

Edgar Dobriban

Wharton Statistics Department
University of Pennsylvania
Academic Research Building, 265 S 37th St
Philadelphia, PA 19104

dobriban@wharton.upenn.edu
<http://statistics.wharton.upenn.edu/profile/dobriban/>
<http://github.com/dobriban/>
<https://twitter.com/EdgarDobriban/>

Employment

University of Pennsylvania

- The Wharton School, Department of Statistics, Assistant Professor, 2017-present
- Secondary appointment in the Department of Computer and Information Science (CIS)
- Member of the graduate group in Applied Mathematics and Computational Science (AMCS)
- Affiliated Faculty, Penn Research in Machine Learning (PRiML)
- Founding member of FINPenn: Center for the Foundations of Information Processing at Penn
- Affiliated Faculty, Warren Center for Network & Data Sciences
- Affiliated Faculty, Analytics at Wharton
- Affiliated Faculty; AI for Business. <https://ai.wharton.upenn.edu/research/>

Education

Stanford University, Ph.D. in Statistics, 2012 - 2017

Ph.D. thesis: "Topics in high-dimensional asymptotics".

Advisor: David Donoho

Mentors: Iain M Johnstone, Stuart K Kim, Art B Owen

Secondary concentration in Electrical Engineering

Princeton University, B.A. in Mathematics with Highest Honors (*Summa cum Laude*), 2012

Certificate in Applications of Computing

Mentors: Jianqing Fan, Robert E Schapire, Amit Singer

Research Interests

- Statistics and machine learning. Main interests:
- the efficient analysis of "big data" using advanced tools, such as those from random matrix theory
 - PCA: [1], [2], [3], [4], [5], [6]
 - multiple testing: [1], [2], [3]
 - high-dimensional regression: [1], [2]
 - invariance-based randomization tests: [1]
- theoretical analysis of modern machine learning, including deep learning
 - data augmentation: [1]
 - weight normalization: [1]

- stochastic gradient descent and flow: [1]
- overparametrization: [1].
- sketching and random projections: [1], [2], [3], [4], [5]
- distributed learning: [1], [2]
- adversarial robustness: [1], [2]
- retraining of ML models [1]
- certain applications in the life sciences:
 - group testing for COVID-19: [1]
 - genome-wide association studies: [1]

Publications

Preprints or Submitted manuscripts

1. E. Dobriban: Consistency of invariance-based randomization tests. <https://arxiv.org/abs/2104.12260>.
2. Y. Yu, Z. Yang, E. Dobriban, J. Steinhardt, Y. Ma: Understanding Generalization in Adversarial Training via the Bias-Variance Decomposition. <https://arxiv.org/abs/2103.09947>.
3. D. Hong, R. Dey, X. Lin, B. Cleary, E. Dobriban: HYPER: Group testing via hypergraph factorization applied to COVID-19. <https://www.medrxiv.org/content/10.1101/2021.02.24.21252394v1>.
4. D. Hong, Y. Sheng, E. Dobriban: Selecting the number of components in PCA via random signflips. <https://arxiv.org/abs/2012.02985>.
5. E. Dobriban, H. Hassani, D. Hong, A. Robey: Provable tradeoffs in adversarially robust classification. <https://arxiv.org/abs/2006.05161>.
6. F. Yang, S. Liu, E. Dobriban, D.P. Woodruff: How to reduce dimension with PCA and random projections? <https://arxiv.org/abs/2005.00511>.
7. E. Dobriban, W.J. Su: Robust inference under heteroskedasticity via the Hadamard estimator. <http://arxiv.org/abs/1807.00347>.

