

Chakraborty, Abhishek

Curriculum Vitae¹

Address: [Department of Statistics](#), Texas A&M University
3143 TAMU, Blocker Building 416C, College Station, TX 77843, USA
Phone: +1-617-866-7381
Email: abhishek@stat.tamu.edu; achakraborty@tamu.edu
Webpage: <https://statistics.wharton.upenn.edu/profile/abhich/>

Education

Doctor of Philosophy (Ph.D.) in Biostatistics, Harvard University. (8/2011 – 5/2016)

Dissertation: *Robust Semi-Parametric Inference in Semi-Supervised Settings.*

Committee: Dr. Tianxi Cai (advisor), Dr. James Robins and Dr. Eric Tchetgen Tchetgen.

Master of Arts (A.M.) in Biostatistics, Harvard University. (8/2011 – 5/2013)

Master of Statistics (M. Stat.), Indian Statistical Institute. (7/2009 – 5/2011)

Specialization: Mathematical Statistics and Probability.

Bachelor of Statistics (B. Stat.), Indian Statistical Institute. (7/2006 – 5/2009)

Employment

- Assistant Professor – Department of Statistics, Texas A&M University. (9/2019 – present)
- Postdoctoral Researcher – Department of Statistics and Department of Biostatistics, Epidemiology and Informatics (DBEI), University of Pennsylvania (UPenn). (8/2016 – 8/2019)
Mentors: Dr. T. Tony Cai (Department of Statistics) and Dr. Hongzhe Li (DBEI).
- Research Fellow (6/2016 – 7/2016) and Graduate Research Assistant and Teaching Fellow (1/2013 – 5/2016) – Department of Biostatistics, Harvard University. (Advisor: Dr. Tianxi Cai)
- Research Intern – [GE Global Research](#), Bengaluru, India. (5/2009 – 7/2009)

Research

Research Interests

Methodology. Semi-supervised inference; Semi-parametric inference with high dimensional and big data; Missing data and causal inference; High dimensional inference; Regularized estimation.

Applications. Discovery research using electronic medical records data; Automated phenotyping; Personalized medicine (treatment selection, treatment effects estimation, risk prediction etc.)

¹Last updated: August 14, 2019.

Others. Concentration inequalities and tail bounds; Empirical processes; Debiasing and sample-splitting; Model misspecification; Non-parametric regression; Sufficient dimension reduction.

Publications and Preprints

1. **Abhishek Chakraborty** and Tianxi Cai (2018). Efficient and Adaptive Linear Regression in Semi-Supervised Settings. *The Annals of Statistics*, 46(4): 1541 – 1572. ([Journal](#), [pdf](#)) ([ArXiv](#))
 - IMS Travel [Award](#) (2016), Joshua Neimark Travel Assistance [Award](#) from AAAS (2014), and finalist for the ASA Nonparametric Statistics Section Student Paper [Award](#) (2015).
2. Sheng Yu, **Abhishek Chakraborty**, Katherine P. Liao *et. al.* (2017). Surrogate Assisted Feature Extraction for High-Throughput Phenotyping. *Journal of the American Medical Informatics Association (JAMIA)*, 24(e1): e143 – e149. ([Journal](#), [pdf](#))
3. David Cheng, **Abhishek Chakraborty**, Ashwin N. Ananthakrishnan and Tianxi Cai (2018+). Estimating Average Treatment Effects with a Double-Index Propensity Score. *Biometrics (under minor revision)*. (Preprint: [arXiv:1702.01349](#), [pdf](#))
 - ASA Health Policy Statistics Section Student Paper [Award](#) (2017).
4. Stephanie F. Chan, Boris P. Hejblum, **Abhishek Chakraborty** and Tianxi Cai (2019). Semi-Supervised Estimation of Covariance with Application to Phenome-wide Association Studies with Electronic Medical Records Data. *Statistical Methods in Medical Research (to appear)*. (Published online: 04/2019, [online version](#), [pdf](#))

Preprints and Submitted Papers (Under Review or Revision/Resubmission)

5. **Abhishek Chakraborty**, Matey Neykov, Raymond J. Carroll and Tianxi Cai (2018+). Surrogate Aided Unsupervised Recovery of Sparse Signals in Single Index Models for Binary Outcomes. *Under revision (resubmission) at Journal of Machine Learning Research (JMLR)*. (Preprint: [arXiv:1701.05230](#), [pdf](#), updated JMLR version of preprint available upon request)
6. **Abhishek Chakraborty**, Preetam Nandy and Hongzhe Li (2018+). Inference for Individual Mediation Effects and Interventional Effects in Sparse High-Dimensional Graphical Models. *Under revision (resubmission) at Annals of Statistics*. (Preprint: [arXiv:1809.10652](#), [pdf](#))
 - IMS New Researcher Travel [Award](#) (2019).
7. Arun K. Kuchibhotla* and **Abhishek Chakraborty*** (2018+). Moving Beyond Sub-Gaussianity in High Dimensional Statistics: Applications in Covariance Estimation and Linear Regression. *Submitted to Bernoulli*. (Preprint: [arXiv:1804.02605](#), [pdf](#)) [*Equal contributors]
8. **Abhishek Chakraborty** and Tianxi Cai (2019+). A Unified Framework for Robust and Adaptive Z-Estimation in Semi-Supervised Settings. *Final preprint in preparation*. (Initial version can be found in my doctoral dissertation at [dash.harvard.edu/handle/1/33493516](#)).

