SHANGHONG XIE

Department of Statistics, University of South Carolina
LeConte College, Columbia, SC 29208
SX2@mailbox.sc.edu
Personal website: `https://shanghongxie.github.io/`

EDUCATION

Columbia University in the City of New York 2019
Ph.D. in Biostatistics
Dissertation Title: Statistical Methods for Constructing Heterogeneous Biomarker Networks
Advisor: Yuanjia Wang

University of Illinois at Urbana-Champaign 2014
M.S. in Statistics

Sichuan University, Chengdu, China 2012
B.S. with highest honors in Statistics

PROFESSIONAL EXPERIENCE

- Assistant Professor, *Department of Statistics*, University of South Carolina 2024 – Present
- Member, *Carolina Autism and Neurodevelopment (CAN) Research Center*, University of South Carolina 2024 – Present
- Associate Professor, *Department of Data Science*, School of Statistics, Southwestern University of Finance and Economics 2023 – 2024
- Assistant Professor, *Department of Data Science*, School of Statistics, Southwestern University of Finance and Economics 2021 – 2023
- Member, *Center of Statistical Research*, Southwestern University of Finance and Economics 2021 – 2024
- Postdoctoral Research Scientist, *Department of Biostatistics*, Mailman School of Public Health, Columbia University 2019 – 2021

RESEARCH INTERESTS

Machine learning; network analysis; graphical model; precision medicine; functional data analysis; causal inference; mediation analysis; variable selection; high dimensional analysis; neuroimaging; biomarker; neurological and psychiatric diseases; mental health; COVID-19

AWARDS AND HONORS

- American Statistical Association (ASA) Mental Health Statistics Section Best Student Paper Award 2020
- International Conference on Health Policy Statistics (ICHPS) Student Travel Award 2018
- NYC Datathon (Data Science Competition) 1st Place Winning Team, among 1000+ participants 2017
- Columbia University Fellowship 2014 – 2017

PUBLICATIONS

An asterisk (∗) is used to indicate corresponding author; An underline is used to indicate students under my supervision; A dagger (†) is used to indicate joint first author, equal contribution.

Peer-Reviewed Journal and Conference Articles

1. McDonnell, E., **Xie, S.**, Marder, K., Cui, F., and Wang, Y. (2024). Dynamic Undirected Graphical Models for Time-Varying Clinical Symptom and Neuroimaging Networks. *Statistics in Medicine*. In Press. (**An earlier version won ASA Statistics in Imaging Section First Prize Student Paper Award**)

2. Shi, B., Liu, Y., **Xie, S.**, Zhu, X., and Wang, Y. (2024). Network-Assisted Mediation Discovery with Neuroimaging Mediators. *Machine Learning for Healthcare Conference*. Proceedings of Machine Learning Research (PMLR). In Press.

3. **Xie, S.***, Zeng, D., and Wang, Y. (2024). Identifying Temporal Pathways Using Biomarkers in the Presence of Latent Non-Gaussian Components. *Biometrics*. In Press.

4. **Xie, S.*** and Ogden, R. T. (2024). Functional Support Vector Machine. *Biostatistics*. In Press.

5. **Xie, S.***, Tarpey, T., Petkova, E., and Ogden, R. T. (2022). Multiple Domain and Multiple Kernel Outcome-weighted Learning for Estimating Individualized Treatment Regimes. *Journal of Computational and Graphical Statistics* 31 (4), 1375-1383.

6. **Xie, S.***, Wang, W., Wang, Q., Wang, Y., and Zeng, D. (2022). Evaluating Effectiveness of Public Health Intervention Strategies for Mitigating COVID-19 Pandemic. *Statistics in Medicine* 41 (9), 3820-3836.

7. COVID-19 Forecast Hub Consortium (2022). Evaluation of Individual and Ensemble Probabilistic Forecasts of COVID-19 Mortality in the US. *Proceedings of the National Academy of Sciences* 119 (15), e2113561119.

