

Yunyi Shen

32-D740
77 Massachusetts Ave., Cambridge, MA, 02139
📞 Contact: 608-400-3982
✉ yshen99@mit.edu
🌐 <https://yunyishen.github.io>

Education

- 06/22–06/26 **Massachusetts Institute of Technology**, Cambridge, MA, United States
PhD, Computer Science (minor in Astronomy)
- 06/22–05/24 **Massachusetts Institute of Technology**, Cambridge, MA, United States
SM, Computer Science
- 08/19–05/21 **University of Wisconsin-Madison**, Madison, WI, United States
MSc, Statistics
- 08/18–08/21 **University of Wisconsin-Madison**, Madison, WI, United States
MSc, Ecology
- 2014 – 2018 **Peking University**, Beijing, China
BSc, Integrated Science (Biology/Physics)

Employment

- 06/25–08/25 **Citadel Securities**, Miami, FL, United States
Quantitative Research intern, alternative data alpha, systematic equity

Research interest

Multimodal probabilistic machine learning that integrates domain knowledge and calibration to solve inverse problems in data-scarce, low-signal-to-noise-ratio settings in biology and physics.

Publications and Preprints

* indicating equal-contribution, " indicating corresponding author.

Preprints/Under review/Workshops

Statistics & Machine learning

Timm Haucke, Lauren Harrell, **Yunyi Shen**, Levente Klein, David Rolnick, Lauren E Gillespie, and Sara Beery. Deep multi-modal species occupancy modeling. *bioRxiv*, pages 2025–09, 2025

Renato Berlinghieri*, **Yunyi Shen***, Jialong Jiang, and Tamara Broderick. Oh SnapMMD! forecasting stochastic dynamics beyond the Schrödinger bridge's end. *arXiv preprint arXiv:2505.16082*, 2025

Hao Sun, **Yunyi Shen**, Jean-Francois Ton, and Mihaela van der Schaar. Reusing embeddings: Reproducible reward model research in large language model alignment without GPUs. *arXiv preprint arXiv:2502.04357*, 2025

Yunyi Shen, Lorenzo Masoero, Joshua G Schraiber, and Tamara Broderick. Double trouble: Predicting new variant counts across two heterogeneous populations. *arXiv preprint arXiv:2403.02154*, 2024. (*under review*)

Yunyi Shen and Claudia Solís-Lemus. Bayesian chain graph LASSO models to learn sparse microbial networks with predictors. *arXiv preprint arXiv:2012.08397*, 2020. (*under review*)

Ecology

Qing Zhao and **Yunyi Shen**. Explicit modeling of density dependence in spatial capture-recapture models. *arXiv preprint arXiv:2412.09431*, 2024. (*revision under review*)

Yunyi Shen", Erik R. Olson, and Timothy R. Van Deelen. Geometric theory on large-scale and local determination of density dependence of a recovering large carnivore population. *arXiv preprint arXiv:2311.14815*, 2023. (*revision under review*)

Hongliang Bu and **Yunyi Shen**. Identify local limiting factors of species distribution using min-linear logistic regression. *arXiv preprint arXiv:2302.09217*, 2023

Astronomy & Astrophysics

Yunyi Shen and Alexander T. Gagliano. Mixture-of-expert variational autoencoders for cross-modality embedding of type Ia supernova data. *arXiv preprint arXiv:2507.16817*, 2025

Yunyi Shen and Alexander T. Gagliano. Variational diffusion transformers for conditional sampling of supernovae spectra. In *7th Symposium on Advances in Approximate Bayesian Inference – Workshop Track*, 2025

Peer reviewed Journals & Conferences

Statistics & Machine learning

Yunyi Shen*, Renato Berlinghieri*, and Tamara Broderick. Multi-marginal Schrödinger bridges with iterative reference refinement. *AISTATS*, 2025 (*Oral*, arXiv:2408.06277)

