

**University of California, Irvine
Statistics Seminar**

***Accommodating Measurement Error in Multivariate
Compositional Count Data with Applications
to Microbiome Research***

**Matt Koslovsky
Assistant Professor of Statistics
Colorado State University**

**4 p.m., Thursday, February 15, 2024
6011 DBH**

The human microbiome is the collection of microorganisms that live on and inside of our bodies. Microbiome data are inherently challenging to analyze due to their high-dimensionality, overdispersion, and zero-inflation. Analysis is further complicated by the steps taken to collect and process microbiome samples. For example, sequencing instruments have a fixed capacity for the total number of reads delivered. It is therefore essential to treat microbial samples as compositional. Another complicating factor of modeling microbiome data is that taxa counts are subject to measurement error introduced at various stages of the measurement protocol. Recently, the Dirichlet-multinomial (DM) distribution and its variants have been used extensively to model microbiome data due to its ability to accommodate the compositional structure of the data as well as overdispersion. A major limitation of the DM distribution is that it is unable to handle excess zeros typically found in practice which may bias inference. In this talk, I will introduce a novel Bayesian zero-inflated DM model for multivariate compositional count data with excess zeros, designed to boost scalability without sacrificing interpretability or imposing limiting assumptions. I will then present extensions to handle high-dimensional regression settings and potential taxonomic misclassification. The performance of the proposed methods is examined through simulation and is further illustrated using human microbiome data.