Chengchun Shi

Assistant Professor @LSE

Columbia House
07555508251
c.shi7@lse.ac.uk
callmespring.githuo.io

All knowledge is, in the final analysis, history
All sciences are, in the abstract, mathematics
All methods of acquiring knowledge are, essentially, through statistics
C. R. Rao

EDUCATION

June. 2014 — **B.S. in Statistics**, *Zhejiang University, Hangzhou, China.*

Aug. 2019 — **Ph.D in Statistics**, *North-Carolina State University, Raleigh, USA.*

- **Dissertation**: On Statistical Learning for Individualized Decision Making with Complex Data
- **Advisor**: Dr. Wenbin Lu & Dr. Rui Song

EMPLOYMENT

Sep. 2019 — **Assistant Professor of Data Science**, *Department of Statistics, London School of Economics and Political Science, London, UK.*

HONORS

Aug. 2021 — **Best Paper Award**, *IJCAI 2021 Reinforcement Learning for Intelligent Transportation Systems Workshop.*

Mar. 2021 — **Royal Statistical Society Research Prize.**

Aug. 2020 — **Finalist in Reinforcement Learning Competition Track in 2020 KDD Cup**, *Ranked 7 out of more than 1000 teams.*

Mar. 2018 — **Institute of Mathematical Statistics Hannan Student Travel Award.**

April 2017 — **Institute of Mathematical Statistics Travel Award.**

April 2017 — **Paige Plagge Citizenship Award, NCSU**, *awarded to a graduate student with an outstanding academic record, who in the judgment of the committee has especially enhanced the life of fellow students with encouragement, generosity and/or humor.*

Feb. 2014 — **Outstanding Undergraduates of Zhejiang Province.**

Feb. 2013 — **Meritorious Winner in 2013 Mathematical Contest in Modelling.**

Oct. 2012 — **National Scholarship**, *2%.*

RESEARCH INTERESTS

- **Reinforcement Learning**
- **Statistical Analysis of Complex Data**

PUBLICATION/ACCEPTED MANUSCRIPT

Dec. 2021 — **Shi, C.***, Wang, X.*, Luo, S., Zhu, H., Ye, J. and Song, R. Dynamic Causal Effects Evaluation in A/B Testing with a Reinforcement Learning Framework, *Journal of the American Statistical Association*, **accepted**.

Dec. 2021 — **Shi, C.**, Xu, T., Bergsma, W. and Li, L. Double Generative Adversarial Networks for Conditional Independence Testing, *Journal of Machine Learning Research*, **22**, 1–32.

Dec. 2021 — **Shi, C.**, Luo, S., Zhu, H. and Song, R. An Online Sequential Test for Qualitative Treatment Effects, *Journal of Machine Learning Research*, **22**, 1–51.

Dec. 2021 — Cai, H.*, **Shi, C.***, Song, R. and Lu, W Deep Jump Learning for Off-Policy Evaluation in Continuous Treatment Settings, *NeurIPS*, 2021 ENAR Distinguished Student Paper Awards.

Sep. 2021 — Li, L., **Shi, C.**, Guo, T. and Jagust, W. Sequential Pathway Inference for Multimodal Neuroimaging Analysis, *Stat*, **accepted**.

Aug 2021 — Wan, R., Zhang, S., **Shi, C.**, Luo, S. and Song, R. Pattern Transfer Learning for Reinforcement Learning in Order Dispatching, *IJCAI RL4ITS Workshop (**best paper**, **spotlight**)*.

July 2021 — **Shi, C.**, Wan, R., Chernozhukov, V. and Song, R Deeply-Debiased Off-Policy Interval Estimation, *ICML (**long talk**, **top 3%** of submissions)*.

July 2021 — **Shi, C.**, Song, R., Lu, W. and Li, R. Statistical Inference for High-Dimensional Models via Recursive Online-Score Estimation, *Journal of the American Statistical Association*, **116**, 1307–1318.

June 2021 — **Shi, C.**, Zhang, S., Song, R. and Lu, W Statistical Inference of the Value Function for Reinforcement Learning in Infinite Horizon Settings, *Journal of the Royal Statistical Society, Series B*, **accepted**.

