Raaz Dwivedi

Cornell raazdwivedi.github.io dwivedi@cornell.edu

Academic appointments

Assistant Professor, Operations Research & Information Engineering (ORIE) *2024—*

Field Member: Applied Math, Computer Science, ORIE, Statistics
Cornell Tech, Cornell University

Visiting Assistant Professor, ORIE, Cornell University *Fall 2023*

FODSI Postdoctoral Fellow, CS, Statistics, EECS *2021–2023*

Harvard University & Massachusetts Institute of Technology (MIT)
Advisors: *Prof. Susan Murphy & Prof. Devavrat Shah*

Education

Ph. D., EECS, University of California (UC), Berkeley *2015—2021*

Advisors: *Prof. Martin Wainwright & Prof. Bin Yu*
Thesis title: *Principled statistical approaches for sampling and inference in high dimensions*

B. Tech., EE, Indian Institute of Technology (IIT), Bombay, India *2010—2014*
Minors in mathematics, Institute Rank 1

Research Interests

My research involves a multi-disciplinary approach to data science and brings together ideas from computer science, electrical engineering, and statistics in collaboration with domain experts. I develop statistical machine learning approaches for data-driven personalized decision-making with research across *causal inference, reinforcement learning, Bayesian inference, random sampling, and high-dimensional statistics.*

Selected Achievements & Awards

- Best Student Paper Award, Statistical Computing & Graphics, American Statistical Association *2022*
- Best Presentation Award, Laboratory of Information & Decision Systems (LIDS) Conference, MIT *2022*
- Certificate of Distinction and Excellence in Teaching (Q Award), Harvard University *2022*
- Foundations of Data Science (FODSI) Postdoctoral Fellowship *2021*
- Outstanding Graduate Student Instructor Award, UC Berkeley *2020*
- Berkeley Fellowship, the most prestigious fellowship for incoming Ph. D. students *2015*
- President of India Gold Medal, IIT Bombay, for the highest GPA in the graduating class *2014*
- All India Rank 10 amongst half a million, IIT Joint Entrance Exam *2010*

Pre-Prints

(⋆ denotes equal contribution and † denotes alphabetical ordering; title is hyperlinked to the online pdf of the paper)

P1. Alberto Abadie†, Anish Agarwal, **Raaz Dwivedi**, Abhin Shah, "Doubly Robust Inference in Causal Latent Factor Models", *arxiv*. *2024*

P2. Jane Dwivedi-Yu, **Raaz Dwivedi**, Timo Schick, "FairPair: A Robust Evaluation of Biases in Language Models through Paired Perturbations", *arxiv*. *2024*

P3. **Raaz Dwivedi**, Katherine Tian, Sabina Tomkins, Predrag Klasnja, Susan Murphy, Devavrat Shah, "Doubly robust nearest neighbors in factor models", *arxiv*. *2023*

P4. **Raaz Dwivedi**, Katherine Tian, Sabina Tomkins, Predrag Klasnja, Susan Murphy, Devavrat Shah, "Counterfactual inference in sequential experimental design", *arxiv*. *2022*

P5. Abhin Shah, **Raaz Dwivedi**, Devavrat Shah, Greg Wornell, "On counterfactual inference with unobserved confounding", *arxiv*. *2022*

Conference Publications

C1. Lingxiao Li, **Raaz Dwivedi**, Lester Mackey, "Debiased Distribution Compression", *International Conference on Machine Learning (ICML).* *2024*

C2. Carles Domingo-Enrich, **Raaz Dwivedi**, Lester Mackey, "Compress then test: Powerful kernel testing in near-linear time", *Conference on Artificial Intelligence and Statistics (AISTATS).* *2023*

C3. **Raaz Dwivedi**, Lester Mackey. "Generalized kernel thinning", *International Conference on Learning Representations (ICLR).* *2022*

C4. Abhishek Shetty, **Raaz Dwivedi**, Lester Mackey. "Distribution compression in near-linear time", *International Conference on Learning Representations (ICLR)*, **Best Student Paper Award, JSM**. *2022*

C5. **Raaz Dwivedi**, Lester Mackey, "Kernel thinning", Extended abstract in *Conference on Learning Theory (COLT). Full version under review in JMLR.* *2021*

