

A Spike and Slab centering Distribution in Dirichlet Process Mixture Models for Gene expression

Sinae Kim¹, David B. Dahl² and Marina Vannucci³

¹ Department of Biostatistics, University of Michigan, Ann Arbor, MI, USA

² Department of Statistics, Texas A&M University, College Station, TX, USA

³ Department of Statistics, Rice University, Houston, TX, USA

Model-based clustering methods using Dirichlet process (DP) mixture models have been proposed to exploit clustering for increased sensitivity in multiple hypothesis testing. Rather than yielding a probability of a hypothesis for each object, existing methods can only provide a ranking of the objects by their evidence for a particular hypothesis. In this work, we adapt the framework on [1] to accommodate point (i.e., sharp) null hypotheses. For that, we use a spike and slab distribution which is a mixture of both a point-mass distribution and a continuous distribution as the centering distribution for the Dirichlet process prior. The method yields probabilities that genes follow the hypotheses of interest, whether those hypotheses be sharp or not. These probabilities not only rank the genes, but their interpretation is very natural and have a variety of uses, for example, in discovery of genes with a specified expected number of false discoveries. We apply our method in gene expression context and show how to simultaneously infer gene clustering and differential gene expression.

References

- [1] Dahl DB, MO Q., Vannucci, M. (2008) Simultaneous Inference for Multiple Testing and Clustering via a Dirichlet Process Mixture Model. *Statistical Modelling: An International Journal* 2:165-175