

# Ying Jin

Email: [yjin@hcp.med.harvard.edu](mailto:yjin@hcp.med.harvard.edu) Website: <https://ying531.github.io>

Address: 180A Longwood Avenue, Boston, MA

## Academic Appointment

**Wojcicki-Troper Postdoctoral Fellow**

2024-2025

Data Science Initiative & Harvard Medical School, *Harvard University*

## Education

**Ph.D. in Statistics**

2019 - 2024

*Stanford University*

Advisors: Emmanuel Candès, Dominik Rothenhäusler

**B.S. in Mathematics**

2015 - 2019

**B.A. in Economics (Finance)**

*Tsinghua University*

## Research<sup>1</sup>

Research interests: Distribution-free inference; Causal inference; Generalizability; Selective inference; Distributional robustness; Replicability; Data-driven decision making.

## Preprints & Under Revision

6. Bai, T. and Jin, Y. (2024). Optimized Conformal Selection: Powerful selective inference after conformity score optimization. [arXiv:2411.17983](https://arxiv.org/abs/2411.17983)
5. Jin, Y.\* and Ren, Z. (2024). Confidence on the Focal: Conformal Prediction with Selection-Conditional Coverage. [arXiv:2403.03868](https://arxiv.org/abs/2403.03868)
4. Jin, Y.\*, Guo, K\*, and Rothenhäusler, D. (2023). Diagnosing the Role of Observable Distribution Shift in Scientific Replications. [arXiv:2309.01056](https://arxiv.org/abs/2309.01056)
3. Jin, Y. and Candès, E. J. (2023). Model-free Selective Inference under Covariate Shift via Weighted Conformal P-values. [arXiv:2307.09291](https://arxiv.org/abs/2307.09291)
2. Jin, Y.\*, Ren, Z.\*, Yang, Z., and Wang, Z. (2022). Policy Learning ‘without’ Overlap: Pessimism and Generalized Empirical Bernstein’s Inequality. [arXiv:2212.09900](https://arxiv.org/abs/2212.09900)
1. Jin, Y.\*, Ren, Z.\*, and Zhou Z. (2022). Sensitivity Analysis under the  $f$ -Sensitivity Models: A Distributional Robustness Perspective. [arXiv:2203.04373](https://arxiv.org/abs/2203.04373)

## Journal Publications

6. Jin, Y. and Rothenhäusler, D. (2023). Modular Regression: Improving Linear Models by Incorporating Auxiliary Data. *Journal of Machine Learning Research (JMLR)*. [arXiv:2211.10032](https://arxiv.org/abs/2211.10032)
5. Jin, Y. and Candès, E. J. (2023). Selection by Prediction with Conformal P-values. *Journal of Machine Learning Research (JMLR)*. [arXiv:2210.01408](https://arxiv.org/abs/2210.01408)

---

<sup>1\*</sup> = equal contribution or alphabetical ordering

4. Jin, Y. and Rothenhäusler, D. (2023). Tailored Inference for Finite Populations: Conditional Validity and Transfer Across Distributions. *Biometrika*. [arXiv:2104.04565](#)
3. Jin, Y.\*, Ren, Z.\*, and Candès, E. J. (2023). Sensitivity Analysis of Individual Treatment Effects: A Robust Conformal Inference Approach. *Proceedings of the National Academy of Sciences (PNAS)*. [arXiv:2111.12161](#)
2. Jin, Y., and Ba, S. (2022). Towards Optimal Variance Reduction in Online Controlled Experiments. *Technometrics*. [arXiv:2110.13406](#) (Internship project at LinkedIn)
1. Jin, Y.\*, Yang, Z.\*, and Wang, Z.\* (2024+). Is Pessimism Provably Efficient for Offline RL?. *Mathematics of Operations Research*. Short version appeared at ICML 2021. [arXiv:2012.15085](#)

