

Objective Bayesian graphical model selection with Bayes' factors based on test statistics and non-local alternative priors

Adarsh Joshi

Texas A&M University and M D Anderson Cancer Center

Joint work: Jianhua Hu, Valen E. Johnson

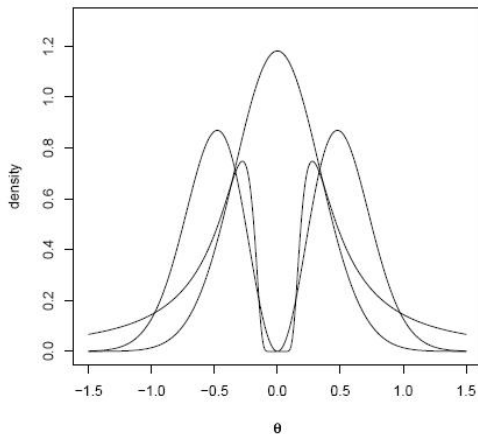
May 15, 2009

Bayes Factors Based on Likelihood Ratio Statistic

- Asymptotic null distribution of the LRS is χ_d^2
- Asymptotic distribution of LRS for *local* alternative is a non-central $\chi_d^2(\delta' \mathbf{C}_{11} \delta)^*$
- Specify prior on the non-centrality parameter \Rightarrow **marginal alternative distribution of LRS**
- Bayes factor obtained as the ratio

* Davidson and Lever 1970

Local and Non-local priors



Source: Johnson and Rossell 2008

Clustering Model: Microarray Data

- Subsets of interacting variables called clusters
- Model Constraints:
 - Only up to third-order interactions permitted
 - Each gene is in at most one cluster
- Two nice properties
 - Interactions in a cluster estimated and tested using only the marginal table of that cluster
 - Estimates and test statistics for marginal tables obtained readily

SNP Data: Considerations

- $\sim 550\text{K}$ SNP locations for ~ 1000 patients
- Clustering model is not adequate
- Clusters can not share variables; tend to form very large clusters which will be rejected due to Occam's Razor
- Model conditional independencies \implies Graphical Models
- Need efficient computing

SNP Data: Graphical Models

- Cliques can share nodes
- Constraints can maintain a rich class of models while drastically reducing computational complexity
- Programming in FORTRAN
- Preselect SNPs in a careful manner