KONSTANTINOS BANOS

Email • LinkedIn Profile • Github Profile • Hamilton, ON, Canada

PERSONAL PROFILE

A self-motivated researcher and Ph.D. candidate in Statistics at McMaster University, with a strong focus on statistical machine learning—particularly probabilistic modeling, Bayesian nonparametrics, natural language processing, and stochastic processes—and broader interests in the theoretical foundations of machine learning and AI. Through my thesis on modeling continuous-time document streams, I developed strong expertise in probabilistic modeling, inference algorithms, and implementation methodologies. My academic excellence and multiple university scholarships demonstrate a sustained commitment to research and innovation, further supported by experience in conference presentations and manuscript preparation.

EDUCATION

Ph.D. Candidate in Statistics Sep 2026 – Present
McMaster University, Hamilton, Ontario, Canada

- Research at the intersection of statistical machine learning, Bayesian nonparametrics, probabilistic modeling, and stochastic processes, with applications in language modeling and quantitative finance.
- Supervisor: Prof. Narayanaswamy Balakrishnan, Distinguished University Professor

Master of Science (M.Sc.) in Statistics Sep 2024 – Aug 2026
McMaster University, Hamilton, Ontario, Canada

- *Thesis: Power-Law Nonparametric Bayesian Models for Continuous-Time Document Streams*
  Supervisor: Prof. Narayanaswamy Balakrishnan, Distinguished University Professor
- GPA: A^+ (First Class with Honours)

Bachelor of Science (B.Sc.) in Statistics and Insurance Science Oct 2019 – Sep 2023
University of Piraeus, Athens, Greece

- *Research Project 1: Likelihood Ratio Order under Exponential Mixtures* (Supervisor: Dr. George Iliopoulos)
- *Research Project 2: Random Censorship* (Supervisor: Dr. George Tzavelas)
- GPA: 9.00/10.00, First Class with Honours (Top 1% of 199 graduates)

AWARDS AND SCHOLARSHIPS

- **McMaster University Graduate Scholarship** ($10,000; 2024–2025): Entrance award at admission to the M.Sc. in Statistics recognizing outstanding academic achievement.
- **McMaster University Departmental Scholarship** ($7,600; 2024–2025): Competitive entrance scholarship awarded upon admission for academic excellence in the Department of

Mathematics and Statistics.

- **McMaster University Research Scholarship** ($8,404.80; 2024–2025): Entrance research award supporting graduate study and early-stage thesis research in statistical machine learning.
- **Academic Excellence Award** (Insurance Union of Greece; €1,000; 2024): National merit distinction recognizing top academic achievement (top 5 student each year). The Insurance Union of Greece identified me as a prospective future partner and included me in its list of outstanding candidates with exceptional academic performance.
- **Academic Excellence Award** (University of Piraeus; €1,500; 2023): Institutional award for graduating with First Class Honours and ranking within the top 3% of the cohort. Included in the University's Dean's List of outstanding academic achievers, promoting exceptional graduates to the professional and research industry.

RESEARCH EXPERIENCE

Graduate Researcher July 2025–Present
Department of Mathematics and Statistics, McMaster University
Advisor: Prof. Narayanaswamy Balakrishnan

- Identified a core limitation of the Dirichlet–Hawkes Process (DHP)—its inability to capture the power-law behavior observed in topic cluster distributions within large-scale document streams.
- Developed the Pitman–Yor–Hawkes Process (PYHP) to address this issue, replacing the Dirichlet prior with a Pitman–Yor process to jointly model text content and temporal dynamics under heavy-tailed clustering.
- Applied the PYHP to a large-scale corpus of over 150,000 news articles from 2017, implementing (in Python) the complete inference pipeline from scratch, including data preprocessing, utility functions, Sequential Monte Carlo algorithms, and visualization tools (e.g., topic word clouds and temporal activity plots). The model successfully identified coherent latent topics and captured realistic bursty dynamics. A complete, documented implementation of the PYHP model will be released on GitHub following thesis submission (code available for review upon request).
- Identified a structural limitation of the PYHP—its tendency to generate multiple clusters for the same topic when an existing cluster becomes inactive—and proposed the Hierarchical Pitman–Yor–Hawkes Process (HPYHP) to resolve this issue through a two-level hierarchical formulation.