8. COVID-19 Forecast Hub Consortium (2022). The United States COVID-19 Forecast Hub Dataset. *Scientific Data* 9, 462.

9. **Xie, S.***, McDonnell, E., and Wang, Y. (2022). Conditional Gaussian Graphical Model for Estimating Personalized Disease Symptom Networks. *Statistics in Medicine* 41 (3), 543-553. (**An earlier version won ASA Mental Health Statistics Section Best Student Paper Award**)

10. **Xie, S.***, Zeng, D., and Wang, Y. (2021). Integrative Network Learning for Multi-modality Biomarker Data. *Annals of Applied Statistics* 15 (1), 64-87.

11. **Xie, S.***, Li, X., McColgan, P., Scahill, R. I., Zeng, D., and Wang, Y. (2020). Identifying Disease-associated Biomarker Network Features Through Conditional Graphical Model. *Biometrics* 76 (3), 995-1006. (**Cover story of *Biometrics* September 2020 issue; An earlier version won the International Conference on Health Policy Statistics (ICHPS) Student Travel Award**)

12. Goldman, J., **Xie, S.**, Green, D., Naini, A., Mansukhani, M. M., and Marder, K. (2021). Predictive Testing for Neurodegenerative Diseases in the Age of Next-generation Sequencing. *Journal of Genetic Counseling* 30, 553-562.

13. Wang, Q., **Xie, S.**, Wang, Y., and Zeng, D. (2020). Survival-Convolution Models for Predicting COVID-19 Cases and Assessing Effects of Mitigation Strategies. *Frontiers in Public Health* 8, 325. (**Our model was used by the Center of Disease Control and Prevention (CDC) for COVID-19 Ensemble Forecast; Our forecasts website:** `https://github.com/COVID19BIOSTAT/covid19_prediction`; **CDC ensemble forecast website:** `https://www.cdc.gov/coronavirus/2019-ncov/covid-data/forecasting-us.html`)

14. Li, X.†, **Xie, S.**†, McColgan, P., Tabrizi, S. J., Scahill, R. I., Zeng, D., and Wang, Y. (2018). Learning Subject-Specific Directed Acyclic Graphs with Mixed Effects Structural Equation Models from Observational Data. *Frontiers in Genetics* 9, 430.

15. Li, X., **Xie, S.**, Zeng, D., and Wang, Y. (2018). Efficient l_0-norm Feature Selection Based on Augmented and Penalized Minimization. *Statistics in Medicine* 37 (3), 473-486.

16. Avissar, M.†, **Xie, S.**†, Vail, B., Lopez-Calderon, J., Wang, Y., and Javitt, D. C. (2018). Meta-analysis of Mismatch Negativity to Simple versus Complex Deviants in Schizophrenia. *Schizophrenia Research* 19, 25-34.

GRANT SUPPORT

- **Principal Investigator** 01/2023 – 07/2024
  "Learning Temporal Causal Network from Biomarker Time Series Data"
  National Natural Science Foundation of China (NSFC), Department of Mathematical and Physical Sciences, Grant No 12201511. Funding rate 17%

- **Principal Investigator** 01/2022 – 12/2022
  "Dynamic Network Learning Using Neuroimaging Data"
  Southwestern University of Finance and Economics Startup Grant

COLLABORATIVE RESEARCH EXPERIENCE

New York State Psychiatric Institute (NYSPI) 11/2016 – 08/2021
Role: Statistician

- Investigated the mental health impact of the COVID-19 pandemic on healthcare workers
- Developed design and statistical analysis plan to investigate the effects of cannabis use with varying concentrations on an NIH R01 grant proposal (funded)
- Designed a clinical trial for schizophrenia patients on an NIH grant proposal (funded)
- Provided statistical support (e.g., power analysis, statistical analysis) for clinicians and fellows
- Designed a study to compare a new short-form survey with a standard long-form for Alzheimer's disease, conducted power analysis and computed sensitivity, specificity, etc
- Conducted meta-analysis for a Schizophrenia study with clinicians

