Ziwei Su

Email: ziwei.su@northwestern.edu Mobile: +1-765-250-0860 Website LinkedIn GitHub

Education

Northwestern University, *McCormick School of Engineering* — Evanston, IL
Ph.D., *Industrial Engineering and Management Sciences* — 2021 - 2026 (expected)
Advisor: Diego Klabjan, collaborated with Matthew Plumlee (Principal Applied Scientist at Amazon)
Major: Applied Statistics and Statistical Learning; Minor: Stochastic Analysis and Simulation

Purdue University, *Department of Statistics* — West Lafayette, IN
Master of Science, *Mathematical Statistics* — 2019 - 2021
Advisor: Raghu Pasupathy

Fudan University, *School of Mathematical Sciences* — Shanghai, China
Bachelor of Science, *Mathematics and Applied Mathematics* — 2015 - 2019
International Exchange, University of Nottingham, Nottingham, United Kingdom — 2018

Skills

Programming Languages: Python, R, MATLAB, C, LaTeX.
Softwares: SAS, Gurobi, Microsoft Office
Learning Frameworks & Packages: PyTorch, Pandas, Scikit-learn.
Natural Languages: English (TOEFL iBT: 114 in Nov 2020), Cantonese, Mandarin.

Research Interest and Expertise

Statistical machine learning, stochastic modeling, computer simulation, and applied probability.

Publication

1. **Z. Su**, M. Plumlee, D. Klabjan. *Train Stochastic Models like GANs: Learning Model Parameters with Consistency and Confidence under Model Discrepancy. Manuscript in preparation.*
   - **"Training stochastic models like GANs"**
   - *Proposed the **first** procedure to learn the input parameter with consistency and confidence under **model discrepancy**, using output-level training data of a real system.*
   - *Proposed a **training procedure based on kernel score and stochastic gradient descent (SGD)**.*
   - *Developed **the first differentiable simulators for queueing systems** in **PyTorch**.*
   - *Applicable in all generative model settings where simulating is easy but the likelihood is intractable.*

2. **Z. Su**, R. Pasupathy, Y. Yeh, P. W. Glynn. *Overlapping Batch Confidence Intervals on Statistical Functionals Constructed from Time Series: Application to Quantiles, Optimization, and Estimation* [pdf]. *To appear* in ACM Transactions on Modeling and Computer Simulation, 2023.
   - **"Overlapping batch is all you need for non-parametric inference."**
   - *Proposed a novel confidence interval procedure for statistical functionals in stationary time series, based on **novel distribution-free analogs of the chi-square and Student's t** random variables in the statistical functional context.*
   - *Produced **significantly better confidence intervals than subsampling or bootstrap** in extensive numerical experiments.*
   - *Applicable in various settings including gradient estimation, quantile estimation, M-estimation, and optimization.*

3. H. Ding, **Z. Su**, X. Liu. *A Modified Multinomial Baseline Logit Model with Logit Functions Having Different Covariates* [pdf]. Communications in Statistics - Simulation and Computation, 2020.
   - *Proposed a new modified multinomial baseline logit model where separate logistic models may be functions of different covariates.*
   - *Outperformed the multinomial baseline logit model and the multinomial sparse group lasso for nominal polychotomous data.*
   - *Analyzed a real data set about an adolescent placement study to demonstrate the flexibility and efficiency of the modified model.*

Selected Project Experience

1. Denoising Diffusion Probabilistic Models (DDPM) for CIFAR-10 and MNIST [Link], Oct 2023 - Dec 2023. *Trained the Denoising Diffusion Probabilistic Models (DDPM) on CIFAR-10 and MNIST dataset for image synthesis. Visualized the generated images for different time steps in the training process. Compared the fine-tuned DDPM model on CIFAR-10 with the score-based generative model and BigGAN. Explored the effect of pre-training with the CIFAR-100 dataset. Explored how group normalization layers and attention layers affect model performance.*

2. A New EM Algorithm for Bayesian Networks Parameter Estimating with Uncertain Data: A Study in COVID-19 Symptom Network [GitHub], Mar 2020 - May 2020. *Studied the inter-relationships among demographics, symptoms, and outcomes for COVID-19 patients with early epidemiological data. Proposed a new Expectation-maximization (EM) algorithm for the expected log-likelihood in Bayesian networks to deal with data uncertainty and estimate network parameters. Developed an interactive network app 'BN for COVID-19' with R-shiny to demonstrate the learned relationships [Link].*

3. A Numerical Exploration of Stochastic Gradient Langevin Dynamics in Thompson Sampling [Link], Oct 2020 - Dec 2020. *Reviewed the regular Thompson sampling algorithms and the stochastic gradient Langevin dynamics (SGLD) by Mazumdar et al. (2020). Compared regular Thompson sampling with SGLD with extensive numerical experiments on Beta-Bernoulli bandits. Suggested using different priors such as spike-and-slab Gaussian-Laplace to improve performance.*

4. Generalization in Deep Learning: Frameworks and Bound Estimation Methods [GitHub], Nov 2021 - Jan 2022. *Reviewed the existing theoretical framework in PAC-Bayes, kernel learning, and distributional robustness to explain the generalization of over-parameterized deep neural networks, and ways to estimate the generalization bound. Leveraged PyTorch to validate the conclusion in Jiang et al. (2021) by training ResNet18 on CIFAR-10 with multiple hyperparameter configurations.*

5. Suicide Rates Modeling [Link], Oct 2020 - Dec 2020. *Leveraged the Poisson count model, Poisson rate model, generalized linear mixed model (GLMM), and generalized estimating equation (GEE) model to study the socio-economic and demographic factors affecting suicide rates with two real-world datasets. Transformed the data, identified the predictors, ran model diagnosis, and estimated the dispersion parameter to fit the quasi-poisson model. Identified that the higher suicide rates are related to lower GDP, lower GDI, male, and elderly people.*

Selected Coursework

Northwestern University: Machine Learning, Deep Learning, Linear Optimization, Non-linear Optimization, Mathematical Statistics, Measure Theory, Advanced Probability, Advanced Stochastic Process, Stochastic Simulation, Statistical Learning, Operations Management, Logistics, Bayesian Inference, Predictive Analytics II (Winter 2024).
Purdue University: Linear Models, Design of Experiment, Intermediate and Advanced Statistical Computation, Stochastic Process, Mathematical Statistics, Intermediate and Advanced Statistical Methods, Machine Learning and Dynamic System, Stochastic Optimization, Convex Optimization, Measure Theory.
Fudan University: Analysis, Linear Algebra, Measure Theory, Functional Analysis, Differential Equation, Numerical PDE, Mathematical Finance, Time Series Analysis, Programming.

Teaching Experience

Instructor, IEMS Bootcamp, Statistics for new PhD Students, Northwestern University — Fall 2022
Teaching Assistant, IEMS 303 Statistics, Northwestern University — Spring, Fall 2023
Grader, STAT 503 Statistical Methods for Biologists, Purdue University — Fall 2020

Conference Presentations

Session Speaker, *Calibration of Inexact Stochastic Simulation Models via Energy Score* [Abstract], INFORMS Annual Meeting, Phoenix, AZ, Oct 2023.
Session Chair, Session Speaker, *Stochastic Simulation Calibration of Inexact Computer Models* [Abstract], INFORMS Annual Meeting, Indianapolis, IN, Oct 2022.