THE UNIVERSITY OF CHICAGO

Department of Statistics

STATISTICS COLLOQUIUM

MATT TADDY

Booth School of Business
University of Chicago

Big Data and Bayesian Nonparametrics

MONDAY, November 2, 2015, at 4:00 PM
Eckhart 133, 5734 S. University Avenue
Refreshments following the seminar in Eckhart 110.

ABSTRACT

Big Data is often characterized by large sample sizes, high dimensions, and strange variable distributions. For example, an e-commerce website has 10-100s million observations weekly on a huge number of variables with density spikes at zero and elsewhere and very fat tails. These properties—big and strange—beg for nonparametric analysis. We revisit a flavor of distribution-free Bayesian nonparametrics that approximates the data generating process (DGP) with a multinomial sampling model. This model then serves as the basis for analysis of statistics—functionals of the DGP—that are useful for decision making regardless of the true DGP. The ideas will be illustrated in the indexing of treatment effect heterogeneity onto user characteristics in digital experiments and in analysis of decision trees employed in fraud prediction. The result is a framework for scalable nonparametric Bayesian decision making on massive data.

For further information and about building access for persons with disabilities, please contact Kirsten Wellman at 773.702.8333 or send email (kwellman@galton.uchicago.edu). If you wish to subscribe to our email list, please visit the following website: https://lists.uchicago.edu/web/arc/statseminars.