Mengxin(Maxine) Yu

Email: mengxiny@princeton.edu
Mobile: +1(609)933-8102

Education

- **Princeton University** — Princeton, NJ, US
  Ph.D. in the Department of Operations Research and Financial Engineering — *Sep. 2018 - now*
  Advisor: *Prof. Jianqing Fan* ***GPA:*** *4.0/4.0*
  Research Interests*: Ranking Estimation and Inference, High-dimensional Statistics (Feature selection/Inference), Statistical Decision Making (Online/Offline Feature-based Dynamic pricing/Multi-arm Bandit/Reinforcement Learning), Factor-adjusted Inference, Robust Statistics, Causal Inference, Time Series Analysis*
- **University of Science and Technology of China** — Hefei, Anhui, China
  B.S. in Statistics — *Sep. 2014 - Jun. 2018*
  GPA: *3.98/4.30* ***Major GPA:*** *4.17/4.30* ***Rank:*** *1st/105*
  Graduated with summa cum laude (Guo Moruo Scholarship, 1%) and best graduation paper award (1%)

Selected Projects

- **Ranking Estimation and Inference under A General Choice Model**:
  - : Investigated in a ranking problem under a general choice set realized from a random hyper-graph.
  - : Proved the consistency of MLE estimator which is minimax optimal.
  - : Developed the asymptotic distribution of the MLE estimator.
  - : Constructed confidence intervals for the ranks of all items simultaneously.
  - : Application: Identify the top-K items. Quantify the uncertainty of the rank of certain item (such as the rank of a team in competition).
- **Strategic Decision-Making in the face of Private Information – Provably Efficient RL with Algorithmic Instruments**:
  - : Investigated in an offline muti-agent reinforcement learning problem under the existence of strategic agents, which leads to the existence of unobserved confounding in reward and state transition functions.
  - : Leveraged actions and states as instrumental variables and used the method of moments to learn the model (reward and transition kernel).
  - : Leveraged 'Pessimistic' (lower confidence bound) to construct our policy, which avoids the curse of the insufficient coverage of offline data.
  - : Presented an upper bound the sub-optimality of our policy by the Rademacher complexity of the function classes of the reward and state transition functions.
  - : Application: college admission with strategic agents, strategic bandits, no-compliance agents in recommendation systems.
- **Are Latent Factor Regression and Sparse Regression Adequate?**:
  - : Proposed a factor augmented linear regression model **(FARM)** which incorporates the latent factor regression model and sparse linear regression model as special cases.
  - : Provided statistical guarantees for the estimation of our model under both light and heavy-tailed noises respectively.
  - : Leveraged our model as the alternative model to test the goodness-of-fit of the latent factor regression and sparse regression models.
  - : Studied the asymptotic behaviors of our test statistics under the null and the test power under the alternative.
  - : Conducted large scale simulation experiments and real data analysis using FRED-MD dataset.
  - : Application: High-dimensional uncertainty quantification of factor structured data (Financial/Gene data etc.).
- **Understanding Implicit Regularization in Over-Parameterized Single Index Model**:
  - : Investigated in a feature selection problem of high-dimensional single index model.
  - : Instead of adding explicit regularizations, combining with over-parameterization, we proposed **a regularization free** method which achieved solutions with **optimal statistical-rates** under both sparse and low rank settings.
  - : Outperforms classical regularized methods in terms of ℓ_2-statistical rate, false discovery rate as well as scalability.
  - : Application: High-dimensional variable selection; Compressed Sensing (e.g. Image Processing).
- **Policy Optimization Using Semiparametric Models for Dynamic Pricing**:
  - : Investigated in a dynamic pricing problem with feature based customer valuation and unknown market noise.
  - : Developed an online price offering policy by leveraging tools of variable selection and kernel regression.
  - : Proved optimal regret upper bounds under regimes with both independent and dependent covariates.
  - : Verified theoretical conclusions by doing large-scale simulations and real data analysis using real-life auto loan dataset provided by Columbia University.
  - : Application: E-commerce policy design; Online (personal) pricing/auction;
- **Statistical Learning of the Worst Regional Smog Extremes with Dynamic Conditional Modeling**:
  - : Proposed a dynamic conditional extreme value modeling method based on time series model.
  - : Theoretically proved asymptotic behaviors of the maximum likelihood estimator of parameters in the model.
  - : Conducted both large-scale simulations and real data analysis of PM2.5 datasets.
  - : Extended the state of art from static model to a dynamic one, which well capture the characteristics of smog data.
  - : Application: Regional extreme-value modeling; Risk control.

