THE UNIVERSITY OF CHICAGO

Department of Statistics
STATISTICS COLLOQUIUM

MICHAEL NEWTON

Departments of Statistics and of Biostatistics and Medical Informatics

University of Wisconsin-Madison

Two Problems in High-Dimensional Statistics: A Specific One on the Analysis of Gene Function, and a General One on Ranking and Selection

MONDAY, April 14, 2014, at 4:00 PM

133 Eckhart Hall, 5734 S. University Avenue

Refreshments following the seminar in Eckhart 110

ABSTRACT

I will investigate two inference problems. The data structure in the first case is a list of genes identified by some genome-wide assay, and the problem is to characterize the functional content of the list using known functional properties of all genes. By modeling the gene-list in terms of latent activities of biological functions, one obtains a compelling multivariate procedure to address the task. The available method ignores what turns out to be an important constraint in the high-dimensional discrete parameter space. I will discuss this constraint and the inference improvements that follow when the constraint is respected. In the second part of the talk I will discuss the general problem of ranking and selection. Data form a collection of unit-specific point estimates coupled with unit-specific quantities that inform the uncertainty in the underlying unit-specific parameters (e.g., standard errors). When uncertainty varies substantially among units, different ranking/selection schemes can give very different inferences about the true ranking. I introduce and investigate a generic empirical Bayesian quantity—the r-value—which leads to rankings that are optimal in the sense of maximizing the joint probability that a unit's parameter is among the top alpha-fraction of parameters and the unit is inferred to be in that top fraction. The gene-set problem is addressed in a paper with Zhishi Wang, Qiuling He, and Bret Larget. The ranking/selection findings are part of work with Nicholas Henderson.