SOMJIT ROY

+1(979) 344-3408 3143 TAMU, 155 Ireland St, College Station, TX 77843, USA
sroy_123@tamu.edu Google Scholar Github ORCiD LinkedIn Homepage

Research Interests and Skill Set

My research focuses on advancing next-generation statistical modeling frameworks that bridge data-driven learning with scientific reasoning. I develop novel methodologies to address complex, real-world problems spanning materials discovery, physics, geophysics, chemical informatics, bioinformatics, and gene expression modeling. Positioned at the intersection of Scientific Machine Learning and Bayesian Modeling *&* Computation, my skills include Scientifically Guided Probabilistic Inference, Physics-Informed Modeling, Variational Inference, Tree-based Models, and Bayesian Optimization; aiming to build interpretable, scalable, and scientifically grounded learning systems.

Education and Training

Texas A&M University, College Station, Texas, USA — 2023 — present
Doctor of Philosophy (Ph.D.) in Statistics — GPA: 4.00/4.00
Advisors: **Bani K. Mallick** (TAMU Statistics) *&* **Debdeep Pati** (UW-Madison Statistics)
Area of Study: Probabilistic Scientific Machine Learning and Computational Bayes

University of Calcutta, Kolkata, West Bengal, India — 2021 — 2023
Master of Science (MSc.) in Statistics (Ranked First Class First) — GPA: 8.15/10.00

St. Xavier's College, Kolkata, West Bengal, India — 2018 — 2021
Bachelor of Science (BSc.) in Statistics (Hons.) — GPA: 8.63/10.00

Experience

Los Alamos National Laboratory — May 2025 — Aug 2025
Graduate Summer Intern — Los Alamos, NM, USA

- **SPINWAVE**: **S**calable **P**hysics-**I**nformed **N**eural Operator for Seismic **WAVE** Modeling.
- Advised by **Kai Gao** *&* **Ting Chen** in Earth and Environmental Sciences (EES-16) division. Developed a scalable framework for seismic wave modeling using Physics-Informed Neural Operators (PINO), for solving the acoustic wave partial differential equation in complex, heterogeneous media.

Tata Electronics Pvt. Ltd. — May 2021 — Aug 2021
Data Science Intern — Bengaluru, KA, India

- **Statistical Analysis and Optimization of Sandblasting *&* Anodizing**.
- Supervised by **Nagasubramanian Kothandaraman** (Tata Electronics Pvt. Ltd.) *&* **Subhamoy Maitra** (ISI, Kolkata). Identified key factors influencing material surface properties (gloss and texture) and optimized process parameters using machine learning architectures, achieving a reduction of two man-days in production time.

Papers and Manuscripts

Convention: ★ → Scientific Machine Learning, ✯ → Bayesian Modeling *&* Computation, ✰ → Variational Inference, ☆ → Bayesian Optimization, and ✯ → Quantum Cryptography.

7. **Roy, S.**, Gao, K., *&* Chen, T. (2025+). SPINWAVE: Scalable Physics-Informed Neural Operator for Seismic WAVE Modeling. (*In preparation*). ★
6. **Roy, S.**, Dey, P., Mallick, B. K., Pati, D., *&* Arróyave, R. (2025+). Multi-Property Materials Discovery using Bayesian Symbolic Trees. (*In preparation*). ★ ✯
5. **Roy, S.**, Jaiswal P., Bhattacharya, A., Pati, D., *&* Mallick, B. K. (2025+). On Frequentist Regret Analysis of Fractional Gaussian Process Thompson Sampling. (*Submitted*). ☆ ✯

4. **Roy, S.**, Dey, P., Pati, D., *&* Mallick, B. K. (2025+). Hierarchical Bayesian Operator-induced Symbolic Regression Trees for Structural Learning of Scientific Expressions. (*Submitted*). arXiv:2509.19710. [SBSS Student Paper Award 2026, American Statistical Association; Selected for refereed oral presentation at ASA's SDSS 2026]. ★ ★

3. **Roy, S.**, Dey, P., Pati, D., *&* Mallick, B. K. (2025+). A Generalized Tangent Approximation Based Variational Inference Framework for Strongly Super-Gaussian Likelihoods. (*Submitted*). arXiv:2504.05431. ★ ★

2. Kumar, A., Maitra, S., *&* **Roy, S.** (2024). Almost Perfect Mutually Unbiased Bases that are Sparse. *Journal of Statistical Theory and Practice* **18**, 61. ★

1. Chaudhury, S., Kumar, A., Maitra, S., **Roy, S.**, *&* Sen Gupta, S. (2022). A Heuristic Framework to Search for Approximate Mutually Unbiased Bases. In Cyber Security, Cryptology, and Machine Learning. (CSCML) 2022. *Lecture Notes in Computer Science*, vol **13301**. Springer, Cham. ★

Open Source Software

On Github: `HierBOSSS` (**Roy, S.** *&* Dey, P., 2025), `TAVIE-SSG` (**Roy, S.** *&* Dey, P., 2025).

