

# Soham Jana

242 Hayes-Healy Center  
University of Notre Dame  
Notre Dame, IN, USA  
Phone: +1 574-631-5503

Updated on: November 11, 2025  
Website: <https://janasoham.github.io>  
Email: sjana2-at-nd-dot-edu

## Research interests

Theoretical and methodological aspects of high-dimensional statistics, mixture modeling, distance based estimators, neural networks.

## 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 Computational Mathematics and Statistics. August 2024 – Current

### **Princeton University, Princeton, NJ, USA**

Postdoc, Department of Operations Research and Financial Engineering June 2022 – July 2024

Hosts: Prof. Sanjeev Kulkarni and Prof. Jianqing Fan

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

Lecturer Spring 2023 and Fall 2023

## Grants and awards

## Professional Development

Kaneb Center Course Design Academy, University of Notre Dame  
*Award Amount: USD 5000*

2024 – 2025

### Preprints (“\*”: Authors list not in alphabetical order)

1. Fan, J., Jana, S., Kulkarni, S., & Yin, Q. (2025). **Factor Informed Double Deep Learning For Average Treatment Effect Estimation**. arXiv preprint arXiv:2508.17136.
2. Chen, X.<sup>†</sup>, Jana, S.<sup>†</sup>, Metzler, C.A., Maleki, A. and Jalali, S.\*, (2025). **Multilook Coherent Imaging: Theoretical Guarantees and Algorithms**. arXiv preprint arXiv:2505.23594. <sup>†</sup> Equal contributions.
3. Tang, S., Jana, S., & Fan, J. (2024). **Factor adjusted spectral clustering for mixture models**. arXiv preprint arXiv:2408.12564. Under major revision at the **Journal of American Statistical Association**.

### Journal publications (“\*”: Authors list not in alphabetical order)

3. Jana, S., Yang, K., & Kulkarni, S. (2025).\* **Adversarially robust clustering with optimality guarantees**. Accepted at the IEEE Transactions on Information Theory. arXiv preprint arXiv:2306.09977.
2. Jana, S., Polianskiy, Y., & Wu, Y. (2025). **Optimal empirical Bayes estimation for the Poisson model via minimum-distance methods**. Information and Inference: A Journal of the IMA, Volume 14, Issue 4, December 2025, iaaf027.
1. Jana, S., Fan, J., & Kulkarni, S. (2025).\* **A provable initialization and robust clustering method for general mixture models** in IEEE Transactions on Information Theory, vol. 71, no. 9, pp. 7176-7207, Sept. 2025.
0. Jana, S., Li, H., Yamada, Y., & Lindenbaum, O. (2023). **Support recovery with Stochastic Gates: theory and application for linear models**. Elsevier Signal Processing (2023), 213, p.109193.
0. Han, Y., Jana, S., & Wu, Y. (2023). **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)
0. Jana, S. & Basu, A. (2019).\* **A characterization of all single-integral, non-kernel divergence estimators**. IEEE Transactions on Information Theory, 65(12), 7976-7984.

### Conference publications (“\*”: Authors list not in alphabetical order)

15. Jana, S., Polyanskiy, Y., Teh, A. & Wu, Y. (2023). **Empirical Bayes via ERM and Rademacher complexities: the Poisson model**. In Conference on Learning Theory 2023 Jul 15, PMLR 195:5199-5235.
14. Han, Y., Jana, S., & Wu, Y. (2021). **Optimal prediction of Markov chains with and without spectral gap**. NeurIPS 2021.
13. Jana, S., Polyanskiy, & Wu, Y. (2020). **Extrapolating the profile of a finite population**. In Conference on Learning Theory 2020 Jul 15 (pp. 2011-2033). PMLR.

### Conferences and invited talks

1. World Meeting of the International Society for Bayesian Analysis June 2026  
Nagoya, JPN
2. IMS Asia Pacific Rim Meeting June 2026  
Hong Kong
3. Joint Statistical Meetings August 2025  
Nashville, TN, USA
4. IMS New Researchers Conference July-August 2025  
Nashville, TN, USA
5. International Webinar on Recent Trends July 2025  
in Statistical Theory and Applications  
Kerala, India
6. International Indian Statistical Association June 2025  
Lincoln, NEB, USA
7. International Indian Statistical Association December 2024  
Cochin, Kerala, India
8. Joint Statistical Meetings August 2024  
Portland, OR, USA
9. University of Notre Dame Statistics Department Seminar February 2024  
Notre Dame, IN, USA
10. University of Wisconsin-Madison Statistics Department Seminar February 2024  
Madison, WI, USA
11. University of Texas at Dallas Statistics Department Seminar January 2024  
Richardson, TX, USA
12. Indian Statistical Institute ISRU Department Seminar July 2023  
Kolkata, West Bengal, India

13. Conference on Learning Theory (COLT) July 2023  
 Bangalore, Karnataka, India
14. Neural Information Processing systems (NeurIPS) December 2021  
 Virtual
15. Conference on Learning Theory (COLT) July 2020  
 Graz, Austria

## Teaching

### **University of Notre Dame**

- Introduction to probability (ACMS 30530) Fall 2024, Fall 2025
- Modern Machine Learning Techniques with Application (ACMS 80870) Spring 2025

### **Princeton University**

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

## Professional services

### **Paper reviews (26)**

- Annals of Statistics (3)  
 Journal of the American Statistical Association (1)  
 IEEE Transactions on Information Theory (5)  
 Journal of the Royal Statistical Society (1)  
 IEEE Wireless Communications Letters (1)  
 IEEE International Symposium on Information Theory (1)  
 Electronic Journal of Statistics (3)  
 Stat - an ISI Journal (1)  
 Algorithmic Learning Theory (7)  
 Bernoulli (1)  
 Statistica Sinica (1)  
 Journal of Statistical Planning and Inference (1)

### **Invited organizational duties at conferences**

- CFE-CMStatistics Conference* December 2025  
*Session organizer: Recent advances in Causal Inference*
- Joint Statistical Meetings* August 2025  
*Session chair: New Advances in Optimization Algorithms for Causal Discovery*
- Joint Statistical Meetings* August 2024  
*Session chair: New Advances in Nonparametric Hypothesis Testing - Part I*  
*Session chair: New Developments in Non-Euclidean Statistics*
- IEEE Conference on Information Sciences and Systems* March 2024  
*Session chair: Machine learning and statistical inference*

**Community Service: Teaching at Math Circle, Notre Dame** Spring 2025  
*Promoting STEM education among school children*

**Yale S&DS M.A. admission 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 at Yale (SAGA)** 2018 – 2021

Treasurer, core committee member and cultural committee head

Objective: organizing socio-cultural events to promote cultural exchanges at Yale

### Other awards

INSPIRE Scholarship, Govt. of India 2012-2017

Indian National Mathematical Olympiad (INMO) merit certificate 2012  
(For ranking among top 75 in INMO)