# Song Mei

367 Evans Hall
Berkeley, CA, 94720

Website: https://stat.berkeley.edu/~songmei/
Email: songmei@berkeley.edu

## Professional Experience

Assistant Professor, University of California, Berkeley

| | |
|---|---|
| Department of Statistics | 07/2020 - present |
| Department of Electrical Engineering and Computer Sciences | 07/2021 - present |

Data Scientist Intern, Google 06/2017 - 09/2017

## Education

| | |
|---|---|
| Ph.D. Computational and Mathematical Engineering, Stanford University | 09/2014 - 07/2020 |
| B.S. Mathematics, Peking University | 09/2010 - 06/2014 |

## Research Areas

- Foundations of deep learning and generative AI: theory, safety, and interpretability.
- Foundations of interactive decision-making: multi-agent and partially observable systems.
- High dimensional statistics: non-convex optimization and variational inference.
- Quantum algorithms: analysis of random combinatorial optimization.

## Grants and Awards

Alfred P. Sloan Research Fellowship. 2025.

NSF CAREER Award. 2024.

Amazon Research Award. 2024.

Google Research Scholar Award. 2024.

Okawa Foundation Research Grant. 2024.

ICSA Junior Researcher Award. 2024.

ONR N00014-24-1-2639. 2024 - 2027.

NSF CCF-2315725. 2023 - 2026.

NSF DMS-2210827. 2022 - 2025.

Stanford Graduate Fellowship. 2014 - 2017.

## Group Members

Kazusato Oko: EECS PhD (2024 - current).

Ruiqi Zhang: Stats PhD, with Peter Bartlett (2022 - current).

Tianyu Guo: Stats PhD, with Michael Jordan (2022 - current).

Licong Lin: Stats PhD, with Peter Bartlett (2021 - current).

Druv Pai: EECS PhD (Collaborator). Yi Ma and Jiantao Jiao (2023 - current).

Yuhang Cai: Math PhD (Collaborator). Peter Bartlett and Michael Lindsey (2021 - current).

Jingfeng Wu: Simons Postdoc (Collaborator). Peter Bartlett and Bin Yu (2023 - current).

## Former Mentee

Leo Zhou: Postdoc at Caltech (2021 - 2024). Assistant professor at UCLA.

Michael Celentano: Berkeley Postdoctoral Miller Fellow (2021 - 2024). Research scientist at OpenAI.

Nikhil Ghosh: Berkeley PhD. With Bin Yu (2019 - 2024). Postdoc at Flatiron institute.

Taejoo Ahn: Berkeley PhD (2018 - 2024). Quant at IMC trading.

Zitong Yang: Berkeley Master, summer intern (2021). PhD at Stanford.

Silas Alberti: LMU Munich Undergrad, research intern (2022). PhD at Stanford.

Ziang Song: Peking University Undergrad, summer intern (2022). PhD at Stanford.

Fan Chen: Peking University Undergrad, summer intern (2023). PhD at MIT.

## Teaching

STAT210B. Theoretical Statistics. Spring 2022, Spring 2023, Spring 2025.

STAT210A. Theoretical Statistics. Fall 2022.

STAT154/254. Modern Statistical Prediction and Machine Learning. Fall 2021. Spring&Fall 2024.

STAT260. Mean Field Asymptotics in Statistical Learning. Spring 2021.

## Professional Services

### Seminar and workshop organization

The Inaugural Workshop on Frontiers in Statistical Machine Learning. 2025.

JSM Invited Session "Theory of deep learning and Generative AI". 2025.

One World Seminar Series on the Mathematics of Machine Learning. 2020 - 2022.

Neyman seminar, University of California, Berkeley. 2020 - 2022.

### Editorial activity

Area Chair: COLT, 2022, 2023, 2024, 2025. COLM 2024, 2025.

Associate Editor: Journal of Machine Learning.

Conferences reviewed: COLT, NeurIPS, ICML, ICLR, AISTATS, ALT, ISIT, FOCS.

Journals reviewed: Annals of Statistics, JASA, JRSSB, Journal of Machine Learning Research, Statistica Sinica, Bernoulli, IEEE Information Theory, Operations Research, Mathematics of Operations Research, Mathematical Statistics and Learning, SIAM Journal on Mathematics of Data Science.

### Department committee

MA committee chair. 2023 - 2025.

MA committee. 2022 - 2023.

PhD admission committee. 2021 - 2022.

Neyman seminar coordinator. 2020 - 2022.

### Grant Proposal Review

NSF Panelist. 2023. NSF Panelist 2024.

## Publications

Citations (Google Scholar): 4615 in total. Date: Jan 26, 2025.

