Soham Jana

240 Hayes-Healy Center
University of Notre Dame
Notre Dame, IN, USA

Updated on: November 12, 2024
Website: https://janasoham.github.io
Email: sjana2-at-nd-dot-edu
Phone: +1 574-631-5503

Research Interests

Theoretical and methodological aspects of high-dimensional statistics, robust estimation, neural networks, causal inference.

Education

PhD. in Statistics and Data Science   May 2022
Yale University, New Haven, CT, USA
Thesis: Learning non-parametric and high-dimensional distributions
via information-theoretic methods
Advisor: Prof. Yihong Wu

Master of Statistics (Hons.) (First class with distinction)   May 2017
Indian Statistical Institute, Kolkata, West Bengal, India
Specialization: Theoretical Statistics
Dissertation: Characterization of single-integral non-kernel divergences
Advisor: Prof. Ayanendranath Basu

Bachelor of Statistics (Hons.) (First class with Distinction)   May 2015
Indian Statistical Institute, Kolkata, West Bengal, India

Work experiences

University of Notre Dame, Notre Dame, IN, USA

Assistant Professor, Department of Applied and   August 2024 – Current
Computational Mathematics and Statistics.

Princeton University, Princeton, NJ, USA

Postdoc, Department of Operations Research and   June 2022 – July 2024
Financial Engineering
Hosts: Prof. Sanjeev Kulkarni and Prof. Jianqing Fan

Researcher, The First Republic Bank Research and   June 2022 – May 2023
Lifelong Learning Program

Lecturer   Spring 2023 and Fall 2023

Preprints ("∗": Authors list not in alphabetical order)

1. Shange Tang, Soham Jana, Jianqing Fan. **Factor adjusted spectral clustering for mixture models**. arXiv preprint arXiv:2408.12564 (2024).

2. Soham Jana, Jianqing Fan, Sanjeev Kulkarni*. **A general theory for robust clustering via trimmed mean**. arXiv preprint arXiv:2401.05574 (2024).

3. Soham Jana, Kun Yang, and Sanjeev Kulkarni*. **Adversarially robust clustering with optimality guarantees**. arXiv preprint arXiv:2306.09977 (2023).

4. Soham Jana, Yury Polyanskiy, and Yihong Wu. **Optimal empirical Bayes estimation for the Poisson model via minimum-distance methods**. arXiv preprint arXiv:2209.01328 (2022). Under major revision at **Information and Inference - a journal of IMA**.

Journal publications ("∗": Authors list not in alphabetical order)

1. Soham Jana, Henry Li, Yutaro Yamada, and Ofir Lindenbaum. **Support recovery with Stochastic Gates: theory and application for linear models**. Elsevier Signal Processing (2023), 213, p.109193.

2. Yanjun Han, Soham Jana and Yihong Wu, **Optimal Prediction of Markov Chains With and Without Spectral Gap**, in IEEE Transactions on Information Theory, vol. 69, no. 6, pp. 3920-3959, June 2023, doi: 10.1109/TIT.2023.3239508. (**Extended from the NeurIPS version with analysis of higher-order Markov chains and different loss functions**)

3. Soham Jana and Ayanendranath Basu.* **A characterization of all single-integral, non-kernel divergence estimators**. IEEE Transactions on Information Theory 65.12 (2019): 7976-7984.

Conference publications ("∗": Authors list not in alphabetical order)

1. Soham Jana, Yury Polyanskiy, Anzo Teh, and Yihong Wu. **Empirical Bayes via ERM and Rademacher complexities: the Poisson model**. In Conference on Learning Theory 2023 Jul 15, PMLR 195:5199-5235.

2. Yanjun Han, Soham Jana, and Yihong Wu. **Optimal prediction of Markov chains with and without spectral gap**. NeurIPS 2021.

3. Soham Jana, Yury Polyanskiy, and Yihong Wu. **Extrapolating the profile of a finite population**. In Conference on Learning Theory 2020 Jul 15 (pp. 2011-2033). PMLR.

Invited Talks

International Indian Statistical Association   December 2024
Cochin, Kerala, India

Joint Statistical Meetings   August 2024
Portland, OR, USA

University of Notre Dame Statistics Department Seminar   February 2024
Notre Dame, IN, USA

University of Wisconsin-Madison Statistics Department Seminar Madison, WI, USA	February 2024
University of Texas at Dallas Statistics Department Seminar Richardson, TX, USA	January 2024
Indian Statistical Institute ISRU Department Seminar Kolkata, West Bengal, India	July 2023
Conference on Learning Theory (COLT) Bangalore, Karnataka, India	July 2023
Neural Information Processing systems (NeurIPS) Virtual	December 2021
Conference on Learning Theory (COLT) Virtual	July 2020

Teaching

University of Notre Dame

Introduction to probability (ACMS 30530)	Fall 2024

Princeton University

Probability and stochastic systems (ORF 309/ENG 309/MAT 380)	Spring 2023
Statistical machine learning (ORF 570)	Fall 2023

Honors and awards

INSPIRE Scholarship, Govt. of India	2012-2017
Indian National Mathematical Olympiad (INMO) merit certificate (For being among top 75 in the country)	2012

Services

Paper reviewer
Annals of Statistics (1)
IEEE Transactions on Information Theory (3)
IEEE International Symposium on Information Theory (1)
Stat - an ISI Journal (1)
Algorithmic Learning Theory(3)
Electronic Journal of Statistics (1)

Organizational duties at conferences

Joint Statistical Meetings *Session chair: New Advances in Nonparametric Hypothesis Testing - Part I* *Session chair: New Developments in Non-Euclidean Statistics*	August 2024
IEEE Conference on Information Sciences and Systems *Session chair: Machine learning and statistical inference*	March 2024

Yale S&DS M.A. admisssion committee   2021
Reviewer: one of the committee members handling over
150 applications and making admission recommendations

Yale S&DS graduate reading group   2020
Co-organizer Scheduled talks and lead discussion sessions

Yale Women in Data Science (WiDS) workshop   2020
Served as a mentor for Yale undergrad students participating
in the WiDS Datathon Challenge 2020

South Asian Graduate and Professional Association   2018 − 2021
at Yale (SAGA)
Treasurer, core committee member and cultural committee head
Objective: organizing socio-cultural events to promote diversity and inclusion at Yale