









RAAZ DWIVEDI

 Harvard/MIT  raazdwivedi.github.io  raaz@seas.harvard.edu  raaz@mit.edu    

ACADEMIC APPOINTMENTS	Postdoctoral Fellow , Computer Science & Statistics, Harvard University and Electrical Engineering (EE) & Computer Sciences (CS), Massachusetts Institute of Technology (MIT) 2021– Advisors: <i>Prof. Susan Murphy & Prof. Devavrat Shah</i>
EDUCATION	Ph.D., EECS , University of California (UC), Berkeley 2015–2021 Advisors: <i>Prof. Martin Wainwright & Prof. Bin Yu</i> Thesis title: <i>Principled statistical approaches for sampling and inference in high dimensions</i> B. Tech., EE , Indian Institute of Technology (IIT), Bombay, India 2010–2014 Minors in mathematics, Institute Rank 1
RESEARCH INTERESTS	My research builds statistically and computationally efficient strategies for data- and computation-driven personalized decision-making. My work develops theory and methods spanning the areas of causal inference, reinforcement learning, Bayesian inference, optimization, and high-dimensional statistics.
SELECTED ACHIEVEMENTS & AWARDS	Best Student Paper Award , Sections on Statistical Computing and Statistical Graphics, American Statistical Association (ASA) 2022 Best Presentation Award , Machine Learning and Statistics Session, Laboratory of Information and Decision Systems (LIDS) Student Conference, MIT 2022 Certificate of Distinction and Excellence in Teaching (Q Award), Harvard University 2022 Institute of Mathematical Statistics (IMS) New Researcher Travel Award, London 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 institute 2014 Best Undergraduate Thesis Award, IIT Bombay 2014 All India Rank 10 amongst half a million, IIT Joint Entrance Exam 2010
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
PRE-PRINTS & WORKING PAPERS	(\star denotes equal contribution and \dagger denotes alphabetical ordering; title is hyperlinked to the online pdf of the paper) P1. Raaz Dwivedi , Katherine Tian, Sabina Tomkins, Predrag Klasnja, Susan Murphy, Devavrat Shah, “Counterfactual inference in sequential experimental design”, <i>arxiv, to be submitted to Annals of Statistics (AoS)</i> . 2022 P2. Abhin Shah, Raaz Dwivedi , Devavrat Shah, Greg Wornell, “On counterfactual inference with unobserved confounding”, <i>NeurIPS workshop, full version to be submitted to AoS</i> . 2022

- P3. Carles Domingo-Enrich, **Raaz Dwivedi**, Lester Mackey, “Compress then test: Powerful kernel testing in near-linear time”, *in conference submission*. 2022
- P4. **Raaz Dwivedi**^{*}, Kelly Zhang^{*}, Prasidh Chhabria, Predrag Klasnja, Susan Murphy, “Assessing personalization by a reinforcement learning algorithm”, *Working paper to be submitted to JRSSB*. 2022
- P5. **Raaz Dwivedi**, Katherine Tian, Sabina Tomkins, Predrag Klasnja, Susan Murphy, Devavrat Shah, “On doubly robust nearest neighbors in factor models”, *Working paper*. 2022
- P6. Nhat Ho^{*}, Koulik Khamaru^{*}, **Raaz Dwivedi**^{*}, Martin J. Wainwright, Michael I. Jordan, Bin Yu, “Instability, computational efficiency, and statistical accuracy”, *under review at JMLR*. 2021
- P7. **Raaz Dwivedi**^{*}, Chandan Singh^{*}, Bin Yu, Martin J. Wainwright, “Revisiting minimum description length complexity in overparameterized models”, *in revision at JMLR*. 2021

