Statistics Winter Workshop: 2025

Computational Methods in Bayesian Statistics

Friday, January 17, 2025

9:00am-5:25pm

J. Wayne Reitz Union Auditorium

Saturday, January 18, 2025

9:00am to 1:15pm

J. Wayne Reitz Union Auditorium

Venue

J. Wayne Reitz Union
655 Reitz Union Drive
University of Florida
Campus Gainesville, FL
32611

Meeting Rooms
Presentations: Auditorium (2nd Floor)
Refreshments: Room 2340
Posters: Room: Room 2335

Parking

The Visitor Welcome Center and Bookstore parking garage is located at the Reitz Union, at the corner of Museum Road and Reitz Union Drive. The garage is an unrestricted pay facility available to all members of the university community. This garage can accommodate 300 vehicles. There are 45 short-term parking spaces located in the garage.

The garage hours of operation are Monday through Friday, 7:30 am to 4:30 pm. Short-term and daily fees apply during this time. The garage may be used during non-operating hours for short-term parking, free of charge.

Program—Day One
Friday, January 17, 2025

8:30 AM   Breakfast (Room 2340)

9:00 AM   Welcome-- Mike Daniels
Professor and Chair, Department of Statistics, University of Florida

9:10 AM   Rajesh Ranganath
Associate Professor, Courant Institute, New York University
Probabilistic inference and large models

9:55 AM   Liyue Shen
Assistant Professor, University of Michigan
Enhance Efficiency of Diffusion-based Generative Models for Solving Inverse Problems via Posterior Sampling

10:40 AM   Break (Room 2340)

10:55 AM   Guanyang Wang
Assistant Professor, Rutgers University
A quantum-inside-quantum Monte Carlo algorithm

11:40 AM   Art Owen
Professor, Stanford University
Practical quasi-Monte Carlo Integration

12:25 PM   Lunch (Room 2340)

1:55 PM   Sifan Liu
Research Scientist, Flatiron Institute
Transport Quasi-Monte Carlo

2:40 PM   Galin Jones
Professor, University of Minnesota
Reliable Metropolis-Hastings Simulations in the High-Dimensional, Large Sample Size Regime

3:25 PM   Break (Room 2340)

3:40 PM   Anirban Bhattacharya
Professor, Texas A&M University
On the Convergence of Coordinate Ascent Variational Inference

4:25 PM   Poster Session (Room 2340)

Program—Day Two
Saturday, January 18, 2025

8:30 AM   Breakfast (Room 2340)

9:00 AM   Lizhen Lin
Professor, University of Massachusetts
   Theoretical and computational foundations of variational Bayes

9:45 AM   Justin Domke
Associate Professor, University of Massachusetts
   Convergence guarantees for variational inference

10:30 AM   Break (Room 2340)

10:45 AM   Yuansi Chen,
Associate Professor, ETH, Zurich, Switzerland
   Regularized Weighted Dikin Walks for Sampling Truncated Logconcave Distributions, Beyond Worst-Case Analysis

11:30 AM   Yves Atchade
Professor, Boston University
   Cyclical MCMC

12:15 PM   Break (Room 2340)

12:30 PM   Brian Trippe
Assistant Professor, Stanford University
   Probabilistic methods for designing functional protein structures

Guest Speakers

Rajesh Ranganath

Associate Professor, Courant Institute of Mathematical Sciences and the Center for Data Science, New York University

Title: Probabilistic inference and large models

Abstract: Probabilistic inference has experienced a resurgence in AI. This resurgence has been fueled by both the use of latent variables to construct generative models like diffusions and the desire to query large generative models like large language models that are expensive to train. This talk will present a few vignettes on the kinds of inference approaches built for and tailored to these types of models.

Biography: Rajesh Ranganath is an associate professor at NYU's Courant Institute of Mathematical Sciences and the Center for Data Science. He is also an affiliate faculty at the Department of Population Health at NYU Medical Center. His research focuses on causal inference, generative models, interpretability, robustness, and machine learning for healthcare. Rajesh completed his PhD at Princeton and BS and MS from Stanford University. Rajesh has won several awards including the Porter Ogden Jacobus Fellowship, given to the top four doctoral students at Princeton University, the Savage Award in Theory and Methods, best paper awards, and a test of time award at AISTATS.

