Stanford University
Department of Statistics

DEPARTMENTAL SEMINAR

4:15pm, Tuesday, January 28, 2014
Sequoia Hall Room 200
Cookies served at 3:45pm, 1st floor Lounge.

Speaker: Art Owen, Stanford University

Title: Data, big and small

Abstract:
Suppose that we have a small data set from exactly the population we want to sample, and with well-measured variables. Additionally there is a much larger data set from approximately the target population with variables similar to those in the target set but perhaps not as well-measured. We present a linear regression method for predictions on a small data set, making cautious use of the larger one. Our method fits linear regressions to both data sets while penalizing the prediction differences of the two models for the smaller population. The result is a shrinkage method similar to those used in small area estimation. Our main result is a Stein-type finding for Gaussian response values. When there are five or more predictors and 10 or more error degrees of freedom, it becomes inadmissible to ignore the large data set. We also develop plug-in and AICc based methods for tuning the penalty parameter. We focus on an L2 penalty but have some results for L1 in the simple location setting.

This is joint work with Aiyou Chen and Minghui Shi of Google Inc.