

# DEBRAJ BOSE

Email: bose@umich.edu | Phone: 734-881-5074

Website: bosedebraj96.github.io | LinkedIn: linkedin.com/in/debrajbose96/

## RESEARCH INTERESTS

---

My research is in statistical genetics, with a focus on developing statistical methods for gene-level inference and causal gene prioritization in human genetic studies. I am also broadly interested in applied statistical modeling, clinical trial design and analysis, and machine learning, with an emphasis on principled inference and real-world data applications.

## EDUCATION

---

**University of Michigan, Ann Arbor** 2019 – 2026 (expected)

PhD in Biostatistics, GPA: 4.0/4.0

**Thesis:** Statistical Methods for Gene-Level Inference and Causal Gene Prioritization in Human Genetic Studies

**Co-advised by:** Prof. Michael Boehnke and Prof. Xiaoquan (William) Wen

**Indian Statistical Institute, Delhi and Kolkata, India**

M.Stat. (Biostatistics Specialization), Overall Score: 79.5% 2017 – 2019

**St. Xavier's College, Kolkata, India**

B.Sc. (Hons) in Statistics, GPA: 8.9/10 2014 – 2017

## PUBLICATIONS

---

- **Bose D**, Fuchsberger C, Boehnke M. Rare-variant association studies: When are aggregation tests more powerful than single-variant tests? *American Journal of Human Genetics* 112(8), 1948-1961 (2025).
- Yin X, Li J, **Bose D**, ... , Morrison J. Assessing the potential causal effects of 1099 plasma metabolites on 2099 binary disease endpoints. *Nature Communications* 16, 3039 (2025).
- Yin X, **Bose D**, Kwon A, ... , Wen X. Integrating transcriptomics, metabolomics, and GWAS helps reveal molecular mechanisms for metabolite levels and disease risk. *American Journal of Human Genetics* 109(10), 1727–1741 (2022).
- Yin X, Chan LS, **Bose D**, ... , Boehnke M. Genome-wide association studies of metabolites in Finnish men identify disease-relevant loci. *Nature Communications* 13, 1644 (2022).
- Ray D, Salvatore M, Bhattacharyya R, ... , **Bose D**, ... , Mukherjee B. Predictions, role of interventions and effects of a historic national lockdown in India's response to the COVID-19 pandemic: data science call to arms. *Harvard Data Science Review*, Suppl. 1 (2020).

## POSTERS AND TALKS

---

- Poster, *American Society of Human Genetics Annual Meeting*, Boston, October 2025.
- Poster, *Eastern North American Region Spring Meeting*, New Orleans, March 2025.
- Contributed talk, *Joint Statistical Meetings*, Portland, August 2024.
- Poster, *American Society of Human Genetics Annual Meeting*, Washington DC, November 2023.
- Poster, *International Chinese Statistical Association Applied Statistics Symposium*, Ann Arbor, June 2023.
- Invited talk, *Graduate Student Working Group, Dept. of Biostatistics, University of Michigan*, Ann Arbor, February 2023.
- Poster, *American Society of Human Genetics Annual Meeting*, Los Angeles, October 2022.

- Poster, *American Society of Human Genetics Annual Meeting*, Virtual Conference, October 2020.
- Poster, *Indian Science Congress*, Mysore, India, January 2016.

## RESEARCH EXPERIENCE

---

### University of Michigan, Ann Arbor

Graduate Research Assistant

Sep 2019 – Present

- Developed method to estimate gene-level posterior probabilities of causality for a given complex disease/phenotype by integrating Genome-Wide Association Studies (GWAS), molecular Quantitative Trait Loci (QTLs), and gene-level features in an **empirical Bayes framework** using an **EM algorithm**. (Manuscript in progress)
- Provided guidelines for practitioners by performing analytic calculations and extensive simulations for rare-variant association analyses on **UK Biobank exome data** ( $n \sim 408K$ , number of rare variants  $\sim 6.7M$ ).
- Collaborated on the **METSIM study** to analyze high-dimensional metabolomics data, using gene-based aggregation methods to detect gene-metabolite associations for  $\sim 1.4K$  metabolites.

