

# Lin Gui

5747 S. Ellis Avenue, Jones 203/204, Chicago, IL, 60637  
Email: [ybnxb@gmail.com](mailto:ybnxb@gmail.com); Website: <https://gl-ybnxb.github.io/>

## EDUCATION

- Ph. D. in Statistics, The University of Chicago, Chicago, USA 2020- Present
- M.S. in Statistics, The University of Chicago, Chicago, USA 2018-2020
- B.S. in Statistics, University of Science and Technology of China, Hefei, China 2014- 2018

## RESEARCH INTERESTS

- Statistical Inference and Multiple Testing
- Concept Control for Score-based Generative Models
- Causal Inference and Machine Learning
- Alignment for Large Language Models

## CODING SKILLS

- R, Python, MATLAB, SQL; PyTorch, Numpy, Pandas

## PUBLICATIONS

- **Aggregating Dependent Signals with Heavy-Tailed Combination Test**  
Lin Gui, Yuchao Jiang, Jingshu Wang  
*Preprint*
- **Concept Algebra for Score-Based Conditional Models**  
Zihao Wang, Lin Gui, Jeffrey Negrea, Victor Veitch  
*NeurIPS 2023*
- **Causal Estimation for Text Data with (Apparent) Overlap Violations**  
Lin Gui, Victor Veitch  
*ICLR 2023*
- **Detecting Multiple Replicating Signals using Adaptive Filtering Procedures**  
Jingshu Wang, Lin Gui, Weijie J. Su, Chiara Sabatti, Art B. Owen  
*The Annals of Statistics 50 (4), 1890-1909*

## RESEARCH

- **A Theoretical and Practical Analysis of the Heavy-Tailed Combination Test for Global Test with Correlated Hypotheses**
  - Undertook comprehensive theoretical evaluations to decipher the intricacies of the state-of-the-art Cauchy combination test and its expansion, termed the heavy-tailed combination test, tailored for the global test with correlated hypotheses.
  - Conducted empirical studies, offering a general practical guideline for the Cauchy/heavy-tailed combination test
  - Enhanced the heavy-tailed combination test into a multiple testing procedure adept at controlling the family-wise error rate (FWER) and introduced a shortcut for this closed testing procedure.
  - Implemented the method on genetic data to address real-world challenges.
  - Estimating the tail probability of the summation of some heavy-tailed random variables with more general correlation structures to confirm the validity of the heavy-tailed combination test in more realistic scenarios. (Ongoing)

- **Concept Algebra for Score-Based Conditional Models**
  - Established a mathematical framework linking representation structures with concepts in text-driven generative models.
  - Demonstrated that the Stein score of the text-controlled distribution is an arithmetically composable representation of the input text.
  - Developed concept algebra as a technique for manipulating the concepts expressed by the model through algebraic manipulation of this representation
  - Showcased this methodology with diverse examples manipulating various concepts.
- **Causal Estimation for Test Data with (Apparent) Overlap Violations**
  - Formulated a formal causal estimand tailored to the causal inference of the text-attribute question, verifying its identifiability under minimal conditions.
  - Provided a computationally efficient estimation of the uncertainty quantification of this causal estimand, supported by theoretical assurances.
  - Conducted empirical evaluations to assess the performance of this estimation approach.
- **Detecting Multiple Replicating Signals using Adaptive Filtering Procedures**
  - Introduced an innovative multiple testing procedure that enhances detection power by adaptively filtering out unlikely candidates of PC nulls.
  - Theoretically established the control of both Family-Wise Error Rate (FWER) and False Discovery Rate (FDR) for this method.
  - Illustrated the application of this AdaFilter method with three case studies: microarray studies of Duchenne muscular dystrophy, single-cell RNA sequencing of T cells in lung cancer tumors and GWAS for metabolomics.

## **CONFERENCES AND PRESENTATIONS**

- **ICML 2023 Workshop SPIGM and SCIS** Jul. 2023
  - Concept Algebra for Score-Based Conditional Models
- **ICLR 2023** May. 2023
  - Causal Estimation for Text Data with (Apparent) Overlap Violations
- **2021 Joint Statistical Meetings** Aug. 2021
  - Detecting Multiple Replicating Signals Using Adaptive Filtering Procedures

## **HONORS & AWARDS**

- Nominee, The 37th. Guo Moruo Scholarship (The highest honor at USTC) 2017
- Winner, Outstanding Student Scholarship, USTC 2016-2017
- Winner, China National Scholarship, USTC 2015