Published or in press

1. M. Dereziński, Z. Liao, E. Dobriban, M. W. Mahoney: Sparse sketches with small inversion bias. <https://arxiv.org/abs/2011.10695>. Conference on Learning Theory (COLT) 2021.
2. R. Liu, E. Dobriban, Z. Hou, K. Qian: Dynamic Load Identification for Mechanical Systems: A Review. <https://link.springer.com/article/10.1007/s11831-021-09594-7>. Archives of Computational Methods in Engineering (2021)
3. L. Lin, E. Dobriban: What causes the test error? Going beyond bias-variance via ANOVA. <https://arxiv.org/abs/2010.05170>. Journal of Machine Learning Research (JMLR)
4. J. Lacotte, S. Liu, E. Dobriban, M. Pilanci: Limiting Spectrum of Randomized Hadamard Transform and Optimal Iterative Sketching Methods. <https://arxiv.org/abs/2002.00864>. NeurIPS 2020
5. X. Wu, E. Dobriban, T. Ren, S. Wu, Z. Li, S. Gunasekar, R. Ward, Q. Liu: Implicit Regularization of Normalization Methods. <https://arxiv.org/abs/1911.07956>. NeurIPS 2020
6. S.X. Chen, E. Dobriban, J.H. Lee: A Group-Theoretic Framework for Data Augmentation. <https://arxiv.org/abs/1907.10905>. Journal of Machine Learning Research, <https://jmlr.org/papers/v21/20-163.html>. NeurIPS 2020 (oral presentation, 1% of submissions selected).

7. E. Dobriban, Y. Sheng: Distributed linear regression by averaging. <http://arxiv.org/abs/1810.00412>. *Annals of Statistics*. 2021, Vol. 49, No. 2, 918-943.
8. Y. Wu, E. Dobriban, S. Davidson: DeltaGrad: Rapid retraining of machine learning models. <https://arxiv.org/abs/2006.14755>. International Conference on Machine Learning (ICML) 2020.
9. A. Ali, E. Dobriban, R.J. Tibshirani: The Implicit Regularization of Stochastic Gradient Flow for Least Squares. <https://arxiv.org/abs/2003.07802>. International Conference on Machine Learning (ICML) 2020.
10. E. Dobriban, Y. Sheng: WONDER: Weighted one-shot distributed ridge regression in high dimensions. <http://arxiv.org/abs/1903.09321>. *Journal of Machine Learning Research* 21(66):1-52, 2020. Short version at ICML 2020.
11. S. Liu, E. Dobriban: Ridge Regression: Structure, Cross-Validation, and Sketching. <https://arxiv.org/abs/1910.02373> and <https://openreview.net/forum?id=Hk1RwaEKwB>. International Conference on Learning Representations (ICLR) 2020. Spotlight presentation (6% of papers selected).
12. E. Dobriban: FACT: Fast closed testing for exchangeable local tests. <http://arxiv.org/abs/1806.10163>. *Biometrika*. Volume 107, Issue 3, September 2020, Pages 761-768.
13. E. Dobriban: Permutation methods for factor analysis and PCA. <http://arxiv.org/abs/1710.00479>. *Annals of Statistics*. Volume 48, Number 5 (2020), 2824-2847.
14. E. Dobriban, S. Liu: Asymptotics for Sketching in Least Squares Regression. <http://arxiv.org/abs/1810.06089>. Short version at NeurIPS 2019.
15. E. Dobriban*, W.E. Leeb*, A. Singer: Optimal prediction in the linearly transformed spiked model. *Annals of Statistics*. Volume 48, Number 1 (2020), 491-513.
16. E. Dobriban, A.B. Owen: Deterministic parallel analysis: an improved method for selecting factors and principal components. *JRSS-B*. Volume 81, Issue 1, 2019, Pages 163-183.
17. L.T. Liu*, E. Dobriban*, A. Singer: ePCA: High-dimensional exponential family PCA. *Annals of Applied Statistics*. Volume 12, Number 4 (2018), 2121-2150.
18. E. Dobriban: Weighted mining of massive collections of p-values by convex optimization. *Information & Inference: A Journal of the IMA*. Volume 7, Issue 2, June 2018, Pages 251-275.
19. E. Dobriban, S. Wager: High-dimensional asymptotics of prediction: ridge regression and classification. *Annals of Statistics*. (2018) 46 (1):, 247-279.
20. E. Dobriban: Sharp detection in PCA under correlations: all eigenvalues matter. *Annals of Statistics*. (2017) 45 (4): 1810-1833.
21. E. Dobriban: Efficient Computation of Limit Spectra of Sample Covariance Matrices. *Random Matrices: Theory and Applications*. 04, 1550019 (2015)
22. K. Fortney*, E. Dobriban, P. Garagnani, C. Pirazzini, D. Monti, D. Mari, G. Atzmon, N. Barzilai, C. Franceschi, A. B. Owen, S. K. Kim: Genome-wide scan informed by age-related disease identifies loci for exceptional human longevity. *PLoS Genetics*. 11(12): e1005728. (2015)
23. E. Dobriban*, K. Fortney, S. Kim, A. B. Owen: Optimal Multiple Testing under a Gaussian Prior on the Effect Sizes. *Biometrika*. (2015) 102 (4): 753-766.
24. E. Dobriban*, J. Fan: Regularity Properties for Sparse Regression. *Communications in Mathematics and Statistics*. (2016) 4: 1.
25. A. Bandeira, E. Dobriban, D. Mixon, W. Sawin: Certifying the Restricted Isometry Property is Hard, *IEEE Transactions on Information Theory*. vol.59, no.6, pp.3448-3450, 2013.