## Conference Publications

4. Yu Gui\*, Jin, Y.\*, and Ren, Z\*. (2024). Conformal Alignment: Knowing When to Trust Foundation Models with Guarantees. *Conference on Neural Information Processing Systems (NeurIPS)*. [arXiv:2405.10301](#)
3. Wang, J., Dong, P., Jin, Y., Zhan, R., and Zhou, Z. (2024). Adaptively Learning to Select-Rank in Online Platforms. *International Conference on Machine Learning (ICML)*. [arXiv:2406.05017](#)
2. Huang, K., Jin, Y., Candès, E. J., and Leskovec, J. (2023). Uncertainty Quantification over Graph with Conformalized Graph Neural Networks. *Conference on Neural Information Processing Systems (NeurIPS), Spotlight*. [arXiv:2305.14535](#)
1. Jin, Y. (2023). Upper bounds on the Natarajan dimensions of some function classes. *IEEE International Symposium on Information Theory (ISIT)*. [arXiv:2209.07015](#)

## Other Collaborations and Pre-PhD Work

3. La Cava W., Orzechowski, P., Burlacu, B., de França, F. O., Virgolin, M., Jin, Y., Kommenda, M., and Moore, J. H. (2021). Contemporary Symbolic Regression Methods and their Relative Performance. *Neural Information Processing Systems Track on Datasets and Benchmarks (NeurIPS)*.
2. Jin, Y., Lu, J., and Wang, Z. (2020). Computational-Statistical Tradeoffs in Inferring Combinatorial Structures of Ising Model. *International Conference on Machine Learning (ICML)*. [PMLR 119:4901-4910](#) (\*Pre-PhD work)
1. Jin, Y., Guo J., Kang, J., and Guo, J. (2020). Bayesian Symbolic Regression. *Proceedings of AAAI Workshop on Statistical Relational Artificial Intelligence (AAAI)*. [arXiv:1910.08892](#) (\*Pre-PhD work)

## Academic Service

### • Seminar/workshop organizing:

- Organizer, *Online Causal Inference Seminar*, September 2021 - Now.  
Duties include inviting speakers and discussants, moderating and hosting the seminars.
- Program committee member, *ICML 2021 Workshop on Reinforcement Learning Theory*, July 2021.
- Co-organizer, *ICLR 2024 Workshop on Machine Learning for Genomics Exploration*, May 2024.
- Area chair, *NeurIPS 2024 Workshop on AI for New Drug Modalities (AIDrugX)*, December 2024.

### • Invited discussion:

- For “Distribution-free inference for regression: discrete, continuous, and in between” by Yonghoon Lee, *International Seminar on Selective Inference*, February 2022.

- For “CAP: A General Algorithm for Online Selective Conformal Prediction with FCR Control” by Changliang Zou, *International Seminar on Selective Inference*, June 2024.
- For “Exploration, Confirmation, and Replication in the Same Observational Study: A Two Team Cross-Screening Approach to Studying the Effect of Unwanted Pregnancy on Mothers’ Later Life Outcomes” by Dylan Small, *Causal Seminar at Harvard Data Science Initiative*, September 2024.
- **Journal referee:** *Journal of the Royal Statistics Society: B*; *Annals of Statistics*; *Biometrika*; *Journal of the American Statistical Association*; *Annals of Applied Probability*; *Journal of Machine Learning Research*; *IEEE Transactions on Information Theory*; *ACM Computing Surveys*; *Statistica Sinica*; *Statistics in Medicine*; *Machine Learning*.
- **Conference reviewer:** *American Causal Inference Conference (ACIC) 2023*; *International Conference on Machine Learning (ICML) 2021-2024*; *Neural Information Processing Systems (NeurIPS) 2021-2024*; *NeurIPS Dataset and Benchmarks Track 2022-2024*; *International Conference on Learning Representations (ICLR) 2022-2024*; *International Conference on Artificial Intelligence and Statistics (AISTATS) 2024*; *AAAI Conference on Artificial Intelligence 2025*.

