

STATISTICS COLLOQUIUM

FRIDAY, FEBRUARY 29, 2008

TALK: 4:00 PM — SCIENCE CENTER RM. 309

RECEPTION: 5:15 PM — SCIENCE CENTER, 7TH FLOOR

Wenguang Sun
Department of Biostatistics
University of Pennsylvania

A New Framework for Large-Scale Multiple Testing: Compound Decision Theory
and Data-Driven Procedures

ABSTRACT

With recent advances in technology, it has become increasingly common in practice to test a large number of hypotheses simultaneously. In this talk, I formulate the large-scale multiple testing problem in a compound decision theoretic framework and discuss oracle and asymptotically optimal data-driven procedures for false discovery rate (FDR) control. My presentation is divided into three parts: the first part develops oracle and adaptive compound decision rules for independent tests, the second part discusses simultaneous testing of grouped hypotheses, and the third part considers large-scale multiple testing under dependency. A key goal is to show that conventional FDR procedures, which are mostly p -value based, can be substantially improved by our new data-driven procedures that adaptively exploit the distributional, structural and external information of the sample. I also discuss results of simulations studies, as well as microarray data analyses from a human immunodeficiency study and a breast cancer study, for illustration of our methods and their comparison with alternative procedures.