Taylor B. Arnold

Curriculum Vitae

Yale University
Department of Statistics
24 Hillhouse Avenue
New Haven, CT 06511-6814

Phone: 508 479-5975
Email: taylor.arnold@yale.edu
Website: `euler.stat.yale.edu/~tba3`

PROFESSIONAL APPOINTMENTS

2015 - Present Yale University, Statistics, Lecturer

2014 - Present AT&T Labs Research, Senior Scientist

EDUCATION

2013 Ph.D. Statistics, Yale University
Thesis: "Inference with Ill-Conditioned Design Matrices"
Advisor: Sekhar Tatikonda
Co-Advisor: John Emerson

2009 M.A. Statistics, Yale University

2007 A.B. Mathematics, Bowdoin College
Thesis: "Convergence Properties of Simplex Optimization Algorithms"
Graduated *cum laude*

RESEARCH INTERESTS

Statistical analysis of petabyte-scale data
develop new techniques, tools, and paradigms for analyzing very large datasets (i.e., "big data"), with a particular focus on streaming, spatiotemporal feeds

Penalized estimation
combine structurally penalized models with modern optimization theory to enforce complex sparsity patterns in high-dimensional inference; applications include image analysis, epidemiology, and spatial point processes

Applications to humanities data
collaborate with humanities scholars to visualize and extract meaning from large text, image, and video corpora; engage in broader debates of Digital Humanities

PUBLICATIONS

Refereed Journal Articles

2015 Arnold, Taylor, Lauren Tilton, Stacey Maples, and Laura Wexler. "Uncovering Latent Metadata in the FSA-OWI Photographic Archive." *Digital Humanities Quarterly*. Accepted August 2015.

2014 Arnold, Taylor and Ryan Tibshirani. "Efficient Implementations of the Generalized Lasso Dual Path Algorithm." *Journal of Computational and Graphical Statistics*. Accepted December 2014. doi: 10.1080/10618600.2015.1043010.

2013 Grippo, Tomas, John Liu, Nazlee Zebardast, Taylor Arnold, Grant Moore, and Robert Weinreb. "Twenty-Four–Hour Pattern of Intraocular Pressure in Untreated Patients with Ocular Hypertension." *Investigative Ophthalmology & Visual Science* 54.1 (2013): 512-517. doi: 10.1167/iovs.12-10709.

2011 Arnold, Taylor, and John Emerson. "Nonparametric Goodness-of-Fit Tests for Discrete Null Distributions." *The R Journal* 3.2 (2011): 34-39.

2011 Emerson, John, and Taylor Arnold. "Statistical Sleuthing by Leveraging Human Nature: A Study of Olympic Figure Skating." *The American Statistician* 65.3 (2011). doi: 10.1198/tast.2011.10165.

2010 Russett, Bruce, and Taylor Arnold. "Who Talks, and Who's Listening? Networks of International Security Studies." *Security Dialogue* 41.6 (2010): 589-598. doi: 10.1177/09670106-10388205.

Books

2015 Arnold, Taylor and Lauren Tilton. *Humanities Data in R: Exploring Networks, Geospatial Data, Images, and Text*. New York: Springer International Publishing, 2015. doi: 10.1007/978-3-319-20702-5.

Manuscripts in Submission

2015 Arnold, Taylor. "An Entropy Maximizing Geohash for Distributed Spatiotemporal Database Indexing." *ACM Transactions on Spatial Algorithms and Systems (TSAS)*. Currently under revision. arXiv:1506.05158 [cs.DB]

2015 Arnold, Taylor, Michael Kane, and Simon Urbanek. "High-Performance I/O Tools for R." *R Journal*. Currently under revision. http://arxiv.org/abs/1510.00041 [stat.CO]

2015 Arnold, Taylor. *Case Studies in Large-Scale Statistical Learning*. Proposal currently under revision for inclusion in the series Springer Texts in Statistics.

2015 Arnold, Taylor. "Sparse Density Representations for Simultaneous Inference on Large Spatial Datasets." *Artificial Intelligence and Statistics* (AISTATS). Submitted September 2015. arXiv:1510.00755 [stat.CO, cs.DS]

2015 Arnold, Taylor, Michael Kane, and Simon Urbanek. "Computing Strategies for Fitting Regressions at Scale." *Journal of Statistical Software*. Submitted October 2015. arXiv:XXXX.XXXXX [stat.CO]

2015 Arnold, Taylor, and Lauren Tilton. "The NLP Pipeline and Deterministic Topic Constructions." *Debates in the Digital Humanities 2017*. University of Minnesota Press. Submitted October 2015.

Conference Proceedings

2012 Arnold, Taylor, "Defining the Charter: Judicial Activism and the Supreme Court of Canada." American Political Science Association 2012 Annual Meeting Paper. Available at SSRN: `http://ssrn.com/abstract=2109074`

2012 Arnold, Taylor. "User-Oriented High-Dimensional Linear Model Estimation." In JSM Proceedings, Statistical Computing Section. Alexandria, VA: American Statistical Association. 2429-2443.

Manuscripts in Preparation

Arnold, Taylor. "An Affine Constrained Fused Lasso for Discrete Spatial Distributions." In preparation for submission to *The Annals of Applied Statistics*.

Arnold, Taylor, Lauren Tilton. "Hierarchical Document Clustering: A Deterministic Method for Organizing Digital Texts." In preparation for submission to *Digital Studies / Le Champ Numérique*.

