

YANRAN LI

(+1) 617-401-5976 [◇ yl5465@cumc.columbia.edu](mailto:yl5465@cumc.columbia.edu) [◇ LinkedIn](#) [◇ GitHub](#)

EDUCATION

Columbia University

Ph.D. in Biostatistics;

New York, New York

08/2023 - 05/2027(Expected)

Harvard University

Master of Science in Biostatistics · GPA: **3.97/4.00**;

Boston, Massachusetts

09/2020 - 03/2023

- Coursework: Probability; Inference; Methods; Statistical Learning; Advanced Data Science; Bayesian Modeling Inference.
- Completed all doctoral core curriculum (Ph.D. in Biostatistics).
- MIT Cross-registered: 6.435 Bayesian Modeling Inference; 6.S982 Clinical Data Learning, Visualization, and Deployments.

Sun Yat-Sen University

Bachelor of Science in Physics · GPA: **3.9/4.0**; Rank: **2/59**;

Guangzhou, China

09/2016 - 06/2020

- *Top Talents in Basic Disciplines Training Program* (outstanding research performance)
- Coursework: Real Analysis; Linear Algebra; C Programming; Optimization; Ordinary/Partial Differential Equation.
- Chinese National Scholarship (**top 1%** 2019); First-class Scholarship for Outstanding Students (top 3% 2019, 2018, 2017).
- Outstanding Graduates, Class of 2020; Excellent Undergraduate Graduation Thesis.

University of California, Berkeley

Exchange Student in Statistics with full scholarship · GPA: **4.0/4.0**;

Berkeley, California

01/2019 - 05/2019

RESEARCH INTERESTS

My research focuses on developing statistical methods that enable rigorous and impactful uses of data to societal questions, which are driven by two key goals: (1) addressing statistical challenges that commonly arise in healthcare decision-making processes, and (2) developing novel machine learning methods to effectively leverage data. Methodologically, I am interested in probabilistic machine learning, spatio-temporal models, ensemble learning, uncertainty quantification and Bayesian methods.

RESEARCH EXPERIENCES

Department of Biostatistics, Harvard T.H. Chan School of Public Health

Boston, Massachusetts

Research Assistant with Prof. Rachel C. Nethery and Prof. Brent A. Coull

➤ Impacts of Census Differential Privacy for Monitoring Health Inequities

06/2021 - 08/2022

- Investigated the impact of the US Census Bureau's new privacy-preserving disclosure avoidance system (DAS) on small-area population counts and studies of health inequities that employ these population counts as denominators.
- Conducted simulation studies using high performance computing systems and real data analyses to evaluate bias induced in racial/socioeconomic inequity estimates from Bayesian spatial disease mapping models due to DAS-affected denominators.
- Led a first-author paper ([published](#) at *Science Advances*), including the novel finding that recent versions of the DAS lead to much more accurate inequity estimates than older versions.

➤ Spatially Adaptive Ensemble Learning with Calibrated Predictive Uncertainty

09/2022 - Present

- Developed and validated Bayesian Nonparametric Ensemble (BNE) models to accurately quantify uncertainty associated with model-predicted high-resolution PM_{2.5} exposure surfaces.
- Integrated PM_{2.5} information from three distinct high-resolution PM_{2.5} exposure models based on out-of-sample prediction at 51 monitoring sites in eastern New England, and performed posterior inference by using BNE models.
- Co-authored paper "Adaptive Ensemble Learning of Spatiotemporal Processes with Calibrated Predictive Uncertainty: A Bayesian Nonparametric Approach" under review at *Journal of the American Statistical Association (JASA)* soon.

➤ Ensemble Learning Method for Different Population Data Sources

07/2022 - Present

- Modified and applied the BNE model for areal spatial data to combine information from multiple population estimate data sources and characterize uncertainty.
- Collected, processed, and merged population estimate datasets from Census Bureau, American Community Survey, Population Estimates Program, WorldPop and Facebook High Resolution Density Layer for the contiguous US.

Laboratory of Computational Physiology, Massachusetts Institute of Technology

Cambridge, Massachusetts

*Research Assistant with Prof. Leo Anthony G. Celi***> Time Dependent Regression on SOFA Score in ICU Mortality**

10/2020 - 08/2021

- Conducted a retrospective cohort study of patients using MIMIC-IV database to evaluate the strength of association between specific organ dysfunction and mortality in Sequential Organ Failure Assessment (SOFA) model for various sub-populations.
- Established and analyzed different regression models for the patients who died in 24 hours and in 168 hours; Applied regression adjustment and factor analysis models to optimize SOFA Score model.
- First-authored paper presented at American Public Health Association's 2022 Annual Meeting.

