

Yanfei Zhou

Address: 701 W Exposition Blvd, Los Angeles, CA

Email: yanfeizh@usc.edu

Mobile: +1-510-325-5896

EDUCATION

- **University of Southern California - Los Angeles, CA** Aug 2021 - Present
Ph.D. in Data Sciences and Operations (Statistics); Advisor: Prof. Matteo Sesia GPA: 4.00/4.00
- **University of Chicago – Chicago, IL** Oct 2018 - Jun 2020
M.S. in Statistics; Advisor: Prof. Weibiao Wu GPA: 3.88/4.00
Thesis: An empirical study on A Fully Online Approach for Statistical Inference Based on SGD
- **London School of Economics and Political Science - London, UK** Sep 2015-Jun 2018
B.S. with First Class Honours in Statistics with Finance GPA: 4.00/4.00

RESEARCH INTEREST

Conformal inference; Machine learning and deep learning; Explainable AI; Natural language processing; Causal inference; Online learning; Statistical genetics and health sciences

PUBLICATIONS, PREPRINTS, AND PRESENTATIONS

- **Training Uncertainty-Aware Classifiers with Conformalized Deep Learning:** Bat-Sheva Einbinder, Yaniv Romano, Matteo Sesia, and **Yanfei Zhou**; <https://arxiv.org/abs/2205.05878> (Accepted to NeurIPS 2022)
Poster presentations at NeurIPS Nov, 2022, in New Orleans, LA
- **Conformalized Early Stopping:** Ziyi Liang, **Yanfei Zhou**, and Matteo Sesia

ACADEMIC PROJECTS

- **Conformal prediction (CP) for model-free predictive uncertainty estimation (Deep learning, Overconfidence):** Developed uncertainty-aware deep learning algorithms that can capture and correct model overconfidence when training black-box deep neural network models (May '22); Constructing a novel early stopping framework that improves data efficiency and achieves higher model predictive power by integrating CP techniques (Ongoing); Designing a CP method that adapts time-dependency and produces tight prediction bands for multiple time-series forecasting problems (Ongoing).
- **Online statistical inference based on Stochastic Gradient Descent (SGD) (Online learning):** Investigated and implemented a fully online learning algorithm that conducts adaptive statistical inference on SGD solutions (May '20)
- **Causal study for retail platforms (causal inference, experiment design):** Collected customers survey data and built interactive visualization tools to analyze the impact of product strategies on retailers' commission ratios (Mar '18, Mar '20)
- **Financial time series (Time series, dimension reduction):** Developed statistical models to forecast UK forward rates by combining nonlinear Principal Component Analysis and time series models; Backtested financial assets (Mar '18)

PROFESSIONAL EXPERIENCE

- **CNA Financial Corporation – Chicago, IL** Aug 2020 - Apr 2021
Data Scientist (Full-time)
 - Conducted change-point detection and trend detection on the insurance submission data using statistical smoothing techniques and two-sample testing; Modelled submission data with Binomial and Poisson rate models.
 - Built machine learning and statistical models to estimate the probability of submitting claims based on characteristics of individual or organizations and tested model robustness.
- **CCC Information Services Inc. - Chicago, IL** Jun 2019 - Sep 2019
Data Scientist (Summer internship)
 - Designed an efficient algorithm to match the car damage data with the casualty data; performed data cleaning, exploratory data analysis, clustering, and correlation studies on the matched data.
 - Proposed and implemented a Empirical-Bayesian based method to quantify the predictive uncertainty of vehicle components' damage; collaborated with the team to provide theoretical backgrounds.

SKILLS SUMMARY

- **Programming languages and platforms:** Python, R, SQL, HTML, CSS; Git, GCP, AWS
- **Statistical modelings:** Data analysis (**'Kaggle' silver medal winner, ranked top 4%/5300+ groups**), Multiple testing, Causal inference with modern machine learning; Large scale optimization, Probability theory and measure theory, Generalized linear models, Time Series, Stochastic process and Stochastic simulations, Bayesian Inference, Dimensionality reduction, Experiment Design, High-dimensional Statistics, text mining; Quantitative finance, Econometrics
- **Languages:** English - Professional, Chinese (Mandarin) - Native