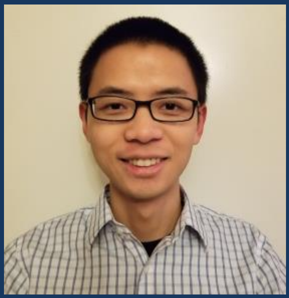
Department of Applied and Computational Mathematics and Statistics Colloquium



Leo Duan

Department of Statistical Sciences

Duke University

Probabilistic Modeling of Structure in Complex Data

In modern applications, the collected data are complex and high dimensional, but the relatively small sample size creates enormous difficulties for building interpretable models. Prior information about the structure (e.g. sparsity,latent low dimensionality) has been proven extremely useful for effective learning of the structure. In this talk, I focus on two aspects of my research: (1) assimilating prior information about parameter constraints; (2) utilizing only a few assumptions for robust clustering of complex data. For (1), the work aims to significantly expand the modeling choices in constraints and tractable distributions, while allowing uncertainty about constraints. The idea is to replace sharp constraints with a Constraint Relaxation prior, creating a shrinkage towards the constrained space. This enables convenient use of posterior sampling toolbox, such as Hamiltonian Monte Carlo, for automatic computation in broad models. For (2), the goal is to avoid the need for specifying mixture kernel for modeling each cluster, while retaining the ability for uncertainty quantification in clustering. I propose to use a pseudo-likelihood based approach, modeling the pairwise distance matrix derived from the data. Through a pair of carefully designed densities, it provides an automatic estimation of the parameters, controlling the distance truncation and the degree of uncertainty. In both works, dramatic gains are shown in real data applications, such as a neuroscience study for finding structure in a population of brain networks.

Thursday, January 11, 2018

4:30 PM – 5:30 PM

127 Hayes-Healy Center

Colloquium Tea   4:00 PM to 4:30 PM   154 Hurley Hall