## Invited and Contributed Talks

- *Conformal Alignment: Knowing When to Trust Foundation Models with Guarantees.*
  5. ENAR Spring Meeting, March 2025.
  4. IMS International Conference on Statistics and Data Science (ICSIDS), December 2024.
  3. Conference on Statistical Learning and Data Science (SLDS), November 2024.
  2. Data Science and Engineering Lab, Michigan State University, October 2024.
  1. Joint Conference on Statistics and Data Science, China, July 2024.
- *Model-Free Selective Inference: From Calibrated Uncertainty to Trusted Decisions.*
  16. Professor Junwei Lu’s Group Meeting, Harvard University, October 2024.
  15. International Conference on Econometrics and Statistics (EcoSta 2024), July 2024.
  14. Statistics & Data Science Department Seminar, University of Pennsylvania, February 2024.
  13. Biostatistics Department Seminar, Columbia University, February 2024.
  12. School of ISyE Seminar, Georgia Tech, February 2024.
  11. Statistics Department Seminar, University of Wisconsin-Madison, February 2024.
  10. Statistics Department Seminar, Virginia Tech, February 2024.
  9. Statistics Department Seminar, Carnegie Mellon University, January 2024.
  8. Statistics Department Seminar, Columbia University, January 2024.
  7. ORIE Department Seminar, Cornell University, January 2024.
  6. Statistics Department Seminar, Harvard University, January 2024.
  5. Statistics, Operations, and Technology Department Seminar, Stern School of Business, New York University, January 2024.
  4. OM&BA Department Seminar, Carey School of Business, Johns Hopkins University, January 2024.
  3. Applied Mathematics Department Seminar, Brown University, December 2023.
  2. MLBoost Online Seminar, January 2024.
  1. Rising Stars in Data Science Workshop, University of Chicago, November 2023.
- *Selection by Prediction with Conformal P-values (and Applications to Drug Discovery).*
  11. Genentech Incorporation, September 2023.
  10. Joint Statistical Meetings, August 2023.

9. Joint Conference on Statistics and Data Science in China, July 2023.
  8. ICSA China Conference, June 2023.
  7. Statistics Seminar, Suzhou University, China, June 2023.
  6. INRIA Causal Inference Group, May 2023.
  5. International Seminar on Selective Inference, May 2023.
  4. International Conference on Design of Experiments (ICODOE), May 2023.
  3. One World Mathematics of Machine Learning Online Seminar, April 2023.
  2. Stanford Statistics Student Seminar, December 2022.
  1. Stanford Statistics Industrial Affiliates Annual Conference, November 2022.
- *Diagnosing Observable Distribution Shift in Effect Generalization with Scientific Replication Data.*
    7. Harvard Applied Statistics Workshop, September 2024.
    6. Joint Statistical Meetings, August 2024.
    5. Inaugural Berkeley-Stanford Workshop on Veridical Data Science (lightning talk), May 2024.
    4. IMS International Conference on Statistics and Data Science (ICSIDS), December 2023.
    3. Prof. Chiara Sabatti's group meeting, Stanford University, November 2023.
    2. Stanford-Berkeley Joint Colloquium Student Seminar, Stanford University, October 2023.
    1. Causality in Practice Conference (Thematic Quarter for Causality), June 2023.
  - *Confidence on the Focal: Conformal Prediction with Selection-Conditional Coverage.*
    1. Hangzhou International Conference on Frontiers of Data Science, China, July 2024.
  - *Policy Learning 'without' Overlap: Pessimism and Generalized Empirical Bernstein's Inequality.*
    6. INFORMS Annual Meeting, October 2024.
    5. Clubear Statistics Organization (Virtual Talk), China, April 2024.
    4. INFORMS Optimization Society Conference (IOS 2024), March 2024.
    3. Workshop on Operations Research and Data Science, Duke University, November 2023.
    2. Online Reinforcement Learning Theory Seminar, May 2023.
    1. Data-Driven Decision Making Seminar, Stanford Graduate School of Business, January 2023.
  - *Adaptively Learning to Rank Items in Online Platforms.*
    1. INFORMS Annual Meeting, October 2023.
  - *Towards Optimal Variance Reduction in Online Controlled Experiments.*
    4. Design and Analysis Conference, Virginia Tech, May 2024.
    3. Doordash Causal Inference and Experimentation Team, May 2023.
    2. INFORMS Workshop on Data Mining and Decision Analytics, October 2022.
    1. Conference on Digital Experimentation (CODE), November 2021.
  - *Sensitivity Analysis under the  $f$ -Sensitivity Models: A Distributional Robustness Perspective.*
    3. Data-Driven Decision Making Seminar, Stanford Graduate School of Business, November 2022.
    2. Stanford Causal Science Conference, November 2022.
    1. ICSA Applied Statistics Symposium (Student Paper Award presentation), June 2022.
  - *Sensitivity Analysis of Individual Treatment Effects: A Robust Conformal Inference Approach.*
    4. UC Berkeley Rising Stars Seminar (Prof. Ahmed Alaa's group), March 2024.
    3. CMStatistics (virtual), December 2022.

