Yuexi Wang

The University of Chicago
Booth School of Business
5807 South Woodlawn Avenue
Chicago, IL 60637

Email: yuexi.wang@chicagobooth.edu
Web: https://yxwang99.github.io

EDUCATION

University of Chicago Booth School of Business *2018-Present*
Ph.D. student in Econometrics and Statistics

University of Chicago *2016-2018*
M.S. in Statistics

Zhejiang University, Hangzhou, China *2012-2016*
B.S. in Mathematics and Applied Mathematics (with honors)

RESEARCH INTERESTS

Deep Learning, Approximate Bayesian Inference, Trees, Machine Learning, Bayesian Nonparametrics

PUBLICATIONS

Data Augmentation for Bayesian Deep Learning
Wang, Y., Polson, N. G., and Sokolov, V. O. (2022)
Bayesian Analysis (accepted)

Variable Selection with ABC Bayesian Forests
Liu, Y., Rockova, V., and **Wang, Y.** (2021)
Journal of the Royal Statistical Association: Series B (Statistical Methodology), 83(3), 453-481.

Uncertainty Quantification for Sparse Deep Learning
Wang, Y. and Rockova, V. (2020)
International Conference on Artificial Intelligence and Statistics, pages 298–308. PMLR.

WORKING PAPER

Adversarial Bayesian Simulation
Wang, Y. and Rockova, V. (2022)

Approximate Bayesian Computation via Classification
Wang, Y., Kaji, T. and Rockova, V. (2021)

Sparse Regularization in Marketing and Economics
Feng, G., Polson, N., **Wang, Y.**, and Xu, J. (2018)

AWARDS

j-ISBA Award for Talks, ISBA New Researcher Travel Award (BAYSM 2022) *June 2022*
ISBA 2022 Travel Award *June 2022*
Arnold Zellner Doctoral Prize, The University of Chicago Booth School of Business *2022*
Ph.D. Program Fellowship, The University of Chicago Booth School of Business *2018 - Present*
Winner of Citadel's Chicago Datathon *May 2017*
Overseas Research Fellowship, Zhejiang University *2015 - 2016*

PRESENTATIONS

2022: Booth E&S Student Seminar (Febuary), TTIC Machine Learning Seminar (April), BAYSM (June), ISBA (June, contributed), JSM (August, contributed), INFORMS (October, contributed)
2021: ISBA (June, contributed), Sparsity in Neural Networks (July), JSM (August, contributed)
2020: AISTATS (August)

REFEREE ACTIVITIES

Statistical Science, ICML 2022, Bayesian Analysis, NeurIPS (2021, 2022), Statistics and Probability Letters

TEACHING ASSISTANTSHIPS

Big Data (MBA elective) *2019 - 2022*
Business Statistics (MBA elective) *2019 - 2022*

SOFTWARE

alphanorm. R package for alpha-norm regularization linear model.
ABCforest. R package for variable selection with ABC forest, available upon request.

SKILLS

R, Python, Matlab, C.
Also experience with big data processing (Scala, Hadoop, Spark, Hive sql).

REFERENCES

Nicholas G. Polson (co-chair)
Robert Law, Jr. Professor of Econometrics
and Statistics
University of Chicago
Booth School of Business
nicholas.polson@chicagobooth.edu

Veronika Ročková (co-chair)
Professor of Econometrics and Statistics
and James S. Kemper Foundation Faculty Scholar
University of Chicago
Booth School of Business
veronika.rockova@chicagobooth.edu

Sanjog Misra
Charles H. Kellstadt Professor of Marketing
University of Chicago
Booth School of Business
sanjog.misra@chicagoboothedu

Chao Gao
Assistant Professor of Statistics
University of Chicago
Department of Statistics
chaogao@uchicago.edu

EXPERIENCE

Data Scientist Intern, Google Research. *June 2021 - Aug 2021*
Data Science Intern, Conversant Media. *June 2020 - Sep 2020*
Research Assistant, Research Computing Center, University of Chicago. *Jan 2017 - Aug 2017*
Research Assistant, Channing Division of Network Medicine, Harvard Medical School. *Oct 2015 - June 2016*

SELECTED COURSEWORK

Statistics

High Dimensional Statistics 1 & 2, Bayesian Nonparametrics, Bayesian Statistics, Topics in Selective Inference, Advanced Statistical Inference 2, Fundamentals of Deep Learning, Time Dependent Data.

Economics

Price Theory 2 & 3, Theory of Income 1, Applied & Advanced Econometrics, Advanced Industrial Organization 1, Topics Information Economics.

Marketing

Foundations of Advanced Quantitative Marketing, Advanced Quantitative Marketing, Applied Bayesian Econometrics.

SELECTED RESEARCH ABSTRACT

Adversarial Bayesian Simulation
job market paper (2022)

In the absence of explicit or tractable likelihoods, Bayesians often resort to approximate Bayesian computation (ABC) for inference. Our work bridges ABC with deep neural implicit samplers based on generative adversarial networks (GANs) and adversarial variational Bayes. Both ABC and GANs compare aspects of observed and fake data to simulate from posteriors and likelihoods, respectively. We develop a Bayesian GAN (B-GAN) sampler that directly targets the posterior by solving an adversarial optimization problem. B-GAN is driven by a deterministic mapping learned on the ABC reference by *conditional* GANs. Once the mapping has been trained, iid posterior samples are obtained by filtering noise at a negligible additional cost. We propose two post-processing local refinements using (1) data-driven proposals with importance reweighing, and (2) variational Bayes. We support our findings with frequentist-Bayesian results, showing that the typical total variation distance between the true and approximate posteriors converges to zero for certain neural network generators and discriminators. Our findings on simulated data show highly competitive performance relative to some of the most recent likelihood-free posterior simulators.

Variable Selection with ABC Bayesian Forests
JRSS-B (2021)

Few problems in statistics are as perplexing as variable selection in the presence of very many redundant covariates. The variable selection problem is most familiar in parametric environments such as the linear model or additive variants thereof. In this work, we abandon the linear model framework, which can be quite detrimental when the covariates impact the outcome in a non-linear way, and turn to tree-based methods for variable selection. Such variable screening is traditionally done by pruning down large trees or by ranking variables based on some importance measure. Despite heavily used in practice, these ad-hoc selection rules are not yet well understood from a theoretical point of view. In this work, we devise a Bayesian tree-based probabilistic method and show that it is consistent for variable selection when the regression surface is a smooth mix of $p > n$ covariates. These results are the first model selection consistency results for Bayesian forest priors. Probabilistic assessment of variable importance is made feasible by a spike-and-slab wrapper around sum-of-trees priors. Sampling from posterior distributions over trees is inherently very difficult. As an alternative to MCMC, we propose ABC Bayesian Forests, a new ABC sampling method based on data-splitting that achieves higher ABC acceptance rate. We show that the method is robust and successful at finding variables with high marginal inclusion probabilities. Our ABC algorithm provides a new avenue towards approximating the median probability model in non-parametric setups where the marginal likelihood is intractable.