Grant support

- NSF DMS 2046874. CAREER: Fast and accurate statistical learning and inference from large-scale data: theory, methods, and algorithms. 2021-2026. PI. https://www.nsf.gov/awardsearch/showAward?AWD_ID=2046874.
- NSF 2031895. NSF-Simons Foundation. Mathematical and Scientific Foundations of Deep Learning. THEORINET: The Transferable, Hierarchical, Expressive, Optimal, Robust, Interpretable NETWORKS. 2020-2025. Co-PI with Rene Vidal (PI). NSF Announcement: https://www.nsf.gov/news/special_reports/announcements/082520.jsp; https://www.nsf.gov/awardsearch/showAward?AWD_ID=2031895. See announcements at <https://blog.seas.upenn.edu> and <https://nsf.gov>. Website: <https://www.minds.jhu.edu/theorinet/>
- NSF 1934960. Harnessing the Data Revolution (HDR) Transdisciplinary Research in Principles of Data Science (TRIPODS) Phase I: FINPenn: Center for the Foundations of Information Processing at the University of Pennsylvania. 2019-2022. Co-PI with Alejandro Ribeiro (PI), Kostas Daniilidis, Robert Ghrist, Saswati Sarkar. https://nsf.gov/awardsearch/showAward?AWD_ID=1934960
- NSF IIS 1837992. BIGDATA: F: Collaborative Research: Moment Methods for Big Data: Modern Theory, Algorithms, and Applications. 2018-2021. Co-PI with Amit Singer (Princeton University), and William Leeb (University of Minnesota). https://www.nsf.gov/awardsearch/showAward?AWD_ID=1837992
- "Wharton Energy Analytics Lab": seed fund from the Data Science and Business Analytics Fund of Analytics at Wharton, which "will support novel opportunities to expand the impact of "big data" analytics". Awarded \$75,000.
- Wharton Dean's Fund for Post-Doctoral Research. 2018. "A small and highly selective program for postdoctoral positions with funding shared equally by the Dean's Office and the sponsoring department." Provides \$42,000 per year for up to two years.
- Google Cloud Platform Education Grant for Students. 2018. Provides Google Cloud Platform credits for students in my class Stat 991: Topics in Deep Learning.

Mentoring

I am fortunate to work or have worked with the following excellent students and postdocs:

Postdoctoral

- Hongxiang (David) Qiu, postdoctoral researcher 2021-2022. Co-mentored with Eric Tchetgen Tchetgen. PhD in Statistics from University of Washington.
- Xianli Zeng, postdoctoral researcher 2021-2022. Co-mentored with Guang Cheng. Post-Doctoral fellow at Shenzhen Research Institute of Big Data. PhD in Statistics and Applied probability from National University of Singapore.
- David Hong, postdoctoral researcher 2019-2021. PhD in EECS from University of Michigan, NSF Mathematical Sciences Postdoctoral Fellow

PhD

- Patrick Chao, PhD in Statistics, 2020-
- Stefanos Pertigkiozoglou, PhD in CIS, 2020- (co-advised with Kostas Daniilidis)
- Evangelos Chatzipantazis, PhD in CIS, 2020- (co-advised with Kostas Daniilidis)

- Xinmeng Huang, PhD in AMCS, 2020- (co-advised with Hamed Hassani)
- Donghwan Lee, PhD in AMCS, 2020- (co-advised with Hamed Hassani)
- Shuxiao Chen, PhD in Statistics, 2019- (co-advised with Zongming Ma and Weijie Su)
- Yue Sheng, PhD in AMCS, 2018-21 (co-advised with Robin Pemantle). Now research scientist at Facebook.

Visiting PhD

- Ruixue Liu, visiting PhD student from Tsinghua University, Fall 2019.

