Debarghya Mukherjee

Address of department
University of Michigan
West Hall
1085 South University Ave
Ann Arbor, MI 48109
e-mail: mdeb@umich.edu

Home Address
1851, Lake Lila Ln
Apt. B7
Ann Arbor, MI 48105
Phone: (917)328-8926

AREAS OF INTEREST

- High dimensional statistical inference
- Theory of machine learning algorithms
- Empirical process theory
- Non-standard asymptotics

EDUCATION

PhD in Statistics **Fall 2017 - present**
University of Michigan, Ann Arbor
Research Area: Inference on high dimensional classification problems with non-standard asymptotics.
Advisors: Prof. Moulinath Banerjee and Prof. Ya'acov Ritov.

Master of Statistics **August 2017**
Indian Statistical Institute, Kolkata, India

Bachelor of Statistics **August 2015**
Indian Statistical Institute, Kolkata, India

PUBLICATION AND PROJECTS

Publications

- **Optimal Linear Discriminators For The Discrete Choice Model In Growing Dimensions**. (Joint work with Moulinath Banerjee and Ya'acov Ritov, resubmitted to Annals of Statistics upon addressing major revisions). (Arxiv link)
- **Two Simple Ways to Learn Individual Fairness Metrics from Data**. (Joint work with Moulinath Banerjee, Yuekai Sun and Mikhail Yurochkin, accepted in the International Conference of Machine Learning, 2020). (Arxiv link)
- **Markovian And Non-Markovian Processes with Active Decision Making Strategies For Addressing The COVID-19 Pandemic**. (Joint work with Hamid Eftekhari, Moulinath Banerjee and Ya'acov Ritov, resubmitted to Biometrics upon addressing major revisions).(Arxiv link)
- **Efficient estimation of treatment effect under regression discontinuity design**. (Joint work with Moulinath Banerjee and Ya'acov Ritov, submitted to Journal of Royal Statistical Society - Series B).

Ongoing research projects

- **Analysis of change plane model in growing dimension**: (Jointly with Moulinath Banerjee and Ya'acov Ritov). In this paper, we showcase some new asymptotic results of retrospective change-plane estimation problem both in growing dimension (i.e. dimension of the change plane) and in high dimension. A striking feature we observed is that traditional squared error loss is no longer rate optimal when dimension starts to increase with sample size especially in presence of heavy-tailed error, whereas solving robust absolute error loss always gives us minimax optimal rate (upto a log factor). (We are currently preparing our final draft which will be submitted to Journal of Machine Learning and Research.)
- **There is no trade-off: enforcing fairness can improve accuracy**: (Jointly with Subha Maity, Yuekai Sun and Mikhail Yurochkin). In this project, we present necessary and sufficient condition of improving accuracy of an automated system under fairness constraint. (will be submitted to International conference of machine learning, 2021.)

PROFESSIONAL EXPERIENCES

IBM *Cambridge, MA, USA.* (Summer internship) **June 2020 - August 2020**

In light of the recent advances, optimal transport distances are widely used as loss functions in minimum distance estimation. Despite its prevalence and advantages, however, OT is extremely sensitive to outliers. In this work we propose an outlier-robust OT formulation. We proceed by deriving a formulation based on cost truncation that is easy to incorporate into modern stochastic algorithms for regularized OT. We demonstrate our model applied to mean estimation under the Huber contamination model in simulation as well as outlier detection on real data. (This work will be submitted to the International Conference of Machine Learning 2021, here is the arxiv link).

Dynamic Digital Technology *Kolkata, India.* (Summer internship) **May 2016 - July 2016**

The broad goal of this project was to detect intrusion by a foreign user/BOT in real time. A dataset was collected over a period of one month on regular users as well as malicious BOTs, which featured attributes like upload bytes, download bytes, HTTP exchanges between a browser and a remote host etc. We used a Bayesian estimation method to determine anomalous web sessions which helps to detect new attacks without having to know their characteristics in advance (Friedman and Singer, 1999). Our method yielded significantly lesser error rates compared to those implemented previously.

TEACHING EXPERIENCES

- Taught a lab and graded **Stat 250**, an introductory course on Statistics for undergraduate students which covers basic statistical and probability concepts and data analysis techniques (loosely based on the book Mind on Statistics, 4th Edition).
- Graded **Stat 510**, **Stat 610** course. The former is a graduate course on statistics primarily based on Chapters 1-5 from the book Statistical Inference by George Casella and Roger L. Berger while the later is based on Theory of point estimation by Eric Lehmann and George Casella.
- Co-guiding two undergraduate students for **Winter 2021 Undergraduate Research Program in Statistics** jointly with Prof. Ya'acov Ritov.

AWARDS

1. Received **Propelling Original Data Science (PODS) Mini-Grants for COVID-19 Research** which is co-written with Hamid Eftekhari, Moulinath Banerjee and Ya'acov Ritov.
2. Received **Outstanding** 1^{st} **year Ph.D. student award** from Department of Statistics, University of Michigan.
3. Received **ISIAA-MRS M.R. Iyer Gold Medal Award** for being the top scorer in Master of Statistics from Indian Statistical Institute.

PRESENTATIONS

1. Presented my work on **Optimal Linear Discriminators For The Discrete Choice Model In Growing Dimension** (Based on this paper) at Joint Statistical Meeting 2019 and was an
2. **Invited speaker** in 13^{th} international conference on Computational and Methodological Statistics, 2020, presented the topic of this paper.
3. Presented my work on **Two Simple Ways to Learn Individual Fairness Metrics from Data** (based on this paper) in International conference on machine learning, 2020.

SERVICES

1. Served as student seminar organizer in both Winter 2020 and Winter 2021 semesters.
2. Refereed papers for Annals of Statistics and Electronic Journal of Statistics.

COMPUTING SKILLS

1. **Platforms**: Windows, MAC and Linux. 2. **Languages**: Matlab, Python, R, Julia.