

Bayesian nonparametric analysis of Kingman's coalescent

STEFANO FAVARO¹, Shui Feng², Paul A. Jenkins³

¹*University of Torino and Collegio Carlo Alberto, Italy*

²*McMaster University, Canada*

³*University of Warwick, UK*

Abstract

Kingman's coalescent is one of the most popular model in population genetics. It describes the genealogy of a population whose genetic composition evolves in time according to the Wright-Fisher model, or suitable approximations of it belonging to the broad class of Fleming-Viot processes. Ancestral inference under Kingman's coalescent has had much attention in the literature, both in practical data analysis, and from a theoretical and methodological point of view. Given a sample of individuals taken from the population at stationarity, most contributions have aimed at making frequentist or Bayesian parametric inference on quantities related to the genealogy of the observed sample. In this talk we propose a Bayesian nonparametric predictive approach to ancestral inference. That is, under the prior assumption that the composition of the population evolves in time according to a neutral Fleming-Viot process, and given the information contained in an initial sample of m individuals taken from the population at stationarity, we estimate quantities related to the genealogy of an additional unobservable sample of size $m' \geq 1$. For instance, how many non-mutant lineages would I expect a time t ago if I enlarged my initial sample by $m' \geq 1$? How many of these non-mutant lineages are associated to the rare genetic types in the initial samples? In the context of ancestral inference, these questions are of great interest because they relate directly to the speed of evolution via the rate of turnover of alleles. As a by-product of our analysis we introduce a class of Bayesian nonparametric estimators which can be thought of as Good-Turing type estimators for ancestral inference. The proposed approach is illustrated through an application to genetic data.