

Shamindra Shrotriya

I strive to collaboratively solve large-scale open problems using a principled statistical workflow.

Education

- 2017-2022 **PhD Statistics and Data Science**, *Carnegie Mellon University, USA.*
 - **Thesis:** *On Some Problems in Nonparametric and Location-Scale Estimation.*
 - **Advisor:** Matey Neykov.
 - **Committee:** Arun Kumar Kuchibhotla, Yang Ning (Cornell), Alex Reinhart, Alessandro Rinaldo, Larry Wasserman.
 - **Interests:** Density estimation, Location-scale estimation, isotonic regression, Bradley-Terry ranking, wildfire prediction.
- 2017-2019 **M.S. Statistics and Data Science**, *Carnegie Mellon University, USA.*
 - **Coursework:** Convex Optimization, Adv. Statistical Inference, Adv. Statistical Computing, Probability Theory, Statistical Machine Learning, Deep Learning.
 - TA of the Year.
 - GPA: 3.92/4.0
- 2015-2016 **M.A. Statistics**, *University of California at Berkeley, USA.*
 - Elizabeth Scott Memorial Award.
 - Outstanding GSI Award.
 - GPA: 3.9/4.0
- 2003-2007 **BCom (Actuarial/Finance)**, *University of New South Wales, Australia.*
 - Graduated with Distinction.
 - UNSW Co-op Scholar in Actuarial Statistics.

Industry Experience

- 2023-Pres. **Principal Data Scientist**, *Walmart, Retail Intelligence, Bentonville, AR.*
 - Building flexible and accurate demand forecasting tools using Bayesian dynamic linear models.
 - Developing the experimental design framework that underlies strategy behind improving our Customer Experience (CX) score.
- 2023 **Faculty, Co-director (Optum Summer Bridges)**, *Carnegie Mellon University, Pittsburgh, PA.*
 - Co-director of 8 week summer program with 30-35 students selected each year nationally with an emphasis on diversity.
 - Co-taught daily data science lectures focused on both parametric and nonparametric methods and their computational implementation in R.
 - Designed and taught a new data engineering course focused on advanced Unix and SQL concepts.
 - Managed TA staff and helped mentor students in meeting deliverables on their data science research projects based on real-world medical data.

- 2014-2015 **Data Science Infrastructure Team Lead**, *freelancer.com*, Sydney, Australia.
- Designed and implemented a prototype of the new A/B testing framework.
 - Co-designed and administered the entire Extract-Transform-Load (ETL) process written with Go and AWS Redshift.
 - Designed and improved the internal metrics monitoring dashboard.
- 2012-2014 **Data Scientist**, *Quantium Consulting*, Sydney, Australia.
- Led the end-to-end development of the behavioural ‘lifestage’ customer classifier for the entire 7 million Woolworths Supermarket customer base.
 - Led the data-driven electronic marketing strategy for Woolworths Life Insurance which included developing scoring models (GLMs) and conducting A/B tests to optimise response rates.
 - Co-designed and developed the National Australia Bank Online Retail Sales Index.
- 2011-2012 **Microinsurance Fellow**, *UN - International Labor Organization*, Pune, India.
- Wrote a report on the best actuarial pricing practices to be undertaken by microinsurance organisations.
- 2007-2011 **Senior Actuarial Consultant**, *PricewaterhouseCoopers*, Sydney, Australia.
- Built visualization dashboards for monitoring key risk metrics for Insurance Australia Group, Australias’ largest private general insurer.
 - Developed key reporting metrics used by Qantas airlines to assess key drivers and trends behind their Qantas Frequent Flyer Program (the largest customer loyalty program in Australia).

Skills

- Proficient R, Python, PyTorch/Tensorflow, scikit-learn/Pyspark/pandas/numpy, SQL (Redshift/Hive), Git/Github.
- Competent Bash, Make, SAS, Unix, Julia, JAX, Docker, L^AT_EX.

