Email: kl411@Duke.edu

Kevin Li

Education

Duke University — Durham, North Carolina
PhD in Statistical Science, GPA: 3.95/4.00 — 2021-2026

Columbia University — New York, New York
B.A. in Mathematics and Statistics, GPA: 3.90/4.00 *Cum Laude (Top 25% of Class)* — 2016–2020

- Sample coursework - Measure theoretic probability, Graduate-Division Linear Algebra, Analysis of Algorithms, Causal Inference, Stochastic Processes, Analysis and Optimization, Numerical Analysis

Publications

[1] **K. Li**, M. Balakirsky, and S. Mak, "Trigonometric quadrature fourier features for scalable gaussian process regression", *To Appear 27th International Conference on Artificial Intelligence and Statistics*, 2023. arXiv: 2310.14544 [stat.ML].

[2] **K. Li** and S. Mak, "Prospar-gp: Scalable gaussian process modeling with massive non-stationary datasets", *Submitted to Journal of Computational and Graphical Statistics*, 2023. arXiv: 2311.08752 [stat.ME].

[3] **K. Li**, S. Mak, J. .-.-F. Paquet, and S. A. Bass, "Additive multi-index gaussian process modeling, with application to multi-physics surrogate modeling of the quark-gluon plasma", *Under Revision at the Journal of the American Statistical Association*, 2023. arXiv: 2306.07299 [nucl-th].

[4] R. Fore, J. Boehme, **K. Li**, J. Westra, and N. tintle, "Multi-set testing strategies show good behavior when applied to very large sets of rare variants", *Frontiers in Genetics*, 2020.

[5] J. Lee, C. Mitelut, H. Shokri, I. Kinsella, N. Dethe, S. Wu, **K. Li**, E. B. Reyes, D. Turcu, E. Batty, Y. J. Kim, N. Brackbill, A. Kling, G. Goetz, E. Chichilnisky, D. Carlson, and L. Paninski, "Yass: Yet another spike sorter applied to large-scale multi-electrode array recordings in primate retina", *bioRxiv*, 2020.

[6] **K. Li** *et al.*, "Implementing and evaluating a Gaussian mixture framework for identifying gene function from TnSeq data ", presented at the 2019 Pacific Symposium On Biocomputing (20% Oral Presentation Acceptance Rate), Jan. 2019, pp. 172–183.

[7] D. Patton, O. Rambow, J. Auerbach, **K. Li**, and W. Frey, "Expressions of loss predict aggressive comments on twitter among gang-involved youth in chicago", *Nature Digital Medicine*, 2017.

Industry and Research Experience

Duke University Statistical Science Department — Durham, North Carolina
Graduate student — August 2021 - Present

- Developed variational inference scheme for estimating probabilistic low dimensional embedding for GP regression. Implemented in Python and Pytorch.
- Developing method to perform scalable non-parametric Instrumental Variable Regression using Interdomain Gaussian Processes and Stochastic Variational Inference.
- Developing scalable state-space model for multivariate time series to forecast the effects of promotional campaigns on various consumer segments. Project in collaboration with marketing research firm 84.51. Implemented in Python/Pytorch.

American Express Credit and Fraud Risk New York, New York
Data Scientist August 2020 - Present

- Leverage novel data sources and statistical methods to drive risk management strategy
- Conduct counterfactual causal inference to quantify return on card member engagement campaigns
- Design and analyze online experiments for new payment collection strategies for international portfolio worth over 10 million dollars

Zuckerman Institute at Columbia University New York, New York
Research Assistant January 2019 - August 2020

- Researched neural spike sorting algorithms under Professor Liam Paninski
- Developed methods for dimension reduction and clustering of high-dimensional neural stimulus data. Implemented stochastic backtracking for sparse-dictionary learning problem. Built probabilistic models of neural noise. Implemented Kalman filter for online estimation of neural signals

Dordt University Biostatistics REU Sioux Center, Iowa
Research Assistant. June 2018 - January 2019

- Conducted research on methods for statistical inference on high-dimensional genetic data under Professor Nathan Tintle
- Conceptualized and implemented Bayesian hiearchal framework for identifying gene function and regularizing estimates of mixture model parameters from high-throughput gene expression experiments
- Investigated properties of aggregating functions for multiple comparison in grouped gene association tests. Developed high-performance populations genetics simulations in C++

Center for Computational Learning Systems at Columbia University New York, New York
Research Assistant June 2017 - December 2017

- Contributed to Gang Violence Prevention Project by independently conducting statistical analysis of NLP and network data of chicago gang members
- Conceptualized variable selection and simultaneous hypotheses testing framework using permutation tests. Developed multilevel model for predicting violence and aggression from twitter network activity.

Skills

- **Programming Languages:** R, Python, C++
- **Machine Learning:** PyTorch, STAN, Keras
- **Big Data:** PySpark, sparkR, SQL, Teradata, Hive

Presentations and Posters

- **PSB 2019 15 minute talk:** Bayesian Hiearchal Models for Tn-Seq Data
- **AICHE 2018 Poster:** Bayesian Regularization for High-Throughput Experiments
- **Emory Barkley Forum Presentation:** Gender Discrimination in Highschool Circuit

Teaching

- **Columbia Academic Services Tutor** at Columbia University Fall 2019- Spring 2020
  Probability Theory (STAT 4203)
- **Columbia Statistics Help Room Tutor** at Columbia University Spring 2020
  Calculus Based Introduction to Statistics (STAT 1202)

Scholarships and Awards

- Columbia Statistics Department Summer Internship Grant 2019
- PSB Conference Travel Award 2019
- Great Communicator Reagan Presidential Foundation Scholarship 2016

Leadership, Volunteering, and Miscellaneous

- Debate Instructor with New York Urban Debate League 2017–2019
  Taught middle school debate teams in the Bronx, organized debate tournaments, provided general academic tutoring
- Manager of Engagement Data Analytics Team at Columbia Spectator 2016–2017
  Developed data reporting pipeline for university newspaper, designed data-driven strategies for digital transformation
- Runner, Hiker, Backpacker, Squash Amateur, Bear Whisperer 1998 - Present
  Finished Marathon, Backpacked Appalachian Trail from Maryland to Vermont, Runner-Up Inter-mural Squash Champion, Survived Bear Intrusion in Tent