

Lorenzo Masoero

EMAIL: lo [dot] masoero [at] gmail [dot] com

WEB: <http://lorenzomasoero.com>

Senior Applied Scientist | Weblab at Amazon

September 2021 - Present

I am a Senior Applied Scientist at Weblab, Amazon's largest A/B testing platform, where I work on advancing experimentation at scale. My work spans across open-ended research, academic paper writing, and the hands-on implementation of large-scale algorithms in production, directly impacting hundreds of millions of customers daily. At Weblab, I collaborate with leading experts in causal inference and drive innovation in:

- **Online Experimentation with Interference:** I have pioneered experimental designs for large-scale testing of interference in online experiments, co-leading the implementation of the first large “multi-randomized” experiments at Amazon. These experiments, unlike classical A/B tests, jointly randomize across multiple units in the experiment (e.g., customers and advertisers). Via these designs, we can capture “spillover effects” that would otherwise be overlooked. For example, these have been implemented to precisely capture the indirect behavioral response of advertisers induced by a change in pricing schemes of ad-auctions.
 - Key papers include:
 - * *Experimental Design in Marketplaces* (Statistical Science)
 - * *Multiple Randomization Designs* (In submission)
 - * *Efficient Switchback Experiments via Multiple Randomization Designs* (Code@MIT 2023)
- **Leveraging Covariates for Enhanced Causal Estimates:** I partnered with engineering teams to design and implement a distributed analytic engine using PySpark, enabling more accurate causal estimates in thousands of daily experiments. This innovation was backed by research on the value of covariates in online A/B testing, contributing to improvements in experiment analytics.
 - Featured paper: *Leveraging Covariate Adjustments at Scale in Online A/B Testing* (KDD CDPD, 2023).

Previously, I completed my PhD at MIT in EECS in August 2021 under the supervision of Professor Tamara Broderick. During my PhD, I worked on Bayesian inference, with an emphasis on the development of efficient and theoretically grounded approaches for optimal design of experiments in genomic studies.

Employment

2024-	Applied Scientist III, Weblab, Amazon.
2022-2023	Applied Scientist II, Weblab, Amazon.
2021	Applied Scientist, Weblab, Amazon.
2020	Applied Science Intern, CoreAI, Amazon.

Education

2016-2021	PhD, Electrical Engineering and Computer Science, Massachusetts Institute of Technology
2015 - 2016	MA in Statistics and Applied Mathematics, with distinction, Collegio Carlo Alberto
2015 - 2016	MA in Quantitative Finance, 110/110 magna cum laude, Università degli Studi di Torino
2012 - 2014	BA in Economics, 110/110 cum laude, Università degli Studi di Torino

Scholarships and Awards

2020	SBSS Best Student Paper Award (ASA)
2020	Bayes Comp Travel Award
2018	BNP@NeurIPS Award
2017	Andrew (1956) and Erma Viterbi Fellowship
2016	Best Graduate Student of the Year (ATLEC)
2015 - 2016	Graduate Allievi Honors Program Scholarship, Collegio Carlo Alberto, Moncalieri
2012 - 2014	Undergraduate Allievi Honors Program Scholarship, Collegio Carlo Alberto, Moncalieri

