

Optimality in Large-Scale Multiple Testing

Tony Cai
The Wharton School
University of Pennsylvania

Abstract

Large-scale multiple testing is commonly used in many applications, from bioinformatics, to machine learning, to econometrics. Conventional multiple testing procedures are based on thresholding the ordered p-values. In this talk, we consider large-scale multiple testing from a compound decision theoretical point of view by treating it as a constrained optimization problem. The solution to this optimization problem yields an oracle procedure. A data-driven test procedure is then constructed to mimic the performance of the oracle and is shown to be asymptotically efficient. In particular, the results show that, although p-value is appropriate for testing a single hypothesis, it fails to serve as the fundamental building block in large-scale multiple testing. Examples will also be discussed.

Monday, December 10, 2007
Friend Center Room 008
4:30 PM

