Duncan Clark

Data Scientist & Researcher

Email : duncan-clark@outlook.com
Phone : 310-254-6345
duncan-clark.github.io

Education

- **University of California, Los Angeles** — Los Angeles, CA
  Doctor of Philosophy in Statistics GPA: 3.89/4.00 — *Sep. 2017 – June. 2022*
- **University College London** — London, United Kingdom
  Combined Bachelors and Masters (MSci) in Mathematics 4.0 GPA equivalent — *Sep. 2011 – June. 2015*

Experience

- **Meta** — New York City, NY
  Research Data Scientist — *Aug. 2022 - Present*
  - **Differentially Private Data Storage**: Advised software engineering team on scope for differentially private aggregation with a meta analysis of SQL queries. Identified common privacy leaks, resulting in improved safeguards to maintain user privacy in the data warehouse.
  - **Privacy Footprint**: Assessed the impact of data minimization efforts. Worked cross functionally with privacy experts and software engineers to rapidly ramp up on the domain. Identified 100s of thousands of engineering hours saved. The impact assessment resulted in strategic investment in data minimization, reducing the cost of privacy compliance.
  - **Privacy Cost Graph**: Metrics development on the cost of engineering hours to ensure Meta infrastructure is privacy compliant. Proposed graph model of the flow of data in the warehouse with respect to privacy cost. Quantified the hours saved due to deidentification efforts, using causal inference techniques to estimate average treatment effects. These robust analyses drove decision makers to invest in high return areas to maximise cost savings.
  - **Differentially Private Neural Network Training**: Ramped up on differentially private deep learning. Implemented a cutting edge differential privacy accountant for the open source private training module Opacus.

  Research Data Scientist Intern — *Jun. 2021 - Sept. 2021*
  - **Classification of semantic data types**: Advised data engineers on sampling methods to achieve unbiased performance estimation. Understood and iterated on in house deep learning model for data type prediction. Proposed simpler alternative model with dramatic training time decrease, whilst improving performance.

  Research Data Scientist Intern — *Jun. 2020 - Sept. 2020*
  - **App performance prediction**: Predicted production app performance regressions with beta testing data. Adapted quickly to Facebook scale, trillion line data, learned internal tools, machine learning workflows. Built, tested and iterated on predictive models using complex, noisy beta testing data. Communicated and collaborated with to software engineers achieving a false positive reduction of 90%.
- **Xafinity Consulting** — Reading, United Kingdom
  Actuarial Analyst — *Sep. 2015 - Sep. 2017*
  - **Pension Scheme Actuarial Work**: General defined benefit pension scheme liability modeling e.g. valuations, severance costs, transfer values, accounting disclosures, assumption derivation.
  - **Pension Insurance**: Selected for specialist pension insurance team due to strong understanding of liability models. Rapid turn around on complex data sets and bespoke scheme rules for a wide range of schemes. Communicated, advised and collaborated with pricing actuaries on multiple $> £1$ billion deals.
  - **Institute Training**: Part qualified IFoA actuary 60% of exams passed (CT 1-8). Strong understanding of foundational finance concepts.

Research & Coursework

University of California, Los Angeles

- **LOLOG Catalog Project**: Verified power of cutting-edge social network model. Requested network data from more than 20 authors, understood context, processed data, fit complex stochastic models, suggested improvements to parameter estimation.
- **Bayesian LOLOG Project**: Developed Riemannian manifold MCMC algorithm with model specific second order information, to provide full Bayesian inference for an intractable model class.

- **Stochastic Network Causal Inference**: Extended state of the art research and developed a general Bayesian framework for causal inference in networks where the network is uncertain.
- **Coursework:**: Machine Learning, Statistical Programming, Theoretical Statistics, Causal Inference, Latent Variable Modelling, Regression and Classification Algorithms, Point Process Models, Neural Networks.

Projects

- **Numerai Data Science Tournament**: Have participated for 2 years in the "Worlds hardest data science tournament". Predictions based on unstructured data are used to power a hedge fund. The hedge fund uses a cryptocurrency stake weighted meta model to generate exceptional returns. Current maintainer of models that are in the top 2% of models on an annual return basis. Deep learning methods primary method for success. Also maintain automatic submission instances through AWS.
- **Strava Team Time Trial Analyses**: Developed algorithms for detecting the position of riders in a cycling team time trial race. Used, GPS, power and heart rate sensor data. Used this to asses the performance and relative strengths and weaknesses of riders in the team.

Programming Skills

- Python, R, SQL, C++

Awards

- **UCLA Department of Statistics:** Outstanding PhD Dissertation Award 2022

Interests

- **Cycling**: Category 1 racing speedster, 2nd Overall 2022 USA Collegiate National Cycling Championships
- **Cricket**: Test Match Special super fan, and all-round cricket statistics geek

Publications

1. Duncan A. Clark and Mark S. Handcock. Comparing the real-world performance of exponential-family random graph models and latent order logistic models for social network analysis. *Journal of the Royal Statistical Society Series A*, 185(2):566–587, April 2022
2. Duncan Clark, James Macinko, and Maurizio Porfiri. What factors drive state firearm law adoption? an application of exponential-family random graph models. *Social Science Medicine*, 305:115103, 06 2022
3. James Macinko, Diana Silver, Duncan A. Clark, and Jennifer L. Pomeranz. The diffusion of punitive firearm preemption laws across u.s. states. *American Journal of Preventive Medicine*, 2023

Under Review:

1. Duncan A. Clark and Mark S. Handcock. An approach to causal inference over stochastic networks, 2021
2. Duncan Clark, James Macinko, and Maurizio Porfiri. A framework for credible prediction of state firearm legislative activity with dynamic network models, 2023