Yinpu Li

Address: Seattle, WA, 98101 | Phone: (202)-344-0121 | Email: yinpuli@icloud.com
LinkedIn:https://www.linkedin.com/in/yinpuli| Website: yinpuli.github.io | Google Scholar: Yinpu Li

EDUCATION

PhD in Statistics, Florida State University, Tallahassee, FL *August 2017 - Dec 2021*

- Research Assistant and Teaching Assistant
- NSF-Mathematical Sciences Graduate Research Fellowship, NSF; Clara Kibler Davis Scholarship, FSU

MS in Statistics, George Washington University, Washington, DC *08/2015 – 05/2017*

BS in Statistics, East China Normal University, Shanghai, China *09/2011 – 06/2015*

- Aegon-Industrial Fellowship for China Elite Academy(Outstanding Student), China

PROFESSIONAL EXPERIENCE

Research Scientist *Nov 2021- Present*

Convoy Inc, Seattle, WA

- Personalization price.

Statistical and Quantitative Science Data Sciences Intern *06/2021 – 08/2021*

Takeda Pharmaceutical, Boston, MA

- Proposed a method that combines traditional statistical models and data-driven Machine Learning techniques to improve predicted survival results in Phase III by utilizing patient-level information from external studies.
- Work resulted in a book chapter for Modern Machine Learning Methodologies in Pharmaceutical Industry (in preparation)..

Mathematical Sciences Graduate Intern *06/2020 – 08/2020*

Statistical Engineering Dept, National Institute of Standards and Technology, Gaithersburg, MD

- Developed a systematic tool of model uncertainty assessment for statistical models in violent crime and ingenious detective work under the framework proposed by NIST scientists.
- Resulted work was presented to the project to NSF-MSGI Virtual Research Cohort Presentations and highlighted at https://orise. orau.gov/nsf-msgi/profiles/li.html

Program Management Intern *11/2013-08/2014*

Headquarters, Fiat Chrysler Automobiles, Shanghai, China

- Actively and cross-functionally coordinated with Engineering, Purchasing, Supplier Delivery, and Product Line and Quality departments from APAC and NAFTA.
- Optimized stored procedures and functions using IBM Db2 SQL to reduce retrieval time with a 50% uplift; Awarded Spotlight Bravo together with the team for increasing FCA profits.

RESEARCH EXPERIENCE

Bayesian Decision Tree Ensembles in Nonparametric Problems (Ph.D. Dissertation) *01/2018 – Present*

Developing appropriate Bayesian methods within the Bayesian nonparametric framework for high dimensional problems which give robust inferences. Developed the theoretical properties and provided sufficient conditions for posterior consistency at close to minimax optimal rate.

- *Automated Gradient Boosting Machine that Adapts to Sparsity (Julia)*
    - Proposed modified algorithms based on gradient boosting machines for classification and regression.
    - Designed and conducted simulation experiments to test theories for improved performance.
    - Achieved 20% ~ 50% accuracy improvement and maintained robustness to irrelevant features and beat xgboost in experiments and selected datasets by designing the modified Bayesian boosting algorithm.
- *Extension to adaptive conditional distribution estimation (R & C++)*
    - Developed computationally convenient/flexible method for conditional density estimation on simulated data and national medical expenditure panel survey (MEPS) data.
    - The two-stage augmentation method conducted the variable selection under high-dimensional regime and increased the accuracy by 45% compared to the most popular Dirichlet Process Mixture models.

    - Extended and elaborated the theoretical properties of BART-based methods and provided sufficient conditions for posterior consistency at close to minimax optimal rate.
    - Developing R package(on-going).
- *Extension to survival models with sub-model shrinkage (R & C++)*
    - Developed two Bayesian models (Modulated BART and CoxBART) to estimate survival time subject to right-censoring in the presence of potential high-dimensional predictors for Mayo Clinic Primary Biliary Cirrhosis data.
    - Validated the effectiveness and robustness of the proposed methods via comprehensive simulation designs. The proposed methods outperform the Cox proportional hazards model, Cox-Linear using LASSO, and random survival forests in accuracy.

Effect of Model Uncertainty and Shape-Constrained Optimization (NSF Graduate Program & NIST)
Proposed an automatic shape-constrained distribution-free estimation method. Estimated the functional curve through non-Bayesian approaches (linear programming and flexible global optimization with simulated annealing) and Bayesian approaches with Dirichlet Process. Addressed model uncertainty quantification of certain stochastic model components and to develop uncertainty pyramids. Created associated algorithms, R packages and web applications for non-statistical users.

- *Application to electronic component failure times in airplanes*
- *Application to image comparison of footwear prints*

PUBLICATIONS AND PRESENTATION

- **Li, Yinpu**, Antonio R. Linero, and Jared S. Murray. "Adaptive Conditional Distribution Estimation with Bayesian Decision Tree Ensembles." *arXiv preprint arXiv:2005.02490 (2020)*.
- Linero, Antonio R., Piyali Basak, **Yinpu Li**, and Debajyoti Sinha. "Bayesian Survival Tree Ensembles with Submodel Shrinkage." *Bayesian Analysis 1, no. 1* (2021): 1-24. *(Highlighted Paper for JSM 2022, Innovations in Bayesian Learning Session).*
- **Li, Yinpu** and Antonio R. Linero , "An Efficient Gradient Boosting Decision Tree Algorithm that Adapt to Sparsity in High-Dimensional Setting ". *Manuscript.*
- Hari Iyer, Steven P. Lund, and **Yinpu Li** , "Effect of Model Uncertainty on Tail Probabilities in Forensic Evidence: A Metrological Perspective". *Manuscript.*
- **Li, Yinpu**, Bradley Hupf, Rachael Liu, and Jianchang Lin, "Predicting Phase III Results by Incorporating Historical Data using BART Extensions". *Book Chapter Manuscript.*
- Invited speaker of ENAR 2020 and NIST Seminar over "BART Models for Nonparametric Problems".