Thanasi Bakis

abakis@uci.edu | GitHub: thanasibakis | Irvine, CA

Education

University of California, Irvine, Ph.D., Statistics (Advisor: Volodymyr Minin) 2022-Present

University of California, Irvine, M.S., Statistics 2020-2022

University of California, Irvine, B.S., Data Science (summa cum laude) 2017-2020

Selected coursework:

- Bayesian Statistics (Stan)
- Data Structures (C++)
- Database Design (PostgreSQL)
- Deep Generative Models (PyTorch)
- Generalized & Longitudinal Linear Models (R)
- Machine Learning (Python)

Publications

Bakis, A. G., & Minin, V. M. (In progress). *Predictive variational inference for probabilistic forecasts of hospitalization time series.*

- Developing a model-agnostic inference framework for probabilistic forecasting of time series, inspired by recent advances in variational inference.
- Applying the method to widely-used forecasting models, enabled by the **ForwardDiff.jl** automatic differentiation library in **Julia**.

Bakis, A. G., Vora, A. A., Araki, T., Jia, T., Galloway, J. G., Jennings-Shaffer, C., … & Minin, V. M. (2025). *Bayesian inference of antibody evolutionary dynamics using multitype branching processes.* arXiv preprint arXiv:2508.09519.

- Implemented a novel branching process model in **Julia** to jointly analyze genetic sequence data from 3,700+ B cells sampled in an immunological experiment—the first application of Bayesian modeling to this large-scale dataset.
- Leveraged state-of-the-art packages **Turing.jl** for Markov chain Monte Carlo-based inference and **DifferentialEquations.jl** for numerical integration.
- Carefully optimized the memory allocation throughout my software to enable parallelization over 80 cores for >1 month of runtime.
- Published all work as an open-source software library & a repository of data analyses to promote reproducibility and extensibility.

Ralph, D. K., **Bakis, A. G.**, Galloway, J., Vora, A. A., Araki, T., Victora, G. D., … & Matsen IV, F. A. (2025). *Inference of germinal center evolutionary dynamics via simulation-based deep learning.* arXiv preprint arXiv:2508.09871.

- Developed a deep learning model in **TensorFlow** for phylogenetic tree data, using the latest research to convert tree data structures to matrix inputs for the neural network without any loss of information.

Work Experience

U.S. Centers for Disease Control and Prevention (CDC), Center for Forecasting and Outbreak Analytics 2024-2025

- Built a pipeline for the iterative development and evaluation of probabilistic forecasting models for the compositional prevalence of SARS-CoV-2 variants, using **Python** and **Polars**.
- Developed a collection of increasingly complex Bayesian multinomial regression models in **NumPyro**, using the pipeline to argue the merits of different model assumptions in various epidemiologically-relevant scenarios.

Stats 170A/B, UC Irvine, Teaching Assistant 2021-2022

- Mentored 15 undergraduate data science teams on end-to-end capstone projects, from data acquisition (**Python**) and database design (**PostgreSQL**) to statistical modeling and visualization (**R**).

iD Tech Camps, Instructor 2018

- Designed and taught an interactive cryptography curriculum in **Python** for middle and high school students.

Honors and Achievements

National Science Foundation, NSF Graduate Research Fellowship 2022
One of 14 Statistics students nationally to secure three years of Ph.D. research funding.

Department of Statistics, UC Irvine, Newcombe Award 2021
Highest qualifying exam score across the Ph.D./M.S. cohort.

CyberPatriot, Air & Space Forces Association, 9th place team nationally, of 2200 2017
Cybersecurity competition for Linux server hardening.