

## Stat 536 Homework 5

Due: 10/3/08

In this question, you will derive an estimator for effective population size in populations where only a small fraction of the total population is allowed to reproduce. This situation is typical in domesticated animals, where reproduction is limited to a few individuals, while the rest serve another purpose. We will assume that the breeding pool is drawn at random from the larger population, with equal numbers of males and females (which is probably not so typical for domesticated animals). Assuming the breeding pool is randomly mating according to HWE assumptions, the effective population size is equal to the size of the breeding pool  $N_b$ , but let's suppose we cannot observe the size and want to estimate it. The key observation that leads to an estimator is that there will be an *excess* of heterozygotes in the offspring, and this excess heterozygosity can be easily measured. In what follows, assume no selfing and distinguish the sexes so that offspring must have one mother and one father.

1. Recall the homework 2 question that claimed when allele frequencies differ in the sexes by amount  $\delta$ , then heterozygote frequency in the offspring is

$$H = 2pq + \frac{\delta^2}{2}$$

where  $p$  and  $q = 1 - p$  are the allele frequencies in the offspring. Previously,  $\delta$  was viewed as fixed. In this context  $\delta$  is a random variable that depends on the sampling process that created the breeding population, so to use this equation, we'll substitute

$$H = 2pq + \frac{E(\delta^2)}{2}$$

Suppose the total population has allele frequencies  $p_0$  and  $q_0$ . By considering the variance in allele proportions in small, random subpopulations (breeding populations) of size  $N_b$  (males and females both number  $\frac{N_b}{2}$ ), show that the above equation can be rewritten as

$$H = 2pq + \frac{p_0q_0}{N_b} \tag{1}$$

2. From equations in the lecture one can show that the ratio of "heterozygosity" across two subsequent generations is

$$\frac{h_{t+1}}{h_t} = \left(1 - \frac{1}{2N_e}\right)$$

where  $N_e$  is the effective population size for any non-Wright-Fisher population. Relate  $h_t$  and  $h_{t+1}$  to expected heterozygosity under HWE and use this observation to eliminate  $p_0$  and  $q_0$  from the right-hand-side of eq. (1), leaving

$$H = 2pq \frac{N_b + \sqrt{N_b^2 + 1}}{N_b - 1 + \sqrt{N_b^2 + 1}}$$

[Note:  $N_e \neq N_b$ . What is  $N_e$  for your population that breeds via a breeding population of size  $N_b$ ?]

3.  $H$  is the observed heterozygosity.  $2pq$  is the HWE predicted heterozygosity. Define

$$D = \frac{H - 2pq}{2pq}$$

as a measure of heterozygote excess and show that

$$N_b = \frac{1}{2D} + \frac{1}{2(D+1)}$$

4. Find an estimate of  $N_b$  and confidence intervals by bootstrap for the following data. Please check your bootstrap datasets carefully and deal sensibly with any problems encountered. Dealing “sensibly” means you solve the problem somehow and you at least make mention of it in your solutions.

|     |     |     |
|-----|-----|-----|
| AA  | Aa  | aa  |
| 135 | 502 | 363 |

5. Critique this estimator. Why might it not be a very good estimator?