Beniamino Green

Email: beniamino.green@tuta.com | Personal Website | Github

Data scientist passionate about quantitative research and numerical methods. Skilled at using Python, R and Rust for quantitative social science research and for developing statistical software.

Education

2021-2022 — Masters in Social Analysis and Research
Brown University
4.0 GPA

2018-2021 — B.Sc. Philosophy, Politics and Economics
University College London
Program of study: Politics and Philosophy
First Class Honors, with Data Science Concentration

Professional Expeience

2022-2024 — Pre-Doctoral Fellow
Yale University
Applying new methods at the intersection of Machine Learning and Causal Inference, with special attention to applications to Medicaid plan management.

Publications:

- Beniamino Green. "Zoomerjoin: Superlatively-Fast Fuzzy Joins". In: *Journal of Open Source Software* 8.89 (2023), p. 5693. DOI: 10.21105/joss.05693. URL: https://doi.org/10.21105/joss.05693
- Chris Frenier, Beniamino Green, Jacob Wallace, et al. "Financial Health Among Louisiana Medicaid Enrollees". In: *JAMA Health Forum* 5.10 (Oct. 2024), e243028. ISSN: 2689-0186. DOI: 10.1001/jamahealthforum.2024.3028. URL: http://dx.doi.org/10.1001/jamahealthforum.2024.3028

Software

- **Zoomerjoin R Package** Rust and R-package for fuzzy-joining datasets with millions or hundreds of billions of rows. Implimented in R and Rust.
- **Cragg R Package** an implementation of the Cragg-Donald and Stock-and-Yogo tests for weak instruments in R. Currently receives between 300-500 downloads per month.

Skills

Programming Languages — R (advanced), Python (proficient), Rust (proficient), Unix Shell / Bash, MC-STAN
Markup Languages — LaTeX,Sweave, R-Markdown, HTML, CSS, Markdown
Prototyping — Familiar with Arduino, Jetson Nano and Raspberry-pi architectures for development
Frameworks and Tools — Git, Linux, Vi/Vim, Office Suite

Relevant Methods Training

Bayesian Statistics — Training in Bayesian statistics with an emphasis on numerical approaches, including coverage of the EM algorithm, Gibbs sampling, MCMC, and importance sampling.

Introduction to Computational Linear Algebra — Fundamental algorithms in computational linear algebra with special focus on numerical stability.

Causal Methods — Observational designs (regression, matching), quasi-experimental methods (instrumental variables, and regression discontinuity designs), and panel-data methods (difference in differences, synthetic control methods).

Recent Applications of Probability and Statistics — Maximum entropy principle for systems and large deviations, bias-variance dilemma in nonparametric inference, and computational methods for estimating graphical models.

Machine Learning — Tree methods, boosting approaches, naive bayes, SVM's and neural networks.

Measurement in Political Science — Foundational measurement and dimension-reduction methods in the social sciences, including PCA, EFA, and item response methods.

Event History Methods — Modeling for time-until-event problems, including Kaplan-Meier product-limit estimates, discrete-time logit models, and Cox proportional-hazards regression models.

Data Analysis — Data analysis and statistics in R, with special emphasis on regression models (logit regression, hierarchical models, MRP), and text analysis.

Applied Data Analysis — Simulation methods, bias-variance dilemma with a focus on regularized regression methods.