

Example Sheet 4

Renewal theory and applications.

1. A cinema in Camford shows one movie at a time, which is changed after a time uniformly distributed in $[0, 3]$ (months). Purchasing the rights to a new movie costs £10,000. How much have they spent after 5 years?

Suppose that a given customer goes to the movies at rate once every two months, but only walks into the cinema if the movie is less than one month old. In 5 years, how many movies has she seen? What if the time to change a movie is exponentially distributed with mean 1.5 months?

2. Suppose that the lifetime of a car is a random variable with density function f . Mr. Smith buys a new car as soon as the old one breaks down or reaches T years. A new car costs $\mathcal{L}c$, while an additional $\mathcal{L}a$ is incurred if the car breaks down before T . In the long-run, how much does Mr. Smith spend on his cars per unit time?

3. Let $(X_t)_{t \geq 0}$ be a renewal process with interarrival times having the Gamma $(2, \lambda)$ distribution. Determine the limiting excess distribution. Determine also the renewal function $m(t)$.

4. A barber takes an exponentially distributed amount of time, with mean 20 minutes, to complete a haircut. Customers arrive at rate 2 per hour, but leave if both chairs in the waiting room are full. Make a Markov chain model on the state space $\{0, 1, 2, 3\}$. Find the equilibrium distribution. Determine the queue length at equilibrium. Using Little's formula, find the average waiting time in the system of a customer.

Population genetics.

5. Let X be a Markov chain and let $A \subset S$. Let T_A be the hitting time of A . Let $h(x) = \mathbb{P}_x(T_A < \infty)$ and let τ_y denote the total time spent at y by the chain before hitting A . Show that $\mathbb{E}_x(\tau_y | T_A < \infty) = [h(y)/h(x)]\mathbb{E}_x(\tau_y)$.

Deduce that in a Moran model, conditionally on fixation of an allele a present initially in i individuals,

$$\mathbb{E}(\tau | \text{fixation}) = N - i + \frac{N - i}{i} \sum_{j=1}^{i-1} \frac{j}{N - j}$$

where τ is the fixation time of allele a .

6. Let $(\Pi_t, t \geq 0)$ denote Kingman's (infinite) coalescent. Let τ denote the coalescence time of Π . Find $\mathbb{E}(\tau)$ and give an expression for $\mathbb{E}(e^{-q\tau})$ for all $q > 0$.

7. Let $(\Pi(t), t \geq 0)$ denote Kingman's n -coalescent, $n \geq 1$. Show directly the consistency property, meaning that $(\Pi(t)|_{[n-1]}, t \geq 0)$ is an $(n - 1)$ -coalescent. *Hint:* fix π a partition of $[n - 1]$ and let π' be a partition obtained by merging two blocks of π , and consider the conditional probability

$$\mathbb{P} \left(\Pi(t + h)|_{[n-1]} = \pi' \mid (\Pi(s), s \leq t), \Pi(t)|_{[n-1]} = \pi \right).$$

8. Consider a Moran model on N individuals. Assume that each individual undergoes a mutation at rate u , and let $\theta = uN$. After a mutation, the type of this individual changes to something completely new (these are the assumptions of the infinite alleles model). Sample two individuals at random (without replacement). Compute the probability of *homozygosity*, i.e., the probability that they have the same allelic type. [Hint: consider the coalescence of their ancestral lineages]. Compare with Ewens sampling formula.

9. A geneticist samples the DNA of two individuals in a population, and wishes to compare the samples. Let Δ be the number of pairwise differences in the two samples: this is the number of base pairs which are different: e.g., in the sequences

$$\begin{array}{cccccc} A & \underline{T} & T & C & \underline{G} & G & A \\ A & \underline{C} & T & C & \underline{C} & G & A \end{array}$$

we have $\Delta = 2$

Assuming individuals are subject to neutral mutations at rate u , show that for all $k \geq 0$,

$$\mathbb{P}(\Delta = k) = (1 - p)^k p, \text{ where } p = \frac{1}{1 + \theta},$$

with $\theta = uN$. Compute $\mathbb{E}(\Delta)$.

10. Write down Ewens' sampling formula (assume that the mutation rate is $\theta > 0$) and the formula for $\mathbb{E}(K_n)$, the expected number of allelic types. For $n = 2$, make a table consisting of all values that (a_1, a_2) can take and their respective probabilities. Check that they sum to one, and check that the formula for $\mathbb{E}(K_n)$ is correct. Looking at this table, can you compute the probability of *homozigosity*, i.e., the probability that two individuals sampled from a population have the same allelic type?

Repeat for $n = 3$ and $n = 4$, each time checking that the probabilities add up to 1 and that the formula for $\mathbb{E}(K_n)$ is correct.

11. Let $\theta > 0$ and let Π denote a random partition of $\{1, \dots, n\}$ having the law of the Ewens sampling formula with parameter θ . Let $K_n = \#$ blocks of Π . Show that given $\{K_n = k\}$, the distribution of Π does not depend on θ . (In statistical terms, K_n is a sufficient statistic for θ .)

12. In 1976, Singh, Lewontin and Felton made a study of the xanthine dehydrogenase locus of 146 individuals from *D. pseudoobscura*, and found the following:

$$a_1 = 7, a_2 = 3, a_3 = 7, a_5 = 2, a_6 = 2, a_8 = 1, a_{11} = 1, a_{68} = 1.$$

Assuming that the population size is roughly 1,000,000, what do you think is the mutation rate per individual? (Keep in mind that this species is diploid so the number of chromosomes is $N = 2,000,000$).

13. (a) Let $(\Pi_t, t \geq 0)$ denote Kingman's n -coalescent. Let L_n denote the total length of the branches of the tree: that is, if τ_i is the first time that there are i lineages, $L_n = \sum_{i=2}^n i(\tau_i - \tau_{i-1})$. Show that $\mathbb{E}(L_n) \sim 2 \log(n)$ and that $\text{Var}(L_n) \leq C$ for some constant $C > 0$. Deduce that

$$\frac{L_n}{\log n} \rightarrow 2$$

in probability.

(b) Consider now the Moran model, and individuals are subject to a mutation at rate $u > 0$ according to the rules of the infinite sites model. Then show that if S_n denotes the number of segregating sites (number of DNA bases where there is variability in the sample) then if $\theta = uN$ is fixed, we have

$$\frac{S_n}{\log n} \rightarrow \theta$$

in probability.