Liyue Shen

Assistant Professor, Electrical Engineering and Computer Science Department, University of Michigan

Title: Enhance Efficiency of Diffusion-based Generative Models for Solving Inverse Problems via Posterior Sampling

Abstract: Diffusion models have recently emerged as powerful generative priors for solving inverse problems. However, training diffusion models is data-intensive and computationally demanding, which restricts their applicability for high-dimensional and high-resolution data such as medical and scientific imaging. In this talk, I will introduce our recent works on how to improve the efficiency (data, time, and memory efficiency) of diffusion-based generative models for solving general inverse problems through posterior sampling. Particularly, I will introduce two plausible solutions we propose to enable learning diffusion priors for solving high-dimensional inverse problems through latent diffusion and patch diffusion models. The results are demonstrated in solving various inverse problems for both natural and medical images including 3D medical image reconstruction, showing the effectiveness of our proposed methods in both model efficiency and model performance. This research open the door to leverage diffusion-based generative models in tackling complex real-world data for addressing various crucial problems in many scientific disciplines.

Biography: Liyue Shen is an assistant professor in the EECS department at the University of Michigan. Prior to that, she received her B.E. degree in Electronic Engineering from Tsinghua University in 2016 and obtained her Ph.D. degree from the Department of Electrical Engineering, Stanford University in 2022. She also spent one year as a postdoctoral research fellow at the Department of Biomedical Informatics, Harvard Medical School. Her research interest is in Biomedical AI, which lies in the interdisciplinary areas of machine learning, computer vision, signal and image processing, biomedical imaging, medical image analysis, and data science. She recently focuses on the generative diffusion models, implicit neural representation learning and multimodal foundation models. She is the recipient of Stanford Bio-X Bowes Graduate Student Fellowship (2019-2022) and was selected as the Rising Star in EECS by MIT and the Rising Star in Data Science by the University of Chicago in 2021. She serves as area chairs for ICLR, MLHC, and helps organize multiple conferences and workshops including CPAL, ISBI, WiML, ML4H. Website: https://liyueshen.engin.umich.edu/

Guanyang Wang

Assistant Professor, Department of Statistics, Rutgers University

Title: A quantum-inside-quantum Monte Carlo algorithm.

Abstract: Classical Monte Carlo estimates expectations through repeated sampling and averaging. With access to a quantum computer, users can achieve a quadratic speedup using quantum amplitude estimation. The first half of this talk will provide background of quantum algorithms, especially those relevant to Monte Carlo estimation. The second half will cover our new quantum-inside-quantum Monte Carlo algorithm for nested expectations. This is a joint work with Mario Szegedy, Jose Blanchet, and Yassine Hamoudi.

Biography: Guanyang Wang is an Assistant Professor in the Department of Statistics at Rutgers University. Guanyang Wang's research centers on Monte Carlo methods, extending to adjacent fields like applied probability and machine learning. Recently, he is also working on quantum computing.

Art Owen

Professor, Department of Statistics, Stanford University

Title: Practical quasi-Monte Carlo Integration

Abstract: Quasi-Monte Carlo (QMC) integration is a method for computing multidimensional integrals. While not always Bayes, it is computational and is at least, Bayes-adjacent. QMC avoids the curse of dimensionality of classical quadrature rules in the same sense that plain Monte Carlo (MC) sampling does. When some randomization is injected into QMC, it is then superior to plain MC in several ways. It has a variance that is o(1/n) asymptotically while not being more than a given multiple of the MC variance at finite n. For favorable integrands, a better convergence rate is seen at practical sample sizes. One could wonder why QMC has not displaced MC. First, there is some modestly increased complexity to use it. Second, while it rarely underperforms plain MC, some skill may be required in order to get dramatic improvements at practical sample sizes. It helps that there are now good software implementations.

