Zhengyan Wan

Email | Google Scholar | Github | Homepage

EDUCATION

East China Normal University **Shanghai, China**

M.S. in Statistics. Advised by Fang Fang Sept 2023 – Jun 2026

- GPA: 3.88/4.0
- Major Courses: Advanced Probability Theory, Advanced Mathematical Statistics, Statistical Learning Theory, Convex Optimization, Survival Analysis, Empirical Processes, Stochastic Analysis

Jilin University **Changchun, China**

B.S. in Statistics Sept 2019 – Jun 2023

- GPA: 3.85/4.0
- Major Courses: Ordinary Differential Equations, Functional Analysis, Time Series Analysis, Multivariate Statistical Analysis, Nonparametric Statistics

PREPRINTS

[1] **Zhengyan Wan**, Yidong Ouyang, Qiang Yao, Liyan Xie, Fang Fang, Hongyuan Zha, and Guang Cheng. "Error Analysis of Discrete Flow with Generator Matching". In: *arXiv preprint* (2025). arXiv: 2509.21906. URL: https://arxiv.org/abs/2509.21906.

[2] **Zhengyan Wan***, Yidong Ouyang*, Liyan Xie, Fang Fang, Hongyuan Zha, and Guang Cheng. "Discrete Guidance Matching: Exact Guidance for Discrete Flow Matching". In: *arXiv preprint* (2025). arXiv: 2509.21912. URL: https://arxiv.org/abs/2509.21912.

[3] **Zhengyan Wan**, Fang Fang, and Binyan Jiang. "High-Dimensional Model Averaging via Cross-Validation". In: *arXiv preprint* (2025). arXiv: 2506.08451. URL: https://arxiv.org/abs/2506.08451.

RESEARCH EXPERIENCE

Error Analysis of Discrete Flow with Generator Matching July 2025 – Sept 2025

https://arxiv.org/pdf/2509.21906

- We systematically develop a stochastic calculus theory tailored to continuous-time Markov chains (CTMCs). We establish a Girsanov-type theorem for CTMC with rigorous proofs. We derive the KL divergence of two path measures of two CTMCs in terms of the integral of the Bregman divergence of two transition rates.
- We establish the non-asymptotic error bound for distribution error in discrete flow models, by taking estimation error into consideration.
- We carefully analyze the stochastic error using empirical process theory, and discuss the choice of early stopping parameter to balance the stochastic error and early stopping error.
- We provide a thorough error analysis for the neural network class with the ReLU activation function, simultaneously analyzing three sources of error, including stochastic error, approximation error and early stopping error.

Discrete Guidance Matching: Exact Guidance for Discrete Flow Matching Jun 2025 – Sept 2025

https://arxiv.org/pdf/2509.21912

- We introduce a guidance framework for discrete flow matching, achieving efficient sampling without approximation.
- We propose to learn the guidance network by minimizing the Bregman divergence. We further propose a regularization technique to utilize the sample from the target distribution.
- We verify the effectiveness of the proposed framework on energy-based sampling through simulations and preference alignment on text-to-image and multimodal understanding benchmark.

High-Dimensional Model Averaging via Cross-Validation Jun 2024 – May 2025

https://arxiv.org/pdf/2506.08451v1

- We propose a high-dimensional model averaging method under a general framework.
- We establish rigorous theoretical guarantees for the proposed estimator, including the non-asymptotic upper

bound, minimax lower bound and asymptotic optimality.

- We introduce a post-averaging one-step debiased estimator and established its Gaussian and Bootstrap approximation results to construct simultaneous confidence intervals.
- We develop a first-order algorithm to solve simplex-constrained optimization problem, with a convergence analysis demonstrating a convergence rate matching the optimal rate of first-order algorithms.

Distributed Model Averaging with Nonignorable Nonresponse - Third author Jan 2025 – Present

- We propose to obtain the weight estimator using penalized squared loss based on IPW-type empirical distribution function under a distributed learning framework with nonignorable nonresponse.
- My contribution: verifying the accuracy of technical proofs & Developing an efficient ADMM algorithm for solving a quadratic programming with simplex constraints

Visiting Experience

The Chinese University of Hong Kong, Shenzhen **Shenzhen, China**

School of Data Science. Advised by Hongyuan Zha Jun 2025 – Sept 2025

Teaching Experience

Teaching Assistant - Course: Stochastic Processes (Graduate-level Course) Sept 2024 – Jan 2025

- Graded weekly assignments and provided feedback to students & Answered student queries

Scholarships

Panshi Scholarship (East China Normal University) 2024

First Class Scholarship of the School of Statistics (East China Normal University) 2024

Second Class Scholarship & College Excellent Student (Jilin University) 2020, 2022, 2023

First Class Scholarship & University Excellent Student (Jilin University) 2021

Competition

Provincial First Prize of the Lanqiao Cup Programming Contest (Python Group) 2022

Honorable Mention Award of the Mathematical Contest in Modeling 2021

Provincial First Prize of the Chinese College Mathematics Competition (Mathematics Group) 2020

Provincial First Prize of the Contemporary Undergraduate Mathematical Contest in Modeling 2020

Skills

Programming languages: R, Python, C, MATLAB, SQL

Languages: Mandarin, English (TOEFL: 82; Scheduled to retake the exam on Oct 18, 2025)