Indiana University Indianapolis
Department of Mathematical Sciences

STATISTICS SEMINAR

12:15pm—1:15pm, Tuesday, October 15, 2024
Zoom Meeting: Meeting ID: 845 0989 4694

Speaker: **Muxuan Liang**
Department of Biostatistics, University of Florida

Title: **Classification under Outcome Misclassification: Reliability Quantification and Partial Identification**

Abstract:

Misclassification of outcomes or labels presents a prevalent challenge in classification problems. In many applications, the underlying outcome may not be directly accessible, while a surrogate outcome subject to misclassification can be observed. Directly using the surrogate outcome may lead to a biased estimation of the optimal classification rule. The sensitivity and specificity of the surrogate outcome at an individual level can be used to remove such bias. However, with limited accessibility of the underlying outcomes, point identification of individual-level sensitivity and specificity is difficult or even impossible. In this work, for classification problems, we assume a range of individual-level sensitivity and specificity as reliability quantification of the surrogate outcomes. With this partial information on sensitivity and specificity, we establish partial identification for the distribution of the underlying outcome as well as the optimal classification rule using the surrogates. Based on this result, we propose a robust classification framework and a novel estimation procedure to estimate a robust classification rule without requiring point identification of individual-level sensitivity and specificity.

Bio:
Dr. Muxuan Liang is an Assistant Professor with the Artificial Intelligence Initiative in the Department of Biostatistics, University of Florida. Dr. Liang obtained his Ph.D. degree in statistics from the University of Wisconsin-Madison in 2018. His research focuses on treatment recommendations based on patient-level information, identifying signals from high-dimensional data, and other novel machine-learning techniques with applications to biomarker identification, cancer surveillance, and digital health. His research goal is to realize data-driven healthcare decision-making through innovative methods combining statistical and machine learning techniques.