Research

- **“How do Advertisers React to Changes in Ad Auction Pricing? Evidence from a Clustered Multiple Randomization Experiment”** (KDD 2024, Workshop on Two-Sided Marketplace Optimization)
- **“Multi-Armed Bandits with Network Interference”**; arXiv preprint, 2405.18621. (Agarwal, Agarwal, M., Whitehouse)
- **“Double trouble: Predicting new variant counts across two heterogeneous populations”**; arXiv preprint, 2403.02154. (Shen, M., Schraiber, Broderick)
- **“Improved prediction of future user activity in online A/B testing”**; arXiv preprint, 2402.03231. (M., Beraha, Richardson, Favaro)
- **“A Nonparametric Bayes Approach to Online Activity Prediction”**; arXiv preprint, 2401.14722. (Beraha, M., Richardson, Favaro)
- **“Multiple Randomization Designs: Estimation and Inference with Interference.”** arXiv preprint, 2401.01264. (M., Vijaykumar, Richardson, McQueen, Rosen, Burdick, Bajari, Imbens)
- **“Leveraging covariate adjustments at scale in online A/B testing”**; 2023 KDD Workshop on Causal Discovery, Prediction and Decision, PMLR 218:25-48, 2023. (M., McQueen, Hains)
- **“Experimental Design in Marketplaces”**; Statistical Science, 38(3): 458-476 (August 2023). (Bajari, Burdick, Imbens, M., McQueen, Richardson, Rosen). Manuscript available at Amazon Science.
- **“Independent Finite Approximations for Bayesian Nonparametric Inference”**; Bayesian Analysis, Advance Publication 1-38 (2023). Manuscript available on arXiv [<https://arxiv.org/pdf/2009.10780.pdf>] (Nguyen, Huggins, M., Mackey, Broderick)
- **“Double trouble: Predicting new variant counts across two heterogeneous populations”**; Learning Meaningful Representations of Life, Neural Information Processing Systems Workshop, 2022. (Shen, M., Schraiber, Broderick). [Preprint available on OpenReview]

- “**Cross-Study Replicability in Cluster Analysis**”. *Statistical Science*, 38(2): 303-316 (May 2023). (M., Thomas, Parmigiani, Tyekucheva, Trippa). Manuscript available on arXiv [<https://arxiv.org/pdf/2202.01910.pdf>]
- “**Multiple Randomization Designs**”; Manuscript available on arXiv [<https://arxiv.org/pdf/2112.13495.pdf>] (Bajari, Burdick, Imbens, M., McQueen, Richardson, Rosen)
- “**Bayesian nonparametric strategies for power maximization in rare variants association studies**”; **Spotlight** at Learning Meaningful Representations of Life, Neural Information Processing Systems, 2021 [[poster](#)]. Manuscript available on arXiv [<https://arxiv.org/pdf/2112.02032.pdf>] (M., Schraiber, Broderick)
- “**The Chicago Police Department dataset**”; Neural Information Processing Systems Track on Datasets and Benchmarks, 2021, [Dataset repository](#), [current release \(vo.1\)](#) and [documentation](#) (Horel, M., Agrawal, Roithmayr, Campbell)
- “**Scaled process priors: Improved predictions and uncertainties for new-feature counts via random scaling in Bayesian nonparametrics**”; Accepted for publication in the *Journal of the American Statistical Association*. Manuscript available on arXiv [[poster](#); <https://arxiv.org/pdf/2106.15480.pdf>]. Featured in ISBA 2021; (Camerlenghi, Favaro, M. and T. Broderick)
- “**More for Less: Predicting and maximizing genetic variant discovery via Bayesian nonparametrics**”; to appear in *Biometrika*. **Best Student Paper Award**, awarded by the American Statistical Association, Section in Bayesian Statistical Science. Featured in SMEEB 2021, ASHG 2020, AABI 2019 [[poster](#); [presentation \(YouTube\)](#)]; Manuscript available on arXiv [<https://arxiv.org/pdf/1912.05516.pdf>] (M., Camerlenghi, Favaro, Broderick)
- “**Posterior representations of hierarchical completely random measures in trait allocation models**”, **Spotlight**, *BNP@NeurIPS2018* [[poster](#)] (M., Camerlenghi, Favaro and Broderick)
- “**Sensitivity of Bayesian inference to data perturbations**”, *AABI 2018* [[poster](#)] (M., Stephenson, Broderick)
- “**Generic finite approximations for practical Bayesian nonparametrics**”, **Spotlight**, *NIPS 2017 Workshop on Advances in Approximate Bayesian Inference* [[poster](#)]. (Huggins, M., Mackey, Broderick)

