

BENJAMIN LEROY

Email: benjaminpeterleroy@gmail.com

Github: github.com/benjaminleroy · Website: benjaminleroy.github.io

EDUCATION

Doctorate of Philosophy, Statistics CARNEGIE MELLON UNIVERSITY <i>Title:</i> Conformal Prediction for Simulation Models <i>Advisor:</i> Chad Schafer	August 2016 - December 2021 (expected)
Masters, Statistics CARNEGIE MELLON UNIVERSITY	August 2016 - May 2017
Bachelor of Arts, Statistics & Applied Mathematics (Double) UNIVERSITY OF CALIFORNIA, BERKELEY	August 2013 - Dec 2015

WORK EXPERIENCE

Nike, <i>Global Sourcing and Manufacturing, Center of Excellence</i> <i>Graduate Data Science Intern</i>	June 2020 - August 2020 Beaverton, OR (remote)
<ul style="list-style-type: none">Assessed Nike sourcing strategies and their impact on factory flexibility<ul style="list-style-type: none">Developed an internal python-package to model and implement a simulation pipelineDesigned forecasting tools and bootstrapping procedures to model potential changes in demandReplicated human allocation methods using linear programmingPresented for and collaborated with stakeholders and a transition team to increase impact of the project	
Lawrence Berkeley National Laboratory, <i>Cosmology Department</i> <i>Science Undergraduate Laboratory Intern</i>	January 2016 - May 2016 Berkeley, CA
<ul style="list-style-type: none">Adapted traditional Bayesian mixture models in pymc3 with truncated Dirichlet process for Supernovae subclass exploration	

SOFTWARE DEVELOPMENT

- EpiCompare.** tidyverse-style R-package for epidemics & epidemic model comparison in a time-invariant framework
- mssim.** private python-package for product demand allocation simulation (for Nike)
- TCpredictionbands.** R-package that implements all methodology in simulation-based prediction bands for tropical cyclones (see Selected Papers)
- golf: a shiny app.** R-based shiny app, showcasing collaborative work with Professor Schlossman analyzing US Open golf tournaments

PAPERS

***LeRoy and Zhao.** “MD-split+: Practical local conformal inference in high dimensions”. ICML 2021’s Work on “Distribution-free uncertainty quantification”. arXiv: [2107.03280](https://arxiv.org/abs/2107.03280).

***Gallagher and LeRoy.** **EpiCompare: A pipeline for epidemic comparison and analysis**, (2021). R package: EpiCompare (github.com/skgallagher/EpiCompare). In progress.

Frisoli, LeRoy, and Nugent. **A novel record linkage interface that incorporates group structure to rapidly collect richer labels**, (2019) IEEE’s International Conference on Data Science and Advanced Analytics (DSAA).

***Dalmassò, Dunn, LeRoy, and Schafer.** **A flexible pipeline for prediction of tropical cyclone paths**, (2019), In: ICML 2019’s 1st Workshop on “Climate Change, how can AI help?”, arXiv: [1906.08832](https://arxiv.org/abs/1906.08832). *equal contributions. R package: TCpredictionbands (github.com/Mr8ND/TC-prediction-bands/)

Dubovik, Starosvetsky, LeRoy, Normand, Admon, Alpert, Ofran, G’Sell, and Shen-Orr. **Immune cellular homeostasis in early life is determined by genetic variants of cellular production and turnover**, (2018). bioRxiv: [256073](https://doi.org/10.1101/256073).

*Equal contribution

PRESENTATIONS

LeRoy and Schafer. “**Conformal Prediction for Simulation Models**”. International Conference of Machine Learning: Workshop on “Distribution-free Uncertainty Quantification”, Virtual, July 2021. Contributed Poster.

LeRoy and Zhao. “**MD-split+: Practical Local Conformal Inference in High Dimensions**”. International Conference of Machine Learning: Workshop on “Distribution-free Uncertainty Quantification”, Virtual, July 2021. Contributed Poster.

