UGA Statistics Colloquium Series

Dr. Ruoxuan Xiong

Assistant Professor in the Department of Data and Decision Sciences (DDS) at Emory University

Thursday, Oct. 22, 2026, 4:30-5:30PM, Caldwell Rm. 204
Refreshments from 4:00-4:30pm in Cohen Room of Brooks

Learning and Inference Under Nonrandom Missingness with AI-Generated Measurements

Abstract:

Missing data are pervasive in modern data systems and are frequently not missing at random. For example, in service settings, a customer's decision to submit a rating may depend on the rating they would have provided. Similarly, clinical notes, reviews, and other unstructured data are often selectively available, influenced by latent patient conditions or user preferences.

This talk considers how AI-generated measurements can improve identification and inference under nonrandomly missing outcomes. Large language models can convert auxiliary unstructured data, such as customer interaction histories, into measurements that inform about missing outcomes. We treat these AI-generated measurements as weak shadow variables that are conditionally independent of missingness given the true outcome and observed covariates. We then introduce two linear programs that provide sharp bounds on key population quantities, such as mean ratings, with the AI-generated measurements tightening these bounds by serving as additional constraints on the observed-data distribution. I will also discuss extensions to nonrandomly missing unstructured covariates, particularly in healthcare and digital platforms, and present methods that integrate observed content and availability patterns to learn latent state representations for improving prediction and sequential decision-making.



Bio:

Ruoxuan Xiong is an Assistant Professor in the Department of Data and Decision Sciences (DDS) at Emory University since Fall 2021, with courtesy appointments in Information Systems & Operations Management and Economics. Before joining Emory, she was a postdoctoral researcher at the Stanford Graduate School of Business from 2020 to 2021. She received her Ph.D. in Management Science and Engineering from Stanford University in 2020 and her B.S. degrees from Peking University in 2014. Her research lies at the intersection of causal machine learning, experimental design, and econometrics, with applications in healthcare, digital platforms, and finance. Her work has appeared in Management Science, the Journal of Econometrics, ICML, NeurIPS, ICLR, KDD, AAAI, ACL, and EMNLP.

Department of Statistics
Franklin College of Arts and Sciences
UNIVERSITY OF GEORGIA