

**MICHIGAN STATE UNIVERSITY**  
Department of Statistics and Probability

## COLLOQUIUM

**Ping Ma**  
University of Georgia

### **Leveraging for big data regression**

Tuesday, October 13  
10:20 a.m. - 11:10 am  
Refreshments 10:00 am  
C405 Wells Hall

#### **Abstract**

Rapid advance in science and technology in the past decade brings an extraordinary amount of data, offering researchers an unprecedented opportunity to tackle complex research challenges. The opportunity, however, has not yet been fully utilized, because effective and efficient statistical tools for analyzing super-large dataset are still lacking. One major challenge is that the advance of computing technology still lags far behind the exponential growth of database.

In this talk, I will discuss an emerging family of statistical methods, called leveraging, to facilitate scientific discoveries using current computing resources. Leveraging methods are designed under a subsampling framework, in which one samples a small proportion of the data (subsample) from the full sample, and then performs intended computations for the full sample using the small subsample as a surrogate. The key of the success of the leveraging methods is to construct nonuniform sampling probabilities so that influential data points are sampled with high probabilities. These methods stand as the very unique development of their type in big data analytics and allow pervasive access to massive amounts of information without resorting to high performance computing and cloud computing.

*To request an interpreter or other accommodations for people with disabilities, please call the Department of Statistics and Probability at 517-355-9589.*