A213 · Bayesian Negative Binomial Softmax Regression for Compositional Sequencing Count Data

Presenter: Seong-Hwan Jun — University of Rochester

Authors: Seong-Hwan Jun, Xiangyi Chen

Session: Day 2 · Fri Oct 2, 2026 · 2:15–4:15 PM

High-throughput sequencing assays generate count vectors that are inherently compositional, making differential expression and differential abundance analysis sensitive to normalization choices and global shifts in the measured system. We consider extit{Negative Binomial Softmax Regression} (NBSR), a Bayesian model-based CoDA framework for compositional sequencing count data. NBSR combines a negative binomial likelihood with a softmax link function that constrains fitted proportions to the simplex. We show that the softmax link is algebraically equivalent to the inverse centered log-ratio (CLR) map applied to the centered linear predictor, establishing NBSR as a generative log-ratio model and unifying additive log-ratio (ALR) and CLR parameterizations through different identifiability constraints. The negative binomial likelihood allows feature-specific, and potentially covariate-dependent, dispersion modeling, providing flexibility beyond multinomial or transformation-based CoDA approaches for overdispersed sequencing counts. We further develop a fully Bayesian formulation fit with Hamiltonian Monte Carlo, incorporating latent factor adjustment for unmeasured sample-level heterogeneity and posterior bias correction for absolute-abundance-oriented differential abundance effects. By applying the bias correction directly to posterior draws, NBSR propagates uncertainty from regression effects, dispersion parameters, latent factors, and the estimated compositional bias into downstream inference. We evaluate NBSR in simulation studies and applications to miRNA-seq and microbiome sequencing data, demonstrating a general framework for CoDA-aware differential abundance analysis of sequencing counts.