R Packages on CRAN: `bayesestdft` (**Roy, S.** *&* Lee, S. Y., 2025), `GoodFitSBM` (Ghosh, S., **Roy, S.**, *&* Pati, D., 2024), `gamblers.ruin.gameplay` (**Roy, S.**, 2022), `YatesAlgo.FactorialExp.SR` (**Roy, S.**, 2021, [Selected for a talk in the useR regional conference in Basel, Switzerland, July 2023]).

Technical Skills

- **Programming Languages**: Python, R, JAVA, C, C++.
- **Statistical *&* Machine Learning Tools**: RStudio, RStan, RShiny, mcmc, coda, PyMC, PyTorch, PyTorch Lightning, TensorFlow, SciPy, scikit-learn.
- **High-Performance *&* Cluster Computing**: Linux/Unix environments, MPI/OpenMP parallelization, CUDA -GPU programming, PyTorch DDP.
- **Version Control *&* Documentation**: Git, Markdown, LaTeX.

Awards and Fellowships

- **SBSS Student Paper Award 2026, American Statistical Association**.
- **NSF Travel Grant *&* TAMU Statistics Department Student Travel Award** (for IISA 2024).
- Awarded **Targeted Proposal Teams (TPT)** grant *&* scholarship by Texas A&M University.
- Selected to attend **CMS^3-FAST Summer School (2024)** in Texas A&M University.
- Awarded the **R.C. Bose Memorial Book Prize (2022)** by Calcutta Statistical Association.
- Awarded the **IASc-INSA-NASI Science Academies Summer Research Fellowship (2022)**.
- Recipient of the **OPHI, University of Oxford - Summer School Grant (2022)**.
- Awarded the **IAOS 2022 Conference and Travel Grant (2022)** by the World Bank.

Talks and Presentations

- **JSM 2026**: Invited Award Presentation for Section on Bayesian Statistical Science (SBSS) Student Paper Award. On *Hierarchical Bayesian Operator-induced Symbolic Regression Trees for Structural Learning of Scientific Expressions.* (Boston, MA, USA).
- **SDSS 2026**: Refereed Invited Talk on *Hierarchical Bayesian Operator-induced Symbolic Regression Trees for Structural Learning of Scientific Expressions.* (Milwaukee, WI, USA).

- **STAT CAFÉ 2025**: On *Bayesian Symbolic Trees for Structural Learning of Scientific Expressions.* (College Station, TX, USA).
- **SIAM 2025**: Invited Talk on *Physics-Informed Neural Operators for Seismic Wave Modeling.* (UT Austin, TX, USA).
- **LANL Student Symposium 2025**: Poster on *Physics-Informed Neural Operators for Seismic Wave Modeling.* (Los Alamos, NM, USA).
- **IISA 2024**: On *Tangent Approximation for Variational Inference in different Exponential Families.* (Kochi, KL, India).

Leadership and Professional Activities

- **Workflow Workshop Organizer (2024—2025)**, as a part of Statistics Graduate Student Association (SGSA), Texas A&M University.
- **Organizing Committee Member (2019—2020)**, as a part of the annual fest ϵpsilon δelta organized by Department of Statistics, SXC Kolkata.

References

- **Bani K. Mallick**
  Department Head; Susan M. Arseven '75 Chair in Data Science and Computational Statistics; University Distinguished Professor; Regents Professor; Director, Center for Statistical Bioinformatics; Director, Texas A&M TRIPODS Data Science Institute (FIDS); Editor-in-Chief, SIAM/ASA Journal of Uncertainty Quantification
  Department of Statistics, Texas A&M University, College Station, TX 77843, USA
  ✉ bmallick@stat.tamu.edu
- **Debdeep Pati**
  Professor, Department of Statistics, University of Wisconsin–Madison, Madison, WI 53706, USA
  ✉ dpati2@wisc.edu