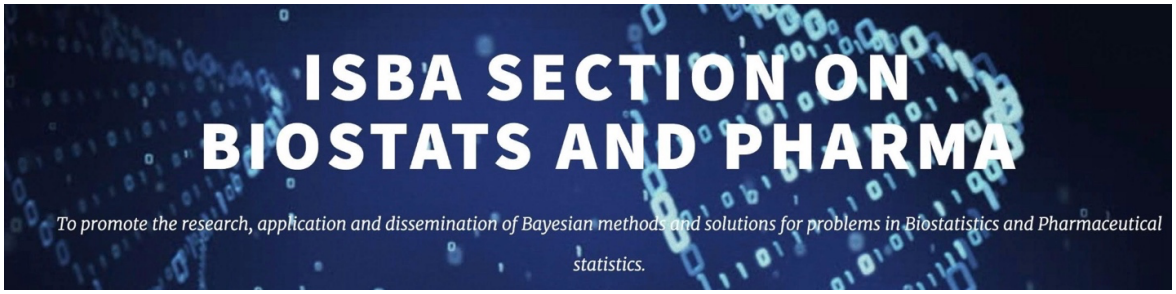

ISBA Biostatistics and Pharmaceutical Section 2024 Webinar Series

Friday, July 19 from noon – 1pm CT

Open Access Zoom Link:
https://us02web.zoom.us/j/88203560209

Sparse Bayesian Group Factor Model for Feature Interactions in Multiple Count Tables Data

By **SHUANGJIE ZHANG** (speaker)
email szhan209@ucsc.edu; juheelee@soe.ucsc.edu
(co-authors: Y. SHEN, I. Chen & J. Lee)

ABSTRACT: Group factor models have been developed to infer relationships between multiple co-occurring multivariate continuous responses. Motivated by complex count data from multi-domain microbiome studies using next-generation sequencing, we develop a sparse Bayesian group factor model (Sp-BGFM) for multiple count table data that captures the interaction between microorganisms in different domains. Sp-BGFM uses a rounded kernel mixture model using a Dirichlet process (DP) prior with log-normal mixture kernels for count vectors. A group factor model is used to model the covariance matrix of the mixing kernel that describes microorganism interaction. We construct a Dirichlet-Horseshoe (Dir-HS) shrinkage prior and use it as a joint prior for factor loading vectors. Joint sparsity induced by a Dir-HS prior greatly improves the performance in high-dimensional applications. We further model the effects of covariates on microbial abundances using regression. The semiparametric model flexibly accommodates large variability in observed counts and excess zero counts and provides a basis for robust estimation of the interaction and covariate effects. We evaluate Sp-BGFM using simulation studies and real data analysis, comparing it to popular alternatives. Our results highlight the necessity of joint sparsity induced by the Dir-HS prior, and the benefits of a flexible DP model for baseline abundances.