Selected Undergraduate Students

- Licong Lin (PKU), Tianle (Tyler) Liu (THU), Jane H Lee (Penn), Sifan (Sophia) Liu (THU)

I have also had the pleasure to collaborate with other great students and postdocs:

- Sangdon Park, PhD student at Penn CIS, advised by Osbert Bastani and Insup Lee
- Souradeep Dutta, postdoctoral research at Penn CIS, advised by Insup Lee
- Yahan Wang, PhD student at Penn CIS, advised by Insup Lee
- Fan Yang, postdoctoral researcher mentored by Jian Ding,
- Yinjun Wu, PhD student at Penn CIS, advised by Susan Davidson
- Soham Dan, PhD student at Penn CIS, advised by Dan Roth

Honors and Awards

- NSF CAREER award 2021.
- Theodore W. Anderson Theory of Statistics Dissertation Award. Stanford Department of Statistics. 2017. Awarded for the best PhD thesis in the theory of statistics from the Department of Statistics at Stanford University in 2017.
- Howard Hughes Medical Institute International Student Graduate Research Fellowship, 2015. "The program supports international students during their third to fifth years of graduate school in the United States. The awardees will receive \$43,000 during each year of the fellowship." Among 45 predoctoral students selected.
- Stanford Department of Statistics Teaching Award, 2013.
- Middleton Miller '29 Prize for best independent work in mathematics, Princeton University Department of Mathematics, 2012.
- Phi Beta Kappa (top 10 % of graduating class), Princeton University, 2012.
- Selected awards in mathematical olympiads: "Traian Lalescu" Prize awarded by the Romanian Mathematical Society to the best student in mathematical olympiads in Romania, 2008. Romanian National Olympiad, 1st place nationally, 2007. Balkan Mathematical Olympiad, Macedonia. Bronze medal, 2008.

Teaching Experience

Teaching at the University of Pennsylvania, 2017-present

- STAT 431: Statistical inference (Spring 2020). Undergraduate level statistical inference. As part this, developed a final project that asked students to review a paper with a strong statistical content related to Covid-19: <https://github.com/dobriban/stat-431-project>.
- STAT 430/510: Probability (Spring 2018, Fall 2018). Undergraduate level probability.
- STAT 991: Topics in Deep Learning (Fall 2018, Spring 2019, Fall 2019). Developed a new seminar course on deep learning. Wrote lecture notes covering standard topics, such as feedforward neural networks, backpropagation, convolutional neural networks (CNNs), recurrent neural networks (RNNs), generative adversarial networks (GANs). Presented in-class computational experiments using Keras. Guided student presentations on various topics, including training methods for deep learning, theory for GANs, robotic vision, neuroscience. This was a success, with broad participation from across the university (including statistics, applied math, biostatistics, computer science, electrical engineering, mechanical engineering, and marketing). In Spring 2019, covered topics in sequential decision-making, ranging from bandits to deep reinforcement learning. Materials are publicly and freely available at <https://github.com/dobriban/Topics-in-deep-learning>.

Publicly Available Software

The computational results from my papers are reproducible, and software to do so is available on my GitHub page, or linked to it are available in the respective papers: <http://github.com/dobriban/>.

- **ePCA**: <http://github.com/ltliu/epca>. Contains the ePCA method for principal component analysis (PCA) of exponential family data. An example is Poisson-modeled count data. Also implements methods for denoising individual datapoints. (with L.T. Liu).
- **EigenEdge**: <http://github.com/dobriban/EigenEdge>. Statistical and computational methods for working with large sample covariance matrices. The SPECTRODE method to compute eigenvalue distributions; methods to find moments and quantiles; optimal linear spectral statistics for PCA; spiked models.
- **pweight**: P-value weighting techniques for multiple hypothesis testing. These can improve power in frequentist multiple testing, in the presence of some prior information about the effects. The iGWAS method for applications to Genome-Wide Association Studies. Available from http://github.com/dobriban/pvalue_weighting_matlab, and archived on CRAN.

Other Experience

Visiting scholar at the program on **Mathematics of deep learning** at the Isaac Newton Institute, Cambridge, UK. Summer 2021. <http://www.newton.ac.uk/event/MDL>. Organisers: Gitta Kutyniok, Peter Bartlett, Anders Hansen, Arnulf Jentzen, Carola-Bibiane Schönlieb.

