JON DAVID MCAULIFFE

18 May 2025

449 Evans Hall
Statistics Department
University of California
Berkeley, CA 94720

VOICE: 510-517-1170
EMAIL: jon@stat.berkeley.edu
WEB: www.stat.berkeley.edu/∼jon/
CITIZENSHIP: USA

ACADEMIC POSITIONS

UNIVERSITY OF CALIFORNIA, BERKELEY
Professor (Adjunct), Statistics Department
July 2018 – Present
Associate Professor (Adjunct), Statistics Department
July 2013 – June 2018
Assistant Professor (Adjunct), Statistics Department
July 2009 – June 2013

UNIVERSITY OF PENNSYLVANIA
Assistant Professor, Statistics Department, Wharton School
Member, Genomics and Computational Biology Graduate Group
July 2005 – June 2009

EDUCATION

UNIVERSITY OF CALIFORNIA, BERKELEY
Ph.D. Statistics, 2005
Dissertation: "Statistical methods for comparing genomes"
Committee: Michael Jordan (chair), Peter Bickel, Lior Pachter

HARVARD UNIVERSITY
M.S. Applied Mathematics, and
B.A. Computer Science, cum laude, 1995

JOURNAL PUBLICATIONS

[1] R. Liu, J. McAuliffe, and J. Regier (2023). "Variational inference for deblending crowded starfields." *Journal of Machine Learning Research* 24(179): 1–36.

[2] S. R. Howard, A. Ramdas, J. McAuliffe, and J. Sekhon (2021). "Time-uniform, nonparametric, nonasymptotic confidence sequences." *Annals of Statistics* 49(2): 1055–1080.

[3] Y. Choi, J. Qu, S. Wu, Y. Hao, J. Zhang, J. Ning, X. Yang, L. Lofaro, D. G. Pankratz, J. Babiarz, P. S. Walsh, E. Billatos , M. E. Lenburg, G. C. Kennedy, J. McAuliffe, and J. Huang (2020). "Improving lung cancer risk stratification leveraging whole transcriptome RNA sequencing and machine learning across multiple cohorts." *BMC Medical Genomics* 13(Suppl 10): 151–165.

[4] S. R. Howard, A. Ramdas, J. McAuliffe, and J. Sekhon (2020). "Time-uniform Chernoff bounds via nonnegative supermartingales." *Probability Surveys* 17: 257–317.

[5] J. Regier, A. C. Miller, D. Schlegel, R. P. Adams, J. D. McAuliffe, and Prabhat (2019). "Approximate inference for constructing astronomical catalogs from images." *Annals of Applied Statistics* 13(3): 1884–1926.

[6] J. Regier, K. Fischer, K. Pamnany, A. Noack, J. Revels, M. Lam, S. Howard, R. Giordano, D. Schlegel, J. McAuliffe, R. Thomas, and Prabhat (2019). "Cataloging the visible universe through Bayesian inference in Julia at petascale." *Journal of Parallel and Distributed Computing* 127: 89–104.

[7] D. M. Blei, A. Kucukelbir, and J. McAuliffe (2017). "Variational inference: A review for statisticians." *Journal of the American Statistical Association* 112(518): 859–877.

[8] H. Saddiki, J. McAuliffe, and P. Flaherty (2015). "GLAD: A mixed-membership model for heterogeneous tumor subtype classification." *Bioinformatics* 31(2): 225–232.

[9] S. T. Jensen, J. Park, A. Braunstein, and J. McAuliffe (2013). "Bayesian hierarchical modeling of the HIV evolutionary response to therapy." *Journal of the American Statistical Association* 108(504): 1230–1242.

[10] S. Nygaard, A. Braunstein, G. Malsen, S. van Dongen, P. P. Gardner, A. Krogh, T. D. Otto, A. Pain, M. Berriman, J. McAuliffe, E. T. Dermitzakis, and D. C. Jeffares (2010). "Long and short term selective forces on malaria parasite genomes." *PLoS Genetics* 6(9): e1001099.

