BENJAMIN LEROY

Lead Data Scientist, Statistics Ph.D.

San Francisco, CA | benjaminpeterleroy@gmail.com | 707-328-0714 | benjaminleroy.github.io | benjaminleroy

PROFESSIONAL SUMMARY

Lead Data Scientist specializing in machine learning, operations research, and uncertainty quantification. Experienced in developing scalable, data-driven models and simulations for complex systems including supply chain resilience, extreme weather, and computer vision.

RELEVANT EXPERIENCE

Lead Data Scientist (Founding)
Oct 2025 – Ongoing

Nudge and Level Labs — San Francisco, CA

- **Engineered** a synthetic data strategy using VLMs for scene generation and pose-consistent rotation; implemented VLM-as-a-judge and human-in-the-loop checks to bypass the need for proprietary multi-camera datasets.
- **Architected** a hybrid system bridging differentiable optimizations with LLM-reasoning for physics-true clothing estimation; designed geodesic-based validation (<4% MAPE) to ensure high-fidelity human mesh estimation.
- **Directed** the strategy for commercializing academic research and developed the end-to-end roadmap for human-clothing physics estimation, transitioning theoretical models into a production-ready product suite.

Senior Data Scientist
Feb 2022 – Sep 2025

Nike, Product Supply Chain - Data Science — Beaverton, OR

- **Causal Inference & Measurement:** Developed a generative counterfactual framework and monotonic regression models to quantify time-to-market costs, identifying ($41M in annual EBIT impact) across 30+ disparate data streams.
- **Supply Chain Optimization:** Led 2–5 person teams to deploy mixed-integer models and rapid tariff prototypes, preserving ($46M in annual EBIT) and enabling agile responses to global trade shifts.
- **Systems Engineering:** Accelerated digital twin simulation runtime 25x via profiling and parallelization; standardized org-wide engineering best practices including CI/CD and GenAI coding standards.
- **Technical Communication:** Authored comprehensive technical reports and thought-leadership papers for project sponsors and developed training and communication tools to bridge knowledge gaps across partner organizations.

PhD Student
Aug 2016 – Dec 2021

Carnegie Mellon University, Statistics and Data Science — Pittsburgh, PA

- **Developed** methods to extend conformal inference (prediction uncertainty) to complex spaces leveraging simulation models. Demonstrated improvements on specified generative models and real world problems.
- **Spearheaded** four self-directed research publications and two R packages and partnered with cross-disciplinary experts on a visual storytelling initiative.
- **Optimized** code for high-performance computing/machine learning on a supercomputer & Azure servers and **launched** departmental computing documentation wiki.

EARLY EXPERIENCE

Graduate Data Science Intern, **Nike (Supply Chain)** — June – Aug 2020
Forecasting, Simulation & Optimization - Supply Chain Resiliance

Research Intern, **Lawrence Berkeley National Lab** — Jan – May 2016
Bayesian Mixture Modeling

TECHNICAL SKILLS

Programming
- Python: *Pandas, PyTorch, NumPy/SciPy, Matplotlib, OR-Tools/GurobiPy, LangChain*
- R: *tidyverse, geosphere, rgeos, ggmap, ggplot*
- Other: SQL · Git · AIMMS

Cloud & AI Platforms
AWS · DataBricks · Azure · GitHub Copilot

Statistical Methods
Uncertainty Quantification · Optimization (Discrete and Continuous) · Economic Cost Modeling · Causal Inference · Data Viz

Machine Learning
Neural Networks · Model Interpretation · Forecasting

Systems & Engineering
CI/CD · Unit & Regression Testing · Version Control · Project Architecture

EDUCATION

Ph.D. in Statistics, 2021
Masters in Statistics, 2017
Carnegie Mellon University

Thesis: Simulator-Enabled Conformal Prediction

2016 – 2021 — Pittsburgh, PA

B.A. in Statistics & Applied Mathematics
University of California, Berkeley

2013 – 2015 — Berkeley, CA

SELECT PAPERS

- **B. LeRoy*** and D. Zhao*, "`Md-split+`: practical local conformal inference in high dimensions," *ICML 2021's Workshop on "Distribution-free uncertainty quantification"*, 2021. arXiv: 2107.03280.
- S. Gallagher* and **B. LeRoy***, "`Epicompare`: a pipeline for epidemic comparison and analysis," 2020. [Online]. Available: https://github.com/skgallagher/EpiCompare/.
- N. Dalmasso*, R. Dunn*, **B. LeRoy***, and C. Schafer, "A Flexible Pipeline for Prediction of Tropical Cyclone Paths," *ICML Workshop: "Climate Change: How can AI Help?"*, 2019. arXiv: 1906.08832.

* equal contribution.