Manuscripts in Preparation (2019+)

9. High-Dimensional M -Estimation with Missing Outcomes: A Semi-Parametric Framework. (With T. Tony Cai and Hongzhe Li) [[Working draft](#), [Slides](#)]

10. High Dimensional Semi-Supervised Regression: Robust and Adaptive Inference and the Multifold Benefits of Unlabeled Data. (With T. Tony Cai and Hongzhe Li) [In preparation]
11. Tail Bounds for Canonical U -Statistics and U -Processes with Unbounded Kernels. (With Arun K. Kuchibhotla) [[Working draft](#)]

Dissertations

12. **Abhishek Chakrabortty** (2016). *Robust Semi-Parametric Inference in Semi-Supervised Settings*. Doctoral dissertation, Harvard University, Graduate School of Arts & Sciences. ([Link](#), [pdf](#))
13. **Abhishek Chakrabortty** (2011). *Association Mapping of Discrete Phenotypes Using Poisson Regression*. Masters thesis, Indian Statistical Institute. (Advisor: Dr. Saurabh Ghosh) ([pdf](#))

Other Publications and Technical Reports

14. Sian Y. Lim, Sara R. Schoenfeld, **Abhishek Chakrabortty et. al.** (2016). Improving Predictive Value of Gout Case Definitions in Electronic Medical Records Using Natural Language Processing: A Novel Informatics Approach. *Arthritis and Rheumatology* 2016, 68 (Suppl. 10).
15. Saurabh Ghosh and **Abhishek Chakrabortty** (2014). A Poisson Regression Model for Association Mapping of Count Phenotypes. *Molecular Cytogenetics*, 7 (Suppl. 1):O1. ([Link](#))
16. Bhaswar B. Bhattacharya, **Abhishek Chakrabortty**, Shirshendu Ganguly and Shyamalendu Sinha (2009). Visual Cryptographic Schemes for Color Images with Low Pixel Expansion. In *Proceedings of the 9th National Workshop on Cryptology 2009 (Surat, India)*: 64-69. ([pdf](#))
17. Semi-Supervised Sliced Inverse Regression (2016). *Technical report*. (With Tianxi Cai)
18. Causal Effects of Treatments with Metformin and Its Intensification with Insulin or Sulfonylureas on the Time to Cardiovascular Events and All-cause Mortality among Patients with Diabetes: An EMR Study (2015). *Technical report*. (With Tianxi Cai and James Robins)

Awards and Distinctions

1. IMS New Researcher Travel [Award](#) (2019) and IMS Travel [Award](#) (2016), awarded by the Institute of Mathematical Statistics (IMS).
2. Joshua E. Neimark Memorial Travel Assistance [Award](#) (2014), awarded by the American Association for the Advancement of Science (AAAS).
3. NBHM Postgraduate Scholarship (2009-11) for masters studies in mathematical sciences, awarded by the National Board for Higher Mathematics ([NBHM](#)), Government of India.
4. Awards for semestral performances (2006-11) at the Indian Statistical Institute ([ISI](#)), Kolkata.
5. M. P. Birla Sponsorship for Higher Studies (2011), a travel fellowship awarded by the M. P. Birla Foundation, India for doctoral studies abroad.
6. National Merit Scholarship (2004 and 2006), awarded by the Government of India for performances in the state secondary, higher secondary and other competitive examinations.
7. Teaching awards (in recognition of efforts as a teaching assistant at Harvard University):

- (a) Certificate of Teaching Excellence (Spring 2013, Fall 2013 and Spring 2014), [awarded](#) by the Graduate School of Arts and Sciences (GSAS), Harvard University.
 - (b) Certificate of Distinction in Teaching for academic year 2012-13, awarded by the Harvard School of Public Health (HSPH) and the Department of Biostatistics.
8. Finalist for the ASA Nonparametric Statistics Section Student Paper [Award](#) (2015).

