Diffusion process models in population genetics, Problems for HT 2007

Alison Etheridge

1. (a) Consider a population, evolving in discrete generations, in which each individual, independently, leaves behind a Poisson number of offspring with parameter μ. Show that if we condition the total population size to be fixed equal to N in each generation, then the conditioned population evolves as a Wright-Fisher model.

   (b) What happens if we subdivide the population into two types? Those of type a have a Poisson number of offspring with parameter $(1+s)\mu$, those of type A have a Poisson number of offspring with parameter μ. If again we condition the total population size to be constant, how does the frequency of a-alleles evolve from generation to generation? (This is one route to the Wright-Fisher model with selection.)

2. Suppose that a population which is divided into two genetic types labelled a and A evolves according to the *Wright Fisher model with selection* so that if the population size is N and there are currently i individuals of type a, then the number of type a individuals in the next generation is binomial with N trials and success probability $\frac{(1+s)i}{(1+s)i+N-i}$. Suppose that the coefficient s is small enough that $Ns \sim \mathcal{O}(1)$ and write $\sigma = Ns$. Mimic the calculation in Example 1.8 to find an approximation for the expected time until absorption of the chain if the current frequency of a-alleles is p. (You may leave your answer as an integral.)

3. In a Moran model with selection and mutation, at exponential rate $\binom{N}{2}$ a pair of genes is sampled from the population. One dies and the other gives birth. If the pair picked are of the same type, then each individual is equally likely to be chosen to reproduce, but if they are of different types then with probability $\frac{1+s}{2}$ it is the type a individual that reproduces. To account for mutation, with probability u_1 the offspring of a type a individual is of type A (otherwise it is of type a) and with probability u_2 the offspring of a type A individual is type a. Find the transition probabilities of the embedded chain.

4. (a) Add a mutation step to the neutral Wright-Fisher model for a population of two genetic types, a and A, so that during reproduction each type a offspring mutates to a type A with probability u and similarly each type A offspring mutates to a type a with probability u. This model has a stationary distribution (although it is not known explicitly). Find a recursion for the expected homozygosity in the population, that is find an expression for the probability that a sample of two individuals taken from the population in generation $n+1$ have the same allelic type in terms of the corresponding probability for a sample of size two from the population in generation n. Let $n \to \infty$ in this recursion to find an expression for the expected homozygosity, F, when the population is in steady state. Now let $\theta = 2Nu$ and let $N \to \infty$ to show that in the infinite population limit

$$F = \frac{1+\theta}{1+2\theta}.$$

(b) Now consider the corresponding Moran model. Incorporate mutation into the lookdown picture by declaring that when a lookdown event takes place, with probability u a mutation takes place and it is equally likely to be on either of the levels involved. Writing as usual $\theta = 2Nu$ show that as $N \to \infty$ the kth level in the lookdown process is hit by mutation events according to a Poisson process with rate $\frac{\theta}{2}$. (In other words the time that you must wait until the next mutation event on level k is exponential with parameter $\frac{\theta}{2}$.) Recalling that taking a sample of size two from the population is equivalent to looking at the first two levels in the lookdown picture, show that the probability of homozygosity in the infinite population limit is the same as we obtained above.

5. In Example 1.9 of lectures, we wrote down an expression for the mean time to absorption for the embedded chain in a neutral Moran model. The aim of this question is to derive that expression.

   (a) Write down a recurrence relation for the mean time, t_i^j, that the chain spends in state j, starting from i, before absorption.

   (b) Solve this recurrence and deduce the expression for k_i, the mean time to absorption starting at i that we wrote down in lectures.

6. We say that a population is *random-mating* if each individual chooses a partner with whom to reproduce at random. Suppose that we are modelling such a population in which a particular gene has just two alleles, a and A, and that the population is so large that genotypic frequencies can be assumed to evolve independently.

   (a) First suppose that the population is *monoecious*. This just means that there is no distinction between males and females. The initial proportion of the different genotypes is

aa	aA	AA
X	$2Y$	Z

   for positive X, Y, Z. Show that after one generation of random mating the frequencies of the different genotypes is in *Hardy-Weinberg equilibrium*. That is it is of the form x^2, $2x(1-x)$, $(1-x)^2$ where $x = (X+Y)^2$ and these proportions are preserved under further random matings.

   (b) Now consider a *dioecious* population. Suppose that the frequencies of genetic types among the male and female subpopulations is X_M, $2Y_M$, Z_M and X_F, $2Y_F$, Z_F respectively. Show that after two generations of random mating, the frequencies of the three genotypes among males and females remain equal in all further generations. For this reason we often make the modelling simplification of ignoring the existence of two sexes.

7. A Galton Watson branching process is a discrete time Markov chain, $\{Z_n\}_{n\geq 1}$, which is often used to model the growth of a population. The evolution is simple. Each individual leaves behind a random number of offspring in the following generation, according to some distribution, independently of all other individuals. Suppose that the mean number of offspring of each individual is a and the variance is σ^2 and write Z_0 for the initial population size.

   (a) What is the expected population size after N generations?

   (b) If we are modelling a very large population, whose size at time zero is NX_0 for some large N and $a \approx 1 + \frac{\mu}{N}$, then find a diffusion approximation for the population size at time Nt in units of size N. Compare to the diffusion obtained in Example 3.8 of lectures. [Hint: In deriving the diffusion approximation, take $\delta t = \frac{1}{N}$.]