2. INFORMS Annual Meeting, October 2022.
  1. Stanford University Causal Inference Group, January 2022.
- *Tutorial on Causal Inference in Networks*. Prof. Tracy Ke's group meeting, Department of Statistics, Harvard University, September 2022.
  - *One Estimator, Many Estimands: Fine-Grained Quantification of Uncertainty using Conditional Inference*. Joint Statistical Meetings, August 2021.
  - *Is Pessimism Provably Efficient for Offline RL?* Online RL Theory Seminar, April 2021.

## Software

- **BSR**, developer, <https://github.com/ying531/MCMC-SymReg>  
Python package for Bayesian Symbolic Regression method in Jin et al. (2020).
- **SRBench**, contributor, <https://github.com/cavalab/srbench>  
Large-scale benchmark for symbolic regression methods in La Cava et al. (2021).
- **condinf**, developer, <https://github.com/ying531/condinf>  
**transinf**, developer, <https://github.com/ying531/transinf>  
R packages for conditional and transductive inference for finite populations in Jin and Rothenhäusler (2023).
- **cfsensitivity**, contributor, <https://github.com/zhimeir/cfsensitivity>  
R package for sensitivity analysis and robust conformal inference of individual treatment effects under unmeasured confounding in Jin, Ren, and Candès (2023).
- **ConfSelect**, developer, <https://github.com/ying531/conformal-selection>  
R package for (Weighted) Conformalized Selection, which conducts calibrated selection of large outcomes with (weighted) conformal p-values in Jin and Candès (2022, 2023).
- **repDiagnosis**, developer, <https://github.com/ying531/repDiagnosis>  
R package implementing Jin, Guo, and Rothenhäusler (2023) for diagnosing replication studies.
- **awesome-replicability-data**, developer, <https://github.com/ying531/awesome-replicability-data>  
Online collection of publicly-available, individual-level datasets of replication study pairs.
- **Rshiny app**, developer, <https://mbzlnj-ying-jin.shinyapps.io/shiny/>  
Online live app for the method in Jin, Guo, and Rothenhäusler (2023) for diagnosing replication studies.

## Teaching Experience

**Guest lecturer** at STATS300C (Theory of Statistics, Instructor: Emmanuel Candès)

**Teaching assistant** at Stanford University: DataSci112 (Principles of Data Science), STATS209 (Introduction to Causal Inference), STATS204 (Sampling), STATS320 (Statistical Methods for Neural Data Analysis), STATS200 (Introduction to Statistical Inference), STATS216 (\*3, Introduction to Statistical Learning), STATS60 (Introduction to Statistical Methods: Precalculus), STATS202 (Data Mining and Analysis), STATS241 (Data-driven Financial Econometrics), STATS240 (Statistical Methods in Finance), STATS305A (Applied Statistics), STATS205 (Nonparametric Statistics)

## Industrial Eexperience

Data Science Applied Research Intern, *LinkedIn Applied Research Team*

June - September 2021

## Honors and Awards

Jack Youden Prize, American Society for Quality's Chemical and Process Industries Division	2024
Ingram Olkin Interdisciplinary Dissertation Award, Department of Statistics, Stanford University	2024
Rising Star in Data Science, University of Chicago	2023
Student Paper Award, ICSA Applied Statistics Symposium	2022
Tom Ten Have Award Runner up, American Causal Inference Conference	2022
D. E. Shaw Zenith Fellowship	2021
Outstanding Graduate Award (2/91), Tsinghua University	2019
Beijing Outstanding Graduate Award	2019
Qualcomm Scholarship for Research	2017
President Scholarship for Comprehensive Excellence (1/91), Tsinghua University	2017
Scholarship for Academic Excellence, Tsinghua University	2015
Scholarship for Freshmen, Tsinghua University	2014
Silver Medal in the 29th China Mathematics Olympics Final	2013