SUBHRAJYOTY ROY

Nationality: Indian
(+91) 80139 76355
roy.s@wustl.edu

Google Scholar
github.com/subroy13
/in/subroy13
Portfolio Website

CURRENT POSITION

- *Postdoctoral researcher* at the Department of Statistics & Data Science in the **Washington University, St. Louis**.

RESEARCH INTERESTS

- Robust and High-dimensional statistical inference.
- Statistical divergences, minimum divergence estimation.
- Spatio-temporal data analysis.
- Robust Machine Learning, AI Safety.

PROFESSIONAL EXPERIENCE

July, 2025- **Postdoctoral Researcher** — **Department of Statistics and Data Science, Washington University at St. Louis**

- **Courses taught:** SDS 4010 Probability (Fall 2025).

July, 2022 - June, 2025 **Principal Information Researcher** — **SysCloud Inc.**

- Continuous system improvement, research and new technology development across the company.
- Implementing a centralized control flow system to reduce resource overconsumption, resulting in 90% reduction in API rate-limit errors.
- Development of SysCloud Sky and Scout, the AI-powered chatbots for customer and internal usage.
- Building company-wide automation and adoption to customized AI solutions using prompting, RAG, AI agents, etc.

May, 2024 -July, 2024 **Mentor** — **Great Learning Education Series Pvt. Ltd.**

- Course instructor for *Mathematics for Artificial Intelligence* module. Received a consistent rating of 4.5+.

July, 2021 - June, 2022 **Data Scientist** — **SysCloud Inc.**

- Developed models for optimized bidding strategy for Google Ads using reinforcement learning.
- Unified, scalable and robust system architecture development of backup, restore and export of any cloud, resulting in business expansion for various SaaS apps (Quickbooks, Hubspot, Salesforce, Slack, Box, etc.).

2020 May - June **Data Scientist Intern** — **General Electric (Aviation)**

- Understanding of the impact of COVID-19 on the aviation industry with an analysis of historical tweets data.

2019 May - June **Data Scientist Intern** — **SysCloud Inc.**

- Developed cost optimization techniques for backup and restore of G-suite and Office 365 workspaces.
- Solving phishing email detection problem using a combination of NLP and Deep learning.

EDUCATION

2021 - 2025 **Ph.D. in Statistics** — **Indian Statistical Institute, Kolkata**
Thesis: Robust Matrix Factorization using the Density Power Divergence and its Applications.
Supervisors: Prof. Ayanendranath Basu and Prof. Abhik Ghosh.

2019-2021 **Masters of Statistics (M.Stat.)** — **Indian Statistical Institute, Kolkata**
Grade: 93.8%, Financial Statistics specialization, Ranked *1st*.

2016 - 2019 **Bachelor of Statistics (B.Stat.)** — **Indian Statistical Institute, Kolkata**
Grade: 93.6%, Ranked *2nd*.

2014 - 2016 **Higher Secondary (WBCHSE)** — **Nimta High School, Kolkata**
Grade: 93.4%, District Rank *2nd*.

2014 **Secondary (WBCSSE)** — **Nimta High School, Kolkata**
Grade: 94.57%, State Rank *20th*.

