Yinpu Li

Ph.D Candidate

Tallahassee, Florida
✉ *yinpu.li@stat.fsu.edu*
yinpuli.github.io

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

Jul. 2020 **Invited Speaker**, *National Institute of Standards and Technology*, Statistical Engineering Division.

- Presented on "BART Models for Nonparametric Problems "

Jun. 2020 – Aug. 2020 **Mathematical Science Graduate Intern**, NSF MSGI.

Proposed the project using non-Bayesian approaches to address model uncertainty quantification of certain stochastic model components and to develop uncertainty pyramids. (i) Developed systematic method of comprehensive model uncertainty assessment for statistical models; (ii) Developed an uncertainty pyramid for a more practically useful and iterpretable treatment of model uncertainty; (iii) Illustrated the develped methodoligy using real life studies with current scientific within the substantive area; (iv) Created associated algorithms, Rpackages and web applications for implementing the approaches.

Mar. 2020 **Session Chair**, *ENAR Spring Meeting*.

- Chaired the session "Bayesian Nonparametric Methods for Causal Inference and Missing Data"

Mar. 2020 **Invited Speaker**, *ENAR Spring Meeting*.

- Presented on "Bayesian Decision Tree Ensembles in Fully Nonparametric Problems" in session "Recent Advances in the Uncertainty Estimation and Properties of Bayesian Additive Regression Trees "

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

Developing appropriate Bayesian methods, model selection and variable selection tools within the Bayesian nonparametric framework for high dimensional problems in which analysts are forced to (i) confront the curse of dimensionality when estimating low-dimensional effects of interest and (ii) are required to make in-principle unverifiable assumptions. Developing flexible Bayesian nonparametric models which give robust inferences in these scenarios is fraught with subtle challenges.

Sep. 2018 – Aug. 2019

Instructor, Florida State University.

- Independently taught STA 1013 Statistics through Examples 3 times a week, 100 minutes/lesson;
- Worked with two sections of students without backgrounds in maths or statistics who took the class to improve skills in applied statistical reasoning;
- Developed instructional plans(the syllabus) for the course and created lecture notes, exercises, calculator guidelines, quiz, homework, and activities;
- Assessed students' progress by grading quizzes, tests, and projects;
- Kept tracking the progress and advised students to maintain good performance and to achieve their goals.
- Brief course description:
  STA 1013 Statistics Through Example is an intro-level applied-statistic course designed for non-scientific major students. The course provides students with a background in applied statistical reasoning. Fundamental topics are covered including graphical and numerical description of data, understanding randomness, central tendency, correlation versus causation, line of best fit, estimation of proportions, and statistical testing. Statistical thinking, relevant ideas, themes, and concepts are emphasized over mathematical calculation. In this class, students learn many of the elementary principles that underlie collecting data, organizing it, summarizing it, and drawing conclusions from it. No programming is included in the course. Instead, all computations are based on the calculator Texas Instrument TI84 plus CE. Guidance on how to use the calculator is involved in the course.

Sep. 2014 – Aug. 2015

Research and Development Team Intern, KangdeXin Composite Material Co., Ltd, Shanghai Branch.

- Constructed and maintained comprehensive database of customer information;
- Co-worked with programmers in pulling dealerships across the country, including supply-side and demand-side contracts, telephone numbers, address and sales records;
- Profiled and developed database to keep companies up to date;
- Mined multiple datasets and conducted data analysis with R, providing support for actionable insights;
- Conducted data analysis, data visualization and predictive modeling to solve complex problems;
- Generated relevant and thought provoking analysis, commentary and visualizations on nationwide key markets.

Nov. 2013 – Aug. 2014

Program Management Team Intern, *Fiat Chrysler Automobiles*.

- Supported the head project manager with strategy planning and process integration, including Jeep programs 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;
- Analyzed operating data to identify trends, patterns which could be used to improve operational efficiency;
- Collaborated with executives, engineers and designers to discuss progress on current initiatives, to present new initiatives, to establish deadlines and to receive feedback from weekly internal and monthly external meetings on a global schedule.

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

- 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.
- Designed, modified and profiled questionnaires based on analyzing results and feedback from co-workers.
- Investigated and analyzed data collected from the marketing department with SAS.

Sep. 2013

- Acquired, cleaned and structured data from multiple sources and analyzed data into draw insightful conclusions, trends and correlations.

2013 **Visiting Undergraduate Student**, *Seoul National University*.

2010 **Leading Founder**, *Xiaoyangren Solar Air Conditioner Co., Ltd.*, Shanghai.

- The 8-th FuXing Challenge Cup National College Students' Extracurricular Academic Science and Technology Contest;
- Silver Medal Award(5%) of 700 finalist teams nationwide; Silver Medal(1%) Shanghai Regional; Winner of 10-Hour Virtual Entrepreneurship Operation Competition;
- Launched the tech venture based on one innovative solar material developed at ECNU; developed business plan and formulated marketing strategies;
- Winner of 3-Day Real-Time Simulation Tactics Competition, where instant decision making was made team-by-team about investment, R&D, supply chain, product line, advitising, marketing etc.

2012

Publications and Projects

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, the lack of smoothness and vulnerability to the curse of dimensionality, would be overcomed.

Aug. 2019

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 Markove Chain dependence type, which has been popular in modeling many discrete mathematics via the probalistic 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.

MATLAB & Python **Methodologies in Face Recognition**.

Sep. 2017 Effective algorithms at facial recognition, including Pricipal Component Analysis, Linear Discriminant Analysis, K-Nearest Neighbor classifier, Simple Projection and 1-vs-1 SVM with Gaussian kernel, were systematically compared in terms of accuracy and efficiency. The procedure involved extracting facial features from an image, recognizing via comparison between the newly extracted features and existing features in data and matching the faces.

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 Extentive 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 Phophet models.

R **Metropolis-Hastings and Parallel Tempering Algorithms in Application of Gaussian Mixtures**.

Sep. 2016 The asymptotic properties of two MCMC algorithms sampling from Gaussian Mixture models, the classical Metropolis-Hasting algorithm and the parallel tempering algorithms, 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 For the interest of investigating if human qualify of tasting could be related to the chemical properties of wine, so that certification & quality assessment and assurance process would be more controlled, different models, including single Classification Tree, Bagging, Random Forest, Support Vector Machine with Linear/Polynomial/Radius 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 Randome Forest technique outperformed other models in prediction accuracy, yet lacked the interpretability.

SAS **Model Standardadization and Improvement-The NCGS Study**.

Sep. 2015 The project was based on the (NCGS) National Coorperative Gallstone Study. To test the relationship between choesterol and some of the potential risk factors, multiple regression analysis was conducted and linear model selection and regularization was 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.

Research Interests

Applications of Bayesian methods to problems in Biostatistics and machine learning.

Bayesian nonparametrics and semiparametrics.

Computational issues associated with the above.

Technical Skills

Programming Julia, R, C++, Python, LaTeX, Matlab

Awards

2020 **NSF-Mathematical Sciences Graduate Fellowshop**, *NSF*.

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)**.

Volunteer Experience

2010 **Volunteer**, *Autistic Volunteers Association*, Shanghai.

- Conducted games and sensory integration training for children with autism;
- Organized and participated in the training camp for children with Down's syndrome;

2012

- One-to-one tutored a child with autism weekly in reading and communication.