

# Yiwei Gong

Tel: 9842098759 Email: [yiwei.gong@utexas.edu](mailto:yiwei.gong@utexas.edu) LinkedIn: [www.linkedin.com/in/yiwei-gong-47b806197](https://www.linkedin.com/in/yiwei-gong-47b806197)

## SUMMARY

PhD candidate with strong Bayesian modeling and scalable inference skills. Proficient in machine learning and predictive analytics. Experienced with hierarchical and dynamic models with projects in clustering, biostatistics, and financial data analysis using Python and R. Currently focusing on Bayesian methods for time series and network analysis.

## EDUCATION

**PhD in Statistics and Data Sciences**, The University of Texas at Austin, TX, **GPA 3.99** Aug 2020 – May 2026 (Expected)

**Master of Statistical Science**, Duke University, NC, **GPA 3.9** 2018 - 2020

**BSc in Economics and Mathematics**, University of Bristol, UK, **First Class Honours** 2015 - 2018

Interdisciplinary Contest in Modelling Meritorious Winner (**international contest, Top 10%**).

## SKILLS

**IT SKILLS:** Python, R; **LANGUAGE SKILLS:** Chinese (**Native**), English (**Full Professional**), German (**Intermediate**)

## WORK EXPERIENCE

**Oncology Bioinformatics Summer Intern, Genentech**, San Francisco CA May 2023- Aug 2023

- Proposed and executed a cutting-edge scalable Bayesian method for tumor mutation phasing by employing hierarchical modeling and variational inference.
- Achieved a computation speed 100-fold faster than the baseline model while maintaining robust model performance across both simulated and real-world datasets.
- Improved adaptability to intricate scenarios, leading to enhanced neoantigen targeting.

**Data Scientist Research Intern/ Duke CFAR Annual Conference Presenter**, Durham, NC May 2019-March 2020

- Inferred the diverse immunogenetic characteristics of B cell repertoires in HIV-1 uninfected individuals; extracted B-Cell DNA fragments, subsequently identifying recombination gene segments and other stochastic components.
- Investigated the interrelationships among these stochastic components, employing both Newton-optimization based MLE and Bayesian methodologies; novel discovery regarding the distribution of the insertion region.
- Developed a comprehensive model for the VDJ recombination process; constructed a hierarchical Bayesian network to estimate the occurrence probability of specific target sequences.

**Machine Learning Research Assistant**, Bristol, UK May 2018 - Aug 2018

- Simulated binary datasets and plotted ROC using Python, illustrated the inefficiency of majority voting.
- Curated, cleaned, and summarized the large, sparse Yelp Dataset, provided degree distributions and local graphs, utilized sparse SVD and t-SNE for high-dimension embedding, clustering and data visualization.

**Teaching Assistant**, multiple locations Jan 2018 - Present

Courses include statistics, data science, machine learning, optimization etc. Lead weekly labs and tutorial sessions.

## PROJECTS, TEAMWORK AND LEADERSHIP EXPERIENCE

**Ongoing Research Elective 1: Detect state changes in brain neuronal networks**, Austin, TX

- Employed forward filtering backward sampling method (FFBS) to learn the time-varying vector autoregression process.
- Detected and visualized channel connectivity and state change points using recurrent switching linear dynamical system (**Presented** at WiML 2022 (NeurIPS affinity workshop)).
- Actively developing a new flexible and interpretable factorial hidden Markov model (FHMM) and proposing a variational Bayes filter leveraging recurrent neural networks (Pytorch) to infer latent states and learn the parameters.

**Ongoing Research Elective 2: Studying the Latent Geometry of Network Data**, Austin TX

Calculating the Ollivier-Ricci curvature to determine the latent manifold of observed network data, designing a posterior predictive checking method and exploring the topological structure of networks.

**Graduate Research Assistant**, Austin, TX

Cleaned and analyzed more than 5 million entries of university-wise asymptomatic covid test data. Formulated a Gaussian mixture model for clustering test results and imputed highly censored data using the MCMC and EM algorithms; analyzed Covid test data, visualized the clusters and tracked the trajectories of multiple test gene counts.

**Stock Index Data Analysis during Pandemics**, Durham, NC

Applied time-varying vector AR processes to analyze major stock indices in three pandemic periods, practiced FFBS for posterior parameter inference with forecasting; latent factor PCA to evaluate the stochastic volatility.