Feb. 2021 — **Shi, C.** and Li, L. Pattern Transfer Learning for Reinforcement Learning in Order Dispatching, *Journal of the American Statistical Association*, **accepted**.

Feb. 2021 — **Shi, C.**, Song, R. and Lu, W. Concordance and Value Information Criteria for Optimal Treatment Decision, *Annals of Statistics*, **49**, 49–75.

July 2020 — **Shi, C.**, Wan, R., Song, R., Lu, W. and Leng, L. Does the Markov Decision Process Fit the Data: Testing for the Markov Property in Sequential Decision Making, *ICML*.

July 2020 — **Shi, C.**, Lu, W. and Song, R. Breaking the Curse of Nonregularity with Subagging — Inference of the Mean Outcome under Optimal Treatment Regimes, *Journal of Machine Learning Research*, **21**, 1–67.

July 2020 — **Shi, C.**, Lu, W. and Song, R. A Sparse Random Projection-based Test for Overall Qualitative Treatment Effects, *Journal of the American Statistical Association*, **115**, 1201–1213.

Oct. 2019 — **Shi, C.**, Song, R., Chen, Z. and Li, R. Linear Hypothesis Testing for High Dimensional Generalized Linear Models, *Annals of Statistics*, **46**, 2671–2703.

Aug. 2019 — **Shi, C.**, Lu, W. and Song, R. On Testing Conditional Qualitative Treatment Effects, *Annals of Statistics*, **47**, 2348–2377.

Feb. 2019 — **Shi, C.**, Lu, W. and Song, R. Determining the Number of Latent Factors in Multi-Relational Learning, *Journal of Machine Learning Research*, **20**, 1–38.

Oct. 2018 — **Shi, C.**, Lu, W. and Song, R. A Massive Data Framework for M-Estimators with Cubic-Rate, *Journal of the American Statistical Association*, **113**, 1698–1709.

July 2018 — **Shi, C.**, Song, R. and Lu, W. Discussion of "Optimal Treatment Allocations in Space and Time for On-Line Control of an Emerging Infectious Disease", *Journal of the Royal Statistical Society, Series C*, **67**, 743–789.

June 2018 — **Shi, C.**, Fan, A., Song, R. and Lu, W. High-Dimensional A-Learning for Optimal Dynamic Treatment Regimes, *Annals of Statistics*, **46**, 925–957.

May 2018 — **Shi, C.**, Song, R., Lu, W. and Fu, B. Maximin Projection Learning for Optimal Treatment Decision with Heterogeneous Individualized Treatment Effects, *Journal of the Royal Statistical Society, Series B*, **80**, 681–702.

June 2017 — **Shi, C.**, Song, R. and Lu, W. Discussion of "Random projection ensemble classification", *Journal of the Royal Statistical Society, Series B*, **79**, 959–1035.

Oct 2016 **Shi, C.**, Song, R. and Lu, W. Robust Learning for Optimal Treatment Decision with NP-Dimensionality, *Electronical Journal of Statistics*, **10**, 2894–2921.

Aug 2016 Zhang, P., Qiu, Z. and **Shi, C.** simplexreg: An R Package for Regression Analysis of Proportional Data Using Simplex Distribution, *Journal of Statistical Software*, **71**, 1–21.

TALKS

Jan. 2022 **Applied Reinforcement Learning (ARL) Seminar (invited)**, *Statistical Inference in Reinforcement Learning, Youtube Link.*

Nov. 2021 **Stat seminar at the University of Warwick, Coventry, UK (invited)**, *Statistical Inference in Reinforcement Learning.*

Oct. 2021 **INFORMS Annual Meeting (invited)**, *Does the Markov Decision Process Fit the Data: Testing for the Markov Property in Sequential Decision Making.*

Sep. 2021 **European Young Statisticians Meeting (invited)**, *Does the Markov Decision Process Fit the Data: Testing for the Markov Property in Sequential Decision Making.*

Sep. 2021 **EuroCIM**, *Testing Mediation Effects Using Logic of Boolean Matrices.*