Visiting faculty and workgroup leader in the IDEAL (Institute for Data, Econometrics, Algorithms, and Learning) **Special quarter on Theory of Deep Learning** for Fall 2020. <https://www.ideal.northwestern.edu/special-quarters/fall-2020/>. Organizers Nathan Srebro (TTIC), Zhaoran Wang (Northwestern University), Dongning Guo (Northwestern University). See materials at <https://github.com/dobriban/ideal-working-group>.

Visiting scholar at the **Special Year on Optimization, Statistics, and Theoretical Machine Learning**, Institute for Advanced Study, Princeton, September 2019-December 2019. http://www.math.ias.edu/sp/Optimization_Statistics_and_Theoretical_Machine_Learning. Organized by Sanjeev Arora.

Visiting scholar at the **Program in Foundations of Deep Learning**, Simons Institute for the Theory of Computing, Berkeley, May-August 2019. <https://simons.berkeley.edu/programs/dl2019>. Organized by Samy Bengio, Alexander Madry, Elchanan Mossel, Matus Telgarsky.

Participant in **Random Matrix Theory Summer School**, Park City Mathematics Institute, Institute for Advanced Studies, June 2017. Organized by Alexei Borodin, Alice Guionnet, and Ivan Corwin.

Participant in **Summer School on Random Matrices**, Michigan, June 2016. <https://web.eecs.umich.edu/~rajnrao/rmtschool/2016/index.html>. Organized by Raj Rao Nadakuditi and Jinho Baik.

Consultant for **Statistical Consulting Class**, Stanford University, Autumn 2013. Provided free statistical consulting to the Stanford University community through the department's weekly consulting sessions.

Participant in **UCLA Logic Summer School**, 2010 (organized by Itay Neeman). Took courses on "Determinacy" and "Computability and Complexity".

Talks and Presentations

Slides for some of my talks are available on my GitHub page, <http://github.com/dobriban/talks>.

12th International Conference on Multiple Comparison Procedures, 2021. Session on "Mean field asymptotics in high-dimensional statistics", organized by Malorzata Bogdan

ICSA Applied Statistics Conference. 2021. Invited by Tracy Ke.

Joint Statistical Meetings (JSM). 2021. Invited presentation on session in deep learning. Invited by Xiaotong Shen.

10th World Congress in Probability and Statistics. 2021.

Virtual INFORMS Healthcare Conference. 2021. Invited by Zhengyuan Zhou.

Youth in High Dimensions. 2021. <http://indico.ictp.it/event/9596/>.

NSF-Simons Science and Mathematics of Deep Learning journal club. 2021.

Econometrics and Statistics seminar at Chicago Booth School of Business. 2021. Invited by Ekaterina Smetanina and Panos Toulis.

Princeton University, IDEAS Seminar, 2021. Invited by Amit Moscovich.

Columbia Econometrics seminar. 2021. Invited by Simon Lee.

UC Davis. Statistics Seminar. 2021. Invited by Xiucui Ding.

ARO MURI "Robust Concept Learning and Lifelong Adaptation against Adversarial Attacks" Meeting, University of Pennsylvania. 2020.

London School of Economics (LSE) Statistics Seminar, 2020. Invited by Xinghao Qiao.

IDEAL (Institute for Data, Econometrics, Algorithms, and Learning) Theory of Deep Learning Seminar Series, TTIC/Northwestern University/University of Chicago.

Wharton Department of Statistics, University of Pennsylvania. 2020.

Stanford University Department of Statistics. 2020. Invited by Paromita Dubey.

University of Wisconsin, Madison, 2020. Invited by Claudia Solis Lemus and Rob Nowak. Talk recording: <https://silo.wisc.edu/edgar-dobriban/>.

Reunion for Program in Foundations of Deep Learning, Simons Institute for the Theory of Computing, Berkeley. 2020. <https://simons.berkeley.edu/workshops/schedule/14838>

Michael Mahoney (UC Berkeley) group meeting. 2020

CMStatistics 2020, London. "Sketching and related methods in regression". Invited by Keith Knight.

Joint Statistical Meetings (JSM). 2020.

Presentation in section on Connections between optimization and Statistics. Invited by Yuxin Chen. Discussant (with Misha Belkin and Daniel Hsu) on section in deep learning, with presentations by Jason Klusowski, Tengyuan Liang, Matus Telgarsky. Invited by Po-Ling Loh.

