HYEBIN SONG

Department of Statistics, The Pennsylvania State University
326 Thomas Building, State College, PA, 16802
email: hps5320@psu.edu | webpage: http://personal.psu.edu/hps5320

EDUCATION

PhD in Statistics, University of Wisconsin-Madison, May 2020

Bachelor of Arts in Applied Statistics, Yonsei University (Rank: 1/72), 2012

EMPLOYMENT HISTORY

2020-	Assistant Professor, The 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

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

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", *Under Review, BioArXiv Preprint*, 2020+.

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

TEACHING EXPERIENCE

Instructor (The Pennsylvania State University)

2020	STAT 414: Introduction to Probability

Teaching Assistant (UW-Madison)

2019	STAT 850: Theory and Application of Regression and Analysis of Variance II
2016	STAT 312: Introduction to Theory and Methods of Mathematical Statistics II
	STAT 424: Statistical Experimental Design
	STAT 324: Introductory Applied Statistics for Engineers
2015, 17	STAT 301: Introduction to Statistical Methods
2014	STAT 371: Introductory Applied Statistics for the Life Sciences

TALKS AND CONFERENCE PRESENTATIONS

Invited Talks

"A Semi-supervised Approach for Protein Function Modeling and Engineering with Large-scale Deep Mutational Scanning Data"

- at Bioinformatics and Genomics Retreat, The Pennsylvania State University, Aug 2020

"Statistical Inference for Large-Scale Data with Incomplete Labels"

- at Statistics Seminar, The Case Western Reserve University, Feb 2020
- at Statistics Seminar, The North Carolina State University, Feb 2020
- at Statistics Colloquium, The Florida State University, Jan 2020
- at Statistics Seminar, The Arizona State University, Jan 2020
- at Statistics Colloquium, The Pennsylvania State University, Dec 2019

"High-dimensional Variable Selection in Positive-Unlabeled Learning"

- at 2019 Workshop on Recent Developments on Mathematical/Statistical approaches in Data Science (MSDAS), UT Dallas, June 2019

Talks

"PUlasso: High-dimensional variable selection with presence-only data."

- at Joint Statistical Meeting (JSM), Vancouver, Jul 2018
- at Systems, Information, Learning and Optimization (SILO) Seminar, UW-Madison, Jan 2018

Conference Presentations

"Statistical Inference in a High-Dimensional Binary Regression Problem with Noisy Responses"

- at Joint Statistical Meeting (JSM), Vancouver, Jul 2019

"PULasso: High-dimensional variable selection with presence-only data"

- at Midwest Machine Learning Symposium (MMLS), Chicago, June 2018

PROFESSIONAL SERVICE

Reviewer for Journal of Machine Learning Research

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.

Languages

- `R, C++, Python`