Aug. 2021 **Joint Statistical Meeting (invited)**, *Testing Directed Acyclic Graph via Structural, Supervised and Generative Adversarial Learning.*

June 2021 **Conference of "Statistical learning methods in modern AI", Xi'an, China (invited)**, *Does the Markov Decision Process Fit the Data? Testing for the Markov Property in Sequential Decision Making.*

June 2021 **Stat seminar at the University of Hong Kong, Hong Kong, China (invited)**, *Statistical Inference in Reinforcement Learning.*

April 2021 **Seminar at Purdue Reinforcement Learning group, West Lafayette, USA (invited)**, *Statistical Inference in Reinforcement Learning.*

Dec. 2020 **CMStatistics, London, UK (invited)**, *Does the Markov Decision Process Fit the Data: Testing for the Markov Property in Sequential Decision Making.*

Nov. 2020 **Applied Reinforcement Learning (ARL) Seminar (invited)**, *Discussion on "Statistically Efficient Offline Reinforcement Learning", Youtube Link, Bilibili Link.*

Aug. 2020 **Stat seminar at Fudan University, Shanghai, China (invited)**, *On Statistical Learning for Individualized Decision Making with Complex Data.*

Aug. 2020 **Joint Statistical Meeting**, *Does the Markov Decision Process Fit the Data: Testing for the Markov Property in Sequential Decision Making.*

Feb. 2020 **EEE seminar at Imperial College London, London, UK (invited)**, *On Statistical Learning for Individualized Decision Making with Complex Data.*

Feb. 2020 **Stat seminar at University of Cambridge, Cambridge, UK (invited)**, *On Statistical Learning for Individualized Decision Making with Complex Data.*

Jan. 2020 **Stat seminar at SHUFE, Shanghai, China (invited)**, *On Statistical Learning for Individualized Decision Making with Complex Data.*

Dec. 2019 **ICSA Applied Statistics Symposium, Hangzhou, Zhejiang (invited)**, *Statistical inference of the Value Function for Reinforcement Learning in Infinite Horizon Settings.*

Nov. 2019 **Biostat seminar at NCSU, Raleigh, NC, USA (invited)**, *On Testing Overall/Conditional Qualitative Treatment Effects.*

Oct. 2019 **Stat seminar at University of York, York, UK (invited)**, *On Statistical Learning for Individualized Decision Making with Complex Data.*

Aug. 2019 **Stat seminar at Shanghai University of International Business and Economics, Shanghai (invited)**, *On Statistical Learning for Individualized Decision Making with Complex Data.*

Aug. 2019 **Joint Statistical Meeting, Denver, CO, (invited)**, *A Sparse Random Projection-based Test for Overall Qualitative Treatment Effects.*

June 2019 **ICSA Applied Statistics Symposium, Raleigh, NC (invited)**, *A Sparse Random Projection-based Test for Overall Qualitative Treatment Effects.*

Sep. 2018 **Biostat seminar at NCSU, Raleigh, NC (invited)**, *On Statistical Learning for Individualized Treatment Regime.*

Mar. 2018 **Poster presentation at NCSU Graduate Student Research Symposium, Raleigh, NC**, *Maximin-Projection Learning for Optimal Treatment Decision with Heterogeneous Individualized Treatment Effects.*

Aug. 2017 **Joint Statistical Meeting, Baltimore, MD**, *On Testing Conditional Qualitative Treatment Effects.*

Aug. 2016 **Joint Statistical Meeting, Chicago, IL**, *Minimax-Angle Learning for Optimal Treatment Decision with Heterogeneous Data.*

April 2016 **Eastern North American Region Meeting, Austin, TX**, *Minimax-Angle Learning for Optimal Treatment Decision with Heterogeneous Data.*

PYTHON & R PACKAGES

Dec. 2021 **CausalRL**, *Dynamic Causal Effects Evaluation in A/B Testing with a Reinforcement Learning Framework*, available on GitHub.

Oct. 2021 **dgcit**, *Double Generative Adversarial Networks for Conditional Independence Testing*, available on GitHub.

Sep. 2021 **DJL**, *Deep Jump Learning for Off-Policy Evaluation in Continuous Treatment Settings*, available on GitHub.