Biography: Art Owen is the Max H. Stein Professor of Statistics at Stanford University. He is best known for inventing the empirical likelihood and for developing and studying randomized quasi-Monte Carlo methods. He is a fellow of the American Statistical Association and the Institute of Mathematical Statistics. He received the 2020 Senior Noether Prize in nonparametric statistics from the ASA and the 2021 Gold Medal from the Statistical Society of Canada and became a SIAM Fellow in 2024.

Sifan Liu

Research Scientist, Center for Computational Mathematics, Flatiron Institute

Title: Transport Quasi-Monte Carlo

Abstract: Quasi-Monte Carlo (QMC) is a powerful method for evaluating high-dimensional integrals. However, its use is typically limited to distributions where direct sampling is straightforward, such as the uniform distribution on the unit cube or the Gaussian distribution. However, its application to more complex target distributions, such as those with unnormalized densities frequently encountered in Bayesian statistics, remains a challenge. We propose an approach that trains a transport map to push forward the uniform distribution to closely approximate an arbitrary target distribution. Our construction of the map is inspired by normalizing flows, with careful attention to parametrization to ensure the low-discrepancy property is preserved when mapping from the unit cube to the target distribution. We present both theoretical and numerical results to demonstrate that this approach, combined with QMC sampling, achieves a root-mean-squared error of order $1/n$, significantly outperforming standard Monte Carlo estimators.

Biography: Sifan Liu is a research scientist in the Center for Computational Mathematics at the Flatiron Institute. She is interested in Monte Carlo sampling and statistical inference. She received her Ph.D. in statistics from Stanford University in 2024. She will be joining the faculty of the statistics department at Duke University in Fall 2025.

Galin Jones

Professor and Director of the School of Statistics, University of Minnesota

Title: Reliable Metropolis-Hastings Simulations in the High-Dimensional, Large Sample Size Regime

Abstract: Monte Carlo experiments produce samples to estimate features of a given distribution. However, simultaneous estimation of means and quantiles has received little attention, despite being common practice. I will describe a multivariate central limit theorem for any finite combination of sample means and quantiles under the assumption of a strongly mixing process and provide a fast algorithm for constructing hyper-rectangular confidence regions having the desired simultaneous coverage probability and a convenient marginal interpretation. A key condition is that the Markov chain enjoys fast convergence, both theoretically and practically. Lower bounds on the convergence rate of Metropolis-Hastings and other accept-reject-based algorithms are developed in both total variation and Wasserstein distances to identify how the simulations will fail so these settings can be avoided, providing guidance on tuning. Particular attention is paid to using the lower bounds to study the convergence complexity of accept-reject-based Markov chains and to constrain the rate of convergence for geometrically ergodic Markov chains in the high-dimensional, large sample size regime.

Biography: I am Professor of Statistics and Director of the School of Statistics at the University of Minnesota. I am also affiliated with the UMN Data Science Initiative. My main research interests are in Markov chain Monte Carlo, statistical theory and methods in both Bayesian and frequentist domains, applications in neuroimaging and the physical sciences. I have also collaborated with a wide range of researchers including psychologists, veterinarians, librarians, and ecologists, among others. I earned a PhD in Statistics from the University of Florida in 2001. I am a fellow of both the American Statistical Association and the Institute for Mathematical Statistics. Past Editor of Journal of Computational and Graphical Statistics.

Anirban Bhattacharya

Professor, Department of Statistics, Texas A&M University

:

Title: On the Convergence of Coordinate Ascent Variational Inference

Abstract: As a computational alternative to Markov chain Monte Carlo approaches, variational inference (VI) is becoming increasingly popular for approximating intractable posterior distributions in large-scale Bayesian models due to its comparable efficacy and superior efficiency. Several recent works provide theoretical justifications of VI by proving its statistical optimality for parameter estimation under various settings; meanwhile, formal analysis on the algorithmic convergence aspects of VI is still largely lacking. In this talk, we will discuss some recent advances towards studying convergence of the popular coordinate ascent variational inference algorithm. We will present some specific case studies and proceed to develop a general framework for studying such questions.

Biography: Dr. Bhattacharya is a professor of statistics at Texas A&M University. He received his PhD from Duke University in 2012. His research interests include methodological, computational, and theoretical aspects of modern Bayesian statistics.

