

HOAI NAM NGUYEN

Houston, TX, 77030

832-758-3848

hn17@rice.edu

Linkedin: [nam-nguyen-rice23](#)

RESEARCH INTERESTS

- **Methodology:** Bayesian statistics, survival models, stochastic processes, Machine Learning
- **Application:** Cancer research

EDUCATION

- **PhD. Statistics, Rice University**, Houston, TX May 2024 (Expected)
Advisors: Dr. Wenyi Wang and Dr. Marek Kimmel
Relevant courses: Applied Stochastic Processes, Multivariate Analysis, Statistical Inferences, Statistical Machine Learning
- **MPhil. Finance, University of Cambridge**, Cambridge, UK May 2016
- **BS. Mathematics, Imperial College London**, London, UK May 2015
- **Online courses:** • IBM Professional Certificate in Data Science • Deep Learning Specialization • Machine Learning • Applied Data Science with Python Specialization • Python for Everybody Specialization • Fundamentals of Computing Specialization

TECHNICAL SKILLS

- **Programming languages:** R, Python, SQL, MATLAB, Mathematica, git, LaTeX • **Libraries/Frameworks:** Numpy, Pandas, Matplotlib, Scikit-learn, Pytorch, ggplot2, Rcpp, Shiny • **Software/Platform:** Presto, Snowflake, Github, GitLab, Unix, Tableau, PowerBI

WORK EXPERIENCES

Indeed Inc

Austin, TX

Product Scientist Intern

May 2022 - Aug 2022

- **Project:** Predicting long-term value of advertisers based on short-term data
 - Conducted thorough literature reviews on the topic of customer lifetime value modeling.
 - Developed highly optimized RFM models on **Presto** to predict 18-month future revenues at both segment and advertiser levels.
 - Using data pulled from **Snowflake**, developed Machine Learning models in **Python** and **R** to predict 18-month future revenues of advertisers. The best model was able to handle extreme observations and detect non-spenders with a 91% accuracy.
 - Communicated the results with stakeholders, including high-level management, through both reports and presentations.
 - **Received a full-time offer upon completion of the internship.**

Department of Statistics, Rice University

Houston, TX

Graduate Teaching Assistant

Aug 2019 - Dec 2020

- Teaching assistance for
 - STAT 615: Regression and Linear Models (Fall 2020).
 - STAT 310: Probability and Statistics (Spring 2020).
 - STAT 413: Introduction to Statistical Machine Learning (Fall 2019).

John von Neumann Institute, Vietnam National University

Ho Chi Minh City, VN

Research Fellow

Aug 2017 - Apr 2019

- **Project:** Time series estimation of index volatility in South East Asian (SEA) stock markets
 - Implemented MCMC and particle filtering in **MATLAB** to estimate various variations of Stochastic Volatility and GARCH models.
 - Optimized and parallelized the implementation to reduce computational time by 75%.
 - Applied the fitted models to stock return data in six SEA stock markets to produce Value-at-Risk (VaR) forecasting.
- **Project:** Modeling yield curves for the Hanoi Stock Exchange (HNX)
 - Conducted literature review on high-dimensional anomaly detection algorithms to filter fixed income data.
 - Evaluated the performance of anomaly detection algorithms on HNX fixed income data with over 200,000 entries.
 - Contributed to an interactive desktop app in **MATLAB** to automate the estimation process of yield curves.
- Teaching assistance for
 - Time Series Analysis (Fall 2018).
 - Introduction to Probability and Statistics (Fall 2017).

7Astar Tutoring Center

Ho Chi Minh City, VN

Instructor

Aug 2016 - Apr 2019

- Prepared high school students for standardized exams for entry into US universities.

VNG Corporation

Ho Chi Minh City, VN

Data Analyst Collaborator

Nov 2017 - Jan 2018

- Explored the use of data science and machine learning tools on the prediction of customer lifetime value in the gaming industry.

RESEARCH EXPERIENCES

MD Anderson Cancer Center

Houston, TX

Graduate Research Assistant

May 2020 - present

- **Project:** Personalized risk prediction for cancer survivors via a Bayesian semi-parametric framework (**Advisor: Dr. Wenyi Wang**)
 - Developed a Bayesian risk prediction model that allows for recurrent cancers and competing risks from multiple cancer types.
 - Implemented the model in **R**, with **Rcpp** being utilized for faster performance.

