Population Genetics Project #4

Krimbas and Tsakas (*Evolution* 25:454-460; 1971) studied the evolution of allele frequencies at two different loci encoding soluble esterases in the olive fruit fly, *Dacus oleae*. The product of locus A catalyzes degradation of acetylcholine. The product of locus B catalyzes degradation of pseudocholine. Laboratory experiments had revealed that homozygotes for the null allele at locus A were more susceptible to dimethoate, an organophosphate insecticide, while there was no detectable selection at locus B.

To see whether selection could be detected in wild populations, Krimbas and Tsakas collected flies form an olive orchard in Greece in 1966, 1967, and 1968. A simplified version of the data they collected is summarized in the following table:

	Year		
	1966	1967	1968
Number of A_1 alleles	206	142	230
Sample size of A alleles	474	312	400
Number of B_1 alleles	292	179	280
Sample size of B alleles	469	281	409

Assume that the changes in allele frequency from year to year are entirely the result of genetic drift, and assume that there is only one generation per year.[1] Using those assumptions and the data above, answer the following questions:

1. What would your estimate of the effective population size be if you based it on allele frequency changes at locus A?

2. What would your estimate of the effective population size be if you based it on allele frequency changes at locus B?

3. Is there evidence that the estimate of effective population size depends on the locus that you select?

[1] There are actually several generations per year, so the estimates for N_e you get below will be substantially smaller than the best estimate of N_e for these data.

4. What is your best estimate of the effective population size?

5. Given that the smallest sample included 281 alleles, we know that there are at least 140 individuals in this population. It's reasonable to presume that the sex ratio in this population is close to 50:50. If we assume that generations are non-overlapping and that the variance in individual reproductive success is roughly binomial, are the observed allele frequency changes plausibly accounted for by genetic drift, or might natural selection be involved?

Hints

- Recall that the variance in allele frequency is

$$\mathrm{Var}(p_{t+1}) = \frac{p_t(1-p_t)}{2N_e} \quad .$$

  Recall also that when we were estimating F-statistics we used a Beta distribution to describe the allele frequency variation among populations. We can use it here to describe allele frequency variation within a population over time. Specifically,

$$\begin{aligned} \mathrm{P}(p_{t+1}|p_t) &\sim \mathrm{Beta}(\nu p_t, \nu(1-p_t)) \\ \nu &= \frac{1-\theta}{\theta} \end{aligned}$$

  Using that approach,

$$\mathrm{Var}(p_{t+1}) = \theta p_t(1-p_t) \quad .$$

  So we can use $1/2\theta$ as an estimate of N_d. In `JAGS` code for one locus

```
p[1] ~ dunif(0, 1)
for (i in 2:n_years) {
  p[i] ~ dbeta(nu*p[i-1], nu*(1-p[i-1]))
}
nu <- ((1.0 - theta)/theta)
theta ~ dunif(0,1)
ne <- 1.0/(2.0*theta)
```

- Use the value of `DIC` that `JAGS` reports to compare a model in which you estimate N_e separately from the data for locus A and locus B and one in which you estimate a single N_e from the combined data.