Lizhen Lin

**Professor and Director, Department of Statistics, University of Maryland**

Title: Theoretical and computational foundations of variational Bayes

Abstract: This talk will provide an overview of recent advancements in both the theory and computation of variational Bayes. A primary focus will be the introduction of a general adaptive variational Bayes (AVB) framework for adaptive statistical inference across a broad class of statistical problems. The AVB framework produces a variational posterior measure that achieves optimal contraction rates adaptively under very general conditions. Furthermore, it ensures model selection consistency when the true model structure exists. In addition to its theoretical advantages, the AVB framework remains computationally feasible, leveraging the inherent computational efficiency of variational methods. We demonstrate the applicability of AVB by applying the general theory to a wide range of statistical models, including deep learning models, and deriving novel adaptive inference results. Time permitting, we will also discuss general amortized variational Bayes and explore neural network-based computational techniques.

Biography: Lizhen Lin is a professor of statistics in the Department of Mathematics at the University of Maryland, where she also currently serves as the director of the statistics program. Her areas of expertise encompasses Bayesian modeling and theory for high-dimensional and infinite-dimensional models, statistics on manifolds, statistical network analysis, and statistical theory of deep neural network models.

Justin Domke

Associate Professor, Department of Statistics, University of Massachusetts-Amherst
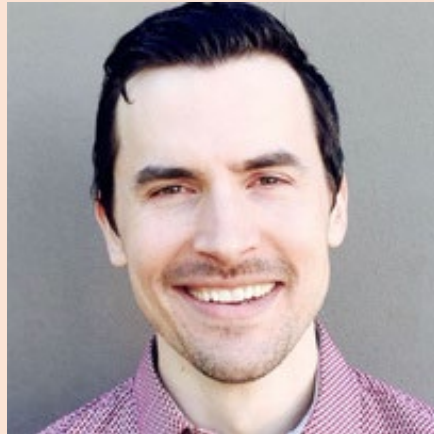
Title: Convergence guarantees for variational inference

Abstract: While black-box variational inference is widely used until recently, there were few guarantees that its stochastic optimization in this talk, I will cover some recent progress on establishing such guarantees. I will first discuss structural properties (convexity, smoothness) gradient variance guarantees. Next, I will discuss how these seem to reveal a gap in the optimization theory landscape. Finally, I will discuss some new guarantees that close that gap and highlight open questions.

Biography: Justin Domke is an associate professor at the University of Massachusetts Amherst. His research focuses on Bayesian inference, particularly the efficiency of variational inference and Monte Carlo methods.

Yuansi Chen

Associate Professor, Seminar for Statistics of ETH D-MATH, Zurich, Switzerland

Title: Regularized Weighted Dikin Walks for Sampling Truncated Logconcave Distributions, Beyond Worst-Case Analysis

Abstract: We study the problem of drawing samples from a log-concave distribution truncated on a polytope, a computational challenge often encountered in Bayesian statistical models with indicator variables, such as probit regression. Building on interior point methods and the Dikin walk for sampling truncated uniform distributions, we analyze the mixing time of regularized weighted Dikin walks. Our contributions are three-fold. First, for a log-concave, log-smooth distribution with condition number κ, truncated on a polytope in n dimensional space defined with m constraints, we prove that the soft-threshold Dikin walk mixes in $O((m + \kappa)n)$ iterations from a warm initialization. It improves upon prior work which required the polytope to be bounded and involved a bound dependent on the radius of the bounded region. Second, we introduce a novel regularized weighted Dikin walk, ridge-Lewis Dikin walk, based on Lewis weights for approximating the John ellipsoid. We show that it mixes in $O((n + \kappa)n)$ iterations from a warm initialization, which yields a sharper bound of $O(n^2)$ for truncated Gaussian sampling when compared to the previous $O(n^{2.5})$ bound. Third, going beyond worst-case mixing time analysis, we demonstrate that soft-threshold Dikin walk can mix significantly faster when only a small number of constraints intersect the high-probability mass of the distribution, improving the $O((m + \kappa)n)$ upper bound to $O(\kappa n + m)$. Additionally, we discuss the per-iteration complexity of both Dikin walks and provide efficient algorithms to generate initializations.