Statistical Learning and Data Science (SLDS) May 27-29 2020. UC Davis. (Cancelled due to Covid-19)

Session on Random Matrix Theory and its Applications. Invited by Haoran Li.

Session on Deep Learning theory. Invited by Tengyu Ma.

International Conference on Learning Representations (ICLR) 2020. Spotlight presentation. https://iclr.cc/virtual/poster_Hk1RwaEKwB.html

54th Annual Conference on Information Sciences and Systems (CISS). Princeton. 2020. Invited by Yuejie Chi. (Conference cancelled due to Covid-19)

University of Massachusetts Amherst. 2020. Invited by Haben Michael. (cancelled due to Covid-19)

Joint Statistical Meetings (JSM). 2019. Invited by Tony Cai.

Microsoft Research NYC. 2019. Invited by Rob Schapire.

Konrad Kording Lab, University of Pennsylvania. 2019.

Harvard Biostatistics Seminar. 2019. Invited by Rajarshi Mukherjee.

Minnesota Data Science Seminar. 2019. Invited by William Leeb.

New Developments in Free Probability and Applications. CRM Montreal. 2019. Invited by Ken Dykema.

Cambridge Statistical Laboratory, 2019. Invited by Richard Samworth.

Alan Turing Institute, UK. 2019. Invited by Mihai Cucuringu.

Simons Institute, Program for Foundations of Data Science, Lightning Talk. 2018.

AMCS Colloquium, University of Pennsylvania. 2018.

Temple University Statistics Department. 2018. Invited by Kuang-Yao Lee and Sanat Sarkar.

UChicago Statistics Department. 2018. Invited by Chao Gao.

Joint Statistical Meetings (JSM). Vancouver, Canada. 2018.

Statistical Learning and Data Science Conference. New York. 2018. Invited by Yuekai Sun.

GDR day, Paris, France. 2017. Invited by Romain Couillet.

IXth workshop on New Developments in Econometrics and Time Series, Rome, 2017. Invited by Marc Hallin.

Georgia Tech Stochastics Seminar, Atlanta, GA. 2017. Invited by Mayya Zhilova and Vladimir Koltchinskii.

Joint Statistical Meetings (JSM), Baltimore. 2017.

Xth International Multiple Comparison Procedures (MCP) conference, Riverside, 2017.

UCLA Department of Statistics. 2017.

NYU Courant Institute. 2017.

Princeton University Department of ORFE. 2017.

Stanford University Department of Statistics. 2017.

University of Michigan Department of Statistics. 2017.

Wharton Department of Statistics, University of Pennsylvania. 2017.

Harvard University Department of Statistics. 2017.

MIT Department of Brain and Cognitive Sciences. 2017.

Columbia Department of Statistics. 2017.

Random Matrix Theory and Probability Seminar, Harvard. 2016. Invited by H.T. Yau and Philippe Sosoe.

Stanford Statistics Department Seminar, Stanford. 2016.

Joint Statistical Meetings (JSM), Chicago. 2016.

3rd ISNPS conference, Avignon. 2016. Invited by Marc Hallin.
 Big Data in Biomedicine, Stanford. 2016.
 IDeAS Seminar, The Program in Applied and Computational Mathematics, Princeton University. 2016.
 Invited by Amit Singer.
 23andme, Palo Alto, CA. 2016.
 Machine Learning reading group, Stanford. 2015.
 IXth International Multiple Comparison Procedures (MCP) conference, Hyderabad, India. 2015.
 Joint Statistical Meetings (JSM), Seattle. 2015.
 Bio-X Interdisciplinary Initiatives Symposium, Stanford. 2015.

