HYEBIN SONG

Department of Statistics, The Pennsylvania State University
414 Thomas Building, State College, PA, 16802
email: hps5320@psu.edu | webpage: https://hsong1.github.io/

Education

PhD in Statistics, University of Wisconsin-Madison, May 2020
Bachelor of Arts in Applied Statistics, Yonsei University, 2012

Employment History

2020-	Assistant Professor, Pennsylvania State University
2014-2020	Research/Teaching Assistant, University of Wisconsin-Madison
2012-2014	Statistician, Bank of Korea, Seoul, South Korea

Publications and Preprints

Publications

† indicates co-first author(s)
* indicates corresponding author(s)

Yang Jiang†, Yingzi Xia†, Ian Sitarik, Piyoosh Sharma, **Hyebin Song**, Stephen Fried*, Edward O'Brien*, A novel class of protein misfolding provides a structural explanation for the origins of stretched-exponential refolding kinetics. Science Advances, 2025.

Hyebin Song†, Stephen Berg†*, Multivariate moment least-squares estimators for reversible Markov chains, *Journal of Computational and Graphical Statistics*, 2024.

Viraj Rana†, Ian Sitarik†, Justin Petucci, Yang Jiang, **Hyebin Song***, Edward O'Brien*, Non-covalent Lasso Entanglements in Folded Proteins: Prevalence, Functional Implications, and Evolutionary Significance, *Journal of Molecular Biology*, 2024.

Stephen Berg†, **Hyebin Song**†*, Efficient shape-constrained inference for the autocovariance sequence from a reversible Markov chain, *Annals of Statistics*, 2023.

Sameer D'Costa†, Emily C. Hinds†, Chase R. Freschlin, **Hyebin Song***, Philip A. Romero*, Inferring protein fitness landscapes from laboratory evolution experiments, *PLOS Computational Biology*, 2023.

Ran Dai, **Hyebin Song**, Rina Foygel Barber*, Garvesh Raskutti. Convergence guarantee for the sparse monotone single index model, *Electronic Journal of Statistics*, 2022.

Yi Ding*, Avinash Rao, **Hyebin Song**, Rebecca Willett, Henry (Hank) Hoffmann, NURD: Negative-Unlabeled Learning for Online Datacenter Straggler Prediction, *MLSys Workshop*, 2022.

Hyebin Song*, Garvesh Raskutti, Rebecca Willett. "Prediction in the presence of response-dependent missing labels", *IEEE Statistical Signal Processing Workshop*, 2021.

Hyebin Song, Bennett J. Bremer, Emily C. Hinds, Garvesh Raskutti, and Philip A. Romero*. "Inferring protein sequence-function relationships with large-scale positive-unlabeled learning", *Cell Systems*, 2021.

Hyebin Song*, Ran Dai, Garvesh Raskutti, Rina Foygel Barber. "Convex and Non-convex Approaches for Statistical Inference with Noisy Labels", *Journal of Machine Learning Research*, 2020.

Yuan Li†, Benjamin Mark†*, Garvesh Raskutti, Rebecca Willett, **Hyebin Song**, David Neiman, "Graph-based regularization for regression problems with alignment and highly-correlated designs", *SIAM Journal on Mathematics of Data Science*, 2020.

Ran Dai, **Hyebin Song**, Rina Foygel Barber*, Garvesh Raskutti, "The bias of isotonic regression", *Electronic Journal of Statistics*, 2020.

Hyebin Song*, Garvesh Raskutti. "PUlasso: High-dimensional variable selection with presence-only data." *Journal of the American Statistical Association*, 2018.

- ASA SLDS Student Paper Competition Winner in 2018, *Statistical Learning and Data Science Section, American Statistical Association*

Preprints

† indicates co-first author(s)
‡ indicates students (co)-advised by me
* indicates corresponding author(s)

Sakshi Arya†, **Hyebin Song**†*, Semi-Parametric Batched Global Multi-Armed Bandits with Covariates, *Submitted, ArXiv preprint*, 2025+.

Kaitlyn Fales‡*, Xurui Zhi‡, **Hyebin Song**, Nicole Lazar, Replicability of Functional Brain Networks: A Study Through the Lens of the Default Mode Network, *Submitted*, 2025+.

