AMAURY LAMBERT, Université Pierre et Marie Curie (Paris 6), Paris

The mutation and allele frequency spectra for the coalescent point process

A *splitting tree* is the genealogical tree with edge lengths associated with a branching population whose individuals have i.i.d. lifespans with general distribution, during which they give birth at constant rate, independently, to copies of themselves. For any fixed time t, we show that individuals alive at t can be ranked in such a way that the coalescence times between consecutive individuals are i.i.d. with specified distribution. The ranked sequence of these branches is called a *coalescent point process*, and encodes all the information about the genealogical structure of the population alive at time t.

When individuals are given DNA sequences, there are two quantities of interest for a sample of n DNA sequences belonging to distinct individuals: the number S_n of *polymorphic sites* (sites at which at least two sequences differ), and the number A_n of *different haplotypes* (distinct sequences). It is standard to assume that mutations arrive at constant rate θ (on germ lines), and never hit the same site on the DNA sequence. For the celebrated Wright–Fisher model with large population size, it is well known that both S_n and A_n grow like $\theta \log n$ as the sample size n grows.

We study the mutation pattern associated to coalescent point processes. Here, S_n and A_n grow *linearly* as n grows, with explicit speed. In addition, we study the *frequency spectrum* of the sample, that is, the numbers of polymorphic sites/haplotypes carried by k individuals in the sample. These numbers are shown to grow also linearly with sample size, and we provide simple explicit formulae for mutation frequencies and haplotype frequencies.