Frisoli, *LeRoy, and Nugent*. “A novel record linkage interface that incorporates group structure to rapidly collect richer labels**”. IEEE: Data Science and Advanced Analytics Conference, Washington, DC, October 2019. Contributed talk.

LeRoy. “**Floating points: living in a discretized, computerized world**”. [useR: Pittsburgh](#), Pittsburgh, PA, August 2019. Invited talk.

Dalmassó, *LeRoy, Dunn, and Schafer*. “A flexible pipeline for prediction of tropical cyclone paths**”. International Conference of Machine Learning: Workshop on “Climate Change, how can AI help?”, Long Beach, CA, June 2019. Contributed Poster.

ADVISING AND CONSULTING (PROJECT MANAGEMENT)

Data Science Initiative (DSI) August 2018 - December 2019
PhD Fellow, Carnegie Mellon University - Department of Statistics and Data Science

- Advised, managed, and taught 3 teams of undergrad students (Finance, Engineering, & Marketing clients)
- Worked directly with corporate partners to ensure smooth project deliveries

Summer Undergrad Research Experience (SURE) May 2018 - July 2018
Graduate Advisor, Carnegie Mellon University - Department of Statistics and Data Science

- Advised a team of undergrad students investigating trends in human trafficking
- Taught summer undergrad students Statistical and Data Science techniques in R

Library Digital Humanities Research (dSHARP) September 2019 - December 2019
PhD Consultant, Carnegie Mellon University Library

- Consulted for CMU History Professor Schlossman under Andrew W. Mellon Digital Humanities Seed Grant
- Assessed US Open tournament player performances and developed story-telling tools in R's *shiny* applications

TEACHING

Instructor

Statistical Computing (Undergrad) Summer 2019

- Updated R-based class by introducing *tidyverse* concepts, OOP, and package development

Statistical Graphics and Visualization (Undergrad) Summer 2017

- Developed course for first summer class, emphasized visualization theory

Teaching Assistant (non-chronological)

Intermediate Statistics (PhD) Spring 2018

Statistics and Machine Learning I/II (Masters) Fall 2018

Experimental Design and Time Series (Masters) Fall 2019

Text Mining (Masters/Undergrad) Spring 2019

Advanced Methods for Data Analysis (Undergrad) Fall 2020

Modern Statistics (Undergrad) Fall 2016

Statistical Graphics and Visualization (Undergrad) Fall 2017 & Spring 2017

Course Development

Statistics and Machine Learning I/II (Masters) Fall 2018

Statistical Graphics and Visualization (Undergrad) Fall 2017

LEADERSHIP

PhD Student Representative, Computing Committee, CMU Statistics & Data Science 2019 - Present

Committee Chair, PhD Committee on Healthcare and Finance, CMU Statistics & Data Science 2018 - 2019

Department

Judge, Pittsburgh Data Jam (high school data science competition) 2018 - 2020

Mentor, CMU Statistics and Data Science Mentorship Program 2018 - 2021

AWARDS

Nike Summer Intern Analytics Team Hackathon (multi-week) - 1st Place	2020
Andrew W. Mellon Digital Humanities Seed Grant (under Professor Schlossman)	2019
IEEE: Data Science and Advanced Analytics Conference, Travel Award	2019
Department Citation, for High Undergraduate Achievement <i>from Statistics Department, UC Berkeley</i>	2016
Berkeley/Amazon Collider <i>by Amazon Logistics</i> (UC Berkeley wide, multi-week competition) - 2nd Place	2015
Cornerstone Research Datathon <i>by Cornerstone</i> (UC Berkeley wide) - 1st Place	2015
Robert F Argella Scholarship (Mathematics)	2013
John M Soares Memorial Economics Scholarship	2013

CODING/TECHNICAL SKILLS

Proficient: Python, R, Git, L^AT_EX

Familiar: Bash, C++, SQL, HTML/CSS