Columbia University Vagelos College of Physicians and Surgeons, Department of Neurology 09/2019 – 08/2021
Role: Statistician

- Conducted statistical analysis to investigate the psychological impact of predictive testing for neurodegenerative diseases using next-generation sequencing panels
- Provided statistical support (e.g., power analysis)

University College London, Institute of Neurology 11/2016 – 09/2021
Role: Collaborator

- Investigated brain connectivities

Columbia University, School of Social Work 01/2016 – 03/2016
Role: Statistician

- Conducted statistical analysis for a 25-year long period longitudinal cardiovascular disease study

TEACHING EXPERIENCE

University of South Carolina, Department of Statistics 2024 – Present
Instructor

- STAT 509: Statistics for Engineers Fall 2024

Southwestern University of Finance and Economics, School of Statistics 2022 – 2024
Instructor

- Reading & Writing Scientific Articles: Advanced Topics in Machine Learning (*Graduate*) Spring 2024
  11 students enrolled
- Machine Learning and Data Mining (*Undergraduate*) Fall 2023
  72 students enrolled
- An Introduction to Machine Learning (*Graduate*) Fall 2023
  63 students enrolled
- Reading & Writing Scientific Articles: Advanced Topics in Machine Learning (*Graduate*) Spring 2023
  12 students enrolled
- Machine Learning and Data Mining (*Undergraduate*) Spring 2023
  40 students enrolled
- Machine Learning and Data Mining (*Undergraduate*) Fall 2022
  2 classes, 121 students enrolled in total
- Reading & Writing Scientific Articles: Advanced Topics in Machine Learning (*Graduate*) Spring 2022
  8 students enrolled
- Machine Learning and Data Mining (*Undergraduate*) Spring 2022
  2 classes, 132 students enrolled in total

Massive Open Online Course (MOOC: XuetangX) 2023
Co-Instructor (Graduate Course)

- Machine Learning

Columbia University, Department of Biostatistics 2015 - 2018
Co-Instructor (Graduate Course)

- Statistical Collaboration for Interdisciplinary Research Spring 2018

Teaching Assistant (Graduate Course)

- Randomized Clinical Trial II Fall 2016
- Generalized Linear Models Fall 2016
- Design of Medical Experiments Spring 2016
- Analysis of Longitudinal Data Fall 2015

University of Illinois at Urbana-Champaign, Department of Statistics 2013 – 2014
Teaching Assistant (Graduate Course)

- Sampling and Categorical Data Spring 2014
- Applied Multivariate Analysis Fall 2013

MENTORING ACTIVITIES

Southwestern University of Finance and Economics, School of Statistics
Advisee

- Hexuan Song (Master thesis) 09/2022 – Present
- Guishan Xiang (Master thesis) 09/2022 – 06/2024
- Anqi Hua (Bachelor thesis) 10/2023 – 06/2024
- Yi Jiang (Bachelor thesis) 10/2023 – 06/2024
- Lu Li (Bachelor thesis) 10/2023 – 06/2024
- Lijuan Guo (Bachelor thesis) 01/2022 – 06/2023
- Silu Liu (Bachelor thesis) 01/2022 – 06/2023
- Qi Yang (Bachelor thesis) 01/2022 – 06/2023

- Yangjie Yin (Bachelor thesis) 01/2022 – 06/2023
- Ruiying Li (Bachelor thesis) 09/2021 – 06/2022
- Mengjie Li (Bachelor thesis) 09/2021 – 06/2022
- Xing Wang (Bachelor thesis) 09/2021 – 06/2022
- Qiyu Wang (Bachelor thesis) 09/2021 – 06/2022

Academic Advisor

- Supervised over 40 undergraduate students, received excellent evaluations 09/2022 – 07/2024