Theses

- “**Improved prediction and optimal sequencing strategies for genomic variant discovery via Bayesian nonparametrics**” — PhD thesis. Supervisor: Professor Tamara Broderick
- “**An asymptotic analysis of Gibbs-type priors**” — Master’s thesis. Supervisors: Professors Pierpaolo de Blasi and Igor Prünster
- “**Econometrics of the Big Data**” — Undergraduate thesis. Supervisor: Professor Alessandro Sembenelli

Skills

- Proficient in Python (numpy, scipy, pandas, matplotlib, scikit-learn), \LaTeX
- Past experience in C++, Matlab, R, RStudio

Talks, Poster sessions and Conference Presentations

2023

- KDD Workshop on Causal Discovery, Prediction and Decision, Long Beach (CA).
- Conference on Online Digital Experimentation, MIT, selected talk.

2022

- Conference on Online Digital Experimentation, MIT, selected talk.

2021

- Collegio Carlo Alberto Statistics Seminar Series, “Improved prediction and optimal sequencing strategies for genomic variant discovery via Bayesian nonparametrics”
- CMS Statistics 2021, “Scaled process priors for Bayesian nonparametric estimation of the unseen genetic variation” [Invited session]
- ISBA: 2021 World Meeting of the International Society for Bayesian Analysis, “Scaled process priors for Bayesian nonparametric estimation of the unseen genetic variation” [Contributed session]
- SMEEB: Stochastic Models and Experiments in Ecology and Biology, “More for less: predicting and maximizing genomic diversity via Bayesian nonparametrics” [Contributed session]

2020

- American Society of Human Genetics meeting, “More for less: predicting and maximizing genomic diversity via Bayesian nonparametrics” [Poster session]
- Learning under complex structure, MIFODS workshop, *Cambridge (MA)*, “More for less: predicting and maximizing genomic diversity via Bayesian nonparametrics” [Poster session]
- Learning under complex structure, MIFODS workshop, *Cambridge (MA)*, “More for less: predicting and maximizing genomic diversity via Bayesian nonparametrics” [Poster session]
- Bayes Comp 2020, *Gainesville (FL)*, “More for less: predicting and maximizing genomic diversity via Bayesian nonparametrics” [Poster session]

2019

- Advances in Bayesian Nonparametric Methods and Its Applications, *Denver (CO)*, *JSM 2019*, “Genomic variety prediction via Bayesian nonparametrics” [Topic-contributed session]
- Advances in Approximate Bayesian Inference, *Vancouver, Canada*, “More for less: Predicting and maximizing genetic variant discovery via Bayesian nonparametrics”
- Statistics and Data Science Conference 2019, *Cambridge (MA)*. “Genomic variety prediction via Bayesian nonparametrics”

- MLxMIT, *Cambridge (MA)*, “Genomic variety prediction via Bayesian nonparametrics”
- LIDS & Stats seminar, *Cambridge (MA)*, “Genomic variety prediction via Bayesian nonparametrics”
- CSAIL-MSR Trustworthy and Robust AI (TRAC) Workshop, *Cambridge (MA)*, “Getting the most bang for your buck: Predicting and maximizing the number of new genetic variants in a future experiment”

2018

- BNP@NeurIPS 2018, Montreal (Canada) “Posterior representations of hierarchical completely random measures in trait allocation models” [**Spotlight**]

Professional Service

2024	Reviewer for KDD, AISTATS, AABI
2023	Reviewer for JMLR, AISTATS, DGM4H@NeurIPS2023
2022	Reviewer for Statistical Science, JMLR, AISTATS
2021	Reviewer for Statistical Science, AISTATS, ICBINB
2020	Reviewer for AAAI 2020, AISTATS 2020
2019	Reviewer for AISTATS 2019, NeurIPS 2019, AABI 2019
2018	Reviewer for BNP@NeurIPS2018