Invited Seminar Talks

1. "Semi-Supervised Inference with Large and High Dimensional Data: A Semi-Parametric Perspective" -
 - (a) Department of Statistics, University of California, Irvine. (2/2019)
 - (b) Department of Statistical Sciences, University of Toronto. (2/2019)
 - (c) Brown Data Science Initiative and Department of Biostatistics, Brown University. (2/2019)
 - (d) Department of Statistics, Florida State University. (2/2019)
 - (e) Department of Statistics, University of Pittsburgh. (2/2019)
 - (f) NYU Center for Data Science (CDS) and Courant Institute of Mathematical Sciences (CIMS), New York University (NYU). (1/2019)
 - (g) Department of Statistics, London School of Economics. (1/2019)
 - (h) Department of Biostatistics, University of Washington, Seattle. (1/2019)
 - (i) Department of Biostatistics, University of California, Berkeley. (1/2019)
 - (j) Department of Statistical Science, Cornell University. (12/2018)
 - (k) Department of Statistics, Texas A&M University. (12/2018)
 - (l) Department of Statistics, University of Florida. (12/2018)
2. "Semi-Supervised Estimation and Inference with Big Data like EMR: Safe and Adaptive Approaches" - Causal Inference and Big Data Summer Institute ([CBD](#)), UPenn. (7/2017)
3. "Semi-Supervised Estimation in Big Data: A Unified Framework for Efficient and Adaptive Semi-Parametric Inference" - [Biostatistics Seminar Series](#), DBEI, UPenn. (3/2017)
4. "Efficient and Adaptive Linear Regression in Semi-Supervised Settings" -
 - (a) PoSI Group Meeting, Department of Statistics, University of Pennsylvania. (9/2016)
 - (b) Department of Statistics, University of Michigan - Ann Arbor. (5/2016)
 - (c) Department of Statistics, Purdue University. (4/2016)
 - (d) Department of Biostatistics Student Seminar Series, Harvard University. (8/2015)

Selected Conference Presentations

1. Upcoming talks (invited):
 - (a) Innovations in Data and Statistical Sciences (INDSTATS) 2019 - IISA Annual Meeting 2019, Mumbai, India. (12/2019)
 - (b) Young Statisticians' Meet (YSM) 2020: Data Science in Action, Kolkata, India. (1/2020).

2. “Surrogate Aided Unsupervised Recovery of Sparse Signals in Single Index Models” -
 - (a) IMS New Researchers Conference, Baltimore, USA. (8/2017) (Invited poster session)
 - (b) Joint Statistical Meetings (JSM), Chicago, USA. (8/2016) (Contributed talk)
3. “Efficient and Adaptive Linear Regression in Semi-Supervised Settings” -
 - (a) JSM, Seattle, USA. (8/2015) (Invited talk in the ASA Nonparametric Statistics Section Student Paper Awards session)
 - (b) International Conference on Robust Statistics, Kolkata, India. (1/2015) (Invited talk)
 - (c) AAAS Annual Meeting, Chicago, USA. (2/2014) (Invited poster session and recipient of the Joshua E. Neimark Travel Assistance Award).

Teaching Experience

1. **Instructor** - Department of Statistics, Texas A&M University:
 - (a) STAT 651 (Statistics in Research I) - Fall 2019.
2. **Instructor** - Department of Biostatistics, Harvard University:
 - (a) Operational Mathematics - Summer 2015 (math camp on real analysis and linear algebra for incoming doctoral students). (Responsibilities: lectures and course materials)
3. **Teaching assistant** (TA) - Department of Biostatistics, Harvard University for the following graduate level courses (responsibilities: weekly sections, office hours and grading):
 - (a) BIOSTAT 244 (Analysis of Failure Time Data) - Spring 2013. (Instructor: Dr. Judith Lok) [Overall TA evaluation score: 4.5/5]
 - (b) BIOSTAT 235 (Advanced Regression and Statistical Learning) - Fall 2013. (Instructor: Dr. Robert Gray) [Overall TA evaluation score: 5/5]
 - (c) BIOSTAT 244 (Analysis of Failure Time Data) - Spring 2014. (Instructor: Dr. Judith Lok) [Overall TA evaluation score: 4.9/5]
 - (d) BIOSTAT 235 (Advanced Regression and Statistical Learning) - Fall 2014. (Instructor: Dr. Robert Gray) [Overall TA evaluation score: 4.8/5]

Other Information

Computing experience: Statistical softwares – R, MATLAB; Programming languages – C, C++; Other softwares – Latex, Microsoft Office; OS – Windows, Linux/Unix; Cluster based computing.

Reviewer services: The Annals of Statistics, Journal of the Royal Statistical Society: Series B (JRSS B), Biometrika, Journal of Machine Learning Research (JMLR), eural Information Processing Systems (NIPS) and Journal of the American Medical Informatics Association (JAMIA).

Academic memberships: IMS (2014 - present), ASA (2014 - present) and AAAS (2013-2016).

Personal information: Citizenship - Indian; Languages spoken - English, Hindi and Bengali; Extracurricular interests - Cricket, Table tennis, Swimming, Music and Movies.

Other links: [Google Scholar](#) profile, [ResearchGate](#) profile, [LinkedIn](#) profile.

Last updated: August 14, 2019.