

Stochastic Models in Mathematical Genetics

Problem Sheet 4

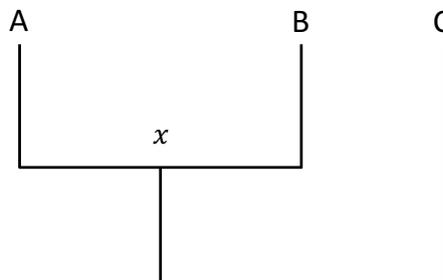
Class 4, Michaelmas Term 2020

1. Consider the number of distinct allelic types H_n in a sample of size n whose history is described by the coalescent with recombination parameter ρ and mutation parameter θ . Mutations occur according to the infinite-sites model. Explain why depending on the two parental allelic types, ancestral recombination events may, or may not, create a novel type in the set of ancestors. By extending the argument from the lecture notes for weeks 1 to 4, to consider a process where lineages are “lost” either by coalescence, mutation, or recombination backward in time, show that the expectation of H_n satisfies

$$1 + \sum_{i=2}^n \frac{\theta}{i-1 + \theta + \rho} \leq \mathbb{E}[H_n] \leq 1 + \sum_{i=2}^n \frac{\theta + \rho}{i-1 + \theta + \rho}.$$

Deduce that as the sample size becomes large, $\frac{\mathbb{E}[H_n]}{\log n}$ is bounded above and below by non-zero constants.

2. In a coalescent with recombination model, suppose a recombination event at a position x is the first to occur in the history of a sample of two sequences:



Choose δ small enough so no other recombination event occurs in $[x - \delta, x + \delta]$. Show that a necessary and sufficient condition for the marginal tree in $[x - \delta, x)$ to be identical (in expected time also) to that in $(x, x + \delta]$ is that the two ancestral lineages A, B produced by the recombination coalesce with each other before either coalesces with C. What is the probability of this?

- (a) Suppose now that a recombination event at x occurs while $j > 1$ ancestors of a sample of size n remain at position x . Deduce that the marginal trees immediately flanking the event are identical if, and only if, the two lineages ancestral to the recombination event coalesce with one another before either coalesces with any other sequence.
- (b) Show that the probability P_j of no tree change satisfies a system of equations recursive in j . Solve this system to obtain a closed form expression for the probability of no tree change.
- (c) Derive the expected number of marginal tree changes along $[0, 1]$ if the overall recombination rate is ρ . Does your answer depend on the recombination rate being constant along the sequence? How does the number of tree changes behave asymptotically, as $n \rightarrow \infty$?

3. Consider a Wright-Fisher model with *selection* and *mutation*. There are two types, A and a in a population of constant size $2N$. Population members choose their parents independently at random, with relative probability 1 of choosing each a type parent and $1 + s$ of choosing each A type parent. *After* a parent is chosen, mutations from type a to type A occur with probability μ_A and from A to a with probability μ_a .

- (a) If the number of individuals in the population who are type A in generation k is Z_k , explain why Z_{k+1} has a Binomial distribution and find the parameters of this distribution in terms of Z_k .
- (b) Mimicking the argument in the lectures, show that as $N \rightarrow \infty$ while $\gamma = 2Ns$, $\theta_A = 4N\mu_A$, $\theta_a = 4N\mu_a$ remain fixed, and setting $X_t = Z_{\lfloor 2Nk \rfloor} / (2N)$, the process X_t will converge to a diffusion process with infinitesimal variance and mean

$$a(x) = x(1 - x)$$

$$b(x) = \gamma x(1 - x) + \theta_A(1 - x)/2 - \theta_a x/2.$$

4. Consider a general Wright-Fisher model where in a constant size population of $2N$ chromosomes, there are two types A and a , and in the k th generation, the proportion of chromosomes carrying the A type is X_k , where $X_0 = x$. Now suppose that in generation $(k + 1)$, each chromosome independently chooses a parent, and that the probability that parent is of type A is given by

$$P(X_k) = X_k + \frac{b(X_k)}{2N} + o(N^{-1}).$$

Show that as $N \rightarrow \infty$, the Markov chain describing the proportion of chromosomes carrying the A type converges to a diffusion process limit after an appropriate rescaling of time, and give the diffusion and drift parameters for the limit process.

5. Consider the diffusion limit of the Wright-Fisher process with mutation and selection considered in Question 3 (i.e., consider the diffusion with this infinitesimal mean and variance).

- (a) Write down the generator of the diffusion.
- (b) Consider the problem of estimating the fixation probability $h(x)$ of the A allele, which has initial frequency x in the population at $t = 0$. (Assume that fixation of either the A or a alleles is certain in finite time.) Write down a differential equation, and boundary conditions, satisfied by $h(x)$.
- (c) Solve the differential equation to give the appropriate fixation probability in terms of a ratio of integrals involving the parameters γ , θ_A , θ_a . Show that these integrals are not defined if either $\theta_A, \theta_a > 1$.
- (d) Supposing $\theta_A, \theta_a < 1$, what is the limiting value of the fixation probability of the A allele as $\theta_a \rightarrow 1$, while θ_A remains fixed? What is the limiting value of the fixation probability of the A allele as $\theta_A \rightarrow 1$, while θ_a remains fixed?
- (e) Adjust your analysis in (5(c)) to obtain the probability that the A allele reaches frequency z before reaching frequency y , where $0 < y < x < z < 1$. If $\theta_a \geq 1$, show that the A allele cannot reach fixation.