

Edgar Dobriban

Wharton Statistics Department
University of Pennsylvania
3730 Walnut Street, Suite 400
Philadelphia, PA 19104

dobriban@wharton.upenn.edu
<http://statistics.wharton.upenn.edu/profile/dobriban/>

Employment

University of Pennsylvania

The Wharton School, Department of Statistics, Assistant Professor, 2017-present
Member of the graduate group in Applied Mathematics and Computational Science (AMCS)
Affiliated Faculty, Penn Research in Machine Learning (PRiML)

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

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

- Statistical methods for "big data", high-dimensional asymptotics, random matrix theory, multiple hypothesis testing, deep learning, scalable machine learning.

Publications

Preprints or Submitted manuscripts

S. Chen, E. Dobriban, J.H. Lee: Invariance reduces Variance: Understanding Data Augmentation in Deep Learning and Beyond. <https://arxiv.org/abs/1907.10905>.

E. Dobriban, Y. Sheng: One-shot distributed ridge regression in high dimensions. <http://arxiv.org/abs/1903.09321>.

E. Dobriban, Y. Sheng: Distributed linear regression by averaging. <http://arxiv.org/abs/1810.00412>.

E. Dobriban: Flexible multiple testing with the FACT algorithm. <http://arxiv.org/abs/1806.10163>.

E. Dobriban, W.J. Su: Robust inference under heteroskedasticity via the Hadamard estimator. <http://arxiv.org/abs/1807.00347>.

Published or in press

- E. Dobriban: Permutation methods for factor analysis and PCA. <http://arxiv.org/abs/1710.00479>. To appear in the *Annals of Statistics*.
- E. Dobriban, S. Liu: A new theory for sketching in linear regression. <http://arxiv.org/abs/1810.06089>. Short version to appear at NeurIPS 2019.
- E. Dobriban*, W.E. Leeb*, A. Singer: Optimal prediction in the linearly transformed spiked model. *Annals of Statistics*. To appear.
- E. Dobriban, A.B. Owen: Deterministic parallel analysis: an improved method for selecting factors and principal components. *JRSS-B*. 2019.
- L.T. Liu*, E. Dobriban*, A. Singer: ePCA: High-dimensional exponential family PCA. *Annals of Applied Statistics*. 2018.
- E. Dobriban: Weighted mining of massive collections of p-values by convex optimization. *Information & Inference: A Journal of the IMA*. 2017, iax013
- E. Dobriban, S. Wager: High-dimensional asymptotics of prediction: ridge regression and classification. *Annals of Statistics*. (2018) 46 (1):, 247-279.
- E. Dobriban: Sharp detection in PCA under correlations: all eigenvalues matter. *Annals of Statistics*. (2017) 45 (4): 1810-1833.
- E. Dobriban: Efficient Computation of Limit Spectra of Sample Covariance Matrices. *Random Matrices: Theory and Applications*. 04, 1550019 (2015)
- 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)
- 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.
- E. Dobriban*, J. Fan: Regularity Properties for Sparse Regression. *Communications in Mathematics and Statistics*. (2016) 4: 1.
- 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.

Author ordering is alphabetical on most publications, except on those marked with the "" symbol, authors are the lead contributors on each publication.*

Grant support

- 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. Amount awarded to Penn: \$1,5000,000.
- 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). Amount awarded to Penn: \$333,000.
- 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.

Mentoring

Postdoctoral

- David Hong, postdoctoral researcher 2019-2021

PhD

- Shuxiao Chen, PhD in Statistics, 2019-
- Yue Sheng, PhD in AMCS, 2018- (co-advised with Robin Pemantle)

Visiting PhD

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

Undergraduate

- Shuo Xie, undergraduate student at Peking University, intern during summer 2019
- Tianle (Tyler) Liu, undergraduate student at Tsinghua University, intern during summer 2019
- Weichen Zheng, undergraduate student at Penn, research during summer 2019
- Jane H Lee, submatriculating student at Penn, research during 2018-2019 academic year
- Sifan Liu, undergraduate student at Tsinghua University, intern during summer 2018 (continuing as PhD student in statistics at Stanford)
- Yijie (Lisa) Zhao, undergraduate student at Penn, research during summer 2018

Honors and Awards

- 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.

Teaching Experience

Teaching at the University of Pennsylvania, 2017-present

- 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). As a follow-up, in Spring 2019, covered topics in sequential decision-making, ranging from bandits to deep reinforcement learning. Materials are 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, <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 **Special Year on Optimization, Statistics, and Theoretical Machine Learning**, Institute for Advanced Study, Princeton, September 2019-May 2020. 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.

Participant in **Summer School on Random Matrices**, Michigan, June 2016. 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>.

CISS. Princeton. 2020. Invited by Yuejie Chi.
MCP Conference. Taiwan. 2019.
ICSA Conference, China. Winter 2019. Invited by Guanming Pan.
Random Matrices and Complex Data Analysis Workshop, Shanghai, 2019. Invited by Jianfeng Yao and Weiming Li.
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.
AMCS Colloquium, University of Pennsylvania. 2018.
Temple University Statistics Department. 2018. Invited by Mihai Cucuringu.
UChicago Statistics Department. 2018.
Joint Statistical Meetings (JSM). Vancouver, Canada. 2018.
Statistical Learning and Data Science Conference. New York. 2018.
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.
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.
Stanford Statistics Department Seminar, Stanford. 2016.
Joint Statistical Meetings (JSM), Chicago. 2016.
3rd ISNPS conference, Avignon. 2016.
Big Data in Biomedicine, Stanford. 2016.
IDeAS Seminar, The Program in Applied and Computational Mathematics, Princeton University. 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

- Reviewer for: *Annals of Statistics* (12), *Annals of Applied Statistics*, *Biometrika*, *Computational Statistics and Data Analysis*, *IEEE Transactions on Information Theory*, *Journal of Computational and Graphical Statistics*, *Journal of Multivariate Analysis*, *Journal of Machine Learning Research* (2), *Proceedings of the National Academy of Sciences* (2), *Scandinavian Journal of Statistics*.
- Seminar organizer, Wharton Statistics Colloquium, Fall 2017
- Seminar organizer, AMCS Seminar, Fall 2018, Spring 2019
- PhD admissions committee, Wharton Statistics Phd program, 2018
- PhD admissions member, AMCS Phd program, 2018
- 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)
- 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.

Conference organization

- Session chair and organizer. JSM. 2019.
- Session chair. Statistical Learning and Data Science Conference. 2018.

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: September 9, 2019