Raaz Dwivedi

(510) 833-1977 raaz@seas.harvard.edu raaz@mit.edu rzrsk.github.io

---

Academic appointments

Post-doctoral Fellow *2021—*

Harvard University, School of Engineering and Applied Sciences, Boston, USA
Massachusetts Institute of Technology, Department of EECS, Cambridge, USA

- Advisors: *Prof. Susan Murphy & Prof. Devavrat Shah*

Education

Ph.D., Electrical Engineering and Computer Sciences *2015—2021*

University of California, Berkeley, USA

- Advisors: *Prof. Martin Wainwright & Prof. Bin Yu*
- Thesis committee members: *Prof. David Aldous & Prof. Peter Bartlett*
- Other major collaborators: *Lester Mackey, Prof. David Madigan, & Prof. Michael Jordan*

B. Tech., Electrical Engineering *2010—2014*

Indian Institute of Technology, Bombay, India

- Advisor: *Prof. Vivek Borkar*
- Graduated with Honors in EE and Minors in Mathematics
- Secured Institute Rank 1 (amongst a thousand)

Research Interests

Theoretical and applied aspects of statistical machine learning and data science with a focus on

- Theory of high-dimensional statistics, and MCMC methods
- Heterogeneity estimation in causal inference
- Reproducible methodologies for interpretable machine learning

Achievements & Awards

Outstanding Graduate Student Instructor Award, UC Berkeley *Berkeley, 2020*

Student Travel Award, NeurIPS 2018 *Canada, 2018*

Oberwolfach Leibniz Graduate Students Travel Grant *Germany, 2017*

Student Travel Award, SAMSI QMC Workshop 2018 *Raleigh-Durham, 2017*

Berkeley Fellowship, the most prestigious fellowship for incoming students *Berkeley, 2015*

President of India Gold Medal, IIT Bombay, for highest GPA in the institute *India, 2014*

Institute Silver Medal, IIT Bombay, for highest Honors GPA in the EE department *India, 2014*

Best B. Tech. Project Award, IIT Bombay *India, 2014*

All India Rank 10 (amongst half a million), IIT Joint Entrance Exam (IIT-JEE) *India, 2010*

All India Rank 46 (amongst a million), All India Engineering Entrance Exam *India, 2010*

Work Experience

Microsoft Research, Research Intern (with Lester Mackey), New England, USA *Summer 2019*

Mist Systems (Juniper Networks), Data Science Intern, Cupertino, USA *Summer 2017*

WorldQuant Research, Senior Quantitative Researcher, Mumbai, India *2014—2015*

Stanford University, Research intern (with Prof. Balaji Prabhakar), USA *Summer 2013*

Ivy Mobility, Data Science Intern, Chennai, India *Winter 2012*

Journal Publications

($\star$ denotes equal contribution, and † denotes alphabetical ordering)

J1. **Raaz Dwivedi***, Yan Shuo Tan*, Briton Park, Mian Wei, Kevin Horgan, David Madigan, and Bin Yu, "Stable discovery of interpretable subgroups via calibration in causal studies", *International Statistical Review (ISR), 2020.*

J2. Nick Altieri†, Rebecca L. Barter, James Duncan, **Raaz Dwivedi**, Karl Kumbier, Xiao Li, Robert Netzorg, Briton Park, Chandan Singh, Yan Shuo Tan, Tiffany Tang, Yu Wang, Chao Zhang and Bin Yu, "Curating a COVID-19 data repository and forecasting county-level death counts in the United States", *Harvard Data Science Review (HDSR), 2020.*

J3. **Raaz Dwivedi***, Nhat Ho*, Koulik Khamaru*, Martin J. Wainwright, Michael I. Jordan and Bin Yu, "Singularity, misspecification, and the convergence rate of EM", *Annals of Statistics (AoS), 2020.*

J4. Yuansi Chen, **Raaz Dwivedi**, Martin J. Wainwright and Bin Yu, "Fast mixing of Metropolized Hamiltonian Monte Carlo: Benefits of multi-step gradients", *Journal of Machine Learning Research (JMLR), 2020.*

J5. **Raaz Dwivedi***, Yuansi Chen*, Martin J. Wainwright and Bin Yu, "Log-concave sampling: Metropolis-Hastings algorithms are fast", *Journal of Machine Learning Research (JMLR), 2019.*

J6. **Raaz Dwivedi**†, Ohad N. Feldheim, Ori Gurel-Gurevich and Aaditya Ramdas, "The power of online thinning in reducing discrepancy", *Probability Theory and Related Fields (PTRF), 2019.*

J7. Yuansi Chen*, **Raaz Dwivedi***, Martin J. Wainwright and Bin Yu, "Fast MCMC sampling algorithms on polytopes", *Journal of Machine Learning Research (JMLR), 2018.*

J8. Vivek Borkar†, **Raaz Dwivedi** and Neeraja Sahasrabudhe, "Gaussian approximations in high dimensional estimation", *Systems & Control Letters, 2016.*

Conference Publications

C1. **Raaz Dwivedi** and Lester Mackey, "Kernel thinning", Extended abstract in *Conference on Learning Theory (COLT), 2021.*

