

Introductory seminar on
“Mathematical Population Genetics”
winter term 2017/18

Prof. Joachim Hermisson
Ilse Höllinger

5 Genetic drift, neutral theory and the coalescent

5.1 Coalescent in the Moran model

In the lecture, the coalescent has been derived based on the Wright-Fisher model. However, the coalescent is by no means restricted to the precise assumptions of the Wright-Fisher model. For example, it can also be derived under the Moran model. In the continuous-time version of the model, birth-death events occur in the population at rate N such that (on average) N events occur during a time unit of length one, which defines a generation in the Moran model. At each birth-death event, one random individual is chosen to reproduce and one individual is chosen to die (with replacement).

Consider a sample of n individuals taken from a population of N individuals in an haploid species.

1. Calculate the coalescence rate $p_{c,1}^{(n)}$ for a common ancestor event occurs the n lineages of the sample.
2. Compute the expected time to the most recent common ancestor for a sample of size n in the Moran model. Also calculate the variance of this time (both on the per-generation time scale).
3. Relate the Moran model and the Wright-Fisher model with a coalescent rate of $\frac{n(n-1)}{2N}$ per generation for N haploid individuals. What is the effective population size of a Moran model with N haploid individuals? Assuming that N time steps in the Moran model relate to one generation in the Wright-Fisher model, we want to obtain a relationship between the coalescence effective population size $N_e^{(c)}$ for the Moran model and the Wright-Fisher model.

5.2 Identity by descent of the X chromosome

In mammals females are carrying two X-chromosomes and males are determined by two different gonosomes, one X and one Y. A daughter always inherits one X chromosome from her mother and one from her father. A son always gets

its unique X chromosome from his mother. The probability that two randomly chosen X-linked alleles (alleles located on the X-chromosome) are descendants from the same ancestor in the previous generation is given by

$$p_{c;1} = \frac{1}{2} \left(\frac{N_f}{N_m} + \frac{N_m}{N_f} \right) \frac{1}{2N_f + N_m} \quad (1)$$

1. Prove this identity.
 *** *Hint*:*** You need to treat males and females in both the pffspring and the parent generation sperately.
2. What is the coalescence effective population size $N_e^{(c)}$ for a sex-ratio of 1:1?

5.3 Neutral fixation probability and effective population size

We have introduced the neutral fixation probability for an ideal Wright-Fisher population. How do we need to adjust the results if we consider natural populations and various consepts of effective population size?

Calculate the neutral fixation probability for:

- an arbitrary offspring distribution.
- fluctuating population size in time $N(t)$.
- Seperate sexes: How does the neutral fixation probability depend on the number of breeding males N_m and breeding females N_f ?

5.4 Relationship with ancestral Human lineages

Modern Humans and Neanderthals co-occurred in Europe until 30 000 years ago. Let us assume that modern Humans and Neanderthals have the same generation time of 20 years and a (joint) effective population size of $N = 3\,400$ individuals. In the recent years, researchers were able to sequence a complete Neanderthal genome from fossil bones.

1. Under the (unrealistic) assumption that modern Humans and Neanderthals freely interbred, what is the probability that (for a given gene copy that is treated like a single locus) either you or your neighbor in class is more closely related to the known Neanderthal sequence than to each other?
2. Now consider a larger sample of modern Humans of size n . What is then the probability that the modern Human coalescent has not yet fully coalesced 30 000 years ago? Evaluate this for a sample sie of $n = (2,5,10)$.
3. Assume now that k modern human lineages are left 30 000 years ago. What is then the probability that (at least) one of the modern Human lineages is more closely related to the sequenced Neanderthal lineage than to (some of) the other modern human lineages?

For this exercise we need the following results, which can also be found in the lecture notes, Chapter 5:

- The waiting time T_k before the first coalescent event in a sample of size k is exponentially distributed with parameter $\binom{k}{2}$ (time is counted in units of $2N_e$ generations).

$$f_{T_k}(t) = \binom{k}{2} e^{-\binom{k}{2}t} \quad (2)$$

- The time $T_{MRC A, n}$ to the most recent common ancestor of a sample of size n has the following distribution:

$$f_{T_{MRC A, n}}(t) = \sum_{k=2}^n \left(\binom{k}{2} e^{-\binom{k}{2}t} \prod_{l=2, l \neq k}^n \frac{l(l-1)}{l(l-1) - k(k-1)} \right) \quad (3)$$