Yunyi Shen, Claudia Solís-Lemus, and Sameer K Deshpande. Estimating sparse direct effects in multivariate regression with the spike-and-slab LASSO. *Bayesian Analysis*, 1(1):1–25, 2024

Hao Sun*, **Yunyi Shen***, and Jean-Francois Ton. Rethinking reward modeling in preference-based large language model alignment. *ICLR*, 2025 (*Oral*, arXiv:2411.04991)

Yunyi Shen'' and Sameer K. Deshpande. Posterior contraction and uncertainty quantification for the multivariate spike-and-slab LASSO. *Journal of Multivariate Analysis*, 105493, 2025

Yunyi Shen*, Hao Sun*, and Jean-François Ton. Active reward modeling in large language model alignment. *ICML*, 2025, arXiv:2502.04354

Jenny Y. Huang, David R. Burt, **Yunyi Shen**, Tin D. Nguyen, and Tamara Broderick. Approximations to worst-case data dropping: unmasking failure modes. *Transactions on Machine Learning Research*, 2025

David R Burt, **Yunyi Shen**, and Tamara Broderick. Consistent validation for predictive methods in spatial settings. *AISTATS*, 2025 (*accepted*, arXiv:2402.03527)

Yunyi Shen and Claudia Solís-Lemus. The effect of the prior and the experimental design on the inference of the precision matrix in Gaussian chain graph models. *Journal of Agricultural, Biological and Environmental Statistics*, pages 1–70, 2024

Ecology

Yunyi Shen. The concept and quantification of diversity. *Nature Reviews Biodiversity*, pages 1–1, 2025 (journal club)

Erik Olson, **Yunyi Shen**, Parker Matzinger, Guido Solano Mora, Evelynand Saborío-R, and Alejandro Azofeifa. Conservation crisis? status of jaguars panthera onca in corcovado national park, costa rica. *Oryx*, 2024 (accepted)

Xiaonan Chen, Jia Tian, Mingzhang Liu, **Yunyi Shen**, Jianping Yu, Feng Liu, Xiaoli Shen, and Li Sheng. Estimating the population size of wild boar (*Sus scrofa*) in Kaihua county, Zhejiang province using camera trap data. *Acta Theriologica Sinica*, 43(5): 523-532, 2023. (in Chinese)

Elyse C. Mallinger, Katy R. Goodwin, Alan Kirschbaum, **Yunyi Shen**, Erin H. Gillam, and Erik R. Olson. Species-specific responses to white-nose syndrome in the Great Lakes region. *Ecology and Evolution*, 13:e10267, 2023

Yunyi Shen, Mingzhang Liu, Dajun Wang, Xiaoli Shen, and Sheng Li. Using an integrative mapping approach to identify the distribution range and conservation needs of a large threatened mammal, the Asiatic black bear, in China. *Global Ecology and Conservation*, 31:e01831, 2021

Yunyi Shen", Erik R Olson, and Timothy R Van Deelen. Spatially explicit modeling of community occupancy using Markov random field models with imperfect observation: Mesocarnivores in Apostle Islands National Lakeshore. *Ecological Modelling*, 459:109712, 2021

Jianping Yu, **Yunyi Shen**, Xiaoyou Song, Xiaonan Chen, Sheng Li, and Xiaoli Shen. Evaluating the effectiveness of functional zones for black muntjac (*Muntiacus crinifrons*) protection in Qianjiangyuan National Park pilot site. *Biodiversity Science*, 27(1):5, 2019. (in Chinese)

Astronomy & Astrophysics

J Arturo Esquivel*, **Yunyi Shen***, Vianey Leos-Barajas, Gwendolyn Eadie, Joshua S Speagle, Radu V Craiu, Amber Medina, and James RA Davenport. Detecting stellar flares in photometric data using hidden Markov models. *The Astrophysical Journal*, 979(2):141, 2025

Miscellaneous

Yunyi Shen and Tamara Broderick. Wild posteriors in the wild. *arXiv preprint arXiv:2503.00239*, 2025

