

Pritam Dey

📍 College Station, TX | ✉ pritam.dey@tamu.edu | 🌐 Website | 🔗 LinkedIn | 🐙 GitHub

Bio

I am a postdoctoral researcher in the Department of Statistics at Texas A&M University, specializing in statistical models for complex datasets motivated by real-world scientific problems. My current research focuses on spatial transcriptomics, multi-omic data integration, variational inference, and symbolic regression. I also have experience in neuroscience, specifically in structural connectomics. My goal is to develop efficient, robust models that advance scientific discovery through methodological and computational innovations.

Work Experience

- Postdoctoral Research Associate, Texas A&M University** Sep 2023–Present
Working in several collaborative teams from the Departments of Statistics and Nutrition to develop Bayesian frameworks and computational tools for spatial transcriptomics, dietary microbiome studies, and multi-omic biomarker discovery.
- Graduate Research Assistant, Duke University** Jan 2019–Aug 2024
Research Assistant developing statistical models for brain structural connectomics, including outlier detection and hierarchical continuous network representation.
- Graduate Teaching Assistant, Duke University** Aug 2019–Aug 2024
Led labs and performed grading for graduate-level courses: Linear Models (Fall 2019), Probability (Summer 2020), Probability and Measure Theory (Fall 2020), Predictive Modelling and Statistical Learning (Fall 2021), Probabilistic Machine Learning (Spring 2023).

Education

- Ph.D. in Statistical Science, Duke University** 2018–2023
 - Advised by Dr. David B. Dunson.
 - Dissertation Title:** Some Advances in Statistical modeling of Brain Structural Connectomes.
- M.S. in Statistical Science, Duke University** 2018–2023
- Master of Statistics, Indian Statistical Institute, Kolkata** 2016–2018
- Bachelor of Mathematics (Hons.), Indian Statistical Institute, Bangalore** 2013–2016
 - Award:** S.H. Aravind Gold Medal for academic performance

Technical Skills

- Statistics:** Bayesian inference, MCMC, Variational Inference, High-dimensional data analysis, Stochastic processes.
- Bioinformatics:** Spatial transcriptomics modeling, multi-omic integration, QIIME2, PICRUSt2, edgeR.

3. **Programming:** Python (Pandas, NumPy, Matplotlib, Scikit-learn), R (tidyverse), SLURM, MATLAB, C/C++, SQL.

Current Research Projects

1. **Bayesian Models for Spatial Transcriptomics:** Developed a Bayesian framework for identifying spatially varying genes (SVGs) in spatial transcriptomics data, accounting for gene co-expression, with efficient algorithm for proposed model in R. Our manuscript is under review at *JASA* and can be shared upon request. We are currently working on several key extensions. This work was done in collaboration with Dr. Rajarshi Guhaniyogi, Dr. Yang Ni and Dr. Bani K. Mallick.
2. **Infant Gut Microbiome and Exfoliome:** Leading statistical analysis of diet composition in terms of the Healthy Eating Index 2015 and its components on gut microbiome and fecally exfoliated host transcriptome on a cohort of 4-year-old infants. This is part of a highly collaborative research work in collaboration with the labs of Dr. Robert Chapkin and Dr. Sharon Donovan.
3. **Multi-omic Biomarkers in Mouse Models:** Designed integrative pipelines combining microbiome, exfoliome, and metabolome data to discover non-invasive fecal biomarkers for aryl hydrocarbon receptor activity in $Apc^{S580/+}/Kras^{G12D/+}$ mice. Preprocessing and integration performed using QIIME2, PICRUSt2, and custom R scripts. Collaboration with Dr. Robert Chapkin's lab.
4. **Tangent Approximation Variational Inference:** Proposed TAVIE, a general-purpose variational inference framework that constructs quadratic lower bounds on log-likelihoods for strongly super-Gaussian models. Achieved scalable updates for GLMs with R implementations; currently extending to spatiotemporal count data. Preprint available on arXiv. Collaboration with Somjit Roy, Dr. Debdeep Pati, and Dr. Bani K. Mallick.
5. **Bayesian Symbolic Regression Trees:** Developed a tree-based Bayesian symbolic regression approach to infer interpretable scientific expressions. The method places priors on tree structures and coefficients, enabling uncertainty quantification in symbolic model discovery. Implemented in R. Our manuscript is under review at *Biometrika*. Joint work with Somjit Roy, Dr. Debdeep Pati, and Dr. Bani K. Mallick.
6. **Ensembles of Mondrian Processes for continuous modeling of Brain connectomes:** Developed methods for continuous modeling of brain structural connectomes as density functions over the cortical surface, supported by theoretical guarantees. Extended to multi-subject modeling, which deconvolutes individual specific variation from common structure. Implemented in R. Two manuscripts under preparation. Joint work with Dr. Zhengwu Zhang and Dr. David B. Dunson.

Past Research Projects

1. **Outlier Detection for Brain Networks:** Developed a model based outlier detection method using influence measures for brain structural connectomics data. Implemented in R and Python. This method successfully detects various outliers in connectome data arising from measurement and preprocessing errors. Joint work with Dr. Zhengwu Zhang and Dr. David B. Dunson.
2. **Interpretable Matching for Causal Inference:** Worked on method development and coding for a large scale matching framework for causal inference.

Publications & Preprints

1. Roy, S., Dey, P., Pati, D., Mallick, B.K. A Generalized Tangent Approximation Framework for Strongly Super-Gaussian Likelihoods. *arXiv Preprint* (2025+).
2. Dey, P., Zhang, Z., Dunson, D.B. Outlier detection for multi-network data. *Bioinformatics* (2022).

3. Gupta, N.R., Dey, P., et al. dame-flame: A Python Library Providing Fast Interpretable Matching for Causal Inference. *arXiv Preprint* (2023).

Manuscripts in Preparation

1. JASPER: Joint Bayesian Analysis of Spatial Expression via Regression, under review at JASA (2025+).
2. Hierarchical Bayesian symbolic regression trees for structural learning, under review at Biometrika (2025+).
3. Noninvasive fecal multi-omic signatures for $\text{Apc}^{\text{S580/+}}/\text{Kras}^{\text{G12D/+}}$ mice, in preparation (2025+).
4. Ensembles of Mondrian Processes for continuous modeling of structural connectomes, in preparation (2025+).

Presentations & Conferences

- | | |
|---|------------------------------|
| 1. Invited talk: IISA Conference , <i>University of Nebraska–Lincoln</i> | <i>Scheduled in Jun 2025</i> |
| 2. Contributed Talk: 2nd Annual Graduate Student Research Conference , <i>Online</i> | 2022 |
| 3. Contributed talk: WNAR Conference , <i>Online</i> | 2022 |
| 4. Poster: Statistical Methods in Imaging Conference , <i>Online</i> | 2021 |

Awards

1. **Rank 1** in National Eligibility Test (NET), conducted by Council of Scientific and Industrial Research (CSIR), India (2018).
2. **S.H. Aravind Gold Medal**, for outstanding performance in Bachelor of Mathematics (Hons.), ISI Bangalore (2016).
3. **KVPY Scholarship**, awarded by Department of Science and Technology, Govt. of India (2013–2018).