

HARVARD UNIVERSITY
DEPARTMENT OF STATISTICS



Tel. 617-495-5496
Fax. 617-496-8057
statdept@stat.harvard.edu
<http://www.stat.harvard.edu/>

Science Center 7th Fl.
One Oxford Street
Cambridge, MA 02138

STATISTICS COLLOQUIUM

MONDAY, FEBRUARY 22, 2016
TALK: 4:15 PM — SCIENCE CENTER 300H
RECEPTION: 3:50 PM

“Randomized Sketch Methods with Statistical Applications: A Survey”

Jelani Nelson

Computer Science

**John A. Paulson School of Engineering & Applied Sciences
Harvard University**

ABSTRACT

In statistics, randomness is typically viewed as a source of noise that corrupts data. Meanwhile in computer science, randomness is often viewed as a resource that algorithm designers can take advantage of to design algorithms that are simple, fast, and/or have low memory footprint. These 'randomized algorithms' work on any data set with no model assumptions, with the caveat that they may output erroneous results with very low probability (over the random bits flipped by the algorithm, and not based on any randomness in the input data).

In particular, recent work coming out of the theoretical science community has developed randomized 'sketch' methods for data processing. These methods compress data to use (often exponentially less) memory while still preserving enough information to support data updates and specific queries. The sketched data can then be processed using less memory and/or more quickly. In this survey talk, I will discuss many recent works which have used randomized sketch methods to speed up algorithms for classical statistical problems, such as (non)linear regression, principle component analysis, and estimating statistical leverage scores.