C6. **Raaz Dwivedi***, Nhat Ho*, Koulik Khamaru*, Martin J. Wainwright, Michael I. Jordan, Bin Yu, "Sharp analysis of Expectation-Maximization for weakly identifiable models", *The 23rd International Conference on Artificial Intelligence and Statistics (AISTATS).* *2020*

C7. **Raaz Dwivedi***, Nhat Ho*, Koulik Khamaru*, Martin J. Wainwright, Michael I. Jordan, "Theoretical guarantees for EM under misspecified Gaussian mixture models", *Advances in Neural Information Processing Systems (NeurIPS).* *2018*

C8. **Raaz Dwivedi***, Yuansi Chen*, Martin J. Wainwright, Bin Yu, "Log-concave sampling: Metropolis-Hastings algorithms are fast", Extended abstract in *Conference on Learning Theory (COLT).* *2018*

C9. Yuansi Chen*, **Raaz Dwivedi***, Martin J. Wainwright, Bin Yu, "Vaidya walk: A sampling algorithm based on the volumetric barrier", *Allerton Conference.* *2017*

C10. **Raaz Dwivedi**, Vivek Borkar, "Removing sampling bias in networked stochastic approximation", *International Conference on Signal Processing and Communications (SPCOM).* *2014*

Journal Publications

J1. **Raaz Dwivedi**, Lester Mackey, "Kernel thinning", Extended abstract in *Journal of Machine Learning Research (JMLR).* *2024*

J2. Raphael Kim, Susobhan Ghosh, Prasidh Chhabria, **Raaz Dwivedi**, Peng Liao, Kelly Zhang*, Predrag Klasnja, Susan Murphy, "Did we personalize? Assessing personalization by an online reinforcement learning algorithm using resampling", *Machine Learning Journal.* *2024*

J3. **Raaz Dwivedi***, Chandan Singh*, Bin Yu, Martin J. Wainwright, "Revisiting minimum description length complexity in overparameterized models", *JMLR.* *2023*

J4. Nhat Ho*, Koulik Khamaru*, **Raaz Dwivedi***, Martin J. Wainwright, Michael I. Jordan, Bin Yu, "Instability, computational efficiency, and statistical accuracy", *accepted with minor revision at JMLR.* *2022*

J5. 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, Bin Yu, "Curating a COVID-19 data repository and forecasting county-level death counts in the United States", *Harvard Data Science Review (HDSR).* *2021*

J6. **Raaz Dwivedi***, Yan Shuo Tan*, Briton Park, Mian Wei, Kevin Horgan, David Madigan, Bin Yu, "Stable discovery of interpretable subgroups via calibration in causal studies", *Int. Statistical Review.* *2020*

J7. **Raaz Dwivedi***, Nhat Ho*, Koulik Khamaru*, Martin J. Wainwright, Michael I. Jordan, Bin Yu, "Singularity, misspecification, and the convergence rate of EM", *Annals of Statistics (AoS).* *2020*

J8. Yuansi Chen, **Raaz Dwivedi**, Martin J. Wainwright, Bin Yu, "Fast mixing of Metropolized Hamiltonian Monte Carlo: Benefits of multi-step gradients", *Journal of Machine Learning Research (JMLR).* *2020*

J9. **Raaz Dwivedi***, Yuansi Chen*, Martin J. Wainwright, Bin Yu, "Log-concave sampling: Metropolis-Hastings algorithms are fast", *Journal of Machine Learning Research (JMLR).* *2019*

J10. **Raaz Dwivedi**†, Ohad N. Feldheim, Ori Gurel-Gurevich, Aaditya Ramdas. "The power of online thinning in reducing discrepancy", *Probability Theory and Related Fields (PTRF).* *2019*

J11. Yuansi Chen*, **Raaz Dwivedi***, Martin J. Wainwright, Bin Yu. "Fast MCMC sampling algorithms on polytopes", *Journal of Machine Learning Research (JMLR).* *2018*

J12. Vivek Borkar†, **Raaz Dwivedi**, Neeraja Sahasrabudhe. "Gaussian approximations in high dimensional estimation", *Systems & Control Letters.* *2016*

Softwares & Methodologies

S1. Carles Domingo-Enrich, **Raaz Dwivedi**, Lester Mackey. Python package "Compress then test" (link).

