Two-group mixture models for multiple testing

Michele Guindani, University of California, Irvine

The simultaneous testing of multiple hypotheses is common to the analysis of high-dimensional data sets. The two-group model, first proposed in (Efron, 2004), identifies significant comparisons by allocating observations to a mixture of an empirical null and an alternative distribution. In the Bayesian nonparametrics literature, many approaches have suggested using mixtures of Dirichlet Processes in the two group model framework. In this talk, we first investigate employing instead mixtures of two-parameter Poisson Dirichlet Processes (2PPD), and show how they provide a more flexible and effective tool for large-scale hypothesis testing. Our model further employs non-local prior densities to allow separation between the two mixture components. We compare the performances of our method for large-scale inference in a simulation study and illustrate its use on a case-control microbiome study of the gastrointestinal tracts in children from underdeveloped countries who have been recently diagnosed with moderate to severe diarrhea. We then consider a problem of model selection over time and space. More specifically, we define a zero-inflated conditionally identically distributed species sampling prior for clustering data to either a null or an alternate process. Spatial dependences are accounted for by means of a Markov random field prior, which allows informing the selection based on inferences conducted at nearby locations. We illustrate the performances of the modeling approach in an application to the monitoring of pneumonia and influenza (P&I) mortality, to detect influenza outbreaks in the continental United States.