Department of Biostatistics, University of California, Berkeley

Berkeley, California

*Research Assistant with Prof. Alan E. Hubbard***> A Data-Adaptive Targeted Learning Approach for Trauma Hospital Outcomes**

07/2019 - 10/2019

- Implemented Targeted Maximum Likelihood Estimator under a causal inference framework to estimate patients' hemostasis and mortality status at different clinical sites.
- Deployed Super Learner (ensemble machine learning method) to select from all possible combinations of algorithms, and constructed a data-adaptive model to estimate the influence of hospitals' volume efficiently.
- Derived variance and confidence interval based on the efficient influence curve: Hospitals with large volume had 13% higher prevalence of hemostasis than those with small volume, but there was no significant effect on either mortality at 6h or 24h.

Department of Physics, Sun Yat-sen University

Guangzhou, China

*Research Assistant with Prof. Daoxin Yao***> Machine Learning for Quantum Many-body Systems**

10/2017 - 10/2018

- Derived *Mean Field Theory* and *Variational Free Energy principle* to deal with the variation of Ising model in statistical physics; Calculated KL divergence and gave the machine learning perspective of Ising model;
- Modelled classical lattice glasses at zero total magnetization and found out the structure factor versus temperature related to symmetry breaking: the critical temperature of such solid-gas phase transition is the same as the Ising transition $T_c/J \approx 2.269$.

PROJECT EXPERIENCES**Analysis on Online Variational Filtering and Parameter Learning**

Cambridge, Massachusetts

Course Project of EECS 6.435, Massachusetts Institute of Technology, Instructor: Tamara Broderick

01/2022 - 04/2022

- Reproduced state estimation and parameter learning in state-space models in *Online Variational Filtering and Parameter Learning*.
- Implemented the standard Variational Autoencoders by training the NN^g (pre-trained decoder) and implemented the online method by training the transition function NN_θ^f (neural network to predict changes).

PUBLICATIONS**I. JOURNAL PAPERS**

- **Y. Li**, B. Coull, N. Krieger, E. Peterson, L. Waller, J. Chen, R. Nethery, "Impacts of Census Differential Privacy for Small-Area Disease Mapping to Monitor Health Inequities", *Sci. Adv.***9**, eade8888(2023). DOI:10.1126/sciadv.ade8888.

II. CONFERENCE PROCEEDINGS

- **Y. Li**, R. Nethery. "Impacts of Census Differential Privacy for Monitoring Health Inequities", *lightning talk presentation*, The Bayesian Young Statisticians Meeting, Montréal, Quebec, Canada (BAYSM 2022).
- **Y. Li**, B. Lam, Z. Chen, Y. Jia, L. Celi, H. Mattie, J. Raffa. "Evaluating how the SOFA score components vary in their contribution to mortality over time", *poster presentation*, American Public Health Association's 2022 Annual Meeting, Boston (APHA 2022).
- **Y. Li**, D. Yao. "Machine Learning for Quantum Many-body Systems", *poster presentation*, American Association of Physics Teachers Summer Meeting, Provo, Utah (AAPT 2019).
- Z. Liang, **Y. Li** (co-first author), et al. "Adversarial Deep Reinforcement Learning in Portfolio Management", *oral presentation*, Sixth Asian Quantitative Finance Conference (AQFC 2018), <https://arxiv.org/abs/1808.09940>.

III. INVITED TALKS

- Impacts of Census Differential Privacy for Small-Area Disease Mapping to Monitor Health Inequities. Harvard Center for Geographic Analysis (CGA) conference, March 2023.

TEACHING EXPERIENCES

Teaching Assistant, BST 215 Linear and Longitudinal Regression

Boston, Massachusetts

Instructor: Prof. Garrett Fitzmaurice

07/2022-08/2022

- Implemented and graded 3 assignments. Held course office hours.

Teaching Assistant, BST 209 ML: Collaborative Data Science in Healthcare

Boston, Massachusetts

Instructor: Prof. Heather Mattie and Prof. Leo Anthony G. Celi

06/2022-07/2022

- Supervised student research teams on semester-long health data science projects. Held course office hours.

HONORS AND AWARDS

- Harvard T.H. Chan School of Public Health Professional Development Support Fund (2022)
- ISBA New Researcher Travel Award, International Society for Bayesian Analysis (ISBA, 2022)
- Silver Medal & Best Software Project: International Genetically Engineered Machine Competition (iGEM 2018)
- Meritorious Winner (First prize): Interdisciplinary Contest in Modeling (ICM 2017)

TECHNICAL STRENGTHS

Programming

Python (TensorFlow, PyTorch, PyMC3, JAX), Linux Shell, Linux Bash, C, Git

Data Analytics

R (dplyr, caret, tidyverse, mgcv), SQL, SAS, Stan

Graphics & Documents

ggplot2, Matplotlib, Seaborn, Shiny, LATEX