[11] M. Braun and J. D. McAuliffe (2010). "Variational inference for large-scale models of discrete choice." *Journal of the American Statistical Association* 105(489): 324–335.

[12] S. M. Sweeney, J. P. Orgel, A. Fertala, J. D. McAuliffe, K. R. Turner, G. A. Di Lullo, S. Chen, O. Antipova, S. Perumal, L. Ala-Kokko, A. Forlino, W. A. Cabral, A. M. Barnes, J. C. Marini, and J. D. San Antonio (2008). "Candidate cell and matrix interaction domains on the collagen fibril, the predominant protein of vertebrates." *Journal of Biological Chemistry* 283(30): 21187–21197.

[13] P. L. Bartlett, M. I. Jordan, and J. D. McAuliffe (2006). "Convexity, classification, and risk bounds." *Journal of the American Statistical Association* 101(473): 138–156.

[14] J. D. McAuliffe, D. M. Blei, and M. I. Jordan (2006). "Nonparametric empirical Bayes for the Dirichlet process mixture model." *Statistics and Computing* 16(1): 5–14.

[15] P. L. Bartlett, M. I. Jordan, and J. D. McAuliffe (2006). "Discussion of 'Support vector machine with applications'." *Statistical Science* 21(3): 341–346.

[16] J. D. McAuliffe, M. I. Jordan, and L. Pachter (2005). "Subtree power analysis and species selection for comparative genomics." *Proceedings of the National Academy of Sciences* 102(22): 7900–7905.

[17] P. Gyaneshwar, O. Paliy, J. McAuliffe, A. Jones, M. I. Jordan, and S. Kustu (2005). "Lessons from *Escherichia coli* genes similarly regulated in response to nitrogen and sulfur limitation." *Proceedings of the National Academy of Sciences* 102(9): 3453–3458.

[18] P. Gyaneshwar, O. Paliy, J. McAuliffe, D. L. Popham, M. I. Jordan, and S. Kustu (2005). "Sulfur and nitrogen limitation in *Escherichia coli* K12: specific homeostatic responses." *Journal of Bacteriology* 187(3): 1074–1090.

[19] J. McAuliffe, L. Pachter, and M. I. Jordan (2004). "Multiple-sequence functional annotation and the generalized hidden Markov phylogeny." *Bioinformatics* 20(12): 1850–1860.

[20] P. L. Bartlett, M. I. Jordan, and J. D. McAuliffe (2004). "Discussion of boosting papers." *Annals of Statistics* 32(1): 85–91.

[21] D. Boffelli, J. McAuliffe, D. Ovcharenko, K. D. Lewis, I. Ovcharenko, L. Pachter, and E. M. Rubin (2003). "Phylogenetic shadowing of primate sequences to find functional regions of the human genome." *Science* 299(5611): 1391–1394.

[22] R. W. Corbin, O. Paliy, F. Yang, J. Shabanowitz, M. Platt, C. E. Lyons, Jr., K. Root, J. McAuliffe, M. I. Jordan, S. Kustu, E. Soupene, and D. F. Hunt (2003). "Toward a protein profile of *Escherichia coli*: Comparison to its transcription profile." *Proceedings of the National Academy of Sciences* 100(16): 9232–9237.

REFEREED PROCEEDINGS

[23] R. Liu, J. Regier, N. Tripuraneni, M. Jordan, and J. McAuliffe (2019). "Rao-Blackwellized stochastic gradients for discrete distributions." *Proceedings of the 36th International Conference on Machine Learning*.

[24] A. Fernandez, K. Kashinath, J. McAuliffe, C. M. Patricola, Prabhat, P. B. Stark, and M. F. Wehner (2018). "A predictive statistical model for tropical cyclone genesis." *Proceedings of the 33rd Conference on Hurricanes and Tropical Meteorology*.