Hyebin Song†, Stephen Berg†*, Weighted shape-constrained estimation for the autocovariance sequence from a reversible Markov chain, *Submitted, ArXiv preprint*, 2025+.

Justin Petucci, Ian Sitarik, Yang Jiang, Viraj Rana, **Hyebin Song***, Edward O'Brien*, Properties governing native state entanglements and relationships to protein function. *Submitted*. 2025+.

Ian Sitarik, Quyen Vu, Justin Petucci, Paulina Frutos, **Hyebin Song**, Edward O'Brien*, A widespread protein misfolding mechanism is differentially rescued by chaperones based on gene essentiality, *Submitted*, 2025+.

Matthew Jensen†, Corrine Smolen†, Anastasia Tyryshkina†, Lucilla Pizzo†,..., **Hyebin Song**, et al.,..., and Santhosh Girirajan†, Genetic modifiers and ascertainment drive variable expressivity of complex disorders, *medRXiv preprint*, 2024+.

Computing

Software

`momentLS` An `R` package to implement moment LS estimators for estimating autocovariance sequence estimation and asymptotic variance of MCMC estimators from reversible Markov chains. Available as a `GitHub` repository (Link: https://github.com/hsong1/momentLS)

`pudms` An `R` package for a streamlined analysis for positive-unlabeled learning for deep mutational scanning datasets. Available as a `GitHub` repository (Link: https://github.com/RomeroLab/pudms).

`PUlasso`. An `R` package for solving PU (Positive and Unlabeled) problem in low or high dimensional setting with lasso or group lasso penalty. Available on `R-CRAN` (Link: https://cran.r-project.org/web/packages/PUlasso/index.html).

`GTV`. An `R` package for graph-based regularization for regression problems with alignment and highly-correlated designs. Available as a `GitHub` repository (Link: https://github.com/hsong1/GTV).

Honors and Awards

Student Research Grants Competition Award, UW-Madison, 2019

ASA SLDS Student Paper Competition Award, Statistical Learning and Data Science Section, American Statistical Association, 2018

Gateway Course Teaching Assistant Award, Department of Statistics, UW-Madison, 2017

GE Scholarship, Fulbright, 2007

Research/ Training Grants

NSF (grant number DMS-2311141): *A shape-constrained approach for non-parametric variance estimation for Markov Chains*, 2023-2026

Role/Amount: Principal Investigator (50%) (co-PI: Stephen Berg (Stat, PSU)) / $256,757

Talks and Conference Presentations

Invited Talks

Department of Mathematics, Statistics, and Computer Science, University of Illinois Chicago, "Weighted shape-constrained estimation with applications to Markov chain autocovariance function estimation", Oct 2024

Department of Mathematics, Applied Mathematics, and Statistics, Case Western Reserve University, "Weighted shape-constrained estimation with applications to Markov chain autocovariance function estimation", Sep 2024

Department of Statistics, Yonsei University, "Weighted shape-constrained estimation with applications to Markov chain autocovariance function estimation", Aug 2024

Department of Statistics, Chung-Ang University, "Weighted shape-constrained estimation with applications to Markov chain autocovariance function estimation", Aug 2024

Joint Statistical Meeting (JSM) 2024, "Weighted shape-constrained estimation with applications to Markov chain autocovariance function estimation", Aug 2024

Korea Statistical Society (KSS) Summer Conference 2024, "Weighted shape-constrained estimation with applications to Markov chain autocovariance function estimation", July 2024

Paul H. Chook Department of Information Systems and Statistics, Baruch College, City University of New York, "Efficient shape constrained inference with applications in autocovariance sequence estimation", Sep 2023

Joint Statistical Meeting (JSM) 2023, "Efficient shape-constrained inference with applications in autocovariance sequence estimation", Aug 2023

ICSA Hong Kong International Conference 2023, “Utilizing Shape Constrained Inference for Estimating Covariance Functions from Stochastic Processes”, July 2023

BayesComp 2023, “Efficient shape-constrained inference for the autocovariance sequence from a reversible Markov chain”, March 2023

Department of Biostatistics, University of Nebraska Medical Center “Efficient shape-constrained inference for the autocovariance sequence from a reversible Markov chain”, Nov 2022

Department of Statistics, George Mason University, “Efficient shape-constrained inference for the autocovariance sequence from a reversible Markov chain”, Oct 2022

