



The University of Mississippi
Department of Mathematics

Statistics Seminar

Dr. Xin Dang

University of Mississippi

Title: Gini Distance Correlation and Feature Selection

11:00-11:50 am, Wednesday, Oct 2, 2019
Hume Hall 321

Abstract: Big data is becoming ubiquitous in the biological, engineering, geological and social sciences, as well as in government and public policy. Building an interpretable model is an effective way to extract information and to do prediction. However, this task becomes particularly challenging for the scenario of big data, which are large scale and ultra-high dimensional with mixed-type features. A common practice in tackling this challenge is to reduce the number of features under consideration via feature selection by choosing a subset of features that are “relevant” and useful. The work in this talk aims at proposing new dependence measure in feature selection. The features having strong dependence with the response variable are selected as candidate features. We propose a new Gini correlation to measure dependence between categorical response and numerical feature variables. Compared with the existing dependence measures, the proposed one has both computational and statistical efficiency advantages that improve the feature selection procedure and therefore the resulting prediction model.

This is joint work with Dao Nguyen and Yixin Chen

All graduate students are encouraged to attend.