(0:n.gen,freqs[i,], col="#44AAAA") }
  close.screen(all.screens=T)
dev.off()

# === Graph allele frequency after 50 generations === #
pdf("drift_hist.pdf", paper="special", height=4*2, width=4*2, onefile=F)
split.screen(c(2,2))
  screen(1)
    hist(freqs.n10 [,n.gen+1], main="N=10", xlab="Allele frequency", xlim=c(0,1))
  screen(2)
    hist(freqs.n100 [,n.gen+1], main="N=100", xlab="Allele frequency", xlim=c(0,1))
  screen(3)
    hist(freqs.n1000 [,n.gen+1], main="N=1000", xlab="Allele frequency", xlim=c(0,1))
  screen(4)
    hist(freqs.n10000[,n.gen+1], main="N=10000", xlab="Allele frequency", xlim=c(0,1))
  close.screen(all.screens=T)
dev.off()
```

popgen_selection.q

```
s.s = c(0.001, 0.01, 0.1)
h = 0.5
n.gen = 100

# === Calculate allele frequency changes === #
for (s in s.s) {
  freqs = rep(NA, n.gen+1)
  af = 0.5
  freqs[1] = 1-af
  for (j in 1:n.gen) {
    omega = 1 - 2*af*(1-af)*h*s - (1-af)*(1-af)*s
    f.het = (1-h*s)*2*af*(1-af)/omega
    f.hom = af*af/omega
    af = f.hom + f.het/2
    freqs[j+1] = 1-af
  }
  assign (paste("freqs.s", s, sep=""), freqs)
}

# === Report allele frequencies after 100 generations === #
for (s in s.s) {
  cat("s=") ; cat(s) ; cat(": ")
  freqs = get(paste("freqs.s", s, sep=""))
  cat(freqs[n.gen+1]) ; cat("\n")
}

# === Graph allele frequency changes === #
pdf("selection_plot.pdf", paper="special", height=4*2, width=4*2, onefile=F)
plot(x=0, y=0, type="n", xlim=c(0,n.gen), ylim=c(0,1), xlab="Generation", ylab="Allele
frequency")
lines (c(0,n.gen), rep(0.5, 2), lty=3, col="#AAAAAA")
for (s in s.s) {
  freqs = get(paste("freqs.s", s, sep=""))
  lines(0:n.gen,freqs, col="#44AAAA")
}
dev.off()
```

Answers

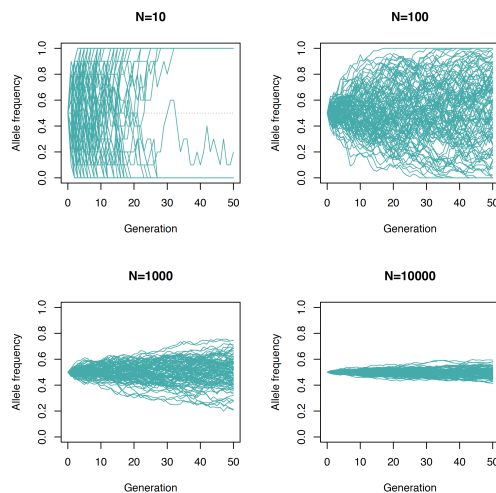
Population Genetics (Sim)

Simulation of genetic drift

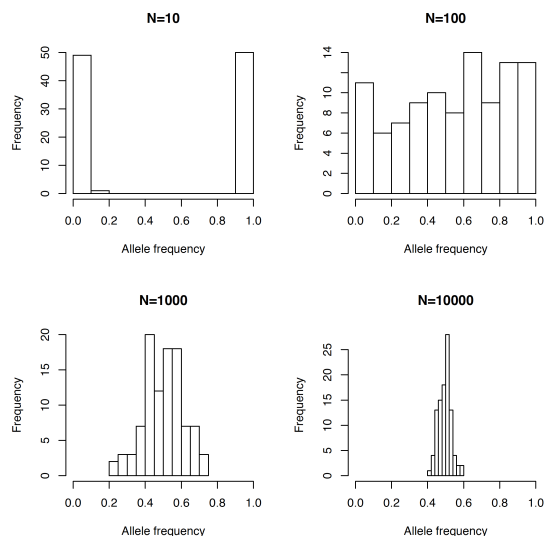
Perform a simulation study on genetic drift. In particular, assume a SNP with two alleles of frequency 50% at generation 0, population sizes of 10, 100, 1000 and 10,000 chromosomes, respectively, and 50 generations. For each population size, perform 100 replications. Graph the allele frequencies by generation for all 100 replications of size 1000 within a single graph. Summarize replication results by generating a histogram and calculate summary statistics for the allele frequencies after 50 generations for each of the three population sizes.

The required R script is placed in the file `popgen_drift.q`. Open this file in a text editor and copy-paste the commands in the R console (you can also use the `source` command in R). Inspect the resulting graphs.

Allele frequency changes



Allele frequencies after 50 generations



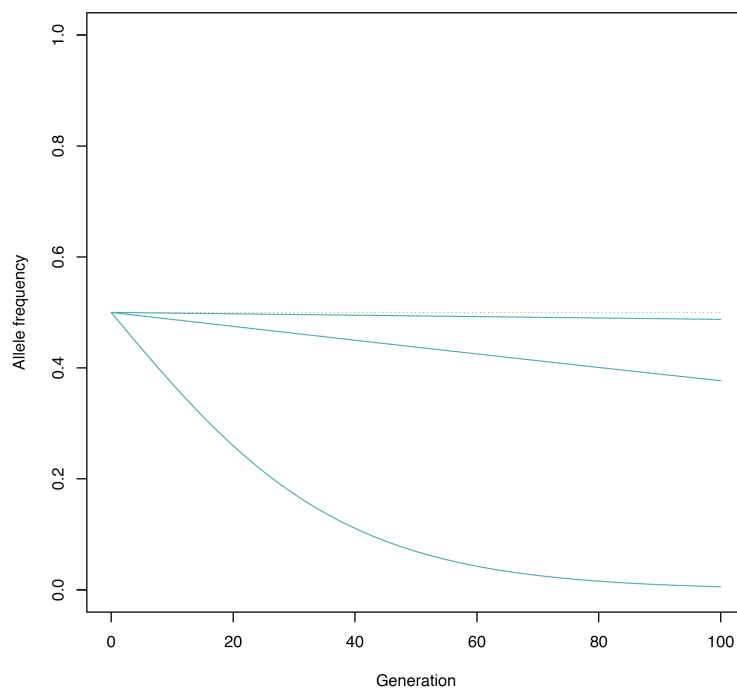
Frequency changes caused by selection

Explore the impact of the strength of selection on the change in allele frequency. To this end, assume selection coefficients s of 0.001, 0.01 and 0.1, respectively, a heterozygote effect of $h=0.5$ and a deleterious allele frequency of 50% at generation 0. Calculate the allele frequencies within the next 100 generations and graph them by generation. Give the allele frequencies after 100 generations for each of the three coefficients.

$$f_{s=0.001}(100) = \underline{\hspace{1cm}} \quad f_{s=0.01}(100) = \underline{\hspace{1cm}} \quad f_{s=0.1}(100) = \underline{\hspace{1cm}}$$

The required R script is placed in the file **popgen_selection.q**. Open this file in a text editor and copy-paste the commands in the R console (you can also use the `source` command in R). Inspect the resulting graph.

Allele frequency changes



Allele frequencies after 100 generations

$$f_{s=0.001}(100) = \underline{0.4875} \quad f_{s=0.01}(100) = \underline{0.3770} \quad f_{s=0.1}(100) = \underline{0.0057}$$