- Performed MCMC on a high performance computing cluster (**Linux**) to estimate model parameters.
- Integrated the new model to the lab's LFSPRO library ([github](#)) and the **Shiny** app ([github](#)).
- Validated the model performance on patient cohorts from multiple research institutions across the US.
- **Project: Validation of risk prediction models on clinical ascertained data (Advisor: Dr. Wenyi Wang)**
 - Evaluated the performance of the lab's LFSPRO library on a clinical family dataset on the basis of AUC and O/E ratio.
 - Implemented interactive visualizations and added functionalities in the **Shiny** app to expedite clinical use.

Department of Statistics, Rice University

Houston, TX

Graduate Research Assistant

Dec 2019 - Jun 2022

- **Project: Mathematical modeling of system biology via multi-type age-dependent branching process (Advisor: Dr. Marek Kimmel)**
 - Derived closed-form expression for the evolution of stem cell population over time under different cell-type-specific dynamics.
 - Performed asymptotic analysis on the probability of extinction and the rate at which the population approaches extinction.
 - Performed simulation in **MATLAB** to confirm theoretical results.

Data to Knowledge (D2K) Lab, Rice University

Houston, TX

Graduate Student Researcher

Aug 2020 - Dec 2020

- **Project: Satellite-based estimation of air pollution response to the COVID-19 pandemic in the US**
 - Implemented a web crawler in **Python** to automatically pull satellite data on the temporal and spatial concentrations of NO₂.
 - Implemented interactive data visualizations in **R/Shiny** to visualize the changes in NO₂ in response to the COVID-19 lockdown.
 - Built Generalized Additive Models (GAM) and broken-stick models to statistically describe the changes in NO₂ over time.

PUBLICATIONS/PREPRINTS

1. **Nam H Nguyen**, Jessica L Corredor, Wenyi Wang et al, *Personalized cancer-type-specific prediction of the second primary malignancy in multi-center cohorts affected by the Li-Fraumeni Syndrome: an external validation study* (in preparation).
2. Jessica L Corredor, Elissa B Dodd-Eaton, Jacynda Woodman-Ross, Ashley Woodson, **Nam H Nguyen** et al, *Performance of LFSPRO TP53 germline carrier risk predictions compared to standard genetic counseling practice on prospectively collected probands* (in preparation).
3. **Nam H Nguyen**, Seung J Shin, Elissa B Dodd-Eaton, Jing Ning, Wenyi Wang, *Personalized risk prediction for cancer survivors: a Bayesian semi-parametric recurrent event model with competing outcomes* (in preparation, **winner of the Best Student Paper Award by the American Statistical Association, 2022**) ([link](#)).
4. **Nam H Nguyen**, Elissa B Dodd-Eaton, Jessica L Corredor et al, *Validating risk prediction models for multiple primaries and competing cancer outcomes in families with Li-Fraumeni syndrome using clinically ascertained data at a single institute*, accepted at the Journal of Clinical Oncology, 2024 ([link](#)).
5. **Nam H Nguyen**, Elissa B Dodd-Eaton, Gang Peng et al, *LFSPROshiny: an interactive R/Shiny app for prediction and visualization of cancer risks in families with deleterious germline TP53 mutations*, JCO Clinical Cancer Informatics, 2024 ([link](#)).
6. **Nam H Nguyen**, Marek Kimmel, *Stochastic Models of Stem Cells and Their Descendants under Different Criticality Assumptions*, Stochastic Models, 2022 ([link](#)).
7. Paul Bui Quang, Tony Klein, **Nam H Nguyen**, and Thomas Walther, *Value-at-risk for south-east asian stock markets: Stochastic volatility vs GARCH*, Journal of Risk and Financial Management, 2018 ([link](#)).

REFERENCES

1. **Wenyi Wang, PhD**
 Professor of Bioinformatics and Computational Biology and Biostatistics, The University of Texas MD Anderson Cancer Center
 Address: T. Boone Pickens Academic Tower, 1400 Pressler Street, Houston, TX, 77030, USA
 Email: wwang7@mdanderson.org
 Phone number: 713-792-5377
 Website: <https://odin.mdacc.tmc.edu/wwang7/>
2. **Marek Kimmel, PhD**
 Professor of Statistics, Department of Statistics, Rice University
 Address: Maxfield Hall 227, Rice University, 6100 Main Street, Houston, TX, 77005, USA
 Email: kimmel@rice.edu
 Phone number: 713-348-5255
 Website: <https://profiles.rice.edu/faculty/marek-kimmel>
3. **Gastón L'Huillier, MS**
 Senior Software Engineering Manager, Indeed Inc
 Address: 1051 E Hillsdale Blvd #800, Foster City, CA, 94404, USA
 Email: glhuillier@indeed.com
 Phone number: 617-418-0212
 LinkedIn: [linkedin.com/in/glhuilli](https://www.linkedin.com/in/glhuilli)