C2. **Raaz Dwivedi***, Nhat Ho*, Koulik Khamaru*, Martin J. Wainwright, Michael I. Jordan and Bin Yu, "Sharp analysis of Expectation-Maximization for weakly identifiable models", *The 23rd International Conference on Artificial Intelligence and Statistics (AISTATS), 2020.*

C3. **Raaz Dwivedi***, Nhat Ho*, Koulik Khamaru*, Martin J. Wainwright and Michael I. Jordan, "Theoretical guarantees for EM under misspecified Gaussian mixture models", *Advances in Neural Information Processing Systems (NeurIPS), Montreál, 2018.*

C4. **Raaz Dwivedi***, Yuansi Chen*, Martin J. Wainwright and Bin Yu, "Log-concave sampling: Metropolis-Hastings algorithms are fast", Extended abstract in *Conference on Learning Theory (COLT), Stockholm, 2018.*

C5. Yuansi Chen*, **Raaz Dwivedi***, Martin J. Wainwright and Bin Yu, "Vaidya walk: A sampling algorithm based on the volumetric barrier", *Communication, Control, and Computing (Allerton), 55th Annual Allerton Conference, 2017.*

C6. **Raaz Dwivedi** and Vivek Borkar, "Removing sampling bias in networked stochastic approximation", *International Conference on Signal Processing and Communications (SPCOM), Bangalore, 2014.*

Pre-Prints

P1. **Raaz Dwivedi***, Chandan Singh*, Bin Yu and Martin J. Wainwright, "Revisiting complexity and the bias-variance tradeoff", *arXiv preprint.*

P2. Nhat Ho*, Koulik Khamaru*, **Raaz Dwivedi***, Martin J. Wainwright, Michael I. Jordan and Bin Yu, "Instability, computational efficiency, and statistical accuracy", *arXiv preprint.*

P3. **Raaz Dwivedi** and Lester Mackey, "Generalized kernel thinning", *arXiv preprint.*

P4. Abhishek Shetty, **Raaz Dwivedi** and Lester Mackey, "Distribution compression in near-linear time", *arXiv preprint.*

Research Talks

T1. Revisiting Minimum Description Length Complexity in Overparameterized Models. *Collaborations on the Theoretical Foundations of Deep Learning.* (Invited talk) *Nov 2021*

T2. Non-asymptotic Guarantees for MCMC and Kernel Thinning. *Finale Doshi-Velez Group Meeting, Harvard University.* (Invited talk) *Oct 2021*

T3. Kernel Thinning. *The Data-Centric Engineering Reading Group (DCE), Alan Turing Institute.* (Invited talk) *Sep 2021*

T4. Kernel Thinning. *Stat 300, Harvard University.* *Sep 2021*

T5. Kernel Thinning. *Monte Carlo Methods & Applications (MCM).* (Contributed talk) *Sep 2021*

T6. Kernel Thinning. *2021 World Meeting of the International Society for Bayesian Analysis (ISBA).* (Contributed talk) *Aug 2021*

T7. Kernel Thinning. *The Bayesian Young Statisticians Meeting (BAYSM) 2021.* (Contributed talk) *Aug 2021*

T8. Kernel Thinning. *Conference on Learning Theory (COLT).* (Contributed talk) *Aug 2021*

T9. Kernel Thinning. *Subset Selection in Machine Learning Workshop, International Conference on Machine Learning (Subset ML, ICML).* (Contributed talk) *July 2021*

T10. Revisiting Complexity and the Bias-Variance Tradeoff: Using Minimum Description Length. *North American School of Information Theory (NASIT).* (Contributed poster) *June 2021*

T11. Revisiting Complexity and the Bias-Variance Tradeoff: Using Minimum Description Length. *Workshop on the Theory of Overparameterized Machine Learning (TOPML).* (Contributed talk) *Apr 2021*

T12. Revisiting Complexity and the Bias-Variance Tradeoff: Using Minimum Description Length. *Stat 212, UC Berkeley.* (Guest Lecture) *Apr 2021*

T13. Subgroup Discovery in Randomized Experiments & Markov Chain Monte Carlo Sampling. *Research Seminar, USC Marshall School of Business.* (Invited talk) *Feb 2021*

T14. Subgroup Discovery in Randomized Experiments & Markov Chain Monte Carlo Sampling. *Statistics Seminar, University of Toronto.* (Invited talk) *Feb 2021*

T15. Subgroup Discovery in Randomized Experiments & Markov Chain Monte Carlo Sampling. *MINDS Symposium on the Foundations of Data Science, John Hopkins University.* (Invited talk) *Feb 2021*

T16. Subgroup Discovery in Randomized Experiments & Markov Chain Monte Carlo Sampling. *Devavrat Shah and Susan Murphy Group Meetings, MIT and Harvard University.* (Invited talk) *Feb 2021*

T17. Subgroup Discovery in Randomized Experiments & Markov Chain Monte Carlo Sampling. *Research Seminar, Microsoft Research New England.* (Invited talk) *Jan 2021*

