



The University of Chicago
Department of Statistics

Seminar Series

NANCY R. ZHANG

Department of Statistics
Stanford University

**Simultaneous Scans of Multiple Sequences
for Shared Variant Intervals**

MONDAY, April 6, 2009 at 4:00 PM
133 Eckhart Hall, 5734 S. University Avenue

Refreshments following the seminar in Eckhart 110.

ABSTRACT

We examine the statistical problem of simultaneous detection in multiple sequences shared signals that may occur in only a fraction of the sequences. The motivation for this study comes from the biological application of detecting recurrent intervals of copy number variation in multiple samples of DNA. We work within the context of the following general statistical model: For each sequence $i = 1, \dots, N$ and position $t = 1, \dots, T$, we assume mutually independent and normally distributed random variables y_{it} with mean values μ_{it} and variances σ_i^2 . The null hypothesis assumes that for every sample i , $\mu_{it} = \mu_i$, whereas the alternative assumes that there exists $J \subseteq \{1, \dots, N\}$ and $1 \leq \tau_1 < \tau_2 \leq T$, such that for each $i \in J$, $\mu_{it} = \mu_{i0} + \delta_i I_{\{\tau_1 < t \leq \tau_2\}}$ where $\delta_i \neq 0$. We propose several statistics for this testing scenario, and derive approximations for their significance level and power. Finally, we discuss the computational schemes and results in applying these simultaneous scan statistics in the detection of DNA copy number variation.

Joint work with David Siegmund.

Key words and phrases: Scan statistics, change-point models, segmentation, DNA copy number, mixture models.