June 2021 **SAVE**, *Sequential Value Evaluation*, available on GitHub.

May 2021 **D2OPE**, *Deeply-Debiased Off-Policy Interval Estimation*, available on GitHub.

Feb 2021 **LOGAN**, *Testing Mediation Effects Using the Logic of Boolean Matrices*, available on GitHub.

June 2020 **TestMDP**, *Testing for the Markov Property in Sequential Decision Making*, available on GitHub.

Nov. 2019 **JQL**, *Jump Q-Learning for Individualized Interval-Valued Dose Rule (version 3.6.9)*, available on R.

Nov. 2018 **ITRLearn**, *Statistical Learning for Individualized Treatment Regime (version 1.0-1)*, available on R.

Sep. 2018 **ITRSelect**, *Variable Selection for Optimal Individualized Dynamic Treatment Regime (version 1.0-1)*, available on R.

April 2018 **BayesSAE**, *Bayesian Analysis of Small Area Estimation (version 1.0-2)*, available on R.

Aug 2016 **Simplexreg**, *Regression Analysis of Proportional Data Using Simplex Distribution (version 1.3)*, available on R.

Apr. 2016 **arleGP**, *Approximated Restricted Likelihood Estimator for Gaussian Process.*

GRANTS

Dec 2022 **EPSRC New Investigator Award EP/W014971/1**, *Statistical Methods in Offline Reinforcement Learning*, £246,075.

April 2021 **LSE New Research Support Fund (PI)**, *Sparse Reinforcement Learning for Optimal Decision Making In Mobile Health*, £19,553.

Mar. 2020 **Eden Catalyst Fund, LSE Eden Center for Education Enhancement (PI)**, *Module-Level Grade Inflation Analysis at LSE*, £2,071.

TEACHING

- PhD Courses
  - ST510 Foundations of Machine Learning 2021 – present
- Master Courses
  - ST445 Managing and Visualizing Data 2019 – present
  - ST455 Reinforcement Learning 2021 – present
- Undergraduate Courses
  - ST202/206 Probability, Distribution Theory and Inference 2020 – 2021
  - Ou(R) Study Groups 2020 – present

STUDENT ADVISING

- Master Students
  - Kevin Li graduated in 2021
    Thesis Title: Reinforcement Learning for Variable Selection
  - Valentin Nachev graduated in 2021
    Thesis Title: Reinforcement Learning in Mobile Health
  - Henry Tse, Priyanshi Gupta, Anuj Srivastava graduated in 2021
    Thesis Title: Federated Learning / Reinforcement Learning in Cancer Data
  - Georgia Stimpson graduated in 2020
    Thesis Title: Double Machine Learning for Optimal Treatment Decision Rules
  - Wen Yun Ong, Wingchi Yip, Warunya Mahaisawariya graduated in 2020
    Thesis Title: Explore Recommender Systems and Their Embeddings
- PhD Students
  - Tao Ma expect to graduate in 2025

DEPARTMENT SERVICES

- Graduate Admissions Committee, Dept. of Statistics, LSE 2020 – present
- Statistics Seminar Committee (co-chair), Dept. of Statistics, LSE 2020 LT & ST
- Data Science Seminar Committee, Dept. of Statistics, LSE 2021 LT & ST
- Faculty Search Committee, Dept. of Statistics, LSE 2020 & 2022

PROFESSIONAL SERVICES

- Associate Editor: Journal of Royal Statistical Society, Series B 2022 – present
- Associate Editor: Journal of Nonparametric Statistics 2022 – present
- Review Service: NeurIPS; ICML; ICLR; Annals of Statistics; Journal of the American Statistical Association; Journal of Royal Statistical Society, Series B; Biometrika; Journal of Machine Learning Research; Annals of Applied Statistics; Biometrics; Statistics Surveys; Statistics

Methods in Medical Research; Biostatistics; Journal of Biopharmaceutical Statistics; POLS ONE; Science China Mathematics; Australian & New Zealand Journal of Statistics; The Canadian Journal of Statistics; Science China Mathematics; Communications in Statistics - Simulation and Computation.