Biography: I am an associate professor in the Seminar for Statistics of ETH D-MATH. Previously, I was an assistant professor in the Department of Statistical Science at Duke University from Spring 2021 to Spring 2024. I was a postdoc fellow at ETH Foundations of Data Science (ETH-FDS) in ETH Zürich under the supervision of Prof. Peter Bühlmann. I obtained my PhD in the Department of Statistics at UC Berkeley in 2019. I was very fortunate to be advised by Prof. Bin Yu.

Yves Atchade

Professor, Department of Mathematics and Statistics, Boston University

Title: Cyclical MCMC

Abstract: Multimodal posterior distributions arise frequently in the applications, but still pose significant numerical challenges, particularly in high-dimensional settings. Commonly use Monte Carlo methods are easily trapped in one mode, and unable to explore the entire distribution. The talk will describe and discuss a scalable Markov chain Monte Carlo scheme based on annealing/tempering for high-dimensional multi-modal distributions. Based on a joint work with Indrajit Kalita, Xinru Liu, Aaron Smith, and Liwei Wang.

Biography: Yves Atchade is Professor in the department of Mathematics and Statistics with a secondary appointment in the Faculty of Computing and Data Science at Boston University. He is fellow of the Institute of Mathematical statistics and has served on the editorial board of several leading statistical journals including the Harvard Data Science Review, the Bernoulli Journal, the Journal of the American Statistical Association -- Theory and Methods, and the Journal of Computational and Graphical Statistics. His current research interests deal mainly with Markov Chain Monte Carlo methods, the statistical properties of Bayesian procedures, and the development of statistical methods for improved environmental prediction.

Brian Trippe

Assistant Professor, Department of Statistics, Stanford University

Title: Probabilistic methods for designing functional protein structures

Abstract: The biochemical functions of proteins, such as catalyzing a chemical reaction or binding to a virus, are typically conferred by the geometry of only a handful of atoms. This arrangement of atoms, known as a motif, is structurally supported by the rest of the protein, referred to as a scaffold. A central task in protein design is to identify a diverse set of stabilizing scaffolds to support a motif known or theorized to confer function. This long-standing challenge is known as the motif-scaffolding problem.

In this talk, I describe a statistical approach I have developed to address the motif-scaffolding problem. My approach involves (1) estimating a distribution supported on realizable protein structures and (2) sampling scaffolds from this distribution conditioned on a motif. For step (1) I adapt diffusion generative models to fit example protein structures from nature. For step (2) I develop sequential Monte Carlo algorithms to sample from the conditional distributions of these models. I finally describe how, with experimental and computational collaborators, I have generalized and scaled this approach to generate and experimentally validate hundreds of proteins with various functional specifications.

Biography: I am an assistant professor at Stanford in the Department of Statistics, with an affiliation in Stanford Data Science. In my research I develop probabilistic machine learning methods to address challenges in biotechnology and medicine. Recently, my focus has been on generative modeling and inference algorithms for protein engineering. Before joining Stanford, I was a postdoctoral fellow at Columbia University in the Department of Statistics, and a visiting researcher at the Institute for Protein Design at the University of Washington.

Poster Winners

Morris Greenburg

PhD Candidate in Statistical Sciences, University of Toronto
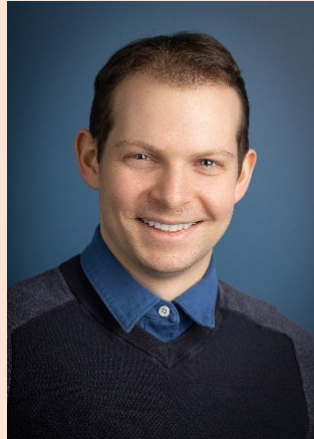
**Title:**   Restricted Search Space Graph MCMC via Birth-Death Processes

