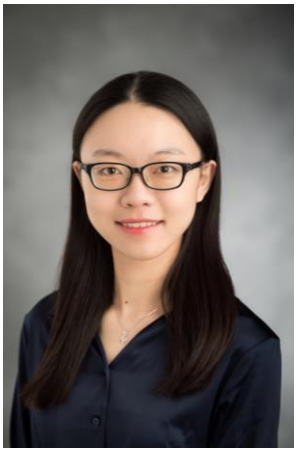
RUTGERS
School of Arts and Sciences

Department of Statistics



Yuqi Gu
Department of Statistics
Columbia University

Generalized Grade of Membership Models: Fast and Sharp Estimation under Local Dependence

Wednesday, November 20th, 2024
12:00 PM
110 Frelinghuysen Road, Hill Center, Room 116

Zoom Meeting: Meeting ID: 969 0606 4706
Password: 745339

https://rutgers.zoom.us/j/96906064706?pwd=ZklvbExpRVBJQ3c5dUhhYTFuR2ZrZz09

Light refreshments will be served in Hill 452, 11:20am

Abstract:

Mixed membership models are popular individual-level mixture models widely used in various fields. This work is motivated by mixed membership models for multivariate categorical data, which is called the Grade of Membership (GoM) model. The rich modeling power of the GoM model comes with challenging estimation issues, especially for high-dimensional polytomous data. Such data take the form of a three-way (quasi-)tensor, with many subjects responding to a large number of items each with a different number of categories. Popular Bayesian inference or maximum likelihood estimation methods are not computationally efficient and lack theoretical guarantees in high dimensions. In contrast, we propose a fast spectral method to flatten the three-way tensor into a ``fat'' matrix and exploit the singular subspace geometry based on matrix SVD for estimation. We introduce the Generalized-GoM models, which encompass a wide range of data distributions and locally dependent noise. For this model family, we establish that our spectral method delivers sharp entrywise estimation error bounds for all parameters. Moreover, we develop a new two-to-infinity singular subspace perturbation theory for arbitrary locally dependent noise under flexible distributions, which is of independent interest. Simulations and applications to data in political voting, population genetics, and single-cell genomics demonstrate our method's superior performance.

Bio: Yuqi Gu is an Assistant Professor of Statistics at Columbia University and also a member of the Data Science Institute. She obtained her PhD in Statistics from the University of Michigan in 2020 and did a one-year postdoc at Duke University before joining Columbia. Her research centers around unobserved latent structures widely present in statistics, machine learning, psychometrics, and other applications. More specifically, she is interested in identifiable deep generative models with latent representations, high-dimensional statistics with latent structures, and latent variable modeling in educational and psychological measurement.