Boston University Statistics Seminar Series

Inference in Perturbation Models, Mixtures and Spatial Scan Processes

Ramani Pilla
Department of Statistics and Department of Biology
Case Western Reserve University
Friday, February 3, 2005, 11:00am - noon  (NOTE REVISED DAY/TIME!)
Mathematics and Computer Science (MCS) Building, Room 149
111 Cummington Street, Boston
Tea and Cookies at 3:30pm in MCS 153

Abstract: In this talk, a general class of models referred to as perturbation models are introduced. These models are described by an underlying “null” model that accounts for most of the structure in data while a perturbation accounts for possible small localized departures. For instance, in the context of finite mixture models, the null model represents a mixture with \( m \) components and the perturbation model represents additional components. In the spatial scan process context, the null density accounts for the background or noise whereas the perturbation searches for an unusual region such as a tumorous tissue in mammography or a target in an image recognition problem. We derive a new test statistic for detecting the presence of perturbation and show that the asymptotic distribution of the test statistic is equivalent to the supremum of a Gaussian process over a high-dimensional manifold (e.g., curve, surface etc.) with boundaries and singularities. A technique for approximating the quantiles of the test statistic via the Hotelling-Weyl volume-of-tube formula is presented.

Fitting mixture models and performing statistical inference on the results is an important but a very challenging problem. A long-pending fundamental question is: how many mixture components? The asymptotic null distribution of the likelihood ratio test statistic is highly complex and very difficult to simulate from in practice. Building on the perturbation theory, inferential methods are developed to address the problem of testing for an arbitrary number of components from smooth families of distributions, including multivariate mixtures. The resulting theory has broad applications including astronomy, astrophysics, particle physics, bioinformatics and genetics. We illustrate the theory in the context of a model problem from high-energy particle physics wherein the goal is to distinguish a signal from random fluctuation in data with a high probability. More information on the particle physics and other application problems is available at http://stat.case.edu/~pillar/PRL/PRL.htm .

For directions and maps, please see http://math.bu.edu/research/statistics/statseminar.html.