Aaron J. S. Schein

www.aaronschein.com

Research Interests

Machine Learning • Bayesian Statistics • Applied Causal Inference • Tensor Methods • Time Series Analysis • Computational Social Science • Political Science • Computational Genomics

Education

PhD Computer Science, University of Massachusetts Amherst 2019
Thesis: *Allocative Poisson Factorization for Computational Social Science*
Advisor: Hanna Wallach.

MS Computer Science, University of Massachusetts Amherst 2017

MA Linguistics, University of Massachusetts Amherst 2012

BA Political Science & Linguistics, University of Massachusetts Amherst 2010

Academic Appointments

Assistant Professor 2022–Present
Department of Statistics & Data Science Institute, University of Chicago

Postdoctoral Fellow 2019–2022
Data Science Institute, Columbia University
Mentors: David Blei, Donald Green

Doctoral Research Assistant 2012–2018
College of Information and Computer Sciences, UMass Amherst
Advisor: Hanna Wallach

Professional Experience

Senior Research Scientist 2021–2022
PredictWise & Ocurate, New York, NY

Research Intern Summers 2014 and 2015
Microsoft Research, New York, NY
Computational Social Science lab

Software Engineering Intern Summer 2013
Google, Mountain View, CA
Machine Intelligence group

Artificial Intelligence Engineer 2011–2013
MITRE Corporation, McLean, VA
Human Language Technologies group

Teaching Experience

Instructor, CICS 191 FYS *Philosophy of A.I. & Theories of Consciousness* Fall 2016
College of Information and Computer Sciences, UMass Amherst

Guest Lecturer, COMS E6998 *Machine Learning with Probabilistic Programming* Fall 2020
Department of Computer Science, Columbia University

Guest Lecturer, CICS 590N *Intro to Numerical Computing with Python* Fall 2016
College of Information and Computer Sciences, UMass Amherst

Teaching Assistant, LING 414 *Intro to Phonetics* Fall 2009
Department of Linguistics, UMass Amherst

Journal articles

S. He*, **A. Schein***, V. Sarsani, and P. Flaherty. A Bayesian nonparametric model for inferring subclonal populations from structured DNA sequencing data. *Annals of Applied Statistics*. * = Equal contribution.

Conference proceedings publications

A. Schein, A. Nagulpally, H. M. Wallach, and P. Flaherty. Doubly non-central Beta matrix factorization for DNA methylation data. *Conference on Uncertainty in Artificial Intelligence (UAI)*, 2021.

A. Schein, K. Vafa, D. Sridhar, V. Veitch, J. Quinn, J. Moffet, D. M. Blei, and D. P. Green. Assessing the effects of friend-to-friend texting on turnout in the 2018 US midterm elections. *The Web Conference (WWW)*, 2021.

A. Schein, S. Linderman, M. Zhou, D. M. Blei, and H. M. Wallach. Poisson-randomized gamma dynamical systems. *Advances in Neural Information Processing Systems (NeurIPS)*, 2019.

A. Schein, Z. Wu, A. Schofield, M. Zhou, and H. M. Wallach. Locally private Bayesian inference for count models. *International Conference on Machine Learning (ICML)*, 2019.

A. Schein, M. Zhou, and H. M. Wallach. Poisson–gamma dynamical systems. *Advances in Neural Information Processing Systems (NeurIPS)*, 2016. Selected for a full oral presentation (among 8% of accepted papers).

A. Schein, M. Zhou, D. M. Blei, and H. M. Wallach. Bayesian Poisson Tucker decomposition for learning the structure of international relations. *International Conference on Machine Learning (ICML)*, 2016.

A. Schein, J. Paisley, D. M. Blei, and H. M. Wallach. Bayesian Poisson tensor factorization for inferring multilateral relations from sparse dyadic event counts. *ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*, 2015.

Workshop proceedings publications

A. Schofield, **A. Schein**, Z. Wu, M. Zhou, and H. M. Wallach. A variational inference approach for locally private inference of Poisson factorization models. *Advances in Approximate Bayesian Inference*, 2018.

A. Schein, Z. Wu, M. Zhou, and H. M. Wallach. Locally private Bayesian inference for count models. *Advances in Approximate Bayesian Inference*, 2017.

A. Schein, P. Flaherty, M. Zhou, D. Sheldon, and H. M. Wallach. Beta Tucker decomposition for DNA methylation data. *NeurIPS Workshop on "Computational Biology"*, 2016. Selected for oral presentation (among 20% of accepted papers).

A. Schein, M. Zhou, D. M. Blei, and H. M. Wallach. Modeling topic-partitioned assortativity and disassortativity in dyadic event data. *NeurIPS Workshop on "Networks in the Social and Information Sciences"*, 2015. Winner of Best Student Poster award (prize: $400).

A. Schein, J. Paisley, D. M. Blei, and H. M. Wallach. Inferring polyadic events with Poisson tensor factorization. *NeurIPS Workshop on "From Graphs to Rich Data"*, 2014.

A. Schein, J. Moore, and H. M. Wallach. Inferring multilateral relations from dynamic pairwise interactions. *NeurIPS Workshop on "Frontiers of Network Analysis"*, 2013.

