“Bayesian Semi parametric Models for Genetic Association Studies of Unrelated Individuals in the Presence of Population Structure”

Nicholas M. Pajewski, PhD Candidate
Division of Biostatistics
Medical College of Wisconsin

There has been considerable discussion about the potential deleterious effects of population structure on association studies. When a sample of unrelated individuals consists of individuals from subpopulations differing in disease risk and allele frequencies, estimates of disease association with a particular locus can be exaggerated or attenuated. Numerous approaches have been proposed to address this without resorting to family-based designs. Foremost amongst these is the Bayesian model-based clustering approach of Pritchard and coauthors (2000, 2001), who proposed modeling the subpopulations, classifying individuals accordingly, and essentially pooling to produce stratified inference. As the focus of research has moved towards genome wide association studies, recent emphasis is on computationally efficient methods of adjustment (Price et.al. 2006, Epstein et.al. 2007, Kimmel et.al. 2007). While these methods are quickly able to handle investigations of hundreds of thousands of markers, they do not necessarily provide a framework for developing more complex models of association, such as those containing environmental covariates or interactions.

We consider a unified Bayesian semiparametric framework for association studies using Dirichlet Process Mixture (DPM) models. The DPM model appropriately integrates out population structure in making association inference. It also uses a nonparametric sparsity prior to incorporate the prior belief that most loci are not phenotypically associated (Dunson et.al. 2008). We illustrate the DPM construction within the case of quantitative traits and case-control designs, and also discuss how it can be easily extended to other outcomes (ordinal, time to event, etc.) Effectiveness of the proposed model and related Markov chain Monte Carlo computations is demonstrated via a simulated study of a quantitative trait based on the International HapMap Project.