Pritam Dey

📍 College Station, TX   ✉ pritam.dey@tamu.edu   🔗 pritamdey.github.io   in pritamdey1906   ⌗ pritamdey

Bio

I am a postdoctoral researcher in the Department of Statistics at Texas A&M University, with a strong background in statistics and machine learning. My research specializes in developing innovative statistical methodologies to tackle critical challenges in analyzing complex, high-dimensional datasets. I contribute to diverse interdisciplinary projects, focusing on Bayesian statistical methods for real-world applications, particularly in spatial transcriptomics and multi-omic data integration. I have also worked extensively with neuroscience data, including structural connectomic modeling. Beyond bioinformatics, the statistical tools and frameworks I develop have broad applicability, supporting the analysis of large-scale, structured data in fields such as artificial intelligence and materials science. My overarching goal is to design computationally efficient and robust models that provide accurate, interpretable results and advance scientific discovery across disciplines. By integrating theoretical advances, computational innovation, and interdisciplinary collaboration, my work bridges methodological innovation with practical impact.

Education

Duke University *PhD in Statistical Science*   *Aug 2018 – Aug 2023*

- **Selected Coursework:** Bayesian Forecasting, Bayesian Nonparametrics, High-dimensional Statistics, Graph Algorithms, Stochastic Calculus.
- **Research:** Statistical modelling and analysis of human brain structural connectomics data. Specifically developed an influence measure based outlier detector for brain networks and currently working on a fast scalable hierarchical framework for continuous connectivity modelling of the brain.

Duke University *Masters in Statistical Science*   *Aug 2018 – Aug 2023*

- **Selected Coursework:** Probability and Measure Theory, Linear Models, Bayesian Modeling, Tensor Methods, Time Series and Dynamical Systems

Indian Statistical Institute, Kolkata *Master of Statistics*   *Aug 2016 – Jun 2018*

- **Selected Coursework:** Statistical Inference, Regression, Large Sample Theory, Nonparametric Inference, Multivariate Statistics, Measure Theory, Statistical Computing, Time Series Analysis, Martingale Theory, Functional Analysis, Brownian Motion, Weak Convergence Theory

Indian Statistical Institute, Bangalore *Bachelor of Mathematics (Hons.)*   *Aug 2013 – Jun 2016*

- **Selected Coursework:** Real and Complex Analysis, Topology, Graph Theory, Differential Geometry, Differential Topology, Differntial Equations
- **Award:** S.H. Aravind Gold Medal

Research Experience

A Bayesian Method for Spatially Varying Gene Detection   *College Station, TX*
Texas A&M University   *Sep 2023 – Dec 2024*

- Developed a Bayesian framework for identifying spatially varying genes (SVGs) in spatial transcriptomics data, accounting for gene co-expression.
- Developed an efficient algorithm for proposed model with fast R code.
- Manuscript submitted to JASA ACS (can be shared upon request).
- Joint work with Dr. Rajarshi Guhaniyogi ↗, Dr. Yang Ni ↗ and Dr. Bani K. Mallick ↗.

Detection of Spatially Varying Networks of Genes   *College Station, TX*
Texas A&M University   *Oct 2024 – Present*

- Extension of the SVG detection method to model spatially varying co-expression networks.
- Two different approaches are being developed.
- R code and manuscript are under preparation.
- Joint work with Dr. Rajarshi Guhaniyogi ↗, Dr. Yang Ni ↗ and Dr. Bani K. Mallick ↗.

Integration of Multi-Omic data for biomarker discovery and discrimination
Texas A&M University

College Station, TX
Sep 2023 – Present

- Designed statistical pipelines for multi-omic integration and biomarker discovery, utilizing QIIME2 and PICRUSt2 in human and mice data.
- Developed a statistical framework for comparing omic datasets in terms of their discriminative power i.e., ability to differentiate among phenotypes.
- Worked with a highly collaborative research team supervised by Dr. Robert Chapkin ↗.
- Two working papers in progress. One to be submitted to Genome Biology in first quarter of 2025.