ICSA 2022 China Conference, “Efficient Autocovariance Estimation and Uncertainty Quantification for Discrete-Time Stochastic Processes”, July 2022

INFORMS Annual Meeting, “Statistical inference for high-dimensional and large-scale data with noisy labels”, Oct 2021

IEEE Statistical Signal Processing Workshop, “Prediction in the Presence of Response-Dependent Missing Labels”, July 2021

Korean International Statistical Society (KISS) Webinar, “Statistical inference for high-dimensional and large-scale data with noisy labels”, Oct 2021

Department of Statistics, Seoul National University, “Statistical inference for high-dimensional and large-scale data with noisy labels”, June 2021

Department of Statistics, Korea University, “Prediction in the Presence of Response-Dependent Missing Labels”, Dec 2020

Department of Statistics, The Case Western Reserve University, “Statistical Inference for Large-Scale Data with Incomplete Labels”, Feb 2020

Department of Statistics, The North Carolina State University, “Statistical Inference for Large-Scale Data with Incomplete Labels”, Feb 2020

Department of Statistics, The Florida State University, “Statistical Inference for Large-Scale Data with Incomplete Labels”, Jan 2020

Department of Statistics, The Arizona State University, “Statistical Inference for Large-Scale Data with Incomplete Labels”, Jan 2020

Department of Statistics, The Pennsylvania State University, “Statistical Inference for Large-Scale Data with Incomplete Labels”, Jan 2020

Workshop on Recent Developments on Mathematical/Statistical approaches in Data Science (MSDAS), University of Texas Dallas, “High-dimensional Variable Selection in Positive-Unlabeled Learning”, June 2019

Joint Statistical Meeting (JSM) 2018, “PUlasso: High-dimensional variable selection with presence-only data”, Jul 2018

Campus Talks or Other Contributed Talks

2024 Women in Statistics and Data Science Conference (WSDS) Conference, "Weighted shape-constrained estimation with applications to Markov chain autocovariance function estimation", Oct 2024

Wartik Weekly Wednesday Genomics Lecture Series (WWWGLS), The Pennsylvania State University, "Uncertainty quantification for parameter estimates from Markov chain Monte Carlo methods", Dec 2023

Stochastic Modeling and Computational Statistics Seminar, The Pennsylvania State University, "Efficient shape-constrained inference for the autocovariance sequence from a reversible Markov chain", Sep 2022

Wartik Weekly Wednesday Genomics Lecture Series (WWWGLS), The Pennsylvania State University, "Learning From Laboratory Protein Evolution Data", Apr 2021

Stochastic Modeling and Computational Statistics Seminar, The Pennsylvania State University, "Prediction in the Presence of Response-Dependent Missing Labels", Nov 2020

Bioinformatics and Genomics Retreat, The Pennsylvania State University, "A Semi-supervised Approach for Protein Function Modeling and Engineering with Large-scale Deep Mutational Scanning Data", Aug 2020

Department of Statistics, University of Wisconsin-Madison, "Statistical Inference for Large-Scale Data with Incomplete Labels", Dec 2019

Systems, Information, Learning and Optimization (SILO) Seminar, University of Wisconsin-Madison, "PUlasso: High-dimensional variable selection with presence-only data", Jan 2018

Conference Poster Presentations

Statistical Foundations of Data Science and their Applications, *a conference in celebration of Jianqing Fan's 60th Birthday*, "Efficient shape-constrained inference for the autocovariance sequence from a reversible Markov chain", May 2023

2019 Joint Statistical Meeting (JSM), "Statistical Inference in a High-Dimensional Binary Regression Problem with Noisy Responses", Jul 2019

Midwest Machine Learning Symposium (MMLS), "PULasso: High-dimensional variable selection with presence-only data", June 2018

Advising

Ph.D in Statistics Current

Kaitlyn Fales (with Nicole Lazar) (Jan 2023 - current; expected, 2026)
Manushi Siriwardana (Jun 2024 - current)
Nathan Weaver (Sept 2024 - current)

Masters in Applied Statistics

Phichchaya Sutaporn (Sept 2023 - current)
Maria A. Angelo (Jan 2024 - current)
Ian (Xurui) Zhi (Jan 2024 - Dec 2024)

