

# Yu “Hugh” ZHENG

352-328-4299 | seanzhengyu1@gmail.com | Gainesville, FL, 32608  
[LinkedIn](#) | [Github](#) | [Personal Website](#)

## EDUCATION

### University of Florida

*Ph.D. in Statistics (Expected to graduate in 2025)*

Gainesville, FL

*Aug. 2021 - present*

- Advisor: Leo L. Duan
- GPA: 3.95 / 4 (Rank: 1 / 11)

### University of Science and Technology of China (Special Class for the Gifted Young)

Hefei, CN

*Bachelor of Science, Mathematics and Applied Mathematics*

*Aug. 2016 - May. 2020*

*Minor in Business Administration*

## SKILLS AND STANDARD TESTS

**Programming Languages:** R (Advanced), Python (Advanced), C/C++ (Advanced)

**Skills and Tools:** SQL, Stan, Git,  $\text{\LaTeX}$ , TensorFlow, Pytorch, JAX

**Languages:** Chinese (Native), English (Fluent), Spanish (Conversational)

## AWARDS

### Mendenhall Award for the best first-year student

Gainesville, FL

*University of Florida*

*Jan. 2023*

### Katzoff Fellowship Award (\$110,000 funds to cover stipend and tuitions)

Gainesville, FL

*University of Florida*

*Jun. 2023 - Dec. 2024*

## WORK EXPERIENCE

### Quantitative Research Intern at Susquehanna International Group, LLP

Jun. 2024 - Aug. 2024

- Conducted in-depth quantitative analysis to develop high-performing features and build profitable trading strategies based on Python, achieving a top-three ranking in single runs during the blast competition of interns.
- Gained experience with financial instruments, market structures, equity options trading, machine learning models, and the application of quantitative methods in financial markets.

## PREPRINT PUBLICATIONS

**Yu Zheng**, and Leo L. Duan. (2023). Blocked Gibbs Sampler with Anti-correlation Gaussian Data Augmentation, with Applications to L1-ball-type Models. [arXiv](#), under review at *Journal of Computational and Graphical Statistics*, June 2024.

**Yu Zheng**, Leo L. Duan, and Arkaprava Roy. (2024). Consistency of Graphical Model-based Clustering: Robust Clustering using Bayesian Spanning Forest. [arXiv](#), under review at *Annals of Statistics*, September 2024.

## RESEARCH PROJECTS

### Clustering Consistency with Forest Process as the Pseudo Posterior

Aug. 2023 - Present

- Developed theories of Bayesian clustering consistency with the forest process model as the pseudo-posterior, demonstrating a feasible approach to bypassing the need for a completely correct specification of the mixture component distribution.
- Extended traditional research by proving posterior consistency not only for the number of clusters but also for partitions, significantly broadening the applicability of the forest process model in clustering analysis.
- Developed a new mathematical technique called *refinement*, offering a novel tool for the theoretical development of asymptotics in Bayesian clustering analysis, with potential implications for large-scale data clustering in financial and other applied settings.

### Anti-correlation Gaussian Data Augmentation ([Github link](#))

Jan. 2023 - Jul. 2023

- Proposed an innovative data augmentation algorithm in the L1-ball-type model, achieving a 30-fold increase in the effective sample size per computing time.
- Proved the geometric ergodicity of the proposed Gibbs sampler, the first such proof for models with a spike-and-slab-type prior, which has broad implications for improving the efficiency of Bayesian inference in high-dimensional spaces.
- Implemented the algorithm using R and Python, integrating the Gibbs sampler with applications in both computational statistics and quantitative finance.

## Neural Network for Alzheimer's Disease Diagnosis ([Github link](#))

May 2022 - Dec 2022

- Developed a novel neural network architecture combining Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Variational Autoencoders (VAEs) with Bayesian modeling for fMRI-based diagnosis of Alzheimer's disease.
- Improved model interpretability and accuracy by incorporating Bayesian uncertainty estimates, providing a potential framework for use in other medical imaging and financial predictive models.

## TEACHING EXPERIENCE

---

### Instructor at the University of Florida

Jan. 2023 - Apr. 2023

- STA 3024: Introduction to Statistics II ([Course material](#))
- Held lectures that help 100+ students make progress in utilizing statistical tools (such as ANOVA, nonparametric methods, contingency tables, and linear and logistic regression) to analyze and quantify real-world problems.

### Teaching Assistant at the University of Florida

Aug. 2021 - Dec. 2022

- STA 6275: Optimization; STA 6166: Statistical Methods in Research I; STA 3032: Engineer Statistics.

### Teaching Assistant at the University of Science and Technology of China

Aug. 2018 - May. 2020

- Mathematical Statistics; Mathematical Analysis B1; Mathematical Analysis B2; Multivariate Calculus.

## CONTRIBUTED PRESENTATIONS

---

### 2023 Joint Statistical Meetings — *Toronto, Canada*

Aug. 2023

- Talk title: Blocked Gibbs Sampling for L1-ball Priors: an Efficient Computation for Structuredly Sparse Models.

### 2024 Theory and Foundations of Statistics in the Era of Big Data — *Tallahassee, Florida*

Apr. 2024

- Gibbs Sampling using Anti-correlation Gaussian Data Augmentation, with Applications to L1-ball-type Models.

## REFERENCES

---

### Dr. Leo L. Duan

- Assistant Professor, Department of Statistics, University of Florida
- Phone: 352-294-1827
- Email: [li.duan@ufl.edu](mailto:li.duan@ufl.edu)

### Dr. Arkaprava Roy

- Assistant Professor, Department of Biostatistics, University of Florida
- Phone: 352-294-5924
- Email: [arkaprava.roy@ufl.edu](mailto:arkaprava.roy@ufl.edu)