### Indian Statistical Institute, Kolkata, India

Master's Project

Aug 2018 – May 2019

Advisor: Prof. Subir K. Bhandari

- Proposed optimal histogram construction in one and multiple dimensions by (1) using **k-means clustering and Voronoi tessellations**, and (2) minimization of mean integrated squared error.
- Developed a novel supervised classification method based on **optimal histogram construction** and demonstrated through simulations across multiple settings that these classifiers achieve performance comparable to optimal Bayes classifier.
- Strengthened expertise in statistical learning theory, density estimation, and algorithmic implementation.

### Reliance Industries Limited, Navi Mumbai, India

Summer Intern

May 2018 – July 2018

- Built NLP-based pipelines in **Python** to extract, summarize, and classify information from quarterly financial reports of major oil and gas competitors, and also automated financial table generation and report summarization, reducing manual analysis time and improving data consistency.
- Collaborated with business analysts to translate unstructured text into structured insights for strategic decision-making.

## AWARDS AND HONORS

---

- Rackham Graduate School Conference Travel Grants 2022–2025
- Sabyasachi Roy Gold Medal award for best M.Stat. project 2019
- Book Grant for exemplary performance in M.Stat. examinations 2018–2019
- INSPIRE scholarship from Government of India 2014–2017
- Bronze medals representing North Kolkata table tennis team in West Bengal State Championships 2009, 2011

## LANGUAGE, PROGRAMMING AND STATISTICAL SKILLS

---

**Programming languages and frameworks:** R, LaTeX, Linux (expert); Python, C++, Git (practitioner).

**Relevant coursework:** Probability, Statistical Inference, Linear Regression, Multivariate Analysis, Longitudinal Data Analysis, Statistical Computing, Clinical Trials, Statistical Genetics, Survival Analysis, Pattern Recognition.

**Languages:** English, Bengali, Hindi.

**Summary of statistical and computational skills:**

- Strong background in analyzing and performing simulations with **large-scale genetic and metabolomic datasets** in high-performance **Linux** computing environments.

- **Statistical modeling and inference in R**, including longitudinal, survival, and clinical trial data analysis.
- Developed visualizations and interactive dashboards using **R Shiny, ggplot2, plotly, and flexdashboard**.

## SERVICE AND LEADERSHIP

---

### Statistics in the Community (STATCOM), University of Michigan

Communications Chair (2024–2025), Member (2023–2025)

- Led and contributed to pro bono data science projects for nonprofit and community organizations.
- Developed interactive **flexdashboard** visualizations for the Oakland Livingston Human Service Agency to understand needs of people experiencing hardship and poverty in Southeast Michigan.
- Built an **R Shiny dashboard** to visualize country-level and state-level visitor demographics and geographic trends for the Ann Arbor Hands-On Museum.
- Led a team project to analyze and summarize door-to-door survey data in two impoverished neighborhoods for The People's Action, producing a research report to support community advocacy efforts.
- Led a team project to design summary analyses and visualizations, including text summaries and word clouds, for program evaluation surveys for the Center for Success literacy initiative.

### Vice-President (2020–2022) and A-Team Member (2019–2025) of University of Michigan Table Tennis Team

- Won team and individual awards for University of Michigan at the Divisional and Regional tournaments.

## REFERENCES

---

Michael Boehnke, PhD

Richard G. Cornell Distinguished University Professor of Biostatistics

University of Michigan, Ann Arbor

boehnke@umich.edu

Xiaoquan (William) Wen, PhD

Professor, Biostatistics

University of Michigan, Ann Arbor

xwen@umich.edu

Christian Fuchsberger, PhD

Research Group Leader

Institute for Biomedicine, Eurac Research, Bolzano, Italy

christian.fuchsberger@eurac.edu

Xianyong Yin, MD

Professor, Epidemiology

Nanjing Medical University, Nanjing, Jiangsu, China

xyyin@umich.edu