Refereed extended abstracts

A. Schein, D. M. Blei, and D. P. Green. Assessing the effects of friend-to-friend texting on turnout in the 2020 US presidential election. *Conference on Digital Experimentation (CODE@MIT)*, 2021

A. Schein, D. M. Blei, and D. P. Green. A pair of large-scale digital field experiments reveal large effects of friend-to-friend texting on voter turnout in the 2018 and 2020 US elections. *Conference on Politics and Computational Social Science (PaCSS)*, 2021.

A. Nagulpally, **A. Schein**, H. M. Wallach, and P. Flaherty. Matrix factorization for DNA methylation data. *Great Lakes Bioinformatics Conference (GLBIO)*, 2021.

A. Schein, M. Zhou, D. M. Blei, and H. M. Wallach. An experimental study of friend-to-friend GOTV text messages in the 2018 midterm elections. *International Conference on Computational Social Science (IC2S2)*, 2020. Winner of the Best Presentation award.

A. Schein, Z. Wu, A. Schofield, M. Zhou, and H. M. Wallach. Locally private Bayesian inference for count models. *International Conference on Computational and Methodological Statistics*, 2019.

A. Schofield, **A. Schein**, Z. Wu, M. Zhou, and H. M. Wallach. Toward practical and locally private inference of topic models. *New Directions in Analyzing Text as Data (TADA)*, 2018.

B. Kim, **A. Schein**, B. Desmarais, and H. M. Wallach. A network model for dynamic textual communications with application to government email corpora. *New Directions in Analyzing Text as Data (TADA)*, 2017.

B. Kim, **A. Schein**, B. Desmarais, and H. M. Wallach. A network model for dynamic textual communications with application to government email corpora. *Political Networks Conference*, 2017.

A. Schein, M. Zhou, D. M. Blei, and H. M. Wallach. Modeling international relations with Bayesian Poisson Tucker decomposition. *International Conference on Computational Social Science (IC2S2)*, 2016.

A. Schein, J. Paisley, M. Zhou, D. M. Blei, and H. M. Wallach. Dynamic Bayesian Poisson tensor factorization. *Conference on Bayesian Nonparametrics*, 2015.

Newspaper articles

A. Schein. Joe Biden's 'virtual ground game' gamble seems to have paid off. *Financial Times Opinion*, November 8, 2020.

Professional Service

Organizer of Columbia Data Science Institute's Distinguished Lecture Series.

Co-organizer of NeurIPS 2021 Workshop "I (Still) Can't Believe It's Not Better!".

Co-organizer of NeurIPS 2020 Workshop "I Can't Believe It's Not Better!".

Co-organizer of NeurIPS 2016 Workshop "Practical Bayesian Nonparametrics".

Co-organizer of NeurIPS 2015 Workshop "Bayesian Nonparametrics: The Next Generation".

Reviewer/PC member for conferences: NeurIPS 2015–2021, ICML 2015–2021, UAI 2015–2021, AISTATS 2015–2021, IJCAI 2015, AAAI 2015.

Reviewer for journals: Annals of Applied Statistics (AoAS), PloS One, PLOS Biology, Journal of Machine Learning Research Transactions on Knowledge and Data Engineering, Physical Review E, American Journal of Political Science, Statistics and Computing, IEEE Transactions on Signal Processing.

Selected Awards and Honors

Data Science Post-Doctoral Fellowship, 2019–2022.
Data Science Institute, Columbia University.

Best Oral Presentation, July 2020.
6th International Conference on Computational Social Science (IC2S2).

NeurIPS Best Reviewer Award, 2019.
Top 10% of reviewers at NeurIPS 2019. Prize: free registration.

Best Student Poster Award, December 2015.
NeurIPS workshop on Networks in the Social and Information Systems. Prize: $400.

First prize in cafe-naming contest, November 2015.
College of Information and Computer Sciences, University of Massachusetts Amherst.
Winning name: *Snack Overflow*. Prize: free cup of coffee.

References

Hanna M. Wallach
Senior Principal Researcher
Microsoft Research, New York, NY
Email: hanna@dirichlet.net
Relationship: Doctoral advisor

David M. Blei
Professor
Columbia University, New York, NY
Departments of Computer Science and Statistics
Email: david.blei@columbia.edu
Relationship: Postdoctoral advisor

Donald P. Green
Burgess Professor of Political Science
Columbia University, New York, NY
Department of Political Science
Email: dpg2110@columbia.edu
Relationship: Postdoctoral advisor

Mingyuan Zhou
Associate Professor
University of Texas Austin
McCombs School of Business
Email: mingyuan.zhou@mccombs.utexas.edu
Relationship: Collaborator and PhD committee member

Patrick Flaherty
Assistant Professor
University of Massachusetts Amherst
Department of Mathematics and Statistics
Email: flaherty@math.umass.edu
Relationship: Collaborator and PhD committee member

David M. Rothschild
Economist
Microsoft Research, New York, NY
Email: david@researchdmr.com
Relationship: Mentor at PredictWise and Ocurate