Homework April 9, 2015 (to be returned on April 21)

We consider a randomly dioecious species with different numbers of males, N_m, and females, N_f. The purpose of this exercise is to derive the expression for the effective population size, N_e.

We define by $\mathcal{G}$ the probability that a random individual is homozygous; by $P_{q,rs}$ the probability that a locus picked at random from two individuals, one of sex r and the other of sex s, descend from the same parent of sex q (the index 1 is for males and 2 for females); $g_{r,s}$ is the probability that two loci picked randomly from individuals of sex r and sex s are the same allele, i.e. identical by state.

1) Derive the equations:

$$\begin{aligned} \mathcal{G}' &= g_{1,2} \\ g'_{r,s} &= \frac{1}{4} \times \left[P_{1,rs} \left(\mathcal{G} + (1-\mathcal{G}) \times \frac{1}{2} \right) + (1 - P_{1,rs})\, g_{11} \right] \\ &+ \frac{1}{4} \times \left[P_{2,rs} \left(\mathcal{G} + (1-\mathcal{G}) \times \frac{1}{2} \right) + (1 - P_{2,rs})\, g_{22} \right] + \frac{1}{2} \times g_{1,2}\,, \end{aligned} \tag{1}$$

where the primes indicate the values at the next generation, after random mating.

We are interested in the limit where both populations N_m and N_f are large, yet possibly different so that $P_{q,rs}$ is a small parameter $O(\epsilon)$ that we shall use as described below. In Wright-Fisher models, homozygosity approaches its asymptotic value (as generations increase) with the rate $1 - 1/2N_e$ so that we can determine the effective population size from the dominant eigenvalue of the dynamics.

2) Define the vector $\boldsymbol{V}^T = (\mathcal{G}, g_{11}, g_{12}, g_{22})$ and write the dynamics (1) as

$$\boldsymbol{V}' = (A_0 + A_1)\, \boldsymbol{V} + \text{const.}\,, \tag{2}$$

where the index of the matrices A indicates the order in ϵ, i.e. A_0 does not contain $P_{q,rs}$ and A_1 is linear in those terms.

We are interested in the asymptotic behavior at long times of homozygosity, i.e. in the eigenvalues

$$(A_0 + A_1)\, \boldsymbol{V} = \lambda \boldsymbol{V}\,. \tag{3}$$

Write down the explicit expressions for the two matrices A_0 and A_1; show that A_0 has eigenvalue unity and determine the corresponding left and right eigenvectors, $\boldsymbol{L}$ and $\boldsymbol{R}$. It will be convenient to normalize them as $\boldsymbol{L} \cdot \boldsymbol{R} = 1$.

3) Use the results above and the small value of ϵ to find perturbatively the dominant eigenvalue $\lambda = 1 - 1/2N_e + \ldots$ of (3). The corresponding right eigenvector is sought as $\boldsymbol{V} = \boldsymbol{R} + \boldsymbol{v}_1 + \ldots$. The eigenvalue equation (3) at the order ϵ^0 is $A_0 \boldsymbol{R} = \boldsymbol{R}$, which is satisfied by construction. Find the eigenvalue equation at the order ϵ and take its scalar product with $\boldsymbol{L}$ to show that

$$-\frac{1}{2N_e} = \boldsymbol{L}^T A_1 \boldsymbol{R}, \tag{4}$$

and write down the explicit expression in terms of the $P_{q,rs}$.

4) For simple generation processes $P_{q,rs} = \frac{1}{N_q}$. Show then that

$$N_e = \frac{4N_m N_f}{N_m + N_f}. \tag{5}$$

5) Show that $N_e \leq N = N_m + N_f$. Estimate N_e/N for $N_m = 0.1 N_f$.