

Yinpu Li

Ph.D Candidate

Tallahassee, Florida
✉ yinpu.li@stat.fsu.edu
📄 [yinpuli.github.io](https://github.com/yinpuli)

Education

2017–Now **Ph.D, Statistics**, *Florida State University*.

2015–2017 **M.S, Statistics**, *George Washington University*.

2011–2015 **B.S, Statistics**, *East China Normal University*.

Experience

Jun. 2020 Mathematical Science Graduate Intern, *NIST*.

- | | |
|-----------|--|
| Aug. 2020 | <ul style="list-style-type: none"> Developed systematic method of comprehensive model uncertainty assessment for models; Investigated uncertainty quantification of tail probabilities and densities and demonstrated the methods with examples regarding failure times and forensic shoeprint comparison; Created associated algorithms, Rpackages and web applications for implementing the approaches. |
|-----------|--|

Jan. 2020 **Research Assistant**, Florida State University.

Now	Developing appropriate Bayesian methods, model selection and variable selection tools within the Bayesian nonparametric framework for high dimensional problems.
-----	--

Sep. 2018 **Instructor**, Florida State University.

-Aug. 2019 ◦ Independently taught STA 1013 Statistics through Examples 3 times a week, 100 minutes/lesson;

Sep. 2014 **Research and Development Team Intern**, KangdeXin Composite Material Co., Ltd,
| Shanghai Branch.

- | | |
|-----------|---|
| Aug. 2015 | <ul style="list-style-type: none"> o Constructed and maintained comprehensive database of customer information; o Co-worked with programmers in pulling dealerships across the country, including supply-side and demand-side contracts, telephone numbers, address and sales records; o Profiled and developed database to keep companies up to date; o Mined multiple datasets and conducted data analysis with R, data visualization and predictive modeling to solve complex problems and to provide support for actionable insights; |
|-----------|---|

Nov. 2013 **Program Management Team Intern**, *Fiat Chrysler Automobiles*.

- | | | |
|-----------|--|--|
| Aug. 2014 | <ul style="list-style-type: none"> Supported the head project manager with strategy planning and process integration, including Jeep programs in APAC execution and product development; Interpreted and documented data information needs; Facilitated the data collection sessions. Analyzed and documented data processes, scenarios and information flow; Organized data of resources of company, customer, products, and requirements from other departments; Analyzed monthly and annual data of each program to ensure programs attain established goals with a clear focus on quality, cost, delivery and complete customer satisfaction; | |
|-----------|--|--|

Jun. 2013 **Data Analyst Intern**, *China Life Insurance, Henan Branch.*

- | | |
|------------------|---|
| <p>Sep. 2013</p> | <ul style="list-style-type: none"> ○ Supported the members in Data and Information Technology Department; ○ Enhanced existing database and analytical models by determining the most effective data collection and calculation method. ○ Presented the reports on weekly meetings with the heads of the marketing department and the sales team. ○ Investigated and analyzed data collected from the marketing department with Excel. |
|------------------|---|

