Yuexi Wang

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

Email: yuexi.wang@chicagobooth.edu
Web: https://ywang99.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

Bayesian Inference, Machine Learning, High-dimensional Inference, Bayesian Nonparametrics

RESEARCH WORK

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

Variable Selection with ABC Bayesian Forests
Liu, Y., Rockova, V., and **Wang, Y.**
Journal of the Royal Statistical Association (Series B) (2021)

Uncertainty Quantification for Sparse Deep Learning
Wang, Y. and Rockova, V.
23rd Conference on Artificial Intelligence and Statistics (2020)

Scalable Data Augmentation for Deep Learning
Wang, Y., Polson, N. G., and Sokolov, V. O. (2019)

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

PRESENTATIONS

2021: ISBA (June, contributed), Sparsity in Neural Networks (July), JSM (August, contributed)

2020: AISTATS (August)

REFEREE ACTIVITIES

Bayesian Analysis, NeurIPS 2021, Statistics and Probability Letters

AWARDS

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*

TEACHING ASSISTANTSHIPS

Big Data (MBA elective) *2019 - 2021*
Business Statistics (MBA elective) *2019 - 2021*

SOFTWARE

alphanorm. R package for alpha-norm regulariztaion 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
Robert Law, Jr. Professor of Econometrics and Statistics
University of Chicago
Booth School of Business
nicholas.polson@chicagobooth.edu

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

Ruey S. Tsay
H.G.B. Alexander Professor of Econometrics and Statistics
University of Chicago
Booth School of Business
ruey.tsay@chicagobooth.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

Uncertainty Quantification for Sparse Deep Learning
AISTATS (2020)

Deep learning methods continue to have a decided impact on machine learning, both in theory and in practice. Statistical theoretical developments have been mostly concerned with approximability or rates of estimation when recovering infinite dimensional objects (curves or densities). Despite the impressive array of available theoretical results, the literature has been largely silent about uncertainty quantification for deep learning. This paper takes a step forward in this important direction by taking a Bayesian point of view. We study Gaussian approximability of certain aspects of posterior distributions of sparse deep ReLU architectures in non-parametric regression. Building on tools from Bayesian non-parametrics, we provide semi-parametric Bernstein-von Mises theorems for linear and quadratic functionals, which guarantee that implied Bayesian credible regions have valid frequentist coverage. Our results provide new theoretical justifications for (Bayesian) deep learning with ReLU activation functions, highlighting their inferential potential.

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.