Abstract: Inferring a directed acyclic graph (DAG) given data is computationally challenging due to graphs existing in a discrete search space that grows super-exponentially with the number of nodes. A promising class of MCMC methods for graph inference addresses scalability by first restricting the search space to a subset of edges (where partial scores can be calculated in advance), and thereafter incrementally expanding the space until a stopping criterion is met. In this work, we estimate lower and upper bounds on the error introduced from current methods that operate on restricted spaces instead of the full space. Building on this, we propose a novel restricted-search MCMC method that reduces these errors. Our method is an adaptive algorithm which allows for either expansion or contraction of the search space throughout the chain. This is determined by a birth-death process, which we derive by extending the detailed balance conditions from previous birth-death process literature to the restricted search space problem. Additionally, we improve upon the computational costs of previous restricted-search methods by including block-matrix operations in expansion steps and memorization in contraction steps. We present extensive simulations that characterize the performance and computational efficiency of our algorithm, contrast this with existing methods, and consider applications in the field of imaging proteomics.

Biography: I am a 4th year PhD candidate, being advised by Kieran Campbell and Radu Craiu. My interests lie at the intersection of Bayesian methods and high-dimensional healthcare data, with current attention to building hierarchical models on single-cell data (especially when we want to infer a graph associated with the set of cell types). I have interest in teaching statistics as well.

Xi Qiao

Postdoctoral Research Associate, Department of Population & Quantitative Health Sciences, Case Western Reserve University, School of Medicine
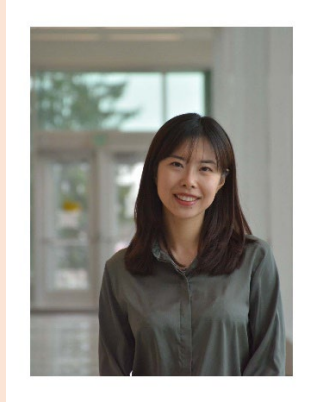
Title: Decoding Microbiome Mediation with ZIMMA: A Bayesian Framework for Enhanced Zero-Inflated Data Analysis

Abstract: Causal mediation analysis plays a crucial role in microbiome research by revealing how dysbiosis or microbial imbalance mediates the effects of interventions or environmental exposures on health outcomes. However, a key challenge in analyzing microbiome data is that excessive zero counts across samples often violate assumptions of traditional models. Despite advancements in handling high-dimensional and compositional microbiome data, few methods have directly addressed this zero-inflation issue. To bridge this gap, we introduce the Zero-Inflated Microbiome Mediation Analysis (ZIMMA) method, a Bayesian high-dimensional mediation model that employs a zero- inflated negative binomial distribution and spike-and-slab priors for sparse mediator analysis, uniquely suited to microbiome mediators. By identifying natural indirect effects (NIE) in the presence of nonlinear zero-inflated count mediators and dissecting the NIE into distinct pathways via mediator abundance and prevalence, ZIMMA offers a novel advancement to the field. Extensive simulations demonstrate its superior performance over existing methods. Its application to human microbiome studies examining the effects of nutrition intake and metabolic syndrome underscores its efficacy in identifying key microbial mediators, offering new insights into the role of the microbiome in health outcomes.

Biography: Xi is currently a postdoctoral research associate in the Department of Population & Quantitative Health Sciences at Case Western Reserve University, School of Medicine, specializing in Bayesian inference, causal mediation analysis, and Omics data analysis. She earned her Ph.D. in Statistics from Western Michigan University in 2023, where she focused on Bayesian high-dimensional mediation analysis and holds a master's degree in Statistics from West Virginia University. In the summer of 2022, Xi completed an internship at Daiichi Sankyo, Inc., concentrating on propensity score methods, Bayesian dynamic borrowing, and other historical control integration techniques.