### Conference Publications

1. **Song Mei**. (2024) U-Nets as Belief Propagation: Efficient Classification, Denoising, and Diffusion in Generative Hierarchical Models. *ICLR, 2025.*

2. Ruiqi Zhang, Licong Lin, Yu Bai, and **Song Mei**. (2024) Negative Preference Optimization: From Catastrophic Collapse to Effective Unlearning. *COLM 2024.*

3. Leo Zhou, Joao Basso, and **Song Mei**. (2024) Statistical Estimation in the Spiked Tensor Model via the Quantum Approximate Optimization Algorithm. *NeurIPS 2024 (Spotlight).*

4. Yuhang Cai, Jingfeng Wu, **Song Mei**, Michael Lindsey, and Peter Bartlett. (2024) Large Stepsize Gradient Descent for Non-Homogeneous Two-Layer Networks: Margin Improvement and Fast Optimization. *NeurIPS 2024.*

5. Tianyu Guo, Wei Hu, **Song Mei**, Huan Wang, Caiming Xiong, Silvio Savarese, Yu Bai. (2023) How Do Transformers Learn In-Context Beyond Simple Functions? A Case Study on Learning with Representations. *International Conference on Learning Representations (ICLR), 2024.*

6. Licong Lin, Yu Bai, and **Song Mei**. (2023) Transformers as Decision Makers: Provable In-Context Reinforcement Learning via Supervised Pretraining. *International Conference on Learning Representations (ICLR), 2024.*

7. Yu Bai, Fan Chen, Huan Wang, Caiming Xiong, and **Song Mei**. (2023) Transformers as Statisticians: Provable In-Context Learning with In-Context Algorithm Selection. *NeurIPS 2023 (Oral).*

8. Hengyu Fu, Tianyu Guo, Yu Bai, and **Song Mei**. (2023) What can a Single Attention Layer Learn? A Study Through the Random Features Lens. *NeurIPS 2023.*

9. Fan Chen, Huan Wang, Caiming Xiong, **Song Mei**, and Yu Bai. (2023) Lower Bounds for Learning in Revealing POMDPs. *International Conference on Machine Learning (ICML), 2023.*

10. Fan Chen, Yu Bai, and **Song Mei**. (2022) Partially Observable RL with B-Stability: Unified Structural Condition and Sharp Sample-Efficient Algorithms. *The International Conference on Learning Representations (ICLR), 2023.*

11. Yu Bai, Chi Jin, **Song Mei**, Ziang Song, and Tiancheng Yu. (2022) Efficient Phi-Regret Minimization in Extensive-Form Games via Online Mirror Descent. *Neural Information Processing Systems (NeurIPS), 2022.*

12. Joao Basso, David Gamarnik, **Song Mei**, and Leo Zhou. (2022) Performance and limitations of the QAOA at constant levels on large sparse hypergraphs and spin glass models. *IEEE Symposium on Foundations of Computer Science (FOCS), 2022.*

13. Ziang Song, **Song Mei**, and Yu Bai. (2022) Sample-Efficient Learning of Correlated Equilibria in Extensive-Form Games. *Neural Information Processing Systems (NeurIPS), 2022.*

14. Theodor Misiakiewicz, and **Song Mei**. (2021) Learning with convolution and pooling operations in kernel methods. *Neural Information Processing Systems (NeurIPS), 2022.*

15. Yu Bai, Chi Jin, **Song Mei**, and Tiancheng Yu. (2022) Near-Optimal Learning of Extensive-Form Games with Imperfect Information *International Conference on Machine Learning (ICML), 2022.*

16. Yu Bai, **Song Mei**, Huan Wang, Yingbo Zhou, and Caiming Xiong. (2021) Efficient and Differentiable Conformal Prediction with General Function Classes. *The International Conference on Learning Representations (ICLR), 2022.*

17. Nikhil Ghosh, **Song Mei**, Bin Yu. (2021) The Three Stages of Learning Dynamics in High-Dimensional Kernel Methods. *The International Conference on Learning Representations (ICLR), 2022.*

18. Ziang Song, **Song Mei**, Yu Bai. (2021) When Can We Learn General-Sum Markov Games with a Large Number of Players Sample-Efficiently? *The International Conference on Learning Representations (ICLR), 2022.*

19. Yu Bai, **Song Mei**, Huan Wang, and Caiming Xiong. (2021) Understanding the Under-Coverage Bias in Uncertainty Estimation. *Neural Information Processing Systems (NeurIPS), 2021.*

20. Zitong Yang, Yu Bai, and **Song Mei**. (2021) Exact Gap between Generalization Error and Uniform Convergence in Random Feature Models. *International Conference on Machine Learning (ICML), 2021.*