Columbia University, Department of Biostatistics
Mentee

- Zexi Cai (PhD student) 09/2022 – Present
  PhD dissertation projects
- Bin Yang (Master student) 01/2021 – 08/2021
  Research project
- Erin Mcdonnell (PhD student, Advisor: Professor Yuanjia Wang) 09/2019 – 08/2021
  Dissertation projects
- Bihui Sun (Master student, Advisor: Professor Yuanjia Wang) 11/2019 – 05/2020
  Practicum project

ACADEMIC SERVICES

Editorial Board

- *Frontiers in Public Health* 2023 – Present

Journal and Conference Review

- *Annals of Applied Statistics*
- *Biometrics*
- *Briefings in Bioinformatics*
- *Computational Statistics and Data Analysis*
- *Journal of the American Statistical Association*
- *Statistics in Medicine*
- *Statistics in Biosciences*
- *Stat*
- *IEEE International Conference on Bioinformatics and Biomedicine*

Grant Review

- Full-time Member, 2024 – Present
  NIH/NINDS Huntington's Disease Biospecimen Resource Access Committee (HD-BRAC)

Conference Service

- Chair, Invited session "New developments in the frontiers of precision medicine and data science", 2024
  7th International Conference on Econometrics and Statistics (EcoSta)
- Member, Student Paper Competition Committee for ASA Mental Health Statistics Section 2023 – 2024
- Organizer, Invited session "Topics in healthcare and biostatistics", R conference in China 2023

- Organizer, Invited session "Novel machine learning methods to advance precision medicine using big biomarker data", ICSA China 2023
- Reviewer, Student Paper Competition, International Conference on Health Policy Statistics (ICHPS) 2022
- Chair, Invited session "Statistical research in rapid response to COVID-19 pandemic: forecasts, risk factors, therapeutics, and vaccine trials", Joint Statistical Meetings (JSM) 2021
- Chair, Topic-contributed session "Topics in clustering", JSM 2018

Departmental and University Committees

Southwestern University of Finance and Economics

- Organizer, Departmental Seminar, School of Statistics 01/2023 – 07/2023
- Member, Postdoctoral Evaluation Committee, School of Statistics 09/2022 – 07/2024
- Member, Master Thesis Committee, School of Statistics 03/2022 – 07/2024
- Member, Bachelor Thesis Committee, School of Statistics 03/2022 – 07/2024
- Member, Graduate Student Admissions Committee, School of Statistics 01/2022 – 07/2024
- Member, Curriculum Committee, School of Statistics 09/2021 – 07/2024

Memberships

- American Statistical Association (ASA)
- International Biometric Society, Eastern North American Region (ENAR)
- International Chinese Statistical Association (ICSA)
- Institute of Mathematical Statistics (IMS)
- New England Statistical Society (NESS)

SOFTWARE

R package 'APML0'
Co-maintainer

- Augmented and penalized minimization method for regularized linear, logistic, and Cox models with ℓ_0 penalty, flexible for ℓ_1, ℓ_2, and network type regularized regression
- Most intensive computation codes written in C++
- Available on CRAN: `https://cran.r-project.org/web/packages/APML0/index.html`
- Downloaded 41,593 times as of 9/2023

R package 'Covariate-dependent-network'
Maintainer

- Estimate covariate-dependent networks through conditional Gaussian graphical model, in which both the mean and precision matrix depend on covariates
- Most intensive computation codes written in C++
- Available on GitHub: `https://github.com/shanghongxie/Covariate-dependent-network`

R package 'INL'
Maintainer

- Integrative network learning for multi-modality data
- Most intensive computation codes written in C++
- Available on GitHub: `https://github.com/shanghongxie/INL`

Matlab toolbox 'OWMKL'
Maintainer

- Outcome weighted multiple kernel learning (OWMKL) for estimating individualized treatment rules
- Available on GitHub: `https://github.com/shanghongxie/OWMKL`

R package 'FSVM'
Maintainer

- Functional support vector machine for classification and regression problems
- Available on GitHub: `https://github.com/shanghongxie/FSVM`