Journal Publications

- **Covariate Assisted Ranking Estimation and Inference** : Joint work with Jianqing Fan, Jikai Hou; forthcoming, available by request.
- **Strategic Decision-Making in the face of Private Information – Provably Efficient RL with Algorithmic Instruments**: Joint work with Jianqing Fan, Zhuoran Yang; forthcoming, available by request.
- **A Theoretical Analysis of Hodges-Lehmann Estimator**: Joint work with Jianqing Fan, Zhipeng Lou; forthcoming, available by request
- **Ranking Estimation and Inference under A General Choice Model**: Joint work with Jianqing Fan, Weichen Wang, Zhipeng Lou; forthcoming, available by request.
- **Are Latent Factor Regression and Sparse Regression Adequate?**: Joint work with Jianqing Fan, Zhipeng Lou; Major Revision Requested by the Journal of the American Statistical Association (JASA), 2022.
- **Policy Optimization Using Semiparametric Models for Dynamic Pricing**: Joint work with Jianqing Fan, Yongyi Guo; Major Revision Requested by the Journal of the American Statistical Association (JASA), 2022.
- **Understanding Implicit Regularization in Over-Parameterized Single Index Model**: Joint work with Jianqing Fan, Zhuoran Yang; accepted by the Journal of the American Statistical Association (JASA), 2022
- **Statistical Learning of the Worst Regional Smog Extremes with Dynamic Conditional Modeling**: Joint work with Zhengjun Zhang, Lu Deng, accepted by Atmosphere, 2020

Presentations

- **Informs Annual Meeting**: October, 2022, Indianapolis, Indiana, USA (upcoming)
- **Joint Statistical Meeting** : August, 2022, Washington D.C., USA (upcoming)
- **ICSA Applied Statistics Symposium**: June, 2022, Gainesville, Florida, USA (upcoming)
- **Statistics in the Big Data Era** : June, 2022, Berkeley, California, USA
- **NYC Optimization Day**: April, 2022, Roosevelt Island, New York, USA
- **Conference on Advances in Bayesian and Frequentist Theory**: April, 2022, Rutgers University, NJ, USA **(Achieved best poster award).**
- **Perspectives in Statistical Modeling and Inference**: December, 2021, The Wharton School, Pennsylvania, USA
- **TOPML Workshop**: April, 2021, Virtual, USA
- **Two Sigma's Inaugural PhD Symposium**: December, 2020, Virtual, USA.

Teaching Experience

- **ORF245: Fundamentals of Statistics**: Spring 2022 (Head AI), Fall 2020, Fall 2019.
- **ORF309-EGR309: Probability and Stochastic Systems**: Spring 2021 (Head AI), Spring 2020.
- **ORF363: Computing and Optimization**: Fall 2021.

Skills

- **Programming**: Python(Numpy/Pandas/Sklearn/Pytorch), R, Matlab, SQL, Kdb+, Latex, C, SAS
- **Probability**: Stochastic Process, Markov Chain, Martingale, Brownian Motion, Ito's Lemma
- **Statistics**: High-dimensional Regression, Non-parametric Statistics, Statistical Machine Learning, Statistical Inference
- **Deep Learning**: DNN(Dropout/Adam/RMSProp), CNN(AlexNet/ResNet/YoLo Alg), RNN(GRU/LSTM)
- **Natural Language Processing**: Word Embeddings, Sequence Modeling, Attention Model, Transformer
- **Reinforcement Learning**: MAB, Policy Gradient, Q-learning, UCB Algorithm, Function Approximation
- **Econometrics**: Time Series Analysis, Factor Model, Portfolio Allocation, Option Pricing(basics)
- **Optimization**: Linear Programming, (Stochastic/Projected) Gradient Descent, Alternating Direction of Multipliers(ADMM)