21. Yu Bai, **Song Mei**, Huan Wang, and Caiming Xiong. (2021) Don't Just Blame Over-parametrization for Over-confidence: Theoretical Analysis of Calibration in Binary Classification. *International Conference on Machine Learning (ICML), 2021.*

22. **Song Mei**, Theodor Misiakiewicz, and Andrea Montanari. (2021) Learning with invariances in random features and kernel models. *Conference of Learning Theory (COLT), 2021.*

23. Behrooz Ghorbani, **Song Mei**, Theodor Misiakiewicz, and Andrea Montanari. (2020) When do neural networks outperform kernel methods? *Neural Information Processing Systems (NeurIPS), 2020.*

24. Behrooz Ghorbani, **Song Mei**, Theodor Misiakiewicz, and Andrea Montanari. (2019) Limitations of lazy training of two-layers neural networks. *Neural Information Processing Systems (NeurIPS), 2019.*

25. **Song Mei**, Theodor Misiakiewicz, and Andrea Montanari. (2019) Mean-field theory of two-layers neural networks: dimension-free bounds and kernel limit. *Conference of Learning Theory (COLT), 2019.*

26. **Song Mei**, Theodor Misiakiewicz, Andrea Montanari, and Roberto I. Oliveira. (2017) Solving SDPs for synchronization and MaxCut problems via the Grothendieck inequality. *Conference of Learning Theory (COLT), 2017.*

## Journal Publications

1. **Song Mei**, and Yuchen Wu. (2023) Deep Networks as Denoising Algorithms: Sample-Efficient Learning of Diffusion Models in High-Dimensional Graphical Models. *IEEE Transactions of Information Theory, 2025.*

2. Fan Chen, **Song Mei**, and Yu Bai. (2022) Unified Algorithms for RL with Decision-Estimation Coefficients: No-Regret, PAC, and Reward-Free Learning. *The Annals of Statistics, 2025.*

3. Michael Celentano, Zhou Fan, and **Song Mei**. (2023) Local convexity of the TAP free energy and AMP convergence for Z2-synchronization. *The Annals of Statistics 51 (2), 519-546.*

4. **Song Mei**, Theodor Misiakiewicz, and Andrea Montanari. (2021) Generalization error of random features and kernel methods: hypercontractivity and kernel matrix concentration. *Applied and Computational Harmonic Analysis, 59, 3-84.*

5. Behrooz Ghorbani, **Song Mei**, Theodor Misiakiewicz, and Andrea Montanari. (2021) When do neural networks outperform kernel methods? *Journal of Statistical Mechanics: Theory and Experiment.*

6. **Song Mei**, and Andrea Montanari. (2021) The generalization error of random features regression: Precise asymptotics and double descent curve. *Communications on Pure and Applied Mathematics, 75 (4), 667-766.*

7. Zhou Fan, **Song Mei**, and Andrea Montanari. (2021) TAP free energy, spin glasses, and variational inference. *The Annals of Probability 49 (1), 1-45.*

8. Behrooz Ghorbani, **Song Mei**, Theodor Misiakiewicz, and Andrea Montanari. (2020) Linearized two-layers neural networks in high dimension. *The Annals of Statistics 2021, Vol. 49, No. 2, 1029-1054.*

9. Behrooz Ghorbani, **Song Mei**, Theodor Misiakiewicz, and Andrea Montanari. (2019) Discussion of "Nonparametric Regression using Deep Neural Networks with ReLU Activation Function". *The Annals of Statistics 48 (4), 1898-1901.*

10. Gerard Ben Arous, **Song Mei**, Andrea Montanari, and Mihai Nica. (2019) The landscape of the spiked tensor model. *Communications on Pure and Applied Mathematics 72 (11), 2282-2330.*

11. **Song Mei**, Andrea Montanari, and Phan-Minh Nguyen. (2018) A mean field view of the landscape of two-layers neural network. *Proceedings of the National Academy of Sciences 115, E7665-E7671.*

12. **Song Mei**, Yu Bai, and Andrea Montanari. (2018) The landscape of empirical risk for non-convex losses. *The Annals of Statistics 46 (6A), 2747-2774.*

13. **Song Mei**, and Pingwen Zhang. (2015) On a molecular based Q-tensor model for liquid crystals with density variations. *Multiscale Modeling and Simulation 13 (3), 977-1000.*

## Review Articles

1. Wenlong Ji, Weizhe Yuan, Emily Getzen, Kyunghyun Cho, Michael I. Jordan, **Song Mei**, Jason E Weston, Weijie J. Su, Jing Xu, and Linjun Zhang. (2025) An Overview of Large Language Models for Statisticians. *arXiv preprint arXiv:2502.17814.*

