

## Statistics and Data Science Seminar

### *New Approaches to Fast Approximate Bayesian Nonparametric Inference*

George Karabatsos (UIC)

**Abstract:** Dirichlet process (DP) mixture models, as well as models with mixture distribution assigned a general Bayesian nonparametric (BNP) prior distribution on the space of probability measures, are widely-applied and flexible models that can provide reliable statistical inferences complex data. For such Bayesian mixture models, in practice, posterior inferences are usually conducted using MCMC, which however, is prohibitively slow for large data sets. Also for such models, prior specification can be non-trivial in practice. As alternatives to MCMC, I consider two new approaches to fast and approximate BNP inference for large data sets. First, I show that if the ordinary least-squares (OLS) estimator of the linear regression coefficients is specified as a functional of the DP posterior distribution, then this functional has posterior mean given by an observation-weighted ridge regression estimator, with ridge (coefficient shrinkage) parameter given by the DP precision parameter; and has a heteroscedastic-consistent posterior covariance matrix. This result is based on the multivariate delta method applied to prior-informed bootstrap distribution approximation to the DP posterior. Second, I consider an approximation to the BNP (infinite) mixture model that I introduced and studied in several articles, defined by ordinal regression mixture weights. The approximate model is defined by a (large) finite mixture, with each component distribution multiplied by a histogram bin indicator function. I show that posterior inference with this approximate BNP model can be conducted by iteratively-reweighted least squares estimation for the mixture weight parameters, and least-squares estimation

Wednesday, October 5 at 4:00 PM in SEO 636

for the component densities, all involving computations that are orders of magnitude faster than MCMC-based inference of the original mixture model. This is also true for a version of the approximate model that is defined by an ordinal regression of DPs. I illustrate the two approximate BNP methods through the analysis of real data sets.

*TBA*

Wednesday, October 5 at 4:00 PM in SEO 636