

JOY Z. ZHANG

U.S. Permanent Resident; Email: zz547@cornell.edu; Phone: 603-339-2958

SUMMARY

Applied Mathematics PhD with dual minors in **Applied Statistics** and **Computer Science**. Specializes in **Bayesian Hierarchical Modeling** and **Non-parametric Regression** to extract insights from noisy biological data. Expert in **R** and **Python** package development, **simulation design**, and **algorithm implementation**, with a strong track record of publication-quality outputs. Technical skillset includes **Deep Learning** and **Reinforcement Learning**, passionate about bridging the gap between classical biostatistics and modern data science.

EDUCATION

Cornell University

Ph.D. student in **Applied Mathematics** advised by **Daniel Kowal**,

Minors in **Applied Statistics** and **Computer Science**

Expected: May 2027

Key graduate coursework (2+ years): Bayesian Statistics & Data Analysis; Theory of Statistics; Linear Models in Statistics; Statistical & Computational Genetics; Reinforcement Learning; Analysis of Algorithms; Optimal Control.

Swarthmore College

Bachelor of Arts in **Mathematics** and **Philosophy with Honors**

Graduated: May 2022

Key courses: Programming in Python, Linear Algebra, Differential Equations, Real Analysis, Complex Analysis, Abstract Algebra, Topology, Algebraic Geometry, Combinatorics, Probability Theory, Topological Data Analysis.

SKILLS

Programming: R, Python; R package development

Biostatistics: Bayesian modeling, generalized linear models, causal inference, missing data

Methods: Simulation studies (power/type I error), bootstrap, Monte Carlo / MCMC, posterior predictive checks, model diagnostics; reproducible reporting

Machine learning & tools: deep learning, reinforcement learning; GitHub, LaTeX, VS Code, HPC

RESEARCH EXPERIENCE

Research Assistant | *Cornell University*

2022–present

Project 1: Semiparametric Bayesian Regression for Support-Aware Outcomes in $[0, 1]$

2025–present

- **Extended semiparametric regression frameworks (SeBR)** with support-aware Monte Carlo algorithms for zero/one-inflated and censored biological outcomes.
- **Designed large-scale simulation studies in R on HPC** to assess operating characteristics (bias, calibration/coverage, robustness) under correlated features.
- **Developed scalable end-to-end workflows** spanning data simulation, statistical inference, validation, and visualization.
- **Built modular, reproducible workflows** for data generation, inference, posterior predictive diagnostics, and visualization; produced publication-quality figures.
- **First-authored the project**, independently translating methodological results into clear figures, tables, and written explanations.

Project 2: Bayesian Phylodynamic Inference of Population Dynamics with Dormancy

2024–2025

- **Led the development of Bayesian phylodynamic methods** to jointly infer latent genealogy, dormancy (seedbank) dynamics, and evolutionary parameters from **whole-genome sequencing (WGS) data**.
- **Designed and implemented MCMC-based inference workflows in BEAST2 (SeedbankTree)**, including reversible-jump proposal kernels and state-dependent substitution models.
- **Executed and monitored large-scale multi-chain MCMC analyses on Cornell BioHPC(HPC)**, incorporating convergence diagnostics, posterior summarization, and scalability considerations.
- **Built reproducible genomics analysis pipelines integrating SNP/gene annotations**, simulation validation, and visualization in **R** (e.g., ggtree).

- **Co-first authored the project**, translating methodological advances into interpretable evolutionary and epidemiological insights from real *Mycobacterium tuberculosis* outbreak genomics data.

Project 3: Semialgebraic Hypothesis Testing with Incomplete U-statistics

2023–2025

- **Conducted large-scale simulation studies** to evaluate statistical tradeoffs in the **Sturman–Drton–Leung (SDL) semialgebraic test**, focusing on p-value stability and power near model boundaries and singularities.
- **Implemented and analyzed randomized incomplete U-statistics** with Gaussian multiplier bootstrap, studying the impact of kernel order, partial summarization, and constraint-set design on computational efficiency and test reliability.
- **Designed parameter-sensitivity experiments** across multinomial and trinomial count models, including exhaustive simulation of distinct datasets under multiple sample-size and regime settings.
- **Built reproducible simulation and analysis workflows in R**, with supporting algebraic computations in **Macaulay2**, enabling systematic benchmarking and result validation.
- **Synthesized simulation results into methodological recommendations** and contributed figures and text to the associated manuscript.

Project 4: Geometric Analysis of Admixed Nodes in the Neighbor-Joining Algorithm

2024

- **Developed a computational framework** to quantify how linear admixture affects **Neighbor-Joining (NJ) phylogenetic reconstruction**, evaluating clustering order, distances, and topological path-lengths.
- **Implemented reproducible simulation pipelines using simulated admixed genetic-distance vectors**, comparing **exact probability calculations** with Monte Carlo estimation and NJ benchmarks.
- **Leveraged polyhedral geometry** to analyze NJ algorithm regions and lower-dimensional projections induced by admixture behavior, enabling precise characterization of algorithmic biases.
- **Built and maintained a public NJ module in Macaulay2**, supplemented with Python utilities, supporting large-scale geometric computations and analysis.
- **Generated figures, tables, and documentation** for reproducible experiments and manuscript support, ensuring clarity and usability of the software module.

PREPRINTS AND PUBLICATIONS

Methodological considerations for semialgebraic hypothesis testing with incomplete U-statistics (with David Barnhill, Marina Garrote-López, Elizabeth Gross, Max Hill, Bryson Kagy, and John A Rhodes), *submitted*, 2025.

Bayesian phylodynamic inference of population dynamics with dormancy (Lorenzo Cappello (co-first author), Wai Tung Lo (co-first author), **Joy Z. Zhang (co-first author)**, Daniel Barrow, Ishani Chopra, Perry Xu, Shenni Liang, Andrew G. Clark, Martin T. Wells, and Jaehee Kim), published in *Proc. Natl. Acad. Sci. U.S.A (PNAS)*, 122 (18) e2501394122, 2025.

Geometric analysis of admixed nodes in the neighbor-joining algorithm (with Wai Tung Lo, Michael Stillman, and Jaehee Kim), *submitted*, 2024.

Friezes over $\mathbb{Z}[\sqrt{2}]$ (with Esther Banaian, Libby Farrell, Amy Tao and Kayla Wright), published in *Involve, a Journal of Mathematics*, 2025.

Topology of augmented Bergman complexes (with Elisabeth Bullock, Aidan Kelley, Victor Reiner, Kevin Ren, Gahl Shemy, Dawei Shen, Brian Sun and Amy Tao), published in *Electronic Journal of Combinatorics*, 2021.

Frobenius templates in Certain 2x2 matrix rings (with Timothy Eller, Jakub Kraus, and Yuki Takahashi), published in *International Journal of Mathematics and Computer Science*, 2020.

Graphs with prescribed radius, diameter, and center (with Kelly Guest, Andrew Johnson, Peter Johnson, William Jones, and Yuki Takahashi), published in *International Journal of Mathematics and Computer Science*, 2020.