S2. Abhishek Shetty*, **Raaz Dwivedi***, Lester Mackey. Python package "Compress++" (link).

S3. **Raaz Dwivedi**, Lester Mackey. Python package "Kernel Thinning" (link).

S4. **Raaz Dwivedi***, Yan Shuo Tan*, Briton Park, Mian Wei, Kevin Horgan, David Madigan, Bin Yu. Python repository "StaDISC" (link).

S5. Yuansi Chen*, **Raaz Dwivedi***, Martin Wainwright, Bin Yu. Python package (with C++ implementation) "Vaidya and John walks" (link).

Selected Invited Talks

Integrating Double Robustness Into Causal Latent Factor Models

- Rising Stars in AI, KAUST *Feb 2024*
- Statistics Seminar, Columbia *Mar 2024*
- Workshop on Statistical Methods for Digital Health, John Hopkins University *Mar 2024*
- Operations Research Seminar, MIT *Apr 2024*
- Online Causal Inference Seminar *May 2024*
- Tom Ten Have Symposium on Mental Health Statistics, Weil Cornell *Scheduled Jun 2024*
- Mini Workshop on Individualized Decisions, Simons/UC Berkeley *Scheduled July 2024*
- ESIF conference on Economics and AI+ML *Scheduled July 2024*
- Joint Statistical Meeting (JSM), Toronto *Scheduled Aug 2024*
- Informs, Seattle *Scheduled Oct 2024*

From HeartSteps to HeartBeats: Personalized Decision-making

-
- Large Scale Learning and Control Workshop, IIT Bombay *Dec 2023*
- AI Seminar, Cornell University *Sep 2023*
- ORIE Industry and Data Science Summit, Cornell University *Sep 2023*
- Statistics and Data Science Seminar, Cornell University *Sep 2023*
- Center for Applied Math Colloquium, Cornell University *Sep 2023*
- Gatsby Unit Seminar, University College London *Feb 2023*
- Statistics and Data Science Seminar, Yale University *Feb 2023*
- Computer Science Seminar, UIUC *Feb 2023*
- Statistics Seminar, UW Madison *Jan 2023*
- Operations, Information, and Technology Seminar, GSB, Stanford University *Jan 2023*
- Statistics and Data Science Seminar, Wharton, University of Pennsylvania *Jan 2023*
- Statistics Seminar, University of Chicago *Jan 2023*
- Statistics and Operation Research Seminar, UNC Chapel Hill *Jan 2023*
- Statistics Seminar, UCLA *Jan 2023*
- Operation Research and Industrial Engineering Seminar, Cornell University *Dec 2022*
- Operation Research and Industrial Engineering Seminar, Cornell Tech *Dec 2022*
- Statistics Seminar, Rutgers University *Nov 2022*
- ISL Colloquium, EE, Stanford University *Nov 2022*
- BLISS Seminar, EECS, UC Berkeley *Nov 2022*

Compress then test: Powerful kernel testing in near-linear time

- Joint Statistical Meeting, Toronto *Jun 2023*
- Monte Carlo Methods Conference, Paris *Jun 2023*
- Computational-Statistical Interplay in Machine Learning Workshop, MIT *May 2023*

Doubly robust nearest neighbors for counterfactual inference

- Causal Inference Workshop, ACM Sigmetrics, Orlando *Jun 2023*
- New England Statistics Symposium, Boston University *Jun 2023*
- Informs Annual Meeting, Indianapolis *Oct 2022*

Counterfactual inference in sequential experiments

- Institute of Mathematical Statistics (IMS) Annual Meeting, London *Jun 2022*
- Learning from Interventions Workshop, Simons Institute, Berkeley *Feb 2022*

Near-optimal compression in near-linear time

- SIAM Conference on Uncertainty Quantification, Atlanta *Apr 2022*
- Statistical learning Workshop, Mathematical Sciences Research Institute, Berkeley *Mar 2022*

Kernel thinning

- Data-Centric Engineering Group, Alan Turing Institute, Virtual *Sep 2021*

Revisiting minimum description length complexity in overparameterized models

- Alg. Info Theory & Machine Learning Symp., Alan Turing Institute, London *Jul 2022*
- Collaborations on the Theoretical Foundations of Deep Learning, Virtual *Nov 2021*

