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 equal contributions
* indicates corresponding author(s)

Sameer D'Costa†, Emily C. Hinds†, Chase R. Freschlin, **Hyebin Song***, Philip A. Romero*, Inferring protein fitness landscapes from laboratory evolution experiments, *Accepted, 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 equal contributions

[*] indicates corresponding author(s)

Hyebin Song[†], Stephen Berg[†*], Multivariate moment least-squares estimators for reversible Markov chains, *Submitted, ArXiv preprint*, 2023+.

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, *Under invited revision*, 2023+.

Stephen Berg[†], **Hyebin Song**[†*], Efficient shape-constrained inference for the autocovariance sequence from a reversible Markov chain, *Under review, ArXiv preprint*, 2022+.

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

TALKS AND CONFERENCE PRESENTATIONS

Invited Talks

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

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

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

TEACHING EXPERIENCE

PHD (CO-) SUPERVISOR

Kaitlyn Fales, Doctoral student (with Nicole Lazar), Department of Statistics (expected, 2026)

PHD COMMITTEE MEMBER

Wenlong , Doctoral student, Department of Statistics (expected, 2024)

Tran Tran, Doctoral student, Department of Statistics (expected, 2024)

Wei Wei, Doctoral student, Bioinformatics and Genomics Program (expected, 2024)

Judith Rodriguez, Doctoral student, Bioinformatics and Genomics Program (expected, 2024)

Shirin Madarshahian, Doctoral student, Department of Kinesiology (graduated, 2022)

RESEARCH PROJECTS (NOT RELATED TO PHD THESIS)

Xinyue Wang, Doctoral student, Department of Statistics (Spring 2022 - Summer 2022)

INSTRUCTOR (PENNSYLVANIA STATE UNIVERSITY)

Graduate level courses

Applied Regression Analysis (PhD core course)	Fall 2021, Fall 2022, Fall 2023
Regression Methods (Grad other majors)	Spring 2023

Undergraduate level courses

Introduction to Mathematical Statistics (junior level Stat major)	Spring 2021, Spring 2022, Spring 2023
Introduction to Probability (junior level Stat major)	Fall 2020

PROFESSIONAL SERVICE

Conference Activity / Participation

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

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

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

Session Chair, “2023 Statistical Network Science with Applications Conference”, 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.

Editorial Service

Reviewer for *Electronic Journal of Statistics, Journal of Machine Learning Research, Annals of Applied Statistics, Journal of Computational and Graphical Statistics, Journal of American Statistical Association, IEEE Transactions on Signal Processing, Stats, Statistical Sinica.*

Department Service

Member, Graduate curriculum committee, 2021 -

Member, PhD Qualifying Exam Committee, 2022

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

Judge, 2023 ASA DataFest, Mar 2023

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

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

Judge, 2021 ASA DataFest, Apr 2021

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

COMPUTING

Software

`pudms` An `R` package for a streamlined analysis for positive-unlabeled learning for deep mutational scanning datasets. Available as a `GitHub` repository.

`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 `CRAN`.

`GTV`. An `R` package for graph-based regularization for regression problems with alignment and highly-correlated designs. Available at my `GitHub` site.