

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, OCTOBER 6, 2014

TALK: 4:15 PM — SCIENCE CENTER RM. 705

RECEPTION: 3:50 PM — SCIENCE CENTER, 7TH FLOOR

“Scalable Bayes”

David Dunson
Department of Statistical Science
Duke University

ABSTRACT

Bayesian methods have great promise in big data sets, but this promise has not been fully realized due to the lack of scalable computational methods. Usual MCMC and SMC algorithms bog down as the size of the data and number of parameters increase. For massive data sets, it has become routine to rely on penalized optimization approaches implemented on distributed computing systems. However, in scientific applications, it is crucial to obtain a good characterization of uncertainty and not simply a point estimate. To scale up Bayes inference to huge sample sizes, we propose directions ranging from running MCMC in parallel for subsets of the data and then combining to approximate MCMC algorithms that replace expensive sampling steps with approximations. In both cases, we obtain strong theoretical guarantees and substantial practical gains in applications. We also provide initial ideas and results for scaling up to data containing massive numbers of observations per subject motivated by neurosciences and genomics.