[25] A. Fernandez, K. Kashinath, J. McAuliffe, Prabhat, P. Stark, and M. Wehner (2017). "Towards a statistical model of tropical cyclone genesis." *Proceedings of the 7th International Workshop on Climate Informatics: CI 2017*.

[26] J. Regier, M. I. Jordan, and J. McAuliffe (2017). "Fast black-box variational inference through stochastic trust-region optimization." *Advances in Neural Information Processing Systems* 30.

[27] A. Miller, A. Wu, J. Regier, J. McAuliffe, D. Lang, Prabhat, D. Schlegel, and R. P. Adams (2015). "A Gaussion process model of quasar spectral energy distributions." *Advances in Neural Information Processing Systems* 28.

[28] J. Regier, J. McAuliffe, and Prabhat (2015). "A deep generative model for astronomical images of galaxies". *Neural Information Processing Systems Workshop: Advances in Approximate Bayesian Inference*.

[29] J. Regier, A. Miller, J. McAuliffe, R. P. Adams, M. Hoffman, D. Lang, D. Schlegel, and Prabhat (2015). "Celeste: Variational inference for a generative model of astronomical images". *Proceedings of the 32nd International Conference on Machine Learning*.

[30] J. Regier, B. Partridge, J. McAuliffe, R. Adams, M. Hoffman, D. Lang, D. Schlegel, and Prabhat (2014). "Celeste: Scalable variational inference for a generative model of astronomical images". *Neural Information Processing Systems Workshop: Advances in Variational Inference*.

[31] A. Braunstein, Z. Wei, S. Jensen, and J. McAuliffe (2009). "A spatially varying two-sample recombinant coalescent, with applications to HIV escape response." *Advances in Neural Information Processing Systems* 21.

[32] D. M. Blei and J. D. McAuliffe (2008). "Supervised topic models." *Advances in Neural Information Processing Systems* 20.

[33] P. L. Bartlett, M. I. Jordan, and J. D. McAuliffe (2004). "Large margin classifiers: convex loss, low noise, and convergence rates." *Advances in Neural Information Processing Systems* 16.

PROFESSIONAL SERVICE

Editorial board, Journal of Machine Learning Research, 2009–2023

Action editor, Journal of Machine Learning Research, 2018–2023

Senior program committee, International Conference on Machine Learning, 2015, 2016, 2017, 2018

Awards committee, International Conference on Machine Learning, 2015

Senior program committee, International Joint Conference on Artificial Intelligence, 2013

Senior program committee, Neural Information Processing Systems, 2009, 2010

Awards committee, Neural Information Processing Systems, 2009

Senior program committee, International Conference on Artificial Intelligence and Statistics, 2009

Chair, UC Berkeley Statistics development and alumni relations committee, 2012-present

Admissions committee, University of Pennsylvania Genomics and Computational Biology Ph.D. program, 2006-2007

UC Berkeley Ph.D. student: Jimmy Butler, expected 2026

UC Berkeley Ph.D. student: Runjing (Bryan) Liu, 2021

UC Berkeley Ph.D. student: Steve Howard, graduated 2020

UC Berkeley Ph.D. student: Ryan Giordano, graduated 2019

UC Berkeley Ph.D. student: Arturo Fernandez, on leave, 2019

UC Berkeley Ph.D. student: Jeffrey Regier, graduated 2016

UT Sydney Ph.D. examiner: Marianne Menictas, graduated 2015

UC Berkeley Ph.D. qualifying and thesis committee: Erez Buchweitz, advanced to candidacy Spring 2025

UC Berkeley Ph.D. qualifying and thesis committee: Keven Laboy, graduated 2018

UC Berkeley master's thesis committee: Daphna Buchsbaum, 2013

UC Berkeley master's thesis committee: Adrienne Hosek, 2013

UC Berkeley undergraduate honors thesis: Daanyal Ahmed Saeed, 2025

UC Berkeley undergraduate honors thesis: Andrew Goldstein, 2015

UC Berkeley undergraduate honors thesis: Jared Park, 2012

UC Berkeley undergraduate honors thesis: Irene Chen, 2012

Penn Ph.D. thesis committee: Zhi Wei, 2008

Penn Ph.D. qualifying examination committee: Qian Liu, 2007

IMS invited session organizer, 2007 Joint Statistical Meetings
Session title: *Machine learning and optimization*

