

STANFORD UNIVERSITY
DEPARTMENT OF STATISTICS
DEPARTMENTAL SEMINAR

4:15 p.m., Tuesday, January 24, 2006
Sequoia Hall Room 200
(Cookies at 3:45 in 1st Floor Lounge)

Nancy Zhang
University of California, Berkeley

Modified BIC for Change-point Problems with Applications to Array-CGH Data

We study the problem of estimating the number of change-points in a data series that is hypothesized to have undergone abrupt changes. First, we focus on the scenario of independent Gaussian data points with changing mean values, and then generalize to the Poisson process with changing rate parameter as well as general exponential families. This can be viewed as a problem in model selection, where the dimension of the model grows with the number of change-points assumed. However, the classic Bayes Information Criterion (BIC) can not be applied because of irregularities in the likelihood function. By asymptotic approximation of the Bayes Factor, we derive the Modified BIC that is theoretically justified for the change-point models that we study.

An example of application as well as a source of inspiration for the Gaussian model is the analysis of array comparative genomic hybridization (array-CGH) data. Array-CGH measures the number of chromosome copies at each genome location of a cell sample, and is useful for finding the regions of genome deletion and amplification in tumor cells. The Modified BIC statistic is tested on array-CGH data sets and compared to existing methods. Variations to the basic change-point model that are inspired by array-CGH data is also discussed.