Department of Mathematical Sciences welcomes

Haiying Wang
University of Connecticut

September 7, 2018

Hosted by:
Prof. Fei Tan

Tea begins at 3:00
in LD 259

Research Topic
begins at 3:30
in LD 229

More Efficient Estimation for Logistic Regression with Optimal Subsamples

ABSTRACT:

Subsampling is a practical technique to extract useful information from massive data. Wang et. al. (2017) developed an Optimal Subsampling Method under the A-optimality Criterion (OSMAC) for logistic regression that samples more informative data points with higher probabilities. This method uses inverses of optimal subsampling probabilities as weights in the log-likelihood function to remove bias. This reduces contributions of more informative data points. In this paper, to improve the estimation efficiency based on an OSMAC subsample, we propose an unweighted estimator is more efficient. In addition, we develop a new algorithm based on Poisson subsampling, which does not require to approximate the optimal subsampling probabilities all at once. This is computationally advantageous when available random-access memory is not enough to hold the full data. Interestingly, asymptotic distributions also show that Poisson subsampling produces more efficient estimator if the sampling rate, the ratio of the subsample size to the full data sample size, does not converge to 0. We also obtain the unconditional asymptotic distribution for the estimator based on Poisson subsampling.

ABOUT THE SPEAKER:

Haiying Wang is an Assistant Professor in the Department of Statistics at the University of Connecticut. He was an Assistant Professor in the Department of Mathematics and Statistics at the University of New Hampshire from 2013 to 2017. He obtained his PhD from the Department of Statistics at the University of Missouri in 2013 and his M.S. from the Academy of Mathematics and Systems Science, Chinese Academy of Sciences in 2006. His research interests include informative subdata selection for big data, model selection, model averaging, measurement error models, and semi-parametric regression.

- IUPUI - WORLD RENOWNED RESEARCHERS EXPLORING MATHEMATICS, STATISTICS & THEIR APPLICATIONS

MATH
COLLOQUIUM