Awards and Honor

- 2020 Finalist of Janice Lee Fenske Memorial Award of The Wildlife Society (TWS)
- 2019 Conservation Leader for Tomorrow (CLfT) Scholarship
- 2017 Peking University Annual Award in Social Works

Travel awards

- 2024 TWS-BWG travel award
- 2023 BayesComp-2023 travel award
- 2023 ISBA travel award for BNP2023

Invited Talks

- 10/2024 "Wolves in the (west) Great Lake regions population, ecosystem, and economy", Quantitative ecology seminar series, University of Washington.
- 08/2022 "Conditional probabilities in some multispecies models", Merging statistical theory and analyses at the interface of microbial and "macrobial" ecology workshop. (NSF-funded) Concordia University, Montréal, Canada.
- 05/2022 "Stellar flare detection using HMM combined with Gaussian process", Stellar Stats Workshop, David A. Dunlap Department of Astronomy and Astrophysics, University of Toronto, Toronto, Canada.
- 05/2022 "Use cases of high throughput computing in statistical ecology and astronomy", HTCondor week, Center of high throughput computing, UW-Madison, Madison, WI
- 03/2022 "HMMs in Astronomy", JSC 270/370 Seminar, Department of Statistical Science, University of Toronto, Toronto, Canada.

Selected Conference Talks

Yunyi Shen, Lorenzo Masoero, Joshua Schraiber, and Tamara Broderick. Double trouble: Predicting new variant counts across two heterogeneous populations. Joint Statistical Meeting 2023, Toronto, Canada (given by coauthor due to visa difficulty); BAYSM:O-2023, CT, USA, 2023

Yunyi Shen, Fan Fan, Sheng Li, and Claudia Solís-Lemus. Direct (conditional) and net (marginal) effects of environmental predictors in multivariate normal and autologistic models for multispecies modeling. 2022 International Statistical Ecology Conference, Cape Town, South Africa, 2022

Yunyi Shen, Dwayne Etter, and Timothy Van Deelen. Age structure effects and population control in urban/suburban white-tailed deer, Chicago, IL 1992-2006. 2020 Midwest Fish and Wildlife Conference, 2020

Software

- R packages
 - mSSL: Collection of multivariate Spike-and-Slab regression models, GitHub
 - CARlasso: Conditional autoregressive LASSO and extensions, CRAN
 - linconGaussR: Sampling multivariate Gaussian under linear constraint, CRAN
 - robustcov: Collection of robust precision estimation, CRAN
- Stan
 - CeleriteQFD: Stellar flare detection using Gaussian process combined with HMM, GitHub

Teaching

- 2023 Spring 6.7830 (Bayesian Modeling and Inference) TA, MIT
- 2018-2019 Chemistry 103-104 TA, UW-Madison
- 2016-2017 Evolution of Human Intelligence TA, Peking University

Services

Reviewing

- Statistics/ML EJS, JMLR, NeurIPS-2023/2025, ICML-2023/2025, AISTATS-2023/2025, AoAS, Bioinformatics, TMLR
- Ecology J. Appl. Ecol., GECCO, Ecology and Evolution, Biological Conservation, Biodiversity & Conservation, Ecosphere

Departmental

- MIT EECS Graduate Student Association Vice President in Academics, 2024
- MIT EECS Graduate Student Association executive board, 2023

Miscellaneous

- Mentoring Student support network member, Undergraduate Network for Increasing Diversity of Ecologists (UNIDE)
- Consulting Statistical consulting, Qianjiangyuan National Park, Zhejiang, China, Summer 2018
Statistical and data management consulting, Corcovado and Piedras Blancas National Parks, Costa Rica, Summer 2023 (in collaboration with Northland College & UW-Madison, WI, USA)

Outreach

- 2025 "After Dark" at MIT museum
- 2021 PBS WI meet the lab "data decoder" program, as "data scientist", see website
- 2018 Peking University Biology Museum, as docent
- 2016 *Bo Wu*, Chinese National Geography Magazine, as editor