Publications and Projects

- R Bayesian Decision Tree Ensembles in Fully Nonparametric Problems**, *In progress*.
 Jan. 2020 | Extend BART to fully-nonparametric problems which are based on modulating an underlying
 Now Poisson process, which allows the well-implemented simple Gibbs sampling algorithms to be derived for BART models for 1) Density Estimation; 2) Nonparametric Survival Analysis; 3) Estimating the Intensity Function of a Spatial Point Process. By taking advantage of the strong theoretical properties of certain BART priors, the posterior concentration is established at near-minimax optimal rates for the problems, adaptively over a large class of function spaces. The methodology is then illustrated on simulated and benchmark datasets.
- C++ & R Adaptive Conditional Distribution Estimation with Bayesian Decision Tree Ensembles**, *First Author*, Submitted, 2020+.
- Julia Bayesian Gradient Boosted Decision Trees that Adapt to Smoothness and Sparsity**.
 Sep. 2018 | A new algorithm based on gradient boosting was implemented in the context of the Bayesian
 sequential additive regression trees framework. By considering sparsity inducing soft decision trees
 where the decisions were treated as probabilistic, two potential shortcomings of tree ensembles,
 Aug. 2019 the lack of smoothness and vulnerability to the curse of dimensionality, would be overcome.
- MATLAB Neural Decoding in Motor Cortex**.
 Jan. 2018 Using observed neural activity from brain cortex in research animals, the Kalman filter model and the Inhomogeneous Poisson model were performed to understand the brain mechanism and made inferences about the external behaviors.
- MATLAB Concentration Inequalities for Markov Chains: A Review and Development**.
 Jan. 2018 The project focused on concentration inequalities for Markov Chain dependence type, which has been popular in modeling many discrete mathematics via the probabilistic method. Marton Couplings technique was applied to develop new methods in theoretical analysis. The results about empirical tail probabilities of estimates by multiple MCMC simulations were tested on the Challenger data, which describes 23 launch experiments on the failure rate of O-ring component at different temperatures.
- Python Methodologies in Face Recognition**.
 Sep. 2017 Effective algorithms at facial recognition, including Principal Component Analysis, Linear Discriminant Analysis, K-Nearest Neighbor classifier, and 1-vs-1 SVM with Gaussian kernel, were systematically compared in terms of accuracy and efficiency.
- R Breast Cancer Analysis**.
 May. 2017 To determine the relevant factors, feature selection was conducted based on the significance level and variance contribution by applying Lasso. A logistic regression model was then fitted with the selected predictors and prediction on the probability of developing tumor was made on new potential breast cancer patients. A survival analysis was then delivered to model the survival time after developing cancer with a log logistic regression.
- R Time Series Analysis and Forecasting - Rossman Drug Store Sales, Kaggle**.
 Jan. 2017 Extensive data analysis was performed on the daily panel data of drugstores in Europe, including seasonal decomposition, trends recognition and autocorrelation calculation. A clustering behavior by geographic location was diagnosed, after handling which would enhance the next-6-weeks-sales prediction accuracy significantly for both Seasonal ARIMA and Prophet models.
- R Metropolis-Hastings and Parallel Tempering Algorithms in Application of Gaussian Mixtures**.
 Sep. 2016 The asymptotic properties of the classical Metropolis-Hastings algorithm and the parallel tempering algorithms, sampling from Gaussian Mixture models, were studied. Theoretical proof and simulation results were provided and both indicated the classical M-H algorithm would be torpidly mixing while the parallel tempering would be rapidly mixing under some mild regularization conditions.
- R Data Mining Techniques in Wine Quality Assessment**.
 Jan. 2016 Different models, including single Classification Tree, Bagging, Random Forest, Support Vector Machine with Linear/Polynomial/Radial kernels, Logistic Regression, KNN and Quadratic Discriminant Analysis, were compared by metrics of prediction accuracy, mean squared error and area over the regression error characteristic curve with the benchmark of multiple linear regression. The Random Forest technique outperformed other models in prediction accuracy, yet lacked the interpretability.

SAS Model Standardization and Improvement-The NCGS Study.

Sep. 2015 The project was based on the (NCGS) National Cooperative Gallstone Study. The relationship between cholesterol and some of the potential risk factors were tested, multiple regression analysis was conducted and linear model selection and regularization were performed. The most significant variables were selected and associations of various characteristics with serum cholesterol among patients with gallstones were described.

R The Impact of VC Involvement on the Corporate Governance of Listed Companies on China's Growth Enterprise Market.

Jan. 2012 |
|
|
|
|
Sep. 2014 |

- Project Leader; received the First-Class Prize in 2014 National Scientific Innovation Challenge.
- Evaluated the 3-year stock performance of Growth Enterprises Market of public companies listed on Shenzhen Stock Exchange and performed regression analysis on whether the VC or PE got involved in the investment ,whether joint investment happened and the proportion of shareholding of each VC as well. The model was optimized with discriminant analysis.

Awards

2017 **Dissertation Research Grant**, *Florida State University*.

2013 **China 100 Elite Youth Outstanding Student Leadership Award**.

2012-2013 **Aegon-Industrial Social Responsibility Scholarship (Merit-based)**, *Shanghai*.

2011-2014 **ECNU Undergraduate Scholarship(Merit-Based)**.