8. In Example 3.8 of lectures we found the expected population size at time t in a diffusion approximation to a population growth model. If the initial poulation size is $M(0)$, find the variance of the population size at time t.

9. Consider a Fisher-Wright diffusion model for the frequencies in a population with two types a and A, which are selectively neutral, in the case when there is strictly positive mutation from a to A and vice versa. Find an expression for the stationary density and sketch it for (a) small and (b) large mutation rates.

10. Find a diffusion approximation for the Moran model with selection and mutation of Question 3 when $u_1 \approx \frac{\mu_1}{2N}, u_2 \approx \frac{\mu_2}{2N}$ and $s \approx \frac{\sigma}{N}$.

    The Moran model can be extended to incorporate *density dependent* selection. So for example, if the current allele frequency is p, then if the pair chosen from the population to reproduce consists of one type a and one type A individual then with probability $\frac{1+s(p)}{2}$ it is the type a individual that splits in two. Assuming that $s(p) \approx \frac{\sigma(p)}{N}$ for some continuous function σ on $[0,1]$, find the corresponding diffusion approximation.

11. (a) Find an expression for the expected homozygosity in a neutral Fisher-Wright diffusion model with mutation.

    (b) Write down an expression for the homozygosity if we introduce selection so that the drift in the diffusion becomes $\sigma p(1-p) - \mu_1 p + \mu_2(1-p)$. Does the homozygosity increase or decrease relative to the neutral value for small σ? What happens as $\sigma \to \infty$?

12. For the neutral Wright-Fisher diffusion with mutation from type A to a, but not vice versa, for which values of the mutation parameter will the proportion of a alleles never be zero?

13. Consider the Fisher-Wright diffusion with selection but no mutation, so the drift is $\alpha p(1-p)$.

    (a) Find an expression for the probability that the diffusion leaves the interval $(0,1)$ through $p=1$ (so that the allele becomes fixed in the population). What happens to this probability as $\alpha \to \infty$?

    (b) Find an expression for the expected time until the diffusion started from $p \in (0,1)$ exits the interval $(0,1)$ and show that it is $\mathcal{O}(\frac{\log \alpha}{\alpha})$ as $\alpha \to \infty$.

    (c) Fix $\epsilon \in (0,1)$. Show that the expected time to exit $(0, 1-\epsilon)$ started from p is $\mathcal{O}(\frac{1}{\alpha})$ as $\alpha \to \infty$.

    (d) Finally consider the deterministic logistic growth curve

    $$\frac{dp(t)}{dt} = \alpha p(t)(1-p(t)).$$

    Show that the time taken to increase from a given level $p(0)$ to $1-\epsilon$ is also $\mathcal{O}(\frac{1}{\alpha})$ in this case.

14. Suppose that a mutant allele of type a arises in an otherwise pure A population. Assume that the a allele has a selective advantage and its frequency is modelled by a Fisher-Wright diffusion. If a selective sweep takes place (that is the proportion of a alleles increased to one), what is the *conditioned* diffusion followed by the frequency of a alleles?

15. Suppose that a *deleterious* mutation a arises in an otherwise pure population of A alleles. If the relative fitnesses of a and A alleles are $1-s:1$ for some positive s, and $s \sim \mathcal{O}(\frac{1}{N})$, write down a diffusion approximation for the frequency of a alleles.

    Suppose now that (in your diffusion model) the a allele eventually fixes. Find the conditioned diffusion for the allele frequencies. How does it compare to the one obtained in the previous question for a selectively *advantageous* mutant?

16. Take the Fisher-Wright diffusion model for a selectively advantageous allele (with no mutation) and condition it to exit the interval $(0,1)$ at 1.

    (a) Suppose that the unconditioned diffusion has scale function $S(x)$ and speed density $m(x)$. Show that the conditioned diffusion has scale function $\tilde{S(x)} = -\frac{1}{S(x)}$ and speed density $\tilde{m(x)} = S^2(x)m(x)$.

    (b) Find an expression for the time spent by the conditioned diffusion in a small interval about its starting point x_0 and show that this is invariant under the mapping $x_0 \mapsto 1 - x_0$.

17. Suppose that a diffusion has generator L (corresponding to drift $\mu(x)$ and variance $\sigma^2(x)$) and speed density $m(x)$. The generator $\hat{L}$ of its time reversal with respect to the speed measure can be obtained formally by an integration by parts:

$$\hat{L}f(y)g(y)m(y)dy = \int f(y)Lg(y)m(y)dy.$$

    Perform this integration (ignoring boundary terms) to find $\hat{L}$.

18. Let $Y_1, Y_2, \ldots, Y_k$ be independent positive random variables with probability density function

$$g_\epsilon(y) = \frac{y^{\epsilon-1}e^{-y}}{\Gamma(\epsilon)}.$$

    Show that $Y = Y_1 + \cdots + Y_k$ has a Gamma distribution with parameter $K\epsilon$ and that the vector $\mathbf{p}$ with components $p_i = \frac{Y_i}{Y}$ has the Dirichlet distribution. Show further that $\mathbf{p}$ and Y are independent.

19. In the infinitely many alleles model (at stationarity),

    (a) What is the expected number of distinct alleles in the population with frequency greater than a?

    (b) What is the expected number of distinct alleles in a sample of size n?

    (c) What is the probability that a random sample of size n contains the oldest allele?

    (d) What is the probability that the oldest allele in a random sample of size n is the jth oldest in the population?

    [Hint: for the first two use the frequency spectrum, for the last two the GEM distribution.]