B.S. in Statistics

Danqi Ding (Jun 2024 - current)

Other Research Projects

Xinyue Wang, Doctoral student, Department of Statistics (Apr 2022 - Aug 2022)

Ph.D Committee Member

Padma Tanikella, Doctoral student, Statistics (expected, 2026)
Judith Rodriguez, Doctoral student, Bioinformatics and Genomics Program (expected, 2025)
Jakob Wiegand, Department of Civil and Environmental Engineering (expected, 2026)
Saurov Hazarika, Doctoral student, Department of Chemistry (graduated, 2025)
Wenlong Yang, Doctoral student, Department of Statistics (expected, 2026)
Tran Tran, Doctoral student, Department of Statistics (graduated, 2024)
Wei Wei, Doctoral student, Bioinformatics and Genomics Program (graduated, 2024)
Pengxiang Zhang, Doctoral student, Department of Civil and Environmental Engineering (graduated, 2024)
Shirin Madarshahian, Doctoral student, Department of Kinesiology (graduated, 2022)

Teaching

Instructor (Pennsylvania State University)

Graduate level courses

Regression Models (PhD core course; STAT 511)	FA2021, FA2022, FA2023, FA2024
Regression Methods (grad other majors; STAT 501)	SP2023, SP2024

Undergraduate level courses

Applied Time Series Analysis	SP2025
Introduction to Mathematical Statistics (junior level Stat major)	SP2021, SP2022, SP2023, SP2025
Elementary Probability (junior level non-Stat major)	SP2024
Introduction to Probability (junior level Stat major)	FA2020

Professional Activities

Professional Service and Committees

Korean International Statistical Society (KISS) Program Chair Elect, Sep 2024 - current

Thematic Lead (Cross-cutting Data Sciences, AI, and ML), NSF National Synthesis Center for Emergence in the Molecular and Cellular Sciences, Sep 2024 - current

Panelist, NSF Proposal Review Panel for the Division of Mathematical Sciences (DMS), Feb 2024

Conference Organization/ Participation

Session Chair, “Nonparametric Statistics”, Keystone State Statistics Symposium at Penn state, Oct 2023.

Session Organizer/Chair, “Recent advances in non-parametric modeling with applications”, JSM 2023, June 2023.

Session Chair, “Dynamic Networks”, Statistical Network Science with Applications Conference at Penn State, May 2023.

Session Chair,“CS6e: Cases and Applications”, 2022 Women in Statistics and Data Science Conference, Oct 2022.

Session Organizer/Chair, “Semi-parametric inference and modeling with shape-constraints”, EcoSta 2022, June 2022.

Journal Referee: Electronic Journal of Statistics, Journal of Machine Learning Research, Annals of Applied Statistics, Journal of Computational and Graphical Statistics, Journal of American Statistical Association, Journal of the Royal Statistical Society: Series B, IEEE Transactions on Signal Processing, Stats, Statistical Sinica.

Department/ University Service (Pennsylvania State University)

Member, Rao Prize Organization Committee, 2024- current

Adviser, Graduate Academic Advisor, 2023- current

Adviser, Undergraduate Advisor, 2023- current

Member, PhD qualifying Exam Committee, 2022- current

Member, Statistics Department Head Search Committee, 2023

Member, Graduate Curriculum Committee, 2021 - 2023

Member, Faculty Search Committee, 2021 - 2022

Organizer, Stochastic Modeling and Computational Statistics (SMAC) Seminar, 2021

Member, Bioinformatics and Genomics PhD Student Recruitment Committee, 2020

Organizer, Statistics Department Colloquium, 2020

Other Professional Service/ Outreach Activities

Judge, 2025 ASA SLDS Student Paper Competition, Jan 2025

Judge, ASA DataFest, Aprl 2021, Mar 2023, Mar 2024

Mentor, 2023 ENVISION:STEM Career Day Supporting Young Women Workshop for grade 6-12 students, Feb 2023

Judge, 2023 Undergraduate Statistics Project Competition (USPROC) competition, Feb 2023, June 2023

Judge, 2021 INFORMS Data Mining and Decision Analytics (DMDA) Workshop Best Paper Competition, Sep 2021

Judge, 2019 UW-Madison Undergraduate Data Challenge, Oct 2019