Independent Research in Scientific Discovery Sep 2025–Present
Hamilton, ON, Canada

- Conducting independent research on transformer-based symbolic regression frameworks to advance automated scientific discovery.
- Building upon the study *Evaluating K-Fold Cross Validation for Transformer-Based Symbolic*

Regression Models to investigate the impact of k-fold cross-validation (KFCV) on model generalization and robustness under small-data constraints.

- Running extensive simulations to examine how dataset size variation and model dimensionality influence the generalization efficiency and stability of KFCV-applied symbolic regression models.
- Quantifying the computational overhead introduced by KFCV for larger datasets and complex transformer architectures, analyzing trade-offs between generalization performance and computational cost.
- Benchmarking KFCV against alternative small-data strategies—such as data augmentation and ensemble learning—to evaluate relative efficiency, scalability, and potential to accelerate data-driven scientific discovery across domains.

Research in Theoretical Statistic Aug 2022 – Dec 2022
Department of Statistics and Insurance Science, University of Piraeus, Athens, Greece
Supervisor: Dr. George Iliopoulos

- Conducted theoretical research on the likelihood ratio order in mixtures of exponential distributions, focusing on establishing conditions under which the family exhibits the monotone likelihood ratio (MLR) property.
- Extended the analytical results derived for the two-component mixture case ($n = 2$) presented in the departmental working paper to the more complex three-component case ($n = 3$).
- Made substantial progress toward formulating a general characterization of the MLR property for mixtures with arbitrary n, involving detailed derivations of positivity conditions for derivative expressions and monotonicity of the likelihood ratio function.
- This work strengthened my foundation in mathematical statistics, order theory, and distributional analysis, laying the groundwork for my later interest in nonparametric Bayesian modeling and stochastic processes.

Research in Mathematical Statistics and Survival Analysis Sep 2022 – Jun 2023
Department of Statistics and Insurance Science, University of Piraeus, Athens, Greece
Supervisor: Dr. George Tzavelas

- Conducted an analytical and computational study on random censoring in the exponential distribution, focusing on parameter estimation, information loss, and efficiency comparison under censored and uncensored settings.
- Derived the likelihood and Fisher information matrices for both the fully observed and randomly censored exponential models, proving that random censoring systematically reduces asymptotic efficiency.
- Constructed and compared maximum likelihood estimators (MLEs) under both regimes, providing asymptotic confidence intervals and validating the results via Monte Carlo simulations implemented in R.
- Applied the methodology to simulated survival data modeling patient retention times in a new Non-Hodgkin lymphoma treatment, highlighting how censoring affects inference accuracy and confidence interval width.

- Proposed an extension of the study by considering dependence between the lifetime and censoring distributions, a direction for future work that could generalize classical survival models.

PUBLICATIONS

- **K. Banos**. "Power-Law Nonparametric Bayesian Models for Continuous-Time Document Streams." (Manuscript in preparation)
- **K. Banos**. "Extensive Simulation of K-Fold Cross-Validation for Transformer Symbolic Regression Models." (Manuscript in preparation).

PRESENTATIONS

- **K. Banos**. "*Hawkes Processes: A Switching-State Marked Hawkes Extension.*" IBM T.J. Watson Research Center, Yorktown Heights, NY, USA, 2025.
- **K. Banos**. "*Regression Models for Count Data with Excess Zeros: A Comparison Using Survey Data.*" Department of Mathematics and Statistics, McMaster University, Hamilton, ON, Canada, 2025.
- **K. Banos**. "*Likelihood Ratio Order in Mixtures of Exponential Distributions.*" Poster presentation, Department of Statistics and Insurance Science, University of Piraeus, Athens, Greece, 2023.
- **K. Banos**. "*Random Censoring and the Exponential Distribution.*" Department of Statistics, University of Piraeus, Athens, Greece, 2023.

RESEARCH INTERESTS

Mathematical Statistics; Stochastic Processes; Nonparametric Bayesian Methods; Bayesian Statistics; Probabilistic and Statistical Machine Learning; Natural Language Processing; Quantitative Finance.

TEACHING EXPERIENCE

Teaching Assistant 2024 – Present
Department of Mathematics and Statistics, McMaster University, Hamilton, ON, Canada

- Lead laboratory sessions for STATS 2B03: Statistical Methods in Science (over 600 students per term), teaching applied statistical methods and data analysis in R.
- Conduct weekly office hours to provide individual mentoring, support student understanding

of statistical concepts, and assist with data interpretation and coding.