JOURNAL
PUBLICATIONS

- J1. Nick Altieri[†], Rebecca L. Barter, James Duncan, **Raaz Dwivedi**, Karl Kumbier, Xiao Li, Robert Netzer, 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
- J2. **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
- J3. **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
- J4. 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
- J5. **Raaz Dwivedi**^{*}, Yuansi Chen^{*}, Martin J. Wainwright, 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, 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, Bin Yu. “Fast MCMC sampling algorithms on polytopes”, *Journal of Machine Learning Research (JMLR)*. 2018
- J8. Vivek Borkar[†], **Raaz Dwivedi**, Neeraja Sahasrabudhe. “Gaussian approximations in high dimensional estimation”, *Systems & Control Letters*. 2016

CONFERENCE
PUBLICATIONS

- C1. **Raaz Dwivedi**, Lester Mackey. “Generalized kernel thinning”, *International Conference on Learning Representations (ICLR)*. 2022
- C2. Abhishek Shetty, **Raaz Dwivedi**, Lester Mackey. “Distribution compression in near-linear time”, *International Conference on Learning Representations (ICLR)*. 2022
- C3. **Raaz Dwivedi**, Lester Mackey, “Kernel thinning”, Extended abstract in *Conference on Learning Theory (COLT)*. Full version under review in *JMLR*. 2021
- C4. **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
- C5. **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
- C6. **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
- C7. Yuansi Chen^{*}, **Raaz Dwivedi**^{*}, Martin J. Wainwright, Bin Yu, “Vaidya walk: A sampling algorithm based on the volumetric barrier”, *Allerton Conference*. 2017

	C8. Raaz Dwivedi , Vivek Borkar, “Removing sampling bias in networked stochastic approximation”, <i>International Conference on Signal Processing and Communications (SPCOM)</i> . 2014
SOFTWARES & METHODOLOGIES	<p>S1. Carles Domingo-Enrich, Raaz Dwivedi, Lester Mackey. Python package “Compress then test” (link).</p> <p>S2. Abhishek Shetty*, Raaz Dwivedi*, Lester Mackey. Python package “Compress++” (link).</p> <p>S3. Raaz Dwivedi, Lester Mackey. Python package “Kernel Thinning” (link).</p> <p>S4. Raaz Dwivedi*, Yan Shuo Tan*, Briton Park, Mian Wei, Kevin Horgan, David Madigan, Bin Yu. Python repository “StaDISC” (link).</p> <p>S5. Yuansi Chen*, Raaz Dwivedi*, Martin Wainwright, Bin Yu. Python package (with C++ implementation) “Vaidya and John walks” (link).</p>
INVITED RESEARCH TALKS	<p>T1. Counterfactual inference in sequential experiments using nearest neighbors. <i>INFORMS Annual Meeting, Indianapolis</i>. Oct 2022</p> <p>T2. Revisiting minimum description length complexity in overparameterized models. <i>Algorithmic Information Theory & Machine Learning Symp., Alan Turing Institute, London</i>. July 2022</p> <p>T3. Counterfactual inference in sequential experimental design. <i>Institute of Mathematical Statistics (IMS) Annual Meeting, Statistical Machine Learning Session, London</i>. June 2022</p> <p>T4. Near-optimal compression in near-linear time. <i>Symp. on Kernel methods for numerical integration, SIAM Conference on Uncertainty Quantification, Atlanta</i>. Apr 2022</p> <p>T5. Near-optimal compression in near-linear time. <i>Stable, Generalizable, & Transferable Statistical Learning Workshop, Mathematical Sciences Research Institute, Berkeley</i>. Mar 2022</p> <p>T6. Counterfactual inference in sequential experimental design. <i>Learning from Interventions Workshop, Simons Institute, Berkeley</i>. Feb 2022</p> <p>T7. Revisiting minimum description length complexity in overparameterized models. <i>Collaborations on the Theoretical Foundations of Deep Learning, Virtual</i>. Nov 2021</p> <p>T8. Kernel thinning. <i>Data-Centric Engineering Group, Alan Turing Institute, Virtual</i>. Sep 2021</p> <p>T9. StaDISC: Stable discovery of interpretable subgroups via calibration. <i>Young Data Scientist Research Seminar, ETH Zurich, Virtual</i>. Sep 2020</p> <p>T10. Veridical data science. <i>ASA Annual Symposium on Data Science & Statistics, Virtual</i>. Jun 2020</p> <p>T11. Statistics meets optimization: Two vignettes. <i>Math & Stats Seminar, IIT Kanpur</i>. Jan 2020</p> <p>T12. Singularity, misspecification, & the convergence rate of EM. <i>AMS Meeting, UC Riverside</i>. Nov 2019</p> <p>T13. Power of gradients and accept-reject step in MCMC algorithms. <i>BIDS, UC Berkeley</i>. Mar 2019</p> <p>T14. Theoretical guarantees for MCMC algorithms, <i>EE Dept, IIT Bombay</i>. Jan 2018</p> <p>T15. Theoretical guarantees for MCMC algorithms, <i>STCS Seminar, TIFR Bombay</i>. Jan 2018</p>
CONTRIBUTED & OTHER RESEARCH TALKS	<p>R1. Generalized kernel thinning. <i>Joint Statistical Meeting (JSM), Washington DC</i>. Aug 2022</p> <p>R2. Counterfactual inference in sequential experimental design. <i>Stats & DS Conference, MIT</i>. Apr 2022</p> <p>R3. Counterfactual inference in sequential experimental design. <i>Econometrics Lunch, MIT</i>. Mar 2022</p> <p>R4. Near-optimal compression in near-linear time. <i>LIDS Student Conference, MIT</i>. Jan 2022</p> <p>R5. Kernel thinning. <i>Monte Carlo Methods & Applications (MCM), Virtual</i>. Sep 2021</p> <p>R6. Kernel thinning. <i>Int. Soc. for Bayesian Analysis (ISBA) World Meeting, Virtual</i>. Aug 2021</p> <p>R7. Kernel thinning. <i>The Bayesian Young Statisticians Meeting (BAYSM), Virtual</i>. Aug 2021</p> <p>R8. Kernel thinning. <i>Joint Statistical Meeting (JSM), Virtual</i>. Aug 2021</p> <p>R9. Kernel thinning. <i>Conference on Learning Theory (COLT), Virtual</i>. Aug 2021</p>

