

Bayesian Methods in Bioinformatics

Veera Baladandayuthapani, Ph.D. Candidate
Texas A & M University

Abstract:

This work is directed towards the development of Bayesian statistical methods for analyzing high dimensional data from biological and genetic experiments, and is presented in two parts:

The first part focuses on new methods to analyze data from an experiment using rodent models to investigate the effect of diet on colon cancer development. In our experiment, various regulatory proteins (biomarkers), known to be precursors of early colon carcinogenesis, are assayed on animals in different diet/treatment groups at various time-points. The biological goals are to understand the relationship among the biomarkers' up and down-regulation as well as complex interplays among the diets/treatments and the various biomarkers at a cellular level. What lends a special structure to the data is that the responses are inherently functional in nature, consisting of observed profiles over a spatial variable nested within a two-stage hierarchy, which we call hierarchical functional data. The problem is cast within the class of hierarchical Bayesian functional models and the question of understanding these up and down-regulatory relationships is viewed as modeling the functional variations and correlations.

The second part focuses on modeling DNA microarray data. Microarray technology enables us to monitor the expression levels of thousands of genes simultaneously and hence to obtain a better picture of the interactions between the genes. In order to understand the biological structure underlying these gene interactions, we present a hierarchical nonparametric Bayesian model based on Multivariate Adaptive Regression Splines (MARS) to capture the functional relationship between genes and also between genes and disease status. The novelty of the approach lies in the attempt to capture the complex nonlinear dependencies between the genes which could otherwise be missed by linear approaches. The Bayesian model is flexible enough to identify significant genes of interest as well as model the functional relationships between the genes. The effectiveness of the proposed methodology is illustrated on leukemia and breast cancer datasets.