# THE UNIVERSITY OF CHICAGO

Department of Statistics

## STATISTICS COLLOQUIUM

## ART B. OWEN

Department of Statistics
Stanford University

## Moment Based Estimation and Inference for Very Large Linear Mixed Effects Models

**FRIDAY, February 3, 2017, at 12:15 PM**

Math-Stat 112 (Stevanovich Center), 5727 S. University Avenue

## ABSTRACT

Mixed effects models with crossed random effects are perhaps the most suitable statistical model for some large e-commerce data sets. Unfortunately, the cost to compute maximum likelihood estimates in these models scales as $O(N^{3/2})$ when there are $N$ observations. A similar problem and rate arises for Bayesian estimation. For instance, the Gibbs sampler takes $O(N^{1/2})$ iterations to converge leading to a total cost once again of $O(N^{3/2})$. We view $O(N)$, or possibly $O(N \log(N))$, as a hard upper bound on the acceptable complexity of algorithms for large data sets.

Our motivating example is a regression data set from Stitch Fix to model product ratings made on a ten point scale. In addition to the fixed effects describing important predictors, there is a random effect by which multiple ratings from the same customer are correlated. Similarly, multiple ratings on the same garment are also correlated.

We suspect that the generalized least squares estimates cannot be fit in $O(N)$ time. We propose a method of moments algorithm to fit a linear model in $O(N)$ time. The model we choose accounts for just one of the two sources of correlation, whichever appears to be most important. It is thus statistically inefficient but computationally efficient. We then construct standard errors for the fixed effect coefficients that do account for both sources of correlation. Correlations can make the effective sample size much smaller than the nominal one. One of our computed variance estimates is more than 100 times what a naive IID calculation would yield.

Joint work with Katelyn Gao (Stanford University). We thank Brad Klingenberg of Stitch Fix for data and helpful discussions.