- Contributed to grading assignments, evaluating examinations, and assisting with invigilation for MATH 2FM3 (Financial Mathematics) and STATS 2D03 (Probability and Statistics).

Private Tutor (Freelance Professional) Sep 2022 – Dec 2024
Athens, Greece

- Provided individualized instruction in estimation theory, hypothesis testing, linear algebra, Bayesian statistics, and regression modeling.
- Offered applied training in statistical computing using R, Python, and SPSS, focusing on problem-solving and conceptual understanding.

PROFESSIONAL EXPERIENCE

Research Internship / Summer School – Principled Scientific Discovery Summer 2025
IBM T.J. Watson Research Center, Yorktown Heights, NY, USA

- Selected for a highly competitive two-week program jointly organized by the Simons Laufer Mathematical Sciences Institute (SLMath) and IBM Research, focused on principled methods for scientific discovery.
- Studied advanced topics including formal methods, symbolic regression, theorem proving, and mechanized model derivation, exploring how computational frameworks can accelerate scientific discovery.
- Conducted data-driven experiments using symbolic regression techniques that successfully rediscovered **Kepler's laws of planetary motion**, illustrating the power of AI-assisted scientific discovery.
- Participated in collaborative workshops integrating formal reasoning with data-driven modeling to generate and verify interpretable scientific hypotheses.

Stock Trader 2020–Present
Hamilton, ON, Canada

- Engaged in active day trading and portfolio management across equities, combining behavioral market analysis with quantitative evaluation.
- Apply technical, fundamental, and order-flow analysis (including Level II and book-map analysis) to identify liquidity zones and detect market-maker driven traps.
- Develop short- and long-term trading strategies based on liquidity dynamics and market structure, emphasizing risk control and adaptive decision-making.
- Maintain detailed trading journals, documenting every position and conducting weekly market review sessions analyzing intraday runners and liquidity shifts.
- Practice dynamic investing alongside active trading, managing a diversified portfolio that has achieved consistent annual returns (44% in 2024; 35% year-to-date in 2025).

TEACHING INTERESTS

Mathematical Statistics; Probability Theory; Stochastic Processes; Bayesian and Nonparametric Bayesian Methods; Statistical and Probabilistic Machine Learning; Regression and Generalized Linear Models; Financial Mathematics and Risk Management; Applied Statistical Computing in R and Python.

UNIVERSITY SERVICE

Undergraduate Mentor 2022–2023
Department of Statistics and Insurance Science, University of Piraeus, Greece

Provided academic and personal mentorship to first-year students, advising on course selection, academic planning, and effective study strategies. Assisted students in transitioning to university life and encouraged engagement with departmental research and extracurricular opportunities.

COMMUNITY ENGAGEMENT

Community Support Volunteer 2018–2024
Galataki, Korinthia, Greece

Actively contributed to community service initiatives focused on elderly care and social inclusion. Provided daily assistance, companionship, and logistical support to senior residents, fostering intergenerational connection and well-being within the local community.

TECHNICAL SKILLS

Programming Languages: R, Python, C, MySQL, PostgreSQL, LaTeX

Machine Learning & AI Frameworks: PyTorch, TensorFlow, scikit-learn, XGBoost, Hugging Face Transformers

Statistical & Bayesian Modeling: Stan, PyMC, brms, glmmTMB, NumPyro, JAGS.

Data Science & Visualization Tools: NumPy, pandas, SciPy, Matplotlib, seaborn, Plotly, and tidyverse (R)

Development & Research Tools: Git, GitHub, Jupyter, Google Colab, VS Code, Overleaf, Linux/Unix environment

Statistical Software: SPSS, PSPP, Stata

Productivity & Design Tools: Microsoft Office Suite (Excel, Word, PowerPoint), Adobe Photoshop

Languages: Greek (native), English (fluent; professional academic proficiency)

LETTER OF RECOMMENDATIONS

Available upon request from the following faculty members:

- Prof. Narayanaswamy Balakrishnan — Distinguished University Professor, McMaster University (bala@mcmaster.ca)
- Prof. George Iliopoulos — Professor, Department of Statistics and Insurance Science, University of Piraeus (geh@unipi.gr)
- Prof. Ben Bolker — Professor of Mathematics & Statistics and Biology, McMaster University (bolker@mcmaster.ca)
- Prof. Noah Forman — Associate Professor, McMaster University (formann@mcmaster.ca)

**** Transcripts and/or related coursework available upon request.**