PUBLICATIONS

Journal Articles

1. **Roy, S.**, Sarkar, A., Ghosh, A., & Basu, A. (2025+). Asymptotic Breakdown Point Analysis for a General Class of Minimum Divergence Estimators. *Bernoulli*. (Accepted). [Link].
2. **Roy, S.**, Basu, A., & Ghosh, A. (2024). Robust Principal Component Analysis using Density Power Divergence. *Journal of Machine Learning Research*, 25(324), 1-40. [Link].
3. **Roy, S.**, Ghosh, A., & Basu, A. (2024). Robust singular value decomposition with application to video surveillance background modelling. *Statistics and Computing*, 34(5), 178. [Link].
4. Pyne, A., **Roy, S.**, Ghosh, A., & Basu, A. (2024). Robust and efficient estimation in ordinal response models using the density power divergence. *Statistics*, 58(3), 481-520. [Link].
5. Bhaduri, R., **Roy, S.**, & Pal, S. K. (2022). Rough-Fuzzy CPD: a gradual change point detection algorithm. *Journal of Data, Information and Management*, 4(3), 243-266. [Link].
6. Ghatak, A., Singh Patel, S., Bonnerjee, S., & **Roy, S.** (2022). A generalized epidemiological model with dynamic and asymptomatic population. *Statistical Methods in Medical Research*, 31(11), 2137-2163. [Link].
7. **Roy, S.**, Sengupta, D., Rudra, K., & Saha, U. S. (2020). Analysis of Pollution Patterns in Regions of Kolkata. *Calcutta Statistical Association Bulletin*, 72(2), 133-170. [Link].
8. Mukherjee, D., Dalal, A., & **Roy, S.** (2019). Feasibility of Transparent Price Discovery in Tea through Auction in India. *Commodity Insights Yearbook (MCX)*, 44-52. [Link].

Preprints

1. Jana, S., **Roy, S.**, Ghosh, A., & Basu, A. (2025). Asymptotic breakdown point analysis of the minimum density power divergence estimator under independent non-homogeneous setups. *arXiv preprint* arXiv:2508.12426. [Link].
2. **Roy, S.**, Basu, S., Ghosh, A., & Basu, A. (2025). Characterization of Generalized Alpha-Beta Divergence and Associated Entropy Measures. *arXiv preprint* arXiv:2507.04637. [Link].
3. Deb, S., Neves, C., & **Roy, S.** (2024). Nonparametric quantile regression for spatio-temporal processes. *arXiv preprint* arXiv:2405.13783. [Link].
4. **Roy, S.** (2024). Trustworthy Dimensionality Reduction. *arXiv preprint* arXiv:2405.05868. [Link].
5. Bhaduri, R., Bonnerjee, S., & **Roy, S.** (2019). Onset detection: A new approach to QBH system. *arXiv preprint* arXiv:1908.07409. [Link].

SEMINARS, CONFERENCES & LECTURES

1. Student Paper Competition at **IISA Conference, 2024** at **Cochin University of Science and Technology**.
   Title: Robust Principal Component Analysis using Density Power Divergence.
2. Contributory talk at **IMS Asia Pacific Rim Meeting, 2024** at **University of Melbourne**.
   Title: A novel and scalable background modelling algorithm from video surveillance data.
3. Invited guest lecture for Robust Statistics course (M.Stat.'22) at **Indian Statistical Institute, Kolkata**.
   Title: A review of Robust Location and Scatter estimators.
4. Prasanta Chandra Mahalanobis memorial lecture at **Indian Statistical Institute, Kolkata**.
   Title: Algorithmic Fairness of Statistical Decision Systems.
5. D. Basu memorial lecture at **Indian Statistical Institute, Kolkata**.
   Title: tSNE: A way to visualize multivariate data.

ADDITIONAL ACTIVITIES

- Maintain a substack newsletter StatWizard with 200+ subscribers.
- Served as a session chair for IMS Asia-Pacifim Rim Meeting 2024 at the University of Melbourne, Australia.
- Reviewer for the journal **Scientific Reports**.

PROJECTS, SOFTWARE AND PACKAGES

2025 — `handanim` - **Contributor & Maintainer** — **Github Link**
A Python library to create whiteboard-style, hand-drawn animations for educational videos, tutorials or data storytelling.

2024 — `Langchain` - **Contributor** — **Github Commits**
Framework for developing applications powered by large language models

2023 — `Multi PDF QueryBot` - **Developer & Maintainer** — **Github Link**
A streamlit application to chat with PDF documents using RAG and Large language models. Uses a locally hosted Chroma vector database, SQLite for metadata storage, and Claude for access to foundational LLM.