R package 'ICATemporalNetwork'
Maintainer

- Temporal causal network learning, adjusting for latent non-Gaussian components and separating the temporal network from the contemporaneous network
- Available on GitHub: `https://github.com/shanghongxie/ICATemporalNetwork`

PRESENTATIONS AND POSTERS

"Identifying Temporal Pathways Using Biomarkers in the Presence of Latent Non-Gaussian Components." International Conference on Econometrics and Statistics (EcoSta), Beijing, China, *Invited session* 07/2024

"Identifying Disease-associated Biomarker Network Features by Integrating Multi-modality Data." Department of Statistics, Virginia Tech, *Invited talk* 08/2023

"Identifying Temporal Pathways Using Biomarkers in the Presence of Latent Non-Gaussian Components." Hangzhou International Conference on Frontiers of Data Science, Hangzhou, China, *Invited session* 08/2023

"Identifying Temporal Pathways Using Biomarkers in the Presence of Latent Non-Gaussian Components." ICSA China, Chengdu, China, *Invited session* 07/2023

"Identifying Disease-associated Biomarker Network Features through Graphical Models." Department of Epidemiology and Biostatistics, University of Georgia, *Invited talk* 04/2023

"Evaluating Effectiveness of Public Health Intervention Strategies for Mitigating COVID-19 Pandemic." New England Statistics Symposium (NESS), Hybrid, *Invited session* 05/2022

"Integrative Network Learning for Multi-modality Biomarker Data." Center for Statistical Science, Tsinghua University, *Invited talk* 11/2021

"Identifying Temporal Pathways Using High-Dimensional Biomarkers." Joint Statistical Meetings (JSM), Virtual, Topic-contributed session 08/2021

"Evaluating Effectiveness of Public Health Intervention Strategies for Mitigating COVID-19 Pandemic." Columbia University Data Science Day, Oral poster session 04/2021

"Integrative Network Learning for Multi-modality Biomarker Data." Department of Biostatistics and Computational Biology & Del Monte Neuroscience Institute, University of Rochester, *Invited talk* 01/2021

"Integrative Network Learning for Multi-modality Biomarker Data." Division of Biostatistics, Department of Public Health Sciences, University of Virginia, *Invited talk* 10/2020

"Survival-Convolution Models for Predicting COVID-19 Cases and Assessing Effects of Mitigation Strategies." Data Science Conference on COVID-19, Presentation session 08/2020

"Conditional Gaussian Graphical Model for Estimating Personalized Disease Symptom Networks." JSM, Virtual, Topic-contributed session 08/2020

"Integrative Network Learning for Multi-modality Biomarker Data." Eastern North American Region (ENAR), Virtual, Topic-contributed session 03/2020

"Statistical Methods for Constructing Heterogeneous Biomarker Networks." Division of Biostatistics, Department of Population Health, New York University School of Medicine, *Invited talk* 11/2019

"Integrative Network Learning for Multi-modality Biomarker Data." ICSA Applied Statistics Symposium, Raleigh, NC, *Invited session* 06/2019

"Estimating Heterogeneous Biomarker Networks and Their Effects on Disease Outcome." JSM, Vancouver, Canada, Topic-contributed session 07/2018

"Learning Subject-Specific Directed Acyclic Graphs (DAGs) from High-Dimensional Biomarker Data." Conference on Statistical Learning and Data Science (SLDS), New York, NY, Poster session 06/2018

"Learning Subject-Specific Directed Acyclic Graphs (DAGs) from High-Dimensional Biomarker Data." ENAR, Atlanta, GA, Poster session 03/2018

"High-dimensional Subject-Specific Network Analysis for Disentangling Genetic Mutation-Phenotype Pathways." ICHPS, Charleston, SC, Poster session 01/2018

CAREER DEVELOPMENT

Neuroimaging Short Courses Harvard University, Martinos Center for Biomedical Imaging

- FreeSurfer Course 04/2017
- Structural and Functional Connectivity via MRI 10/2016