Tangent Approximation based Variational Inference for Generalized Linear Models
Texas A&M University

College Station, TX
Nov 2023 – Present

- Extension of a variational inference framework based on Tangent Approximation of the likelihood as proposed by Jaakkola and Jordan (2000) to different types of Generalized linear regression frameworks.
- Implemented with R and C++.
- One working paper in progress to be submitted to ICML 2025.
- Joint work with Somjit Roy ↗, Dr. Debdeep Pati ↗ and Dr. Bani K. Mallick ↗.

Bayesian Modeling of Operator-Induced Descriptors Using Symbolic Regression Trees
Texas A&M University

College Station, TX
July 2024 – Present

- Symbolic Regression model for Materials science applications with the goal of obtaining descriptors with interpretable mathematical expressions as a combination of basic physical properties (atomic weight, activation energy, etc.) and mathematical operators ($+$, $-$, $\times$, $\div$, exp, log, etc.)
- Implemented with R.
- One working paper in progress.
- Joint work with Somjit Roy ↗, Dr. Debdeep Pati ↗ and Dr. Bani K. Mallick ↗.

A fast scalable continuous representation framework for structural connectomics data
Duke University

Durham, NC
Oct 2021 – Present

- Motivated by the need for an efficient and fast multi-resolution representation of the structural connectome, we developed a scalable continuous representation of the connectome using Mondrian Processes.
- Developed theory and a hierarchical extension of Mondrian processes for continuous connectome representation based on density estimation.
- Developed an efficient and fast implementation in R.
- Two working papers in progress.
- Joint work with Dr. Zhengwu Zhang ↗ and Dr. David B. Dunson ↗.

Outlier Detection for Multi-Network Data
Duke University

Durham, NC
Jun 2019 – Jan 2022

- Motivated by the presence of large number of outliers in structural connectomics data we developed a model based outlier detection method based on an influence measure.
- Implemented this method in R and Python.
- Paper published in *Bioinformatics*
- Joint work with Dr. Zhengwu Zhang ↗ and Dr. David B. Dunson ↗.

dame-flame: A Python Library Providing Fast Interpretable Matching for Causal Inference
Duke University

Durham, NC
Oct 2018 - Dec 2019

- Python package based on a large-scale causal inference model.
- Contributed to the creation and debugging.

Teaching Experience

Teaching Assistant *Durham, NC*
Duke University *Aug 2019 - Aug 2024*

- ○ Worked as TA in several graduate level courses.
- ○ Led labs, created assignments and helped with grading.
- ○ 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).

Skills

Statistics: Bayesian modeling and inference, High-dimensional data analysis, Markov Chain Monte Carlo (MCMC), Variational Inference, Machine Learning, Stochastic Processes, Multivariate Time Series Analysis.

Bioinformatics: Spatial Transcriptomics data modeling, Multi-omic modeling and data analysis, Bioinfomatics software such as QIIME2 and edgeR.

Programming: Python (Pandas, NumPy, Matplotlib, Scikit-learn, etc.), R (tidyverse), SLURM, MATLAB, C/C++, SQL with strong proficiency in R and Python.

Awards and Recognition

Rank 1 in National Eligibility Test *Kolkata, India*
conducted by Council of Scientific and Industrial Research (CSIR) *2018*

S. H. Aravind Gold Medal *Bangalore, India*
for outstanding performance in Bachelor of Mathematics (Hons.) *2016*

KVPY Scholarship *Kolkata, India*
awarded by Department of Science and Technology (Govt. of India) *2013 - 2018*

Publications and Preprints

Outlier detection for multi-network data Jun 2022
Pritam Dey, Zhengwu Zhang, David B Dunson
10.1093/bioinformatics/btac431 ↗

dame-flame: A Python Library Providing Fast Interpretable Matching for Causal Inference 2021
Neha R. Gupta, Vittorio Orlandi, Chia-Rui Chang, Tianyu Wang, Marco Morucci, ***Pritam Dey***, Thomas J. Howell, Xian Sun, Angikar Ghosal, Sudeepa Roy, Cynthia Rudin, Alexander Volfovsky
doi.org/10.48550/arxiv.2101.01867 ↗

Talks and Posters

WNAR Conference, Contributed Talk *Online, 2022*

2nd Annual Graduate Student Research Conference, Contributed Talk *Online, 2022*

Statistical Methods in Imaging Conference, Poster *Online, 2021*