Publications

Papers

1. Bong, H., Li, W., Shrotriya, S., & Rinaldo, A. (2020). Nonparametric Estimation in the Dynamic Bradley-Terry Model. In *AISTATS (Online)*.
2. Li, W., Shrotriya, S., & Rinaldo, A. (2022). sup-norm Bounds of the MLE in the BTL Model under General Comparison Graphs. *Uncertainty in Artificial Intelligence (UAI)*.
3. Dalmaso, N., Shrotriya, S., & Reinhart, A. (2019). Predictive Inference of a Wildfire Risk Pipeline in the United States. *NeurIPS 2019 Workshop on Tackling Climate Change with Machine Learning*.

Under Review (submitted)

1. Shrotriya, S., & Neykov, M. (2022). *Revisiting Le Cam’s Equation: Exact Minimax Rates over Convex Density classes*.
2. Shrotriya, S., & Neykov, M. (2022). *Uniform Location Estimation on Convex Bodies*.
3. Shrotriya, S., & Neykov, M. (2022). *Adversarial Sign-Corrupted Isotonic Regression*.
4. Fogliato, R., Shrotriya, S., & Kuchibhotla, A. K. (2021). *maars: Tidy Inference under the “Models as Approximations” Framework in R*.

Competitions

1. Bong, H., Li, W., & Shrotriya, S. (2019). Efficient Estimation of Distribution-Free Dynamics in the Bradley-Terry Model. *Carnegie Mellon Sports Analytics Conference (Reproducible Research Winner)*.
2. Barter, R., & Shrotriya, S. (2016). Integrated Data Analysis for Early Warning of Lung Failure. *ODBMS.org (Geisinger Competition Winner)*.

Awards and Honors

- 2021 rstudio::global(2021) Diversity Scholar. RStudio
- 2020 NGC Wildfire Research Scholar. American Australian Association
- 2020 TA of the Year. Carnegie Mellon University
- 2019 NeurIPS Climate Change Workshop Travel Award.
- 2019 CMSAC Best Paper Award. Carnegie Mellon University
- 2017 Outstanding Graduate Student Instructor. University of California, Berkeley
- 2016 Elizabeth Scott Memorial Award. University of California, Berkeley
- 2016 Best Paper and Competition Winner. Geisinger Health Collider Project
- 2012 Microinsurance Fellowship. UN - International Labor Organization
- 2007 Associate of the Institute of the Actuaries Australia.
- 2003 Co-op Industrial Scholarship in Actuarial Studies. University of New South Wales, Australia
- 2003 Council Tertiary Scholarship. Parramatta Council, Sydney, Australia
- 2002 Entry Award Scholarship in Engineering (declined). University of Sydney, Australia
- 2002 Australian Students Prize for Academic Excellence. Australian Federal Government
- 2002 Premier's Award for Academic Excellence. NSW Government, Australia
- 2002 University Admissions Index (UAI) 99.90 (top 0.1% in State).

Presentations

Workshops

- 2019-12-15 **Predictive Inference of a Wildfire Risk Pipeline in the United States (Spotlight)**, *NeurIPS 2019 Climate Change Workshop*, Vancouver, BC.
- 2019-11-02 **Efficient Estimation of Distribution-free dynamics in the Bradley-Terry Model**, *CMSAC Reproducible Research Competition*, Pittsburgh, PA.

Posters

- Dec 2018 **Predictive Inference of a Wildfire Risk Pipeline in the United States**, *NeurIPS 2019 Climate Change Workshop*, Vancouver, BC.

Dec 2018 **Efficient Convex Estimation of the Time Varying Bradley-Terry Model**, *COPTS conference*, Pittsburgh, PA.

Talks

Jul 2021 **maars: Tidy Inference under misspecified statistical models in R**, *useR! 2021: The R Conference (Regular Talk)*, Virtual.

Dec 2018 **Introduction to the Tidyverse**, *STAT 36-350*, Pittsburgh, PA.

Dec 2018 **Functional Connectivity in iEEG Data**, *Advanced Data Analysis Presentation*, Pittsburgh, PA.

Jul 2016 **Predicting COPD in pneumonia patients**, *Geisinger Collider Project*, Berkeley, CA.

Research Experience

2018-2019 **Advanced Data Analysis (ADA) Project**, *Carnegie Mellon University*, Pittsburgh, PA.