2023 — `callgrid-reader` - **Developer & Maintainer** — **NPM Link**
A typescript utility package to read, parse and analyze callgrind files for profiling analysis of codes

2023 — `decompy` - **Developer & Maintainer** — **PyPI Link**
Python package for multiple robust matrix and tensor decomposition algorithms

2022 — `roufcp` - **Developer & Maintainer** — **PyPI Link**
Python package implementing a gradual changepoint detection algorithm using rough-fuzzy set theory

2021 — `rsvddpd` - **Developer & Maintainer** — **CRAN Link**
R package to perform Robust SVD using Density Power Divergence

2020 — **COVID-19 Tracker: A Dashboard for prediction and control assistance**
Developed a dashboard to visualize the COVID-19 situation in different states of India. It dynamically estimates the time-varying reproduction rate and prediction of COVID-19 spread for various lockdown scenarios for district levels in India, based on a modified eSIR model. The project was funded by the Indian Insititute of Management, Vishakhapatnam.

2019 Class Project — **Analysis of Climber's challenges in scaling Mount Rainier** — **Report Link**
Modelling of climber's success rate in climbing Mount Rainier based on weather modelling. Modelling of climber's team sizes for insights about mountaineering-based local economy.

2019 Class Project — **An Introduction of Causal Inference using Direct Acyclic Graph** — **Report Link**
Understanding of the setup and assumptions for causal inference. Estimating the effect of intervention using frontdoor and backdoor criterion, simulations and applications to real-world datasets to estimate causality.

2018 — **Chatbot using Seq2Seq Attention Mechanism** — **Github Link**
It is a chatbot created with a seq2seq neural network with a basic attention mechanism, trained with the Cornell Movie Corpus dataset.

ACCOLADES

2024 — **Nominated for the Student Paper Award** — **IISA Conference 2024**

2024 — **AWS Cloud Practitioner Certification** — **Amazon Web Services (AWS)**

2023 — **AIR 6 at National Eligibility Test** — **Council of Scientific and Industrial Research (CSIR), Govt. of India**

2022 — **Nominated for P. C. Mahalanobis Memorial Gold Medal** — **Indian Statistical Institute, Kolkata**

2022 — **J. K. Ghosh Memorial Gold Medal** — **Indian Statistical Institute, Kolkata**

2016-2021 — **Dean's list of toppers** — **Indian Statistical Institute, Kolkata**

2020 — **Scaled score 324 at GRE General Test** — **Educational Testing Service (ETS)**

2020 — **Nikhilesh Bhattacharya Memorial Gold Medal** — **Indian Statistical Insititute, Kolkata**

2020 — **D. Basu Memorial Gold Medal** — **Indian Statistical Insititute, Kolkata**

2019 — **93 percentile at GRE Mathematics Subject Test** — **Educational Testing Service (ETS)**

2018 — **1st Quartile at Simon Marais Mathematical Competition** — **Australian Mathematical Science Insititute**

2016 — **AIR 1 in Undergraduate Entrance Exam** — **Insititute of Mathematics and Application, Bhubaneshwar**

SKILLS

Programming — R, Python, SQL, PHP, JavaScript (NestJS, NodeJS, React, NextJS).
Technologies — AWS (S3, Lambda, ECS, Sagemaker, Bedrock), Docker, Langchain, Langflow.
Tools — MS-Office, LaTeX

REFERENCES

- **Ayaendranath Basu.** Professor, Indian Statistical Institute, Kolkata. ayanbasu@isical.ac.in
- **Abhik Ghosh.** Associate Professor, Indian Statistical Institute, Kolkata. abhik.ghosh@isical.ac.in
- **Diganta Mukherjee.** Professor, Indian Statistical Institute, Kolkata. diganta@isical.ac.in
- **Soudeep Deb.** Associate Professor, Indian Institute of Management, Bangalore. soudeep@iimb.ac.in
- **Vijay Krishna.** CEO, SysCloud Inc. vijay@syscloudsoftware.com

I certify that the information provided above is true and accurate to the best of my knowledge

— Subhrajyoty Roy