Professional Service and Activities

- On the review panel for the National Science Foundation DMS Statistics program (2020).
- Meta-reviewer/Area chair for International Conference on Machine Learning (ICML) 2021
- Reviewer for: *Annals of Statistics* (17), *Annals of Applied Statistics*, *Bernoulli*, *Biometrika* (2), *Conference on Learning Theory (COLT)* 2020, *Computational Statistics and Data Analysis*, *IEEE Transactions on Information Theory* (2), *International Conference on Machine Learning (ICML)* 2020, *Journal of Computational and Graphical Statistics*, *Journal of Multivariate Analysis*, *Journal of Machine Learning Research* (3), *JRSS-B* (2), *Mathematical and Scientific Machine Learning* 2020, *NeurIPS* 2020, *Proceedings of the National Academy of Sciences* (2), *Scandinavian Journal of Statistics*, *STAT*.
- Seminar organizer: Wharton Statistics Colloquium, Fall 2017; AMCS Seminar, Fall 2018, Spring 2019
- Community Strengthening Committee, Wharton Statistics department. 2020.
- Committee on Lawrence D. Brown Distinguished Lecture, Wharton Statistics department. 2021.
- PhD admissions committee: Wharton Statistics Phd program, 2018, 2020; AMCS Phd program, 2018, 2019, 2020
- PhD thesis committee member: Yezheng Li (AMCS, Advised by Honghze Lee), Peter Ballen (CIS, Advised by Sudipto Guha), Shaokun Li (Statistics, Advised by Tony Cai), Yebiao Jin (AMCS, Advised by Kent Smetters), Saeed Sharifimalvajerdi (Statistics, Advised by Aaron Roth),
- Written Preliminary Examination (WPE) II committee: Carlos Esteves, CIS PhD student advised by Kostas Daniilidis. Yinshuang Xu, CIS PhD student advised by Kostas Daniilidis.
- Contributed to the ASA statement "Comment on Statistics, as a Discipline and Practice, in AI Research and Development", lead by Sarah Kalicin. Link on ASA website: <http://www.amstat.org/asa/files/pdfs/POL-Comment-on-Stats-as-a-Discipline-and-Practice-in-AI-RandD.pdf>
- Teaching Continuing Education course "Machine Learning and Deep Learning," with Annie Qu and Xiao Wang, for the ASA Professional Development Program.
- Teaching Invited Short Course "Deep Learning in Statistics" with Annie Qu and Xiao Wang, at Statistical Learning and Data Science (SLDS) 2020. (Postponed due to Covid-19)
- Member of ASA, IMS, Bernoulli Society, SIAM, ACM, IEEE, AMS, INFORMS

Conference organization

- Session chair and organizer. Session on "Sketching and random projection methods for modern data analysis" at CMStatistics 2021. Invited speakers: Michal Derezhinski, Mert Pilanci, Lorenzo Rosasco, Tengyao Wang.

- Moderating a "Brainstorm and Discussion" Session on Robustness at the workshop on The Analytical Foundations of Deep Learning: Interpretability and Performance Guarantees at the C3DTI <https://c3dti.ai/events/workshops/foundations-of-deep-learning/>. YouTube video: <https://www.youtube.com/watch?v=p4JpfmW5PNQ>
- Co-organizer with Kostas Daniilidis of a workshop on "Equivariance and Data Augmentation", Sept 2020. Invited speakers include: Christine Allen-Blanchette, Fabio Anselmi, Pratik Chaudhari, Taco Cohen, Carlos Esteves, Jane H Lee, Chelsea Finn, Haggai Maron, George J. Pappas, Danilo J. Rezende, Christopher Ré, Alejandro Ribeiro, Tess E Smidt, Greg Valiant, Kilian Q. Weinberger. See <https://sites.google.com/view/equiv-data-aug/home>. YouTube playlist of the videos: <https://www.youtube.com/playlist?list=PL7VfmM00wN1J5PndlhiMwjvqDTA6m1Z9i>.
- Co-organizer of minisymposium on "High-dimensional PCA in the High-noise Regime" at the First SIAM Conference on Mathematics of Data Science (MDS20). Co-organized with Amit Singer and William Leeb. Invited speakers: Laura Balzano, Tracy Ke, Anru Zhang (Moved online due to Covid-19)
- Session chair and organizer. Session on "Spectral methods for massive data", JSM 2019. Invited speakers: Boaz Nadler, Arup Bose, Miles Lopes, Matan Gavish.
- Session chair. Statistical Learning and Data Science Conference. 2018. Session organizer: Yuekai Sun.

Media Coverage

My research on extreme human longevity, done jointly with Art Owen, Kristen Fortney, Stuart Kim and our biology collaborators, published in *PLoS Genetics* 11(12): e1005728. (2015), has been covered by the BBC and by 28 *online media outlets*, including:

New Scientist: <http://newscientist.com/article/dn28688-four-genes-discovered-that-will-help-you-live-beyond-100/>

TIME: <http://time.com/4153835/live-longer-genetic-clues/?xid=homepage>

US News: <http://usnews.com/news/articles/2015-12-17/4-genes-that-will-help-you-live-past-100>

Last updated: June 5, 2021