

STATISTICS COLLOQUIUM
JOINT
COMPUTER SCIENCE

WEDNESDAY, SEPTEMBER 07, 2011
TALK: 4:00 PM – MAXWELL DWORKIN G115

“Statistics and Computation in the Age of Massive Data”

Michael I. Jordan
University of California, Berkeley
Department of Statistics and Department of EECS

ABSTRACT

There are many issues remaining to be addressed, or even formulated, at the interface of statistics and computation. One way to capture the current state of affairs is the following: If we view data as a resource, how can it be that in many practical problems of interest we find ourselves embarrassed by being given too much data? Our inferential procedures use polynomial amounts of time and space but that doesn't suffice; we need to be able to guarantee that on a fixed computational budget the statistical risk decreases as the number of data points grows (without bound).

In this talk I present three vignettes that describe different lines of attack on massive data problems: one involving the bootstrap, another involving matrix completion algorithms and the third involving phylogenetic analysis in the regime of large numbers of taxa.

Speaker: Michael I. Jordan is the Pehong Chen Distinguished Professor in the Department of Electrical Engineering and Computer Science and the Department of Statistics at the University of California, Berkeley. His research in recent years has focused on Bayesian nonparametric analysis, probabilistic graphical models, spectral methods, kernel machines and applications to problems in statistical genetics, signal processing, computational biology, information retrieval and natural language processing. Prof. Jordan is a member of the National Academy of Sciences, a member of the National Academy of Engineering and a member of the American Academy of Arts and Sciences. He is a Fellow of the American Association for the Advancement of Science. He has been named a Neyman Lecturer and a Medallion Lecturer by the Institute of Mathematical Statistics. He is a Fellow of the ACM, the IMS, the IEEE, the AAAI and the ASA.