



ANNOUNCES A  
COLLOQUIUM

**Dr. Yanyuan Ma**

*Texas A&M University  
Department of Statistics*

will speak on

## **Efficient Semiparametric Distribution Estimation in Mixed Samples**

**Time: 3:00 – 4:00 PM**

**Date: Thursday, November 11, 2010**

**Place: Alter Hall 744**

### **Abstract**

We study efficient nonparametric estimation of distribution functions of several scientifically meaningful sub-populations. The data consists of a mixture of these sub-populations where the sub-population identifier is missing. Only the probabilities of each observation belonging to the sub-populations are available. The problem arises from kin-cohort study and quantitative trait locus (QTL) analysis where the interest often lies in estimating the cumulative distribution function of a trait given a specific genotype. However, the genotypes in the relatives in the kin-cohort study or the QTL genotypes in a QTL study are not directly observed. The distribution of the trait outcome is then a mixture of several genotype-specific distributions. We derive the complete class of consistent estimators and identify the efficient one which reaches the semiparametric efficiency bound. We construct the efficient estimator using a simple procedure that remains consistent even if several components of the estimator are mis-specified. In addition, we inspect two commonly used nonparametric maximum likelihood estimators (NPMLEs) in these problems and show the surprising results that the NPMLE in one form is highly inefficient, while in the other form is inconsistent. We provide simulation procedures to illustrate the theoretical results and demonstrate the proposed methods through two real data examples.