

# Liyi Zhang

2900 Broadway, New York, NY 10027 | lz2574@columbia.edu | (0086)137-0198-0337

## EDUCATION

**Columbia University, New York, NY**

*Columbia College Class of 2021*

GPA: 3.85 on 4.0 scale | *Double Majors in Statistics and Applied Mathematics*

- *Relevant past coursework*
  - *Math*: Modern Analysis I, Analysis and Optimization, Linear Algebra, Calculus III & IV, ODE
  - *Statistics*: Probability Theory, Statistical Inference, Bayesian Statistics, Stochastic Processes, Linear Regression Models, Statistical Computing and Intro to Data Science
  - *Computer Science*: Data Structures, NLP, Optimization for Machine Learning
- *Relevant current coursework*
  - Modern Analysis II, Foundations of Graphical Models, Seminar in Applied Mathematics
- *Other academic certifications*
  - Machine Learning – Professor John Paisley, ColumbiaX on edX
  - Neural Networks and Deep Learning; Hyperparameter Tuning, Regularization, and Optimization; Convolutional Neural Networks – Professor Andrew Ng, deeplearning.ai on Coursera

## PUBLICATION

Antonio Moretti, **Liyi Zhang**, Itsik Pe'er. **Variational Combinatorial Sequential Monte Carlo in Bayesian Phylogenetic Inference**. *Machine Learning in Computational Biology (MLCB)*, 2020, **Oral Presentation**.

## RESEARCH EXPERIENCE

**Columbia Dept. of Computer Science – The Pe'er Lab**

*Undergraduate Researcher | Jan. 2020 – present*

- Build probabilistic model for phylogenetic inference by sampling on discrete tree spaces with Combinatorial Sequential Monte Carlo, and using sampling computations with Variational Inference for model learning, which is a novel approach. This method explores higher probability spaces on the primates dataset, and samples meaningful phylogenetic trees. Python and TensorFlow are used.

**Columbia Dept. of Statistics – Professor Andrew Gelman**

*Undergraduate Researcher | May 2020 – present*

- Develop MCMC method in R and Stan to enable sampling in discrete spaces in phylogenetic models by isolating discrete models and using stacking of predictive distributions for model combination.

## PROJECT EXPERIENCE

**Foundations of Graphical Models – Professor David Blei**

*In-Class Final Project | Dec. 2020 – present*

- Implemented a Bayesian Neural Network (BNN) using TensorFlow with customized layer and training. Test robustness of BNN by implementing adversarial perturbations of the MNIST image data and compare performance across BNN and seven deep learning models.

**Optimization for Machine Learning – Professor Satyen Kale**

*In-Class Final Project | Oct. – Nov. 2019*

- Implemented Stochastic Variance Reduced Gradient Descent to optimize non-convex loss functions, in a team of three. Rewrote SVRG in Python and compared its performance with Stochastic Gradient Descent, using two and three-layer neural nets and the MNIST, CIFAR10, STL, and FashionMNIST datasets.

## CAREER EXPERIENCE

**QTG Capital Management, Shanghai, China**

*Quantitative Researcher Intern | June – August 2019*

- Performed analysis and cleaning on datasets on financial products, conducted clustering analyses using the K-Means algorithm for risk control, built factor models and performed feature engineering on financial and commodities futures to develop and back-test trading strategies.

## SKILLS AND INTERESTS

**Skills:** Python (+TensorFlow), R, Stan, LaTeX, Java

**Language:** Mandarin Chinese: native; English: fluent; Latin: intermediate reading ability

**Interests:** reading and writing, such as introducing math in machine learning on Medium; playing piano and clarinet, currently working on Bach's Goldberg Variations; distance running, hiking, and biking