

Mathematical Genetics, Fall 2015, HW 5

Instructions: Do 2 of the 3 problems.

1. Let S_n be the number of segregating sites in a sample of size n . Assume infinite sites mutation model.

(a) Show that

$$P(S_n = 0) = \frac{(n-1)!}{(\theta+1)(\theta+2)\cdots(\theta+n-1)}.$$

Hint: Think of what has to happen in the coalescent tree to get all haplotypes in the sample identical.

(b) Derive the recursion

$$P(S_n = k) = \frac{\theta}{(\theta+n-1)}P(S_n = k-1) + \frac{n-1}{(\theta+n-1)}P(S_{n-1} = k).$$

Hint: Condition on the first event back in time from the sample.

2. Consider the following set of haplotypes (with their frequencies in the sample in parentheses on the right). Assume that the ancestral allele at a site is the most frequent allele at that site.

- a) T G C G A (1)
- b) T G T G T (1)
- c) A A T G T (2)
- d) T G C G A (1)
- e) T G C T T (1)

(a) Calculate Watterson's $\hat{\theta}_W$ and Tajima's $\hat{\theta}_T$ estimates of θ .

(b) Find the site frequency spectrum.

(c) Compute Tajima's D statistic. Note: The estimated variance in the denominator is given by

$$\widehat{Var}\left(D_n - \frac{S_n}{a_n}\right) = c_n S_n + d_n S_n (S_n - 1),$$

where

$$c_n = \frac{1}{a_n} \left[\frac{n+1}{3(n-1)} - \frac{1}{a_n} \right]$$

and

$$d_n = \frac{1}{a_n^2 + b_n} \left[\frac{2(n^2 + n + 3)}{9n(n-1)} - \frac{n+2}{na_n} + \frac{b_n}{a_n^2} \right]$$

(d) Does your value of D lead you to accept or reject the null hypothesis of a standard Wright-Fisher model?

3. Consider a 2-allele Wright-Fisher model with population size N . If the two alleles are A and a, let

A_t = number of A's in the population in generation t , and

X_t = fraction of A's in the population in generation t .

(a) Calculate $Var(A_{t+1}|A_t = k)$.

(b) Explain why $Var(A_{t+1}|X_t = \frac{k}{N}) = Var(A_{t+1}|A_t = k)$.

(c) What is $Var(A_{t+1}|X_t = x)$?

(d) Find $Var(X_{t+1}|X_t = x)$. (Your answer should involve $1/N$.)

(e) Use (d) to define the variance effective size, $N_e^{(v)}$. (Be sure to specify what you did to get this from (d).)