
Sundez Keles, M.A.University of California-Berkeley
Division of Biostatistics***Binding site detection by incorporating structural information***

Rapid developments in the sequencing technology in combination with microarray technology enable scientists to study gene expression and gene regulation at a genomic level. The identification of transcription factor binding sites using statistical modeling is an important component of such large scale studies. Among the methods developed for this purpose are the ones that utilize finite multinomial mixture models with specific implementations as MEME and Gibbs Motif Sampler. Although there are continuing attempts to enhance these models, there has been little or no effort to incorporate transcription factor specific information into these models. In this talk, I will present an extension of such mixture models that takes into account the characteristics of the true DNA binding sites. In particular, the parameter space of binding sites (which are represented as position weight matrices) are constrained. These constraints could be as simple as specific base conservations at specific locations or more global constraints on the information profile. I will present simulation studies and data analysis that illustrate the utility of such a model in identifying real binding sites.

Tuesday, February 25th
4:00 p.m.Room 1221
Computer Sciences / Statistics Building
1210 West Dayton Street