T18. Non-asymptotic Guarantees for Markov Chain Monte Carlo. *Flatiron Institute Seminar.* (Invited talk) *Jan 2021*

T19. Subgroup Discovery in Randomized Experiments & Markov Chain Monte Carlo Sampling. *Statistics Seminar, University of Washington.* (Invited talk) *Jan 2021*

T20. Subgroup Discovery in Randomized Experiments & Markov Chain Monte Carlo Sampling. *Operations Research and Statistics Group Seminar, MIT Sloan.* (Invited talk) *Jan 2021*

T21. New Perspectives on Old Problems in Causal Inference and MCMC Sampling. *Statistics Seminar, UC Irvine.* (Invited talk) *Jan 2021*

T22. StaDISC: Stable discovery of interpretable subgroups via calibration. *Young Data Scientist Research Seminar, ETH Zurich.* (Invited talk) *Sep 2020*

T23. Veridical Data Science and the PCS Framework. *ASA Annual Symposium on Data Science and Statistics (SDSS).* (Invited talk) *Jun 2020*

T24. Statistics Meets Optimization: Two Vignettes on The Intersection, *Department of Mathematics and Statistics, IIT Kanpur, India.* (Invited talk) *Jan 2020*

T25. Singularity, misspecification and the convergence rate of Expectation-Maximization. *Fall Western Sectional Meeting of the AMS, UC Riverside.* (Invited talk) *Nov 2019*

T26. Power of gradients and accept-reject step in MCMC algorithms. *BIDS Statistics and Machine Learning Forum, UC Berkeley.* (Invited Talk) *Mar 2019*

T27. Log-concave sampling: Metropolis Hastings algorithms are fast. *Conference on Learning Theory (COLT) 2018, Stockholm, Sweden.* (Conference Poster) *Dec 2018*

T28. Log-concave sampling: Metropolis Hastings algorithms are fast. *Jerusalem Joint Statistical Event, Israel.* (Contributed talk) *Dec 2018*

T29. Theoretical guarantees for EM under misspecified Gaussian mixture models. *Neural Information Processing Systems (NeurIPS) 2018, Montreál, Canada.* (Conference Poster) *Dec 2018*

T30. Theoretical Guarantees for MCMC Algorithms, *Department of Electrical Engineering, IIT Bombay, India.* (Invited talk) *Jan 2018*

T31. Theoretical Guarantees for MCMC Algorithms, *School of Technology and Computer Science Seminar, TIFR Bombay, India.* (Invited talk) *Jan 2018*

T32. The power to two choices in reducing discrepancy, *SAMSI QMC Opening Workshop, Raleigh-Durham, Duke University.* (Contributed Poster) *Aug 2017*

Teaching Experience

Graduate Student Instructor, UC Berkeley *2018—2019*

- *EECS 189, Spring 2018*: Introduction to Machine Learning, taught by Prof. Anant Sahai and Prof. Jennifer Listgarten. *Co-led* the content development (homeworks, discussions and exams) in a team of 20+ TAs in a class of 350+ students.
- Spring 2019 for *STAT 154*: Modern Statistical Prediction and Machine Learning taught by Prof. Bin Yu. Helped in *redesigning* the class along with one other TA, Yuansi Chen, for a class of 140+ students.
- **Guest lectures**: (i) *STAT 154, Spring 2019*: "Boosting" (ii) *EECS 189, Fall 2019*: "An Introduction to Ensemble Methods: Bagging, Random Forest, Boosting".

Teaching Assistant, IIT Bombay and MHRD, India *2011—2014*

- 9-time TA for undergraduate courses on *Calculus, Linear Algebra, Differential Equations, and Electromagnetism*. Responsible for weekly discussions (40 students) besides exam grading.
- TA for an online course on *Linear Algebra* (for 400 undergraduate colleges) organized by Ministry of Human Resource Development (MHRD) of Government of India.

Academic Services

Mentoring Activities

- BAIR PhD Buddy Program for *incoming graduate students*, UC Berkeley *2020—2021*
- BAIR UG Mentoring Program for *undergraduates*, UC Berkeley *2017—2021*
- Institute Student Mentoring Program for *incoming undergraduates*, IIT Bombay *2013—2014*
- EE Academic Mentoring Program for *sophomores and juniors*, IIT Bombay *2012—2014*
- Intensive Mentoring Program for selected *incoming undergraduates*, IIT Bombay *2012—2013*

Reviewing Activities

- Journal of Machine Learning Research (JMLR) (4 Papers) *2020-*
- IEEE Transactions on Information Theory (4 Papers) *2020-*
- Bernoulli (1 Paper) *2021-*
- International Conference on Machine Learning (ICML) *2019, 2020*
- Neural Information Processing Systems (NeurIPS) *2019, 2020*
- Conference on Learning Theory (COLT) *2019*
- Foundations of Computer Science (FOCS) *2018, 2020*
- Symposium on Discrete Algorithms (SODA) *2019*
- AAAI Conference on Artificial Intelligence *2020*

EECS Graduate Admissions Student Committee, UC Berkeley *2018, 2019, 2020*