StaDISC: Stable discovery of interpretable subgroups via calibration

- Young Data Scientist Research Seminar, ETH Zurich, Virtual *Sep 2020*
- ASA Annual Symposium on Data Science & Statistics, Virtual *Jun 2020*

Singularity, misspecification, & the convergence rate of EM

- Math & Statistics Seminar, IIT Kanpur *Jan 2020*
- AMS Special Sections Meeting, UC Riverside *Nov 2019*

Theoretical guarantees for MCMC algorithms

- BIDS Seminar, UC Berkeley *Mar 2019*
- EE Seminar, IIT Bombay *Jan 2018*
- STCS Seminar, TIFR Bombay *Jan 2018*

Contributed & Other Research Talks

Counterfactual inference in sequential experiments

- Informs APS Meeting, Nancy, France *Jun 2023*
- Statistics and data science conference (SDSCON), MIT *Apr 2022*
- Econometrics Lunch, MIT *Mar 2022*

Near-optimal compression in near-linear time

- LIDS Student Conference, MIT, **Best presentation award** *Jan 2022*

Generalized kernel thinning

- Joint Statistical Meeting (JSM), Washington DC *Aug 2022*

Kernel thinning

- Monte Carlo Methods & Applications (MCM), Virtual *Sep 2021*
- International Society for Bayesian Analysis (ISBA) World Meeting, Virtual *Aug 2021*

- The Bayesian Young Statisticians Meeting (BAYSM), Virtual *Aug 2021*
- Joint Statistical Meeting (JSM), Virtual *Aug 2021*
- Conference on Learning Theory (COLT), Virtual *Aug 2021*
- Subset Selection, International Conference on Machine Learning (ICML), Virtual *Jul 2021*

Revisiting complexity and the bias-variance tradeoff: Using minimum description length

- Theory of Overparameterized Machine Learning (TOPML) Workshop, Virtual *Apr 2021*

Converging fast and slow: Statistics vs optimization

- BAIR and BDD Retreat, Berkeley, Virtual *Aug 2020*

Log-concave sampling: Metropolis Hastings algorithms are fast

- Joint Statistical Meeting (JSM), Washington DC *Dec 2018*

Vaidya walk: A sampling algorithm based on the volumetric barrier

- Allerton Conference *Oct 2017*

Contributed Poster Presentations

Compress then test: Powerful kernel testing in near-linear time

- Conference on Artificial Intelligence and Statistics (AISTATS) Conference, Spain *Apr 2023*

Doubly robust nearest neighbors for counterfactual inference

- American Causal Inference Conference (ACIC), Austin *May 2023*

On counterfactual inference with unobserved confounding

- American Causal Inference Conference (ACIC), Austin *May 2023*
- NeurIPS Causality for Real world impact workshop, New Orleans *Nov 2022*

Counterfactual inference in sequential experiments

- Cornell ORIE Young Researchers Workshop, Ithaca *Oct 2022*
- Royal Statistical Society (RSS) Conference, Aberdeen, Scotland *Sep 2022*
- Synthetic Control Methods Workshop, Data X, Princeton University *Jun 2022*
- American Causal Inference Conference (ACIC), UC Berkeley *May 2022*
- Symposium for Mathematical Sciences (SMaSH), Harvard University *May 2022*
- Statistics and data science conference (SDSCON), MIT *Apr 2022*

Near-optimal compression in near-linear time

- Royal Statistical Society (RSS) Conference, Aberdeen, Scotland *Sep 2022*

Generalized kernel thinning

- Advances in Approximate Bayesian Inference (AABI), Virtual *Feb 2022*

Revisiting minimum description length complexity in overparameterized models

- North American School of Information Theory (NASIT), Virtual *Jun 2021*

Theoretical guarantees for EM under misspecified Gaussian mixture models

- Neural Information Processing Systems (NeurIPS), Montreál, Canada *Dec 2018*

Log-concave sampling: Metropolis Hastings algorithms are fast

- Conference on Learning Theory (COLT), Stockholm, Sweden *Jul 2018*

On power of two choices in reducing discrepancy

- SAMSI Workshop, Duke University, Raleigh *Aug 2017*

PhD Students

Albert Gong, First Year, CS *2024–*

Kyuseong Choi, Third Year, Statistics *2024–*

Minor Committee Member

Brian Cho, Second Year, ORIE *2024–*

Teaching Experience

T1. Causal Inference (ORIE 7790), *Cornell University* *Spring 2024*

T2. Instructor: Statistical Principles (ORIE 6700), *Cornell University* *Fall 2023*

