

Lorenzo Masoero

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

WEB: <http://lorenzomasoero.com>

Senior Applied Scientist III | Amazon Weblab

September 2021 - Present

I am a Senior Applied Scientist at Weblab, Amazon's largest A/B testing platform, where I work on advancing online experimentation at scale. My work spans open-ended research, academic paper writing, and hands-on implementation of large-scale algorithms in production, directly impacting hundreds of millions of customers daily.

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 jointly randomize across multiple units (e.g., customers and advertisers) to detect spillover effects that would otherwise be overlooked. This work has been implemented at scale to capture indirect behavioral responses of advertisers to ad-auction pricing changes.

Enhanced Causal Estimates via Covariate Adjustments: A/B tests are notoriously noisy. To improve causal effect estimation, I leverage unit-level side information through regression models. I led the design and implementation of a distributed analytic engine using PySpark for analyzing large volumes of experimental data, enabling more accurate causal estimates across thousands of daily experiments.

Traffic Prediction: I developed Bayesian nonparametric methods to better predict future traffic in online experiments, improving experiment planning and resource allocation.

Previously, I completed my PhD at MIT in EECS under Professor Tamara Broderick, focusing on Bayesian nonparametric methods for optimal experimental design 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)
2012 - 2016	Allievi Honors Program Scholarship, Collegio Carlo Alberto, Moncalieri

Research

- **“Carryover detection in switchback experimentation”**; CauScien Workshop, NeurIPS 2025. (Missault, M.); Manuscript available at [openreview](#)
- **“Optimizing duration of online experiments via Bayesian early termination”**; Code@MIT 2025. (M., Gualavisi); Manuscript available at [Amazon Science](#)
- **“Rethinking Power in Online Experimentation: Connecting Frequentist and Bayesian Approaches to Inform Experiment Design”**; Code@MIT 2025. (Gualavisi, Kessler, M.). Manuscript available at [Amazon Science](#)
- **“Robust and efficient multiple-unit switchback experimentation”** Code@MIT 2025. (Missault, M., Delbé, Richardson, Imbens). Manuscript available at [Amazon Science](#). Working paper available at [arXiv:2506.12654](#)
- **“Online activity prediction via generalized Indian buffet process models”** (pre-print; Beraha, M., Favaro, Richardson). Manuscript available at [arXiv:2505.1964](#)
- **“Measuring direct and Indirect Impacts in a Multi-Sided Marketplace: Evidence from a Clustered Multiple Randomization Experiment”** CODE@MIT2024, KDD 2024, Workshop on Two-Sided Marketplace Optimization; (Hut, Imbens, M., Vijaykumar). Manuscript available at [Code@MIT 2024 proceedings](#)
- **“Multi-Armed Bandits with Network Interference”**; NeurIPS 2024. (Agarwal, Agarwal, M., Whitehouse). Manuscript available at [arXiv](#)
- **“Double trouble: Predicting new variant counts across two heterogeneous populations”**; arXiv preprint, 2403.02154. (Shen, M., Schraiber, Broderick). Manuscript available at [arXiv](#); earlier version accepted at Learning Meaningful Representations of Life, Neural Information Processing Systems Workshop, 2022. (Shen, M., Schraiber, Broderick) (manuscript available at [OpenReview](#))
- **“Improved prediction of future user activity in online A/B testing”**; arXiv preprint, 2402.03231. (M., Beraha, Richardson, Favaro). Manuscript available at [arXiv](#)
- **“A Nonparametric Bayes Approach to Online Activity Prediction”**; arXiv preprint, 2401.14722. (Beraha, M., Richardson, Favaro). Manuscript available at [arXiv](#)
- **“Multiple Randomization Designs: Estimation and Inference with Interference.”** arXiv preprint, 2401.01264. (M., Vijaykumar, Richardson, McQueen, Rosen, Burdick, Bajari, Imbens). Manuscript available at [arXiv](#)
- **“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). Manuscript available at [PMLR](#)
- **“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). (Nguyen, Huggins, M., Mackey, Broderick). Manuscript available at [arXiv](#)
- **“Cross-Study Replicability in Cluster Analysis”**. Statistical Science, 38(2): 303-316 (May 2023). (M., Thomas, Parmigiani, Tyekucheva, Trippa). Manuscript available at [arXiv](#)

- “**Multiple Randomization Designs**”; (Bajari, Burdick, Imbens, M., McQueen, Richardson, Rosen). Manuscript available at [arXiv](#)
- “**Bayesian nonparametric strategies for power maximization in rare variants association studies**”; **Spotlight** at Learning Meaningful Representations of Life, Neural Information Processing Systems, 2021 [[poster](#)]. (M., Schraiber, Broderick). Manuscript available at [arXiv](#)
- “**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. (Camerlenghi, Favaro, M. and T. Broderick). Manuscript available at [arXiv](#) [[poster](#)]. Featured in ISBA 2021
- “**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. (M., Camerlenghi, Favaro, Broderick). Manuscript available at [arXiv](#) [[poster](#); [presentation \(YouTube\)](#)]. Featured in SMEEB 2021, ASHG 2020, AABI 2019
- “**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). Manuscript available at [NIPS Workshop](#)

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

2025

- Conference on Online Digital Experimentation, MIT. Three selected papers: two selected talks, one poster.
- NeurIPS; Poster presentation at <https://sites.google.com/view/causcienc>.

2024

- Conference on Online Digital Experimentation, MIT. Three selected papers: two selected talks, one poster.
- KDD 2024 Workshop on Two Sided Marketplace Optimization, selected talk (“**Measuring direct and Indirect Impacts in a Multi-Sided Marketplace: Evidence from a Clustered Multiple Randomization Experiment**” Manuscript available at [Code@MIT 2024 proceedings](#))
- NeurIPS; Poster presentation (“**Multi-Armed Bandits with Network Interference**”; NeurIPS 2024. (Agarwal, Agarwal, M., Whitehouse). Manuscript available at [arXiv](#))

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, DGM ₄ H@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