Lin Gui

5747 S. Ellis Avenue, Jones 203/204, Chicago, IL, 60637

Email: ybnbxb@gmail.com; Website: https://gl-ybnbxb.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

- Alignment for Large Language Models

- Causal Inference and Machine Learning

- Statistical Inference and Multiple Testing

- Controllable Generative Models

PUBLICATIONS

- **Aggregating Dependent Signals with Heavy-Tailed Combination Test**

  Lin Gui, Yuchao Jiang, Jingshu Wang

  Preprint

  - *We undertook comprehensive theoretical and empirical evaluations to decipher the intricacies of the state-of-the-art Cauchy combination test and its extension, termed as the heavy-tailed combination test, tailored for the global test with correlated hypotheses. On top of that, we offered a general practical guideline about when the method should be applied and how much power gain can be expected.*

- **Concept Algebra for Score-Based Text-Controlled Generative Models**

  Zihao Wang, Lin Gui, Jeffrey Negrea, Victor Veitch

  Neural Information Processing Systems (**NeurIPS**), 2023

  - *We established a mathematical framework linking representation structures with concepts in text-driven generative models. We demonstrated that the Stein score of the text-controlled distribution is an arithmetically composable representation of the input text, and developed concept algebra as a technique for manipulating the concepts expressed by the model through algebraic manipulation of this representation.*

- **Causal Estimation for Text Data with (Apparent) Overlap Violations**

  Lin Gui, Victor Veitch

  International Conference on Learning Representations (**ICLR**), 2023

  - *We formulated a formal causal estimand tailored to the causal inference of the text-attribute question, and verified its identifiability under minimal conditions. We provided a computationally efficient estimation of the uncertainty quantification of this causal estimand, supported by theoretical assurances.*

- **Detecting Multiple Replicating Signals using Adaptive Filtering Procedures**

  Jingshu Wang, Lin Gui, Weijie J. Su, Chiara Sabatti, Art B. Owen

  The Annals of Statistics (**AOS**) 50.4 (2022), 1890-1909

  - *We introduced an innovative multiple testing procedure that enhances detection power by adaptively filtering out unlikely candidates of PC nulls, and theoretically established the control of both Family-Wise Error Rate (FWER) and False Discovery Rate (FDR) for this method.*

RESEARCH (ONGOING)

- **Human Preference Alignment for Large Language Models**
  - Developing methods to improve large language models aligned with human preference
  - Designing better reward models and investigating the theoretical and empirical performances
- **A Theoretical and Practical Analysis of the Heavy-Tailed Combination Test for Global Test with Correlated Hypotheses**
  - 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.

CODING SKILLS

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

CONFERENCES AND PRESENTATIONS

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

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

TEACHING EXPERIENCE

- STAT 22000: Statistical Methods and Applications   Winter 2021, Spring 2021, Autumn 2021
- STAT 27420: Introduction to Causality with Machine Learning   Autumn 2022
- STAT 24630: Causal Inference Methods and Case Studies   Spring 2022