

JON DAVID MCAULIFFE

8 December 2011

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 POSITION

UNIVERSITY OF CALIFORNIA, BERKELEY
Assistant Professor (Adjunct), Statistics Department
July 2009 – present

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] A. Braunstein, S. T. Jensen, and J. McAuliffe. “Bayesian coalescent modeling of the HIV evolutionary response to therapy.” Submission under revision.
- [2] D. M. Blei and J. D. McAuliffe. “Supervised topic models.” Submission under revision.
- [3] 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.
- [4] 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.

- [5] 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.
- [6] 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.
- [7] 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.
- [8] P. L. Bartlett, M. I. Jordan, and J. D. McAuliffe (2006). “Discussion of ‘Support vector machine with applications’.” *Statistical Science* 21(3): 341–346.
- [9] 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.
- [10] 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.
- [11] 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.
- [12] J. McAuliffe, L. Pachter, and M. I. Jordan (2004). “Multiple-sequence functional annotation and the generalized hidden Markov phylogeny.” *Bioinformatics* 20(12): 1850–1860.
- [13] P. L. Bartlett, M. I. Jordan, and J. D. McAuliffe (2004). “Discussion of boosting papers.” *Annals of Statistics* 32(1): 85–91.
- [14] 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.
- [15] 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**

- [16] 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.
- [17] D. M. Blei and J. D. McAuliffe (2008). “Supervised topic models.” *Advances in Neural Information Processing Systems* 20.
- [18] 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.
- [19] B. I. P. Rubinstein, J. McAuliffe, S. Cawley, M. Palaniswami, K. Ramamohanarao, and T. P. Speed (2003). “Machine learning in low-level microarray analysis.” *ACM Special Interest Group for Knowledge Discovery in Databases: Explorations* 5(2): 130–139.

**PROFESSIONAL
SERVICE**

Editorial board, *Journal of Machine Learning Research*, 2010–

Program committee (area chair), *Neural Information Processing Systems (NIPS)*, 2009, 2010

Program committee (area chair), *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2009

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

Admissions committee member, *University of Pennsylvania Genomics and Computational Biology (Penn GCB) Ph.D. degree program*, 2006–7

Penn Ph.D. thesis committee member, student: Zhi Wei (defended March 2008)

Penn Ph.D. qualifying committee member, student: Qian Liu, 2007

Penn *Statistics Department seminar organizer*, 2006–2007

**JOURNAL AND
TEXTBOOK
REVIEWING**

The 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
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**

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

GRANTS

Granting agency: Center for AIDS Research, University of Pennsylvania
Project title: *Spatially varying evolutionary models of viral escape response*
Principal investigators: JDM and Prof. 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**

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

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, UC 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.
Statistical arbitrage and machine learning. Automated trading systems.

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.