

Yuling Yao

162 5th Ave · New York NY 10010 · 646-908-6510
yyao@flatironinstitute.org · <https://yulingyao.com>

Employment

Flatiron Research Fellow, 2021–present *New York, NY*
Flatiron Institute, Center for Computational Mathematics.

Education

Columbia University, 2015–2021 *New York, NY*
Ph.D. Statistics May 2021
Advisor: Andrew Gelman
Dissertation: Toward a scalable Bayesian workflow.

Tsinghua University, 2011–2015 *Beijing, China*
B.S. Mathematics, with distinction Jul 2015
B.Ec. Economics Jul 2015

Research interests

- My ultimate goal is to develop a scalable Bayesian workflow for real problems. Past applications include the lead fallout in Paris, arsenic diffusion in groundwater, and Covid-19 mortality in Bangladesh.
- But better applied statistics needs better methodologies. To that end, I investigate statistical and machine learning methods for model evaluation and aggregation. Some recent focuses are on cross-validation, stacking, and Bayesian causal inference.
- But to facilitate complex methods further needs scalable and diagnosable computing. Hence, I develop algorithms for fully Bayesian and approximate computations. Recently, I am interested in combining Monte Carlo methods with sophisticated numerical tricks, including importance sampling, simulated tempering and annealing, and free energy estimation.

Published papers

1. P. Barnwal, **Y. Yao** (co-first author), Y. Wang, N. A. Juy, S. Raihan, M. A. Haque, A van Geen. (2021). "Assessment of excess mortality and household income in rural Bangladesh during the Covid-19 pandemic in 2020." *JAMA Network Open*, in press.
2. Y. Su, D. Lien, **Y. Yao**. (2021). "Economic growth and happiness in China: A Bayesian multi-level age-period-cohort analysis based on the CGSS data 2005-2015.", *International Review of Economics and Finance*.
3. **Y. Yao**, G. Pirš, A. Vehtari, A. Gelman. (2021). "Bayesian hierarchical stacking: Some models are (somewhere) useful." *Bayesian Analysis*, in press.
4. A. Gelman, **Y. Yao**. (2020). "Holes in Bayesian statistics." *Journal of Physics G*.
5. A. van Geen, **Y. Yao**, T. Ellis, A. Gelman. (2020). "Fallout of lead over Paris from the 2019 Notre-Dame Cathedral fire." *Geohealth*.
6. **Y. Yao**. (2020). "Bayesian aggregation." *Wiley StatsRef*.
7. A. Vehtari, D. Simpson, **Y. Yao**, A. Gelman. (2018). "Limitations of 'limitations of Bayesian leave-one-out cross-validation for model selection'." *Computational Brain & Behavior*.

8. Y. Yao, A. Vehtari, D. Simpson, A. Gelman. (2018). "Yes, but did it work?: Evaluating variational inference." In *International Conference on Machine Learning*.
9. Y. Yao, A. Vehtari, D. Simpson, A. Gelman. (2018). "Using stacking to average Bayesian predictive distributions (with discussions and rejoinder)." *Bayesian Analysis*.
10. M. Marsman, D. Schonbrodt, D. Morey, Y. Yao, A. Gelman, E-J. Wagenmakers. (2017). "A Bayesian bird's eye view of 'replications of important results in social psychology'." *Royal Society Open Science*.

Preprints submitted for publications

1. Y. Yao, A. Vehtari. (2021+). "Make cross-validation Bayes again." under review.
2. Y. Yao, A. Vehtari, A. Gelman. (2021). "Stacking for non-mixing Bayesian computations: The curse and blessing of multimodal posteriors." *Journal of Machine Learning Research*, conditional acceptance.
3. Y. Yao, R. Mozumder, B. Bostick, B. Mailloux, C. Harvey, A. Gelman, A. van Geen. (2021+). "Making the most of imprecise measurements: Changing patterns of arsenic concentrations in shallow wells of Bangladesh from laboratory and field data." under review.
4. Y. Yao, C. Cademartori, A. Vehtari, A. Gelman. (2020+). "Adaptive path sampling in metastable posterior distribution." under review.
5. A. Gelman, A. Vehtari, D. Simpson, C. Margossian, B. Carpenter, Y. Yao, P-C. Bürkner, L. Kennedy, J. Gabry, M. Modrák. (2020+). "Bayesian workflow." under review.
6. A. Vehtari, D. Simpson, A. Gelman, Y. Yao, J. Gabry. (2019+). "Pareto smoothed importance sampling." under review.
7. O. Chang, Y. Yao, D. W-King, H. Lipson. (2019+). "Ensemble model patching: A parameter-efficient variational Bayesian neural network." preprint.

Open source softwares

- loo. (R package, *co-author*) 2017-present
Efficient approximate leave-one-out cross-validation using Pareto smoothed importance sampling.
- Stan. (*core developer team*) 2017-present
A state-of-the-art platform for statistical modeling and high-performance statistical programming.

Professional service

- *Refereed articles for journals:*
Annual of Applied Statistics, Journal of Computational and Graphical Statistics, Journal of Nonparametric Statistics, Neural Processing Letters, Philosophy of Science.
- *Reviewed articles for conferences:*
NeurIPS 2018, 2021; ICLR 2019; ICML 2019.

Conference presentation

- *Oral presentations* in ICML 2018; JSM 2018, 2019; StanCon 2018, 2019.
- *Invited talk* in CMStatistics 2020, Flatiron Institute 2021, Google Research New York 2021, Queensland University of Technology EC Bayes Seminar Series 2021.