
Modeling population structure under hierarchical Dirichlet processes

Lloyd Elliott*¹

¹Oxford University – United Kingdom

Abstract

We propose a Bayesian nonparametric model to infer population admixture, extending the hierarchical Dirichlet process to allow for correlation between loci due to linkage disequilibrium. Given multilocus genotype data from a sample of individuals, the proposed model allows inferring and classifying individuals as unadmixed or admixed, inferring the number of subpopulations ancestral to an admixed population and the population of origin of chromosomal regions. Our model does not assume any specific mutation process, and can be applied to most of the commonly used genetic markers. We present an MCMC algorithm to perform posterior inference from the model and we discuss some methods to summarize the MCMC output for the analysis of population admixture. Finally, we demonstrate the performance of the proposed model in a real application, using genetic data from the EDAR gene, which is considered to be ancestry-informative due to well-known variations in allele frequency as well as phenotypic effects across ancestry. We also conduct a simulated experiment and show that our algorithm outperforms parametric methods.

*Speaker