THE UNIVERSITY OF CHICAGO | DEPARTMENT OF STATISTICS

PhD Dissertation Proposal Presentation

Jimmy Lederman

Department of Statistics

The University of Chicago

"Count-Based Data Augmentation for Flexible Probabilistic Modeling of Nonstandard Data"

December 10, 2025, at 2:00 PM

Jones 111, 5747 S. Ellis Ave.

Abstract

Building probabilistic models for complex data often requires choosing from a small set of distributions that permit computationally tractable inference, at the cost of neglecting important properties of the data. This proposal seeks to expand the probabilistic modeling toolkit by developing computationally tractable models for data with nonstandard dispersion, bounded support, and compositional structure. First, to model count data which exhibits conditional underdispersion---where latent structure renders data more regular than Poisson--- we introduce a new class of models that assume observed data are order statistics (e.g., minimum, median, maximum) of some discrete parent distribution, focusing particularly on the Poisson and negative binomial distributions. Second, we build models for bounded support data based on a novel family of distributions on the unit interval (0,1) that arises by randomizing the parameters of a standard beta distribution with three dependent negative binomial random variables. Third, we study the scaled Dirichlet distribution as a flexible tool for connecting covariates and spatiotemporal structure to compositional and nonnegative count data. In each case, we repurpose underutilized tools from the statistics literature into practical probabilistic modeling components. Through a shared reliance on Poisson latent variable representations, we derive general data augmentation schemes that are modular with existing tools and enable parameter estimation and posterior inference in a wide range of hierarchical models. Overall, this work introduces some of the first computationally tractable Bayesian models for several important but previously underserved data types.