	R10. Kernel thinning. <i>Subset Selection, International Conference on Machine Learning (ICML), Virtual.</i> Jul 2021
	R11. Revisiting complexity and the bias-variance tradeoff: Using minimum description length. <i>Theory of Overparameterized Machine Learning (TOPML) Workshop, Virtual.</i> Apr 2021
	R12. Converging fast and slow: Statistics vs optimization. <i>BAIR and BDD Retreat, Berkeley, Virtual.</i> Aug 2020
	R13. Log-concave sampling: Metropolis Hastings algorithms are fast. <i>Jerusalem Stat. Event.</i> Dec 2018
	R14. Vaidya walk: A sampling algorithm based on the volumetric barrier. <i>Allerton Conference.</i> Oct 2017
CONTRIBUTED POSTER PRESENTATIONS	C1. On counterfactual inference in factor models with nearest neighbors. <i>Cornell ORIE Young Researchers Workshop, Ithaca.</i> Oct 2022
	C2. Counterfactual inference in sequential experimental design. <i>Royal Statistical Society (RSS) Conference, Aberdeen, Scotland.</i> Sep 2022
	C3. Near-optimal compression in near-linear time. <i>RSS Conference, Aberdeen, Scotland.</i> Sep 2022
	C4. Counterfactual inference in sequential experimental design. <i>Synthetic Control Methods Workshop, Data X, Princeton University.</i> Jun 2022
	C5. Counterfactual inference in sequential experimental design. <i>American Causal Inference Conference (ACIC), UC Berkeley.</i> May 2022
	C6. Counterfactual inference in sequential experimental design. <i>Symposium for Mathematical Sciences (SMaSH), Harvard University.</i> May 2022
	C7. Counterfactual inference in sequential experimental design. <i>MIT Statistics & Data Science Conference (SDSCon), MIT.</i> April 2022
	C8. Generalized kernel thinning. <i>Advances in Approx. Bayesian Inference (AABI), Virtual.</i> Feb 2022
	C9. Revisiting complexity and the bias-variance tradeoff: Using minimum description length. <i>North American School of Information Theory (NASIT), Virtual.</i> Jun 2021
	C10. Log-concave sampling: Metropolis Hastings algorithms are fast. <i>Conference on Learning Theory (COLT), Stockholm, Sweden.</i> Dec 2018
	C11. Theoretical guarantees for EM under misspecified Gaussian mixture models. <i>Neural Information Processing Systems (NeurIPS), Montréal, Canada.</i> Dec 2018
	C12. The power to two choices in reducing discrepancy, <i>SAMSI Workshop, Duke University, Raleigh.</i> Aug 2017
TEACHING EXPERIENCE (TASHIP)	T1. Sequential Decision Making (STAT 234), <i>Harvard University.</i> Gave four guest lectures and supervised several half-semester long research projects. 2022
	T2. Modern Statistical Prediction and Machine Learning (STAT 154), <i>UC Berkeley.</i> Gave one guest lecture, helped redesign the class, and developed new discussion sections, homeworks, and exams. 2019
	T3. Introduction to Machine Learning (EECS 189), <i>UC Berkeley.</i> Co-head for the content developments in team of 10+ TAs, helped design discussion sections, homeworks, and exams. 2018
	T4. Linear Algebra, Calculus, Differential equations (MA 105, 106, 108, 207), <i>IIT Bombay.</i> 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), <i>Harvard University</i> Apr 2022
	L2. Offline off-policy reinforcement learning (STAT 234) <i>Harvard University.</i> Feb 2022
	L3. Revisiting complexity and the bias-variance tradeoff (STAT 212) <i>UC Berkeley.</i> Apr 2021
	L4. Introduction to ensemble methods in machine learning (EECS 189), <i>UC Berkeley.</i> Oct 2019
	L5. Introduction to boosting methods (STAT 154), <i>UC Berkeley.</i> Apr 2019