Wei Yuan

PhD student, Department of Statistics, Rutgers University

Title: Markov chain Monte Carlo without evaluating the target: an auxiliary variable approach.
Authors Wei Yuan and Guanyang Wang

Abstract: In sampling tasks, it is common for target distributions to be known up to a normalizing constant. However, in many situations, even evaluating the unnormalized distribution can be costly or infeasible. This issue arises in scenarios such as sampling from the Bayesian posterior for tall datasets and the 'doubly-intractable' distributions. In this paper, we begin by observing that seemingly different Markov chain Monte Carlo (MCMC) algorithms, such as the exchange algorithm, PoissonMH, and TunaMH, can be unified under a simple common procedure. We then extend this procedure into a novel framework that allows the use of auxiliary variables in both the proposal and the acceptance–rejection step. Several new MCMC algorithms emerge from this framework that utilize estimated gradients to guide the proposal moves. They have demonstrated significantly better performance than existing methods on both synthetic and real datasets. Additionally, we develop the theory of the new framework and apply it to existing algorithms to simplify and extend their results.

Biography: I am a 4th year PhD student in the Department of Statistics at Rutgers University. My research interests are MCMC, diffusion model and uncertainty quantification in GenAI.

Zihan Zhu

Postdoctoral Scholar, Department of Epidemiology and Biostatistics, Case Western Reserve University, School of Medicine

Title: Phylogeny-Guided Selection of Microbiome Feature using Bayesian Zero-inflated Multinomial Logit Normal Regression

Abstract: The human microbiome regulates essential functions such as digestion, immunity, and even brain activity. Disruptions in this balance are related to diseases such as cancer, obesity, autoimmune disorders, and mental health issues, which offer potential for transformative healthcare insights. To facilitate these studies, next-generation sequencing techniques are used, in which bacterial DNA sequence counts serve as a proxy for bacterial abundance in a sample by quantifying the number of reads assigned to each specific taxon. The distribution of sequencing read data is often modeled using zero-inflated count data analysis, where the mean is regressed on host variables to establish microbiome-host associations. In a Bayesian hierarchical framework, these multivariate count data are typically modeled using a multinomial distribution, with compositional proportions following a Dirichlet distribution as the prior. However, a Dirichlet multinomial model does not offer a flexible dependency structure and does not consider the sequence similarity, which is crucial for accurately modeling relationships between microbial taxa. Furthermore, the relationship between Dirichlet parameters and observed counts is not straightforward, which makes interpretation less intuitive. To address these gaps, we propose a new Bayesian model that promotes joint selection of related taxa, informed by predefined phylogenetic relationships.

In the proposed model, we develop a zero-inflated log-normal prior that incorporates phylogenetic correlation. This prior can not only address the frequent occurrence of zeros in microbiome data but also enforces a zero covariance between present and absent taxa, effectively reducing noise from irrelevant taxa and selecting phylogenetic-correlated features. Furthermore, we model the relationship between pseudo-counts and covariates via log-normal regression, with the regression coefficients assigned an independent log-Cauchy prior. This super heavy-tailed prior enhances the model's ability to capture significant effects while also sharpening the distinction between signal and noise in the posterior distribution, making it easier to differentiate between the two. To implement our model, we develop a Metropolis-within-Gibbs sampler. We employ the between-within method for taxa selection. To identify key covariates that contribute to variations in microbial composition, we compute the Jensen-Shannon distance among the posterior empirical distributions of the regression coefficients and cluster the regression coefficients using the partition around medoids algorithm.

The advantages of the proposed model are demonstrated through simulation studies and real data analysis. In the simulation studies, we consider two data-generating approaches: 1) a zero-inflated negative binomial model, which is straightforward and provides a clear ground truth for performance evaluation, and 2) the gunifrac method, which is based on real microbiome data and generates more realistic microbiome datasets. In real data analysis, we apply the proposed model to analyze tumor microbiome composition in patients with pancreatic adenocarcinoma (PDAC) with short-term survival (STS) and long-term survival (LTS).

Biography: Zihan Zhu is a Postdoctoral Scholar in the Department of Epidemiology and Biostatistics at Case Western Reserve University School of Medicine. He holds a Ph.D. in Statistics from the Department of Mathematics at the University of Arizona. His research focuses on high-dimensional Bayesian statistics, spatial statistics, and microbiome data analysis, with an emphasis on developing advanced statistical methods for complex biological data.

Sponsors

UF Department of Statistics

UF College of Liberal Arts and Sciences

UF Office of Research