LONGTIAN SHI

+86 15639073546 ◇ Shenzhen, Guangdong, China
shilt2022@mail.sustech.edu.cn; LinkedIn; Personal Website

EDUCATION

Southern University of Science and Technology(SUSTech), China **SEP 2022 - Expected JUN 2026**

- **Major**: **Statistics(Supervised by Chair Professor Qi-Man Shao)**
- **Major GPA**: **3.91/4.00 (2/44)** with **Relevant Coursework**: Topics in Probability and Statistics(100, PhD Level), Statistical Learning(100), Python Program(100), Time Series Analysis(99), Mathematical Statistics(99), Statistical Linear Model(98), Nonparametric Statistics(97), Bayesian Statistics(95), Multivariate Statistical Analysis(94), Mathematical Analysis III(90), Advanced Linear Algebra(90)
- Minor: Finance (Minor GPA: 3.88/4.00). Financial Investment(96), Marketing(95), Economics(94), Management Information System(94), Accounting(94), Advanced Operations Research

Summer Session, University of California, Davis **Expected JUN 2025 - AUG 2025**
Coursework: ECN 140 Econometrics and MAT 127C Real Analysis(100, A+)

PUBLICATIONS&MANUSCRIPTS

1. **Shi, L.**, Shi, Y., Fu, Y., Jiang, F., Ma, Y. (2025+). ***The Prize Premium in Publishing Timelines***. Under Review of the *Journal of Informetrics (JOI)*.
2. Zhang, X.†, **Shi, L.**†, Zhao, H. (2025+). ***A Novel Empirical Bayes Method for Genetic Fine-mapping with GWAS Summary Statistics***. Manuscript to be submitted.
   †These authors contributed equally to this work.

RESEARCH EXPERIENCE

Research on Debiased Machine Learning and Semi-parametric Inference **FEB 2025 - PRES**

- Supervised by Assistant Professor Molei Liu, Peking University. Collaborating with Dr. Doudou Zhou.
- **Responsible for theoretical derivation, simulation investigation, and real data application into how kernel smoothing could be used in statistical inference of debiased/double machine learning**. Parameters of interest include TPR, FPR, ROC and AUC, accommodating three data sources: human-labeled (gold standard), AI-labeled surrogates, and an unlabeled target set, under mild assumptions and conditions. Via replacing an indicator function as a regularized kernel function, and other debiasing techniques like Neyman Orthogonality and cross fitting, we successfully showed that the debiased cross-fitting estimators are $\sqrt{n}$-consistent under the transfer learning setting with multiple covariate shifts between datasets. The doubly robustness is also illustrated theoretically. In the simulation, the density ratio is estimated based on posterior probabilities after sample splitting. Our framework will address AI-driven problems like how to evaluate the predictive performance of AI tools as surrogates for gold standard labels in Electronic Health Records(EHR).

Project on Applying Statistical Learning Methods on sedaDNA Datasets **DEC 2024 - PRES**

- Led by Professor Rasmus Nielsen at Department of Statistics, University of California, Berkeley.
- Responsible for implementing high-dimensional sparse PCA, multiple testing correction, linear mixed model, and other novel regularization methods like uniLasso, on **sedaDNA**(Sedimentary Ancient DNA) allele frequency data and the environmental metadata, working to uncover evolutionary patterns and to predict future environmental changes. The population bias is corrected by including the principle components in the regularization methods. After tracing back to the BAM files and matching the identified SNPs to the organisms, several species, mostly bacteria, were found whose genetic variants could probably be caused by the mean annual temperature and annual precipitation.

Research on Empirical Bayes Methods for Genetic Fine-mapping **JUN 2024 - SEP 2024**

- Supervised by Professor Hongyu Zhao, Department of Biostatistics, Yale University. On-site internship.
- Responsible for developing **E**mpirical-**B**ayes for **F**ine **M**apping, namely **EBFM**, a biostatistical method for enhancing identification of diseasing-causing genetic features using GWAS summary statistics. The proposed method EBFM utilizes the **spike-and-slab prior and posterior maximization** for estimating the genetic architecture and then captures the Credible Sets of SNPs based on the PIP(Posterior Inclusion Probability) via **greedy search**. The greedy search is constructed on the correlation score and is rectified based on this particular setting. EBFM is more powerful with a lower FDR(False Discovery Rate) by capturing more credible sets with fewer SNPs in each of them, yielding a higher replication rate, precision-recall rate, reproduction rate, etc., in both simulation and real-data studies via the data of European's and African's BMI, UK Biobank and 1KGP. The simulation and real-data applications(BMI data of different ancestries) aim to compare EBFM's performance with existing popular methods such as SuSiE and CARMA.

Projects on Applied Causality in Computational Social Science **AUG 2023 - APR 2025**
Department of Statistics and Data Science, SUSTech, Shenzhen, China

- Held by Assistant Professor Yifang Ma in SUSTech. Two manuscripts have been produced - one submitted to JOI and the other is ready for submission. We mainly research on how prizes can influence the review time and citation style of winners' publications and how to conduct statistical inference on winners' citation networks.
- Responsible for data curation and analysis of resources on OpenAlex. I also conducted the Fixed Effect regression analysis, DID Event Studies, and synthetic control methods, for several causal models of interest. Conceptualization of variables like academic experience is included as well. Some results of our works show that in many academic fields, significant reduction of review time of winners' works is discovered between before and after the prize, compared with their coauthors. The review time is also found to be negatively correlated with several numerical variables, including the academic expereicne and citation number of the publication. Finally, journal prestige is a vital indicator of such reduction. The citation style, on the other hand, shows similar patterns.

SKILLS&LANGUAGES

Coding: Python&R(Specialized), SQL(Proficient); Competent in LaTeX, STATA and Linux
English: Fluent, TOEFL(109, Speaking 24), GRE in progress. **Chinese(Mandarin):** Native

COMPETITIONS&AWARDS&HONORS

Gold Medal for iGEM(International Genetically Engineered Machine) Competition **2024**
Second Prize of The Chinese Mathematics Competitions **2023**
Second Prize of China Undergraduate Mathematical Contest in Modelling **2023**
National Second Prize of National English Competition for College Students 2023&2024
First Prize(<1%) of University Merit Student Scholarship **2023&2024**
University Student Elite(<0.25%), Outstanding Student & Student Leader(<1%) 2023&2024
University Top 10 Volunteer Candidate (Annually Over 100 Hours of Volunteering) 2023

EXTRA-CURRICULAR ACTIVITIES

President of the Students' Union and College Peer Tutor of Shuli College **MAR 2023 - SEP 2024**
The First Level Athletic of Land Rowing in China SEP 2024
Member of College Basketball Team SEP 2022 - PRES
Hobbies: Photography, Music, Billiards, Basketball, Table Tennis, Movie.