- Advised by: Prof. Max G'Sell and Prof. Avniel Singh Ghuman.
- Investigated the dynamic functional connectivity in human epilepsy patients using iEEG data.
- Successfully presented oral defense of research work.

2017 **Research Associate**, *University of California*, Berkeley, CA.

- Advised by: Prof. Bin Yu and Prof. Ben Brown.
- Investigated the statistical properties of the iterative Random Forests (iRF) algorithm.
- Co-developed the Python implementation of the iRF algorithm.
- Helped complete a successful four-year NSF BIGDATA grant proposal for this project.

2016 **Geisinger Collider Project**, *University of California*, Berkeley, CA.

- Joint work with Rebecca Barter (UC Berkeley).
- Investigated using Electronic Medical Record data to determine whether a pneumonia patient will develop Chronic Obstructive Pulmonary Disease (COPD).
- Winner - Best paper award and overall competition.

Teaching Experience

Head Teaching Assistant

- 2020 **STAT 36-350 (Statistical Computing)**, *Carnegie Mellon University*, Pittsburgh, PA.
- Instructor: Prof. Peter Freeman.
 - Developed R programming course materials.
 - Managed 9 TAs and grading via Gradescope/Canvas, held office hours.
- 2019 **STAT 36-350 (Statistical Computing)**, *Carnegie Mellon University*, Pittsburgh, PA.
- Instructor: Prof. Peter Freeman.
 - Developed R programming course materials.
 - Managed 7 TAs and grading process, held office hours.
- 2018 **STAT 36-700 (Intermediate Theoretical Statistics)**, *Carnegie Mellon University*, Pittsburgh, PA.
- Instructor: Prof. Larry Wasserman.
 - Wrote HW solutions, helped with HW/exam design.
 - Managed other TAs and grading process, held office hours.
- 2016 **STAT133 (Computing with Data)**, *University of California*, Berkeley, CA.
- Instructor: Prof. Gaston Sanchez.
 - Managed other TAs and grading process, Held weekly R tutorial sessions.
 - Winner - Outstanding Graduate Student Instructor award.

Teaching Assistant

- 2021 **rstudio::global(2021)**, *RStudio*, Pittsburgh, PA.
- Instructor: Prof. Mine Çentikaya-Rundel.
 - Materials: <https://wtf-teach.netlify.app/team.html>.
 - Managed zoom questions and breakout room discussions.
- 2019 **STAT 36-750 (Graduate Statistical Computing)**, *Carnegie Mellon University*, Pittsburgh, PA.
- Instructor: Prof. Alex Reinhart.
 - Wrote HW solutions, graded 300+ Github Pull Requests, held office hours.
- 2018 **STAT 36-350 (Statistical Computing)**, *Carnegie Mellon University*, Pittsburgh, PA.
- Instructor: Prof. Ryan Tibshirani.
 - Reviewed course materials, held office hours.
- 2017 **STAT 36-401 (Modern Regression)**, *Carnegie Mellon University*, Pittsburgh, PA.
- Instructor: Prof. April Galyardt.
 - Reviewed course materials, held office hours.

Service

Reviewing

- 2020 **Program Committee**, *NeurIPS 2020 Workshop, Tackling Climate Change with Machine Learning*, (Held Virtually).
- Peer-reviewed workshop papers.

- 2020 **Program Committee**, *ICLR 2020 Workshop, Tackling Climate Change with Machine Learning*, Addis Ababa, Ethiopia.
- Peer-reviewed workshop papers.

Software

I enjoy using and contributing to open-source scientific software. I've co-developed the following software packages in R and python.

- 2021 **Co-creator of the maars R package.**
- Joint work with Riccardo Fogliato and Arun Kumar Kuchibhotla
Tidy Inference under the 'Models as Approximations' Framework in R
 - <https://shamindras.github.io/maars/>
- 2017 **Co-developer of the iRF python package.**
- Python package for the Iterative Random Forests (iRF) algorithm to detect predictive and stable high-order interactions
 - <https://github.com/Yu-Group/iterative-Random-Forest>

References

Available upon request.