

Title: Nonparametric Bayesian Methods for Genetic Inference

The problem of inferring the haplotype phases, linkage disequilibrium patterns, chromosomal recombination events, and population structures from genetic polymorphism data is essential for understanding the origin and characteristics of genome variations. Statistical genetic methodologies developed so far mostly address these problems separately using specialized models ranging from the coalescence and mixture models for haplotype inheritance, to HMMs and renewal processes for recombination, but many of these approaches ignore the inherent uncertainty in the genetic complexity (e.g., the number of genetic founders of a population) of the data, and the close statistical and biological relationships underlying different problems. In this talk, I discuss our recent work on a new class of nonparametric Bayesian models motivated by the Dirichlet process, for solving haplotype phasing, LD modeling and demographic inference in an OPEN ancestral space based on a unified statistical framework. This approach offers to a compact and natural representation of the population structure and inheritance processes underlying genetic polymorphism, and leads to competitive performance on a wide range genetic inference tasks. I will present experimental results on both simulated and HapMap SNP data, and compare our method with existing ones on haplotype inference, recombination hotspot prediction, and structural map estimation.

If time permits, I will also briefly introduce our recent work on maximum entropy discrimination Markov networks, and temporal exponential random graph models.