T3. Instructor: Statistical RL for real life (one week; link), *CDT Summer School, Missenden* *Jul 2023*

T4. TA: Sequential Decision Making (STAT 234), *Harvard University*. Gave four guest lectures and supervised several half-semester long research projects. *Spring 2022*

T5. TA: Modern Statistical Prediction and Machine Learning (STAT 154), *UC Berkeley*. Gave one guest lecture and helped in redesign of the class. *Spring 2019*

T6. TA: Introduction to Machine Learning (EECS 189), *UC Berkeley*. Co-head for the content developments in team of 10+ TAs, helped design discussion sections, homeworks, and exams. *Spring 2018*

T7. TA: Linear Algebra, Calculus, Differential equations (MA 105, 106, 108, 207), *IIT Bombay*. Taught teaching sections and several voluntary help sessions that were often attended by 200+ students. *2011–2014*

Guest Lectures

L1. Regret analysis of posterior sampling (*3 lectures*, STAT 234), *Harvard University* *Apr 2022*

L2. Offline off-policy reinforcement learning (STAT 234) *Harvard University*. *Feb 2022*

L3. Revisiting complexity and the bias-variance tradeoff (STAT 212) *UC Berkeley*. *Apr 2021*

L4. Introduction to ensemble methods in machine learning (EECS 189), *UC Berkeley*. *Oct 2019*

L5. Introduction to boosting methods (STAT 154), *UC Berkeley*. *Apr 2019*

Academic Services

Undergraduate Research Mentoring

- UC Berkeley, One student that led to a co-authored journal publication *2020—2021*
- Harvard, Two students with three co-authored submissions in preparation *2022—*

Institutional Mentoring

- MIT Institute for Data, Systems, & Society (IDSS) Postdoc Mentors for two *PhD* students *2022—*
- UC Berkeley Artificial Intelligence Research (BAIR) Buddies for two *incoming PhD* students *2020—2021*
- UC Berkeley BAIR Mentoring Program for five *undergraduates* *2017—2021*
- IIT Bombay Student Mentoring Program (ISMP) for twelve *incoming undergraduates* *2013—2014*
- IIT Bombay Academic Mentoring Program (DAMP) for four *sophomores & juniors* *2012—2014*
- IIT Bombay Intensive Mentoring Program for thirty *undergraduates* *2012—2013*

Committees

- Member, Committee on Equality and Diversity, IMS *2022—*

Scientific Meetings

- Organizer and chair, Informs Session on Causal inference and reinforcement learning *2023*
- Organizer and chair, Informs Session on Statistical Methods for Healthcare *2023*
- Mentor, Let-All Mentoring Session, Learning Theory Mentorship Workshop *2023*
- Moderator, Panel Discussion on Mentoring, New Researcher Conference Statistics, Toronto *2023*
- Chair, New Researchers Group Session, IMS Annual Meeting *2022*
- Chair, Statistical Machine Learning Session, IMS Annual Meeting *2022*

- Mentor, Summer Institute on Just-in-Time Adaptive Interventions via MRTs *2021*

Graduate Admissions

- ORIE Graduate Admissions Committee, Cornell *2023*
- EECS Graduate Admissions Committee, MIT *2021*
- EECS Graduate Admissions Committee, UC Berkeley *2018—2020*

Reviewing Activities

- *Journals*:MOR, OR, Sto. Sys., AOS, JMLR, IEEE-IT, JRSSB, Bernoulli, HDSR, Stats-Comp., SIAM, Jour. of Causal Inference, ISR, JCGS, ACM
- *Conferences*: NeurIPS (Area Chair), COLT, ICML, AISTATS, FOCS, STOC, SODA, AAAI, UAI, SIGMETRICS

Work Experience

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

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

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

Stanford University, Research Intern with Prof. Balaji Prabhakar, USA *2013*

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