

JINGYANG LYU

jlyu55@wisc.edu | [\[Home page\]](#) | [\[LinkedIn\]](#) | [\[GitHub\]](#)

EDUCATION

University of Wisconsin-Madison, WI, USA

09/2022 - 06/2027

- *Ph.D. in Statistics* (advised by: Yiqiao Zhong). GPA: 4.00/4.00
- *Minor in Computer Science*. GPA: 4.00/4.00
- **Honors & Awards:** Frontiers in Statistical Machine Learning Travel Award (2025, the Institute of Mathematical Statistics (IMS)). NESS Student Poster Award (2025, New England Statistical Society).

University of Chicago, IL, USA

09/2020 - 06/2022

- *M.S. in Statistics* (advised by: Wei-Biao Wu). GPA: 4.00/4.00
- **Honors & Awards:** Tuition Scholarship Increase (2021)

Tsinghua University, Beijing, China

08/2016 - 06/2020

- *B.Eng. in Industrial Engineering*. GPA: 3.71/4.00 (Rank: 7/63)
- *Minor in Statistics*. GPA: 3.94/4.00
- **Honors & Awards:** Outstanding Academic Performance & Progress Scholarship (2019), Excellent Student Leadership Award (2018), Excellent Social Worker Scholarship (2018)

INDUSTRY EXPERIENCE

United Airlines, Chicago, IL

05/2025 - 08/2025

Statistics Analyst, *Internship*

- Joined the Analytics & Innovation group, focusing on developing a short-term customer engagement forecasting model (customer lifetime value prediction) for the MileagePlus[®] loyalty program.

RESEARCH INTERESTS

- ◇ High-dimensional Statistics
- ◇ Statistical Foundations of Deep Learning
- ◇ Bias in Large Language Models
- ◇ Applied Statistics

PUBLICATIONS

1. **Lyu, J.**, Zhou, K., Zhong, Y. (2025). A statistical theory of overfitting for imbalanced classification. *ArXiv preprint arXiv:2502.11323*.
2. Chen, F., **Lyu, J.***, Wang, T., Sze, N. N. (2023). Exploring the association between quantified road safety target attributes and their success: An empirical analysis from OECD countries using panel data. *Journal of safety research*, 85, 296-307.
3. Chen, F., Zhu, Y., Zu, J., **Lyu, J.**, Yang, J. (2022). Appraising road safety attainment by CRITIC-ELECTRE-FCM: a policymaking support for Southeast Asia. *Transport policy*, 122, 104-118.
4. Chen, F., **Lyu, J.***, Wang, T. (2020). Benchmarking road safety development across OECD countries: An empirical analysis for a decade. *Accident Analysis & Prevention*, 147, 105752.

SELECTED RESEARCH EXPERIENCE

A Statistical Theory of Overfitting for Imbalanced Classification.

05/2023 - 01/2025

Supervised by Prof. Yiqiao Zhong

- Characterized overfitting of minority classes in deep learning by deriving asymptotic margin distribution from the last layer, in both separable and non-separable regimes.
- Revealed the impact of overfitting on generalization and model calibration, verified by extensive data analysis including convolutional neural networks and large language models with PyTorch.

- Propose a margin-adjustment approach to improve the performance of imbalanced classification with theoretical guarantee, especially for highly imbalanced datasets.

A Debiased Version for Online Covariance Matrix Estimation in SGD. 09/2021 - 12/2022

Supervised by *Prof. Wei-Biao Wu*

- Improved the estimator of online SGD covariance matrix with a efficient single-pass algorithm.
- Derived the error bound for the proposed debiased estimator in serveral different settings.

THESIS

Online Bootstrap Confidence Intervals for Stochastic Gradient Descent (SGD). 06/2022

Supervised by *Prof. Wei-Biao Wu*, Department of Statistics, University of Chicago

- Conducted statistical inference for SGD by using Bootstrap perturbed estimates, with mathematical details added and improved for the proofs of a previous study.
- Performed simulation experiments for regressions models and real data analysis by using R.

Fault Diagnosis of High-Speed Train Based on Imbalanced Learning. 06/2020

Supervised by *Prof. Yan-Fu Li*, Department of Industrial Engineering, Tsinghua University

- Proposed an algorithm, i.e., Recursive Different-Error-Costs Support Vector Machine (R-DEC-SVM), which can handle highly imbalanced classification tasks, implemented in R.
- Outperformed previous methods on a train operation dataset in terms of fault detection.

TALKS AND PRESENTATIONS

-
- | | |
|--|---------|
| • <i>The Inaugural Workshop on Frontiers in Statistical Machine Learning (FSML)</i> , Vanderbilt | 2025.08 |
| • <i>Midwest Machine Learning Symposium (MMLS 2025)</i> , UChicago | 2025.06 |
| • <i>The 38th New England Statistics Symposium (NESS 2025)</i> , Yale | 2025.06 |
| • <i>Midwest Optimization & Statistical Learning Conference 2025</i> , Northwestern | 2025.05 |
| • <i>6th Annual Data Science Research Bazaar</i> , UW–Madison | 2025.03 |
| • <i>Math Machine Learning seminar MPI MIS + UCLA</i> , UCLA | 2025.03 |
| • <i>Institute for Foundations of Data Science Ideas Forum</i> , UW–Madison | 2024.11 |
| • <i>100th Transportation Research Board (TRB) Annual Meeting</i> , Virtual | 2021.01 |

TEACHING

Teaching Assistant at University of Wisconsin–Madison

- | | | |
|------------|---|-------------|
| • STAT 324 | Introductory Applied Statistics for Engineers | Fall 2022 |
| • STAT 240 | Data Science Modeling I | Spring 2023 |
| • STAT 340 | Data Science Modeling II | Fall 2023 |
| • STAT 601 | Statistical Methods I | Spring 2024 |
| • STAT 709 | Mathematical Statistics | Fall 2024 |
| • STAT 340 | Data Science Modeling II | Spring 2025 |
| • STAT 709 | Mathematical Statistics | Fall 2025 |

TECHNICAL STRENGTHS & OTHERS

Programming Languages	Python, R, Matlab, Java, C, SQL, SAS
Software	Gurobi, CPLEX, AnyLogic, Plant Simulation
Tools	L ^A T _E X, Git, AWS, Linux