Stanford University
Department of Statistics

Departmental Seminar

4:15pm, Tuesday, September 27, 2011
Sequoia Hall Room 200

Cookies served at 3:45pm, 1st Floor Lounge.

Speaker: Art Owen, *Stanford University*

Title: Bootstrapping r-fold tensor data

Abstract:

The famous Netflix data is a sparsely sampled table with rows for customers and columns for movies (or vice versa). Both movies and rows are naturally modeled as random effects. Bootstrapping such data is problematic: no proper bootstrap can exist, according to a theorem of Peter McCullagh. Resampling rows and columns independently is effective though slightly conservative.

Computerized data gathering frequently produces data sets with three-way or even higher order data tables. We present a bootstrap for such tensor valued data. Our version uses independent weights instead of multinomial ones. It remains mildly conservative. Poisson weights are close to the original bootstrap, but binary weights have computational and statistical advantages. Under certain conditions a single bootstrap replicate suffices to give a variance estimate. We apply our method to show that Facebook users in the UK make longer comments than those in the US when using their phone. The opposite pattern holds for comments made via computer.

This is joint work with Dean Eckles, Facebook.