Penn Statistics Department seminar organizer, 2006–2007

JOURNAL AND TEXTBOOK REVIEWING

Annals of Statistics
Bayesian Analysis
Bioinformatics
Computational Statistics and Data Analysis
Electronic Journal of Statistics
Genome Research
IEEE/ACM Transactions on Computational Biology and Bioinformatics
IEEE Transactions on Pattern Analysis and Machine Intelligence
Journal of the American Statistical Association
Journal of Computational and Graphical Statistics
Journal of Machine Learning Research
Neural Computation
Proceedings of the National Academy of Sciences
Quantitative Finance
Springer Verlag
Statistical Science
Statistics and Computing
Systematic Biology

CONFERENCE REVIEWING

Neural Information Processing Systems (NIPS), 2003, 2004, 2006, 2007, 2008, 2011, 2012, 2013
International Conference on Machine Learning (ICML), 2006, 2007, 2008, 2010, 2011, 2012, 2013
Artificial Intelligence and Statistics (AISTATS), 2007, 2010, 2011, 2012, 2013
Uncertainty in Artificial Intelligence (UAI), 2013
Workshop on Algorithms in Bioinformatics (WABI), 2007
Conference on Computational Learning Theory (COLT), 2004
Conference of ACM SIG for Information Retrieval (ACM-SIGIR), 2001

GRANT REVIEWING

Department of Energy, Advanced Scientific Computing Research, Early-Career Grants, 2015

GRANTS

Granting entity: Intel Corporation
Project title: Intel Parallel Computing Center, as part of the
National Energy Research Scientific Computing Big Data Center
Principal investigators: Prabhat (LBNL), JDM
Award: $152K in FY 2017

Granting entity: U.S. Department of Energy
Project title: MANTISSA
Principal investigators: Prabhat (LBNL), JDM, Michael Mahoney (UC Berkeley)
Award: $500K in FY 2014; $700K in FY 2015; $600K in FY 2016

Granting entity: Center for AIDS Research, University of Pennsylvania
Project title: Spatially varying evolutionary models of viral escape response
Principal investigators: JDM and Shane Jensen (U. Penn. Statistics)
Award: $36,144 in FY 2006

AWARDS AND FELLOWSHIPS

Lehmann citation for outstanding Ph.D. thesis in theoretical statistics, 2005
University of California Regents fellowship, 2000–2004
Outstanding graduate student instructor, UC Berkeley, 2003
Harvard College scholarship, 1991–1995
Harvard College Dean's list, 1991–1995
Outstanding teaching fellow, Harvard course evaluation guide, 1994
Harvard College advanced standing, 1991

INVITED TALKS

Statistical Challenges in Modern Astronomy VI, Pittsburgh, PA	Jun 2016
Information Theory and Applications Workshop, San Diego, CA	Feb 2015
University of California, Berkeley, Applied Mathematics Seminar	Dec 2014
Information Theory and Applications Workshop, San Diego, CA	Feb 2013
Stanford University, Biostatistics Dept	May 2010
University of California, Berkeley, Statistics Dept	Dec 2009
University of Southern California, Information and Operations Management Dept	Dec 2008
INFORMS annual meeting, machine learning and statistics session	Nov 2007
Introductory Overview Lecture (IOL) on machine learning and bioinformatics, Joint Statistical Meetings, Salt Lake City, UT	Jul 2007
University of Pennsylvania, Computer and Information Sciences Dept	Mar 2007
Symposium on the Interface, Pasadena, CA	May 2006
New York University, Information, Operations, and Management Sciences Dept	Apr 2006
University of Pennsylvania, Center for Bioinformatics	Mar 2006
University of California, Los Angeles, Statistics Dept	Feb 2005
Yale University, Statistics Dept	Feb 2005
Harvard University, Statistics Dept	Feb 2005
University of Washington, Genome Sciences Dept	Feb 2005
University of Chicago, Statistics Dept	Feb 2005
University of Pennsylvania, Statistics Dept	Jan 2005
Carnegie Mellon University, Statistics Dept	Jan 2005

