

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 19, 2015
TALK: 4:15 PM — SCIENCE CENTER 300H
RECEPTION: 3:50 PM

“Moment Based Methods for Large Crossed Unbalanced Random Effects Models”

[Art B. Owen](#)

Department of Statistics
Stanford University

ABSTRACT

Modern electronic commerce generates enormous data sets with an unbalanced crossed random effects structure. The factors are things like customer IDs, URLs, product IDs (SKUs), cookies, IP addresses, news stories, tweets, and query strings, among others. These variables could be treated as plain categorical variables that just happen to have a large number of levels. But many of the specific levels are evanescent. It is more realistic to treat them as random effects and seek conclusions that apply to the distributions from which these variables have been sampled. The result is a model with a tangle of correlations that we find defeats maximum likelihood as well as MCMC approaches to Bayesian computation. Both end up with costs that are superlinear in the size of the data.

We propose a plain method of moments approach for settings like this where the data size has outstripped the computational resources. Moment methods are easy to parallelize and they make only very weak assumptions. We get estimates for variance components, and plug-in formulas for their variances using sample estimates of the relevant kurtoses. When there are N observations spanning R distinct rows and C distinct columns we keep the cost to $O(N)$ time and $O(R+C)$ space, using upper bounds where necessary.

This is joint work with Katelyn Gao, Stanford University.