THE UNIVERSITY OF CHICAGO | DEPARTMENT OF STATISTICS

# Master's Thesis Presentation

Daniel Purizhansky
Department of Statistics
The University of Chicago

## "Generative Models of Heterogeneity in Compositional Microbiome Data and the Batch-Corrected Latent Space"

November 9, 2026, at 10:00 AM
Jones 111, 5747 S. Ellis Avenue

Abstract

Compositional microbiome studies are difficult to analyze with traditional statistical inference due to sparsity, high dimensionality, and potentially complex covariance structures among individual taxa. Moreover, individual studies often contain too few independent observations to conduct robust statistical inference, motivating pooling to learn shared biological signals. We use variational autoencoders to model batch effects and estimate a posterior batch-corrected latent space, though we acknowledge that because study membership is correlated with biology, batch signal cannot necessarily be fully removed. This is intended to facilitate inference on biological variation while accounting for confounding effects introduced by study-specific sequencing and data processing.

We investigate multiple choices of model architecture, prior distributions, loss terms penalizing latent-space separation, temperature controls, and random effect amortizations. We find that tree-based likelihood models generally outperform corresponding flat alternatives in prior-generative fidelity and that attention-based batch-residual mappings achieve competitive generative performance. Although tree-based decoders improve generative fidelity, we observe relatively modest additional benefits from incorporating flexible conditional covariance structures. We hypothesize that nonlinear VAE latent-to-mean mappings already capture substantial dependence among taxa, reducing the incremental value of explicit covariance modeling. In particular, Dirichlet-based likelihoods demonstrate strong prior-generative fidelity despite their restricted conditional covariance structures.

---

For information about building access for persons with disabilities, please contact Keisha Prowoznik at 773.702-0541 or send an email to kprowoznik@uchicago.edu. If you wish to subscribe to our email list, please visit the following website: https://lists.uchicago.edu/web/info/statseminars.