Justin D. N. Weltz

justin.weltz@duke.edu | +1 917-653-3771

EDUCATION

Duke University, PhD, Statistical Science, Durham, NC   Aug. 2019–Present
Thesis: "Optimizing the Study and Assistance of Hard-to-Reach Populations" (*working title*)
Advisors: Eric Laber and Alexander Volfovsky

Pomona College, Bachelor of Arts, Mathematics Major, Claremont, CA   Aug. 2015–May 2019
GPA: 3.96/4.00; Phi Beta Kappa (2018)

RESEARCH INTERESTS

Social network sampling, model-based reinforcement learning, experimental design, algorithmic fairness, large language models and quantitative methods in literary studies

SELECTED RESEARCH

Duke University   Nov. 2020–Present
- Advance methodology used to estimate the size and characteristics of understudied groups such as undocumented workers, intravenous drug users and unhoused people
- Develop network sampling methods to study and assist hidden, at-risk populations using reinforcement learning
- Create tools to measure the impact of interventions on infectious diseases, marketing promotions and other phenomena that disseminate over social networks
- Formulate efficient algorithms to identify optimal experimental designs with heteroskedastic noise

Pomona College   May 2018–May 2019
- Simulated the effect of over-policing using a multistage stochastic process
- Analyzed the false positive rates of logistic models trained on this simulated data to study the effect of over-policing on fairness in algorithms

PUBLICATIONS & WORKING PAPERS

Weltz, J., Volfovsky, A., Laber, E.B. "Reinforcement learning respondent driven sampling." *Manuscript in progress.*

Weltz, J., Volfovsky, A., Laber, E.B. "Improving hidden population size estimation with auxiliary information." *Submitted to AISTATS* (2024).

Weltz, J., Fiez, T., Volfovsky, A., Laber, E., Mason, B., Nassif, H., & Jain, L. "Experimental designs for heteroskedastic variance." *NeurIPS* (2023).

Weltz, J., Volfovsky, A., Laber, E.B. "Reinforcement learning methods in public health." *Clinical Therapeutics* (2022).

Weltz, J., Hardin, J. "Over-policing and fairness in machine learning." Undergraduate thesis (2019).

TALKS & POSTERS

"Experimental designs for heteroskedastic variance,"   Dec. 2023
     Contributed Poster at the Conference on Neural Information
     Processing Systems (NeurIPS)

"Improving hidden population size estimation with auxiliary information,"   Oct. 2022
> Invited Talk at the International Conference on Advances in Interdisciplinary
> Statistics and Combinatorics (AISC)

"Improving hidden population size estimation with auxiliary information,"   June 2022
> Contributed Poster at the International Society for Bayesian Analysis
> (ISBA) World Meeting; received **Best Poster Award**

TEACHING & MENTORSHIP

Teaching Assistant *(course selection forthcoming)*   Spring 2024

Graduate Mentor   Spring 2021–Present

Teaching Assistant, STA642: Time Series and Dynamic Models   Fall 2021

Teaching Assistant, STA721: Linear Models   Fall 2020

Teaching Assistant, STA102: Introduction to Biostatistics   Spring 2020

PROFESSIONAL EXPERIENCE

Amazon, New York, NY   June 2022–April 2023
Applied Research Scientist Intern, Machine Learning Team, Amazon Prime
- Produced a series of reinforcement learning algorithms in Python to efficiently identify the optimal characteristics of prime membership advertisements through strategic sequential testing
- Adapted these models to complex data settings, including sparse signals and heteroskedastic noise
- Created realistic simulation structures to test best-advertisement identification algorithms ("best-arm identification" in the multi-armed bandit literature) against A/B testing and other competitors in high-dimensional data contexts

Amazon, New York, NY   June–Aug. 2021
Applied Research Scientist Intern, Sponsored Products
- Designed a reinforcement learning algorithm to better select advertisements sourced by a variety of machine learning algorithms using query features and partially observed customer behavior signals
- Implemented the infrastructure for evaluating the advantages of modifications to advertisement sourcing algorithms on downstream outcomes in Spark
- Ran off-line experiments testing the utility of my algorithm over a week of Amazon search query traffic and found that it was consistently capturing more than 3% of the high-quality ads missed by the current method

TECHNICAL SKILLS

Languages: Python, Stata, Scala, Matlab, R
Databases: MySQL, Spark
Familiar: Java, HTML