2. Minshuo Chen, **Song Mei**, Jianqing Fan, Mengdi Wang. (2024) An overview of diffusion models: Applications, guided generation, statistical rates and optimization. *National Science Review, nwae348.*

## Preprints and Submissions

1. Yuhang Cai, Kangjie Zhou, Jingfeng Wu, **Song Mei**, Michael Lindsey, Peter L. Bartlett. (2025) Implicit Bias of Gradient Descent for Non-Homogeneous Deep Networks. *arXiv preprint arXiv:2502.16075.*

2. Tianyu Guo, Hanlin Zhu, Ruiqi Zhang, Jiantao Jiao, **Song Mei**, Michael I. Jordan, and Stuart Russell. (2025) How Do LLMs Perform Two-Hop Reasoning in Context? *arXiv preprint arXiv:2502.17814.*

3. Kazusato Oko, Licong Lin, Yuhang Cai, and **Song Mei**. (2025) A Statistical Theory of Contrastive Pre-training and Multimodal Generative AI. *arXiv preprint arXiv:2501.04641.*

4. Tianyu Guo, Druv Pai, Yu Bai, Jiantao Jiao, Michael I. Jordan, and **Song Mei**. (2024) Active-Dormant Attention Heads: Mechanistically Demystifying Extreme-Token Phenomena in LLMs. *arXiv preprint arXiv:2410.13835.*

5. Chongyu Fan, Jiancheng Liu, Licong Lin, Jinghan Jia, Ruiqi Zhang, **Song Mei**, and Sijia Liu. (2024) Simplicity Prevails: Rethinking Negative Preference Optimization for LLM Unlearning. *arXiv preprint arXiv:2410.07163.*

6. Taejoo Ahn, Licong Lin, and **Song Mei**. (2022) Near-optimal multiple testing in Bayesian linear models with finite-sample FDR control. *arXiv preprint arXiv:2211.02778.*

7. Hui Xu, **Song Mei**, Stephen Bates, Jonathan Taylor, and Robert Tibshirani (2023) Uncertainty Intervals for Prediction Errors in Time Series Forecasting. *arXiv preprint arXiv:2309.07435.*

8. Michael Celentano, Zhou Fan, Licong Lin, and **Song Mei** (2023) Mean-field variational inference with the TAP free energy: Geometric and statistical properties in linear models. *arXiv preprint arXiv:2311.08442.*

## Invited Talks

2025: UC Berkeley (Simons Institute of Computing, MoDL meeting). Indian Statistical Institute for the 4th Winter School on Deep Learning, Plenary talk. ASA SLDS webinar.

2024: UC Davis (Peter Hall Memorial Workshop), UC Berkeley (LTV talk of Simons Institute), Yale University (FDS seminar), Harvard University (Statistics seminar), Joint Statistical Meeting, ICSA China Meeting (ICSA Junior Researcher Award Session), Northwestern University (IDEAL workshop), Bocconi University (ELLIS workshop), Georgia Institute of Technology (Statistics seminar), University of Pennsylvania (PriML seminar), UC Davis (Statistics seminar).

2023: Princeton University (Machine learning seminar), Stanford University (Stanford Berkeley joint colloquium), UC Berkeley (Statistics seminar), UCLA (Statistics seminar), Columbia University (Berkeley–Columbia Meeting in Engineering and Statistics).

2022: Allerton (58th Annual Allerton Conference on Communication, Control, and Computing), Stanford University (Statistics seminar), UC Berkeley (Workshop at the Simons Institute).

2021: UC Berkeley (Applied math seminar), UC Berkeley (Workshop at the Simons Institute), USC (Probability seminar), Harvard University (Probability seminar), Princeton University (EE seminar), Cornell University (SCAN seminar), Peking University (CAM seminar), Purdue University (Mathematical Data Science Seminar), UC Berkeley (BLISS seminar).

2020: Georgia Institute of Technology (Stochastic seminar), UC Berkeley (Statistics seminar), NYU (MAD seminar), Columbia University (Statistics seminar), UCLA (Statistics seminar), UC Davis (Statistics seminar), Yale University (Statistics seminar), University of Pennsylvania (Statistics seminar), University of Chicago (Statistics seminar).

Pre-2019: USC (Statistics seminar), NYU (special MaD seminar), Cornell University (Annual Young Researchers Workshop), École Normale Supérieure (Golosino Workshop), Hangzhou (International Conference on Frontiers of Data Science), UC Santa Barbara (KITP Workshop), Duke University (Workshop in Operations Research and Data Science), Joint Statistical Meeting, UC Berkeley (Probability seminar), Vilnius (IMS Annual Meeting on Probability and Statistics), Stanford University (Linear Algebra and

Optimization seminar), Peking University (Operation Research seminar), Stanford University (Bay Area Scientific Computing Day).