Undergraduate Research Mentoring

- UC Berkeley, One female student that led to a journal publication
- Harvard, Two students including one female with three journal submissions in preparation

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 *struggling sophomores & juniors* 2012—2014
- IIT Bombay Intensive Mentoring Program for thirty *underprivileged undergraduates* 2012—2013

Committees

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

Scientific Meetings

- 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

- EECS Graduate Admissions Committee, MIT 2021
- EECS Graduate Admissions Committee, UC Berkeley 2018—2020

Reviewing Activities

- *Journals*: JMLR, IEEE-IT, JRSSB, Bernoulli, HDSR, Stats & Comp., SIAM, MOR, Jour. of Causal Inference
- *Conferences*: COLT, ICML, AISTATS, NeurIPS, FOCS, STOC, SODA, AAAI

REFERENCES

SUSAN MURPHY
Professor, CS & Statistics
Harvard University
(Postdoc Advisor)
✉ samurphy@fas.harvard.edu
🏠 people.seas.harvard.edu/~samurphy

DEVAVRAT SHAH
Professor, EECS
MIT
(Postdoc Advisor)
✉ devavrat@mit.edu
🏠 devavrat.mit.edu

LESTER MACKEY
Principal Researcher
Microsoft Research New England
Adjunct Professor, Stanford University
✉ lmackey@stanford.edu
🏠 web.stanford.edu/~lmackey

MARTIN WAINWRIGHT
Professor, EECS
MIT
(Ph. D. Advisor)
✉ mjwain@mit.edu
🏠 people.eecs.berkeley.edu/~wainwrig

BIN YU
Professor, EECS & Statistics
UC Berkeley
(Ph. D. Advisor)
✉ binyu@berkeley.edu
🏠 binyu.stat.berkeley.edu