TEACHING EXPERIENCE

STATISTICS 215B, UNIV. OF CALIFORNIA, BERKELEY. Second graduate course in applied statistics. Professor, Spring 2013-16, 2018, 2019, 2023-25.

STATISTICS 272, UNIV. OF CALIFORNIA, BERKELEY. Statistical consulting seminar. Professor, Spring 2021 and 2022.

STATISTICS 151B, UNIV. OF CALIFORNIA, BERKELEY. Advanced undergraduate course in modern statistical prediction and machine learning. Professor, Spring 2010, 2011, and 2012.

STATISTICS 471, UNIV. OF PENNSYLVANIA. Advanced undergraduate course in modern statistical prediction and machine learning. Professor, Spring 2007.

STATISTICS 991, UNIV. OF PENNSYLVANIA. Graduate seminar on ensemble methods in prediction. Professor, Spring 2006.

STATISTICS 431, UNIV. OF PENNSYLVANIA. First undergraduate course in statistical methods. Professor, Fall 2005 and 2006.

STATISTICS 215A, UNIV. OF CALIFORNIA, BERKELEY. First graduate course in applied statistics. Graduate Student Instructor, Fall 2002. Professor: David Freedman.

APPLIED MATH 107, HARVARD. Undergraduate graph theory and combinatorics. Teaching Fellow, Spring 1995. Professor: Leslie Valiant.

COMPUTER SCIENCE 182, HARVARD. Automated theorem proving in first-order logic; state-space search. Head Teaching Fellow, Fall 1994. Professor: Barbara Grosz.

COMPUTER SCIENCE 51, HARVARD. Introduction for majors, second semester. Teaching Fellow, Spring 1993 and 1994. Co-author, coursebook chapter on C++. Professor: Harry Lewis.

COMPUTER SCIENCE 50, HARVARD. Introduction for majors, first semester. Teaching Fellow, Fall 1993. Professor: Margo Seltzer.

INDUSTRY EXPERIENCE

VOLEON CAPITAL MANAGEMENT. Co-founder, Chief Investment Officer, July 2007–present.
Systematic investment management using modern statistical machine learning.

VERACYTE. Scientific advisory board member, 2013–2019.
Statistical methods for clinical molecular cytology.

EFFICIENT FRONTIER. Consulting research statistician, 2003–2004.
Predictive modeling of keyword-search ad click-through rates.

AFFYMETRIX. Research statistician, Summer 2003.
Genotyping using oligonucleotide microarrays. Support vector machine methods for inferring genotypes from probe intensities in a single-nucleotide polymorphism chip assay.

VINDIGO. Research statistician, Summer 2000.
Mobile computing. Minimum description length approach to optimal asymmetric compression for hand-held computers; utility-based optimization of dynamically generated navigational directions.

AMAZON.COM. Research statistician, Apr 1999–Mar 2000.
Personalized recommendations. Community-based item rankings;
high-dimensional model-based clustering; design of experiments for live website.

D. E. SHAW & CO. Quantitative analyst, Mar 1996–Mar 1999.
Vice President, international equity arbitrage.
Statistical equity arbitrage models. Portfolio optimization. Design of experiments for execution strategies. Third-market trading systems.

JUNO ONLINE (D. E. SHAW & CO.). Computer scientist, Aug 1995–Feb 1996.
Online services. Web advertising delivery and authentication; optimization of ad display schedule.

QUINCY FUTURES MANAGEMENT. Quantitative analyst, Summer 1994.
Futures trading. Covariance analysis; intraday currency trading strategies.