Arnold, Taylor, Michael Kane, and Simon Urbanek. "Tools for the Datalake: Indicies and Fast Queries over Large Deliminated Files." In preparation for submission to *Computational Statistics*.

TEACHING EXPERIENCE

Intructor (with full course responsibility)

Data Mining and Machine Learning (STAT 365/665) - Yale, Spring 2016
Linear Models (STAT 312/612) - Yale, Fall 2015
Introduction to Statistics (MAT 107) - Connecticut College, Spring 2011
Introduction to Statistics (STAT 107) - Yale, Summer 2010

Teaching Assistant

Statistical Consulting (STAT 627) - Yale, Fall 2009, Spring 2010
Data Analysis (STAT 361/661) - Yale, Fall 2010
Statistical Case Studies (STAT 325/625) - Yale, Fall 2010
Statistical Computing (STAT 662) - Yale, Spring 2011
Introductory Data Analysis (STAT 230/530) - Yale, Spring 2011, Spring 2012
Introduction to Statistics (STAT 103/503) - Yale, Fall 2012

INVITED TALKS

2016 Culture Analytics and User Experience Design, Institute for Pure and Applied Mathematics, UCLA, April 2016

2015 "Using R in Humanities Research", Institute for Liberal Arts Digital Scholarship (ILiADS), Hamilton College, July 2015

2015 "High Performance Data I/O", New England Statistics Symposium, University of Connecticut, April 2015

2015 "Oh the Places We'll Go", Department Seminar, Amherst College, February 2015

2013 "The Genlasso Package", Connecticut R Users Group, New Haven, CT, February 2013

2012 "Historical Data with Photogrammar", University Library Colloquium, Columbia University, March 2012

2011 "The Best Guess: Comparing Bernoulli Estimators", Department Seminar, Bowdoin College, November 2011

CONFERENCE ACTIVITY

2016 Joint Statistical Meetings, "What to do with messy data? Four case studies." Chair and Organizer. Invited Session, 30 July - 4 August.

2014 Digital Humanities, "Photogrammar: Organizing Visual Culture through Geography, Text Mining, and Statistical Analysis." 7-12 July.

2012 American Studies Association Annual Meeting, "Photogrammar in Puerto Rico: Reading the FSA's 1930s Visual Archive with Twenty-first-Century Visualization Tools." 15-18 November.

2012 Joint Statistical Meetings, "User Oriented High Dimensional Linear Model Estimation." Contributed Talk, 28 July - 2 August.

2012 Joint Statistical Meetings, "Statistical Sleuthing by Leveraging Human Nature." Invited Poster, 28 July - 2 August.

2012 useR! Meeting, "Package hdlm: Fitting High Dimensional Linear Models." Contributed Talk, 12-15 June.

2012 American Historical Association Annual Meeting, "From Archive to Interdisciplinary Tool: Transforming Our Image of the FSA-OWI Photograph Collection." Contributed Talk, January 5-8.

2011 American Studies Association Annual Meeting, "Digital Humanities Lightning Shorts." Contributed Talk, 20-23 October.

2011 useR! Meeting, "Nonparametric Goodness-of-Fit Tests for Discrete Null Distributions." Contributed Talk, 16-18 August.

AWARDS

2013 Yale Thesis Writing Prize

2011-2015 NEH Digital Start-Up Grant, Photogrammar, Co-Director

2008-2013 Yale Graduate Fellowship

ADVISING

2015-2016 David Marcano, Senior Thesis Advisor, Yale Statistics
2015 Yutaro Yamada, Undergraduate Summer Intern, Yale Graphics Group

PROFESSIONAL SERVICE

2015-2016 American Statistical Association, Section Program Chair
2015 Biometrika, Referee
2014-2015 Journal of Statistical Software, Referee
2013-2015 Neurology, Referee
2012-2014 The R-Journal, Referee

UNIVERSITY SERVICE

2009-2013 Graduate Student Assembly, Statistics Representative
2010-2011 Yale Tribunal, Member
2010-2011 Yale Graduate School Executive Committee, Member

OTHER POSITIONS

2013-2014 Traveler's Insurance, Research and Development, Manager
2011 Visiting Instructor, Connecticut College
2007-2008 IBM, Global Business Services, Consultant

SELECTED SOFTWARE

genlasso: Path algorithm for generalized lasso problems. R package.
`cran.r-project.org/web/packages/genlasso`
glmgen: Generalized lasso implementations. C library and R package.
`github.com/statsmaths/glmgen`
iotools: I/O tools for streaming. R package.
`cran.r-project.org/web/packages/iotools`
iosub: Fast selection of data subsets. C library and R package.
`github.com/statsmaths/glmgen`
dgof: Discrete Goodness-of-Fit Tests. R package.
`cran.r-project.org/web/packages/dgof`

PROFESSIONAL MEMBERSHIPS

American Statistical Association (ASA)
Association for Computing Machinery (ACM)
American Mathematical Society (AMS)

REFERENCES

John Emerson
24 Hillhouse Avenue
New Haven, CT 06511
john.emerson@yale.edu

Michael Kane
30 Thomas Street
New York, NY 10007
kane@att.research.com

Sekhar Tatikonda
18 Hillhouse Avenue
New Haven, CT 06511-6607
sekhar.tatikonda@yale.edu

Ryan Tibshirani
229B Baker Hall
Pittsburgh, PA 15213
ryantibs@cmu.edu

Laura Wexler
100 Wall Street
New Haven, CT 06511-6607
laura.wexler@yale.edu