

Mario Beraha

CONTACT INFORMATION	Dipartimento di Economia e Statistica, Università di Torino Corso Unione Sovietica 218bis, Torino, Italy	mario.beraha@unito.it
CURRENT POSITION	Research Associate Università di Torino Advised by Prof. Stefano Favaro, supported by ERC grant NBEB-SSP	Sept 2022 - ongoing
EDUCATION	Ph.D., Data Science and Computation Università degli Studi di Bologna and Politecnico di Milano Dissertation Advisor: Prof. Alessandra Guglielmi	Oct. 2018 - Nov. 2022
	M.S., Mathematical Engineering Politecnico di Milano	Oct. 2015 - Apr. 2018
	B.S., Mathematical Engineering Politecnico di Milano	Oct. 2012 - Sep. 2015
VISITING PERIODS	Dept. of Statistical Science, University College London. Invited by Prof. Jim E. Griffin	March-May 2022
	Team Statify, Inria Grenoble Rhone Alpes. Invited by dr. Julyan Arbel	Oct. 2020
	Dept. of Mathematics, Aalborg University. Invited by Prof. Jesper Møller	Sep. 2020
AWARDS	Junior travel award for the conferences: BNP 12 (Oxford, UK, 2018, accommodation), BNP Networking (Nicosia, Cyprus, 2022, 400USD), ISBA 2022 World Meeting (Montreal, Canada, 2022, 500USD), BNP 13 (Puerto Varas, Chile, 2022, 1000USD)	
SOFTWARE	Co-editor and maintainer of BayesMix : a C++ library for posterior simulation for general Bayesian mixture models. BayesMix features 30 mixture models and 5 MCMC algorithms. Thanks to the use of static and runtime polymorphism based on the CRTP pattern, new users with little familiarity with mixture models can extend BayesMix with minimal coding effort.	
	Editor and maintainer of the Python package pybmix : a user-friendly interface to BayesMix .	
TEACHING EXPERIENCE	Teaching Assistant , Politecnico di Milano Bayesian Statistics (prof A. Guglielmi) Fondamenti di Statistica e Segnali Biomedici (prof Sangalli, Guglielmi)	Fall 2021, 2022 Spring 2021, 2019
	Tutor , Politecnico di Milano Bayesian Statistics (prof A. Guglielmi) Bayesian Statistics (prof F.A. Quintana)	Fall 2019, 2020 Fall 2018
	Teaching Assistant , Ecole Centrale Supélec, Paris Saclay Artificial Intelligence (prof C. Hudelot)	Fall 2017
SERVICE AND OUTREACH	Referee for: Annals of Applied Statistics, Bayesian Analysis, Biometrics, Computational Statistics and Data Analysis, Statistical Methods and Applications	

SKILLS

Language: Italian (mother tongue), English (fluent, C1), French (B1)

Programming Languages: C++, Python, R, C, Matlab, LaTeX.

Software & Frameworks: Stan, JAGS, Apache Spark, JAX, Tensorflow, Pytorch, AWS S3/EC2, GCP, MondoDB, Redis, Docker

PUBLICATIONS

*: equal contribution

JOURNAL ARTICLES, INTERNATIONAL CONFERENCES PROCEEDINGS, AND DISCUSSIONS

7. Pegoraro*, M., **Beraha***, M. (2022), *Projected Statistical Methods for Distributional Data on the Real Line with the Wasserstein Metric*. [Journal of Machine Learning Research \[arXiv:2101.09039\]](#)
6. **Beraha**, M., Argiento, R., Møller, J., Guglielmi, A (2022), *MCMC computations for Bayesian mixture models using repulsive point processes*. [Journal of Computational and Graphical Statistics \[arXiv:2011.06444\]](#)
5. **Beraha**, M., Pegoraro, M., Peli, R., Guglielmi, A (2021), *Spatially dependent mixture models via the Logistic Multivariate CAR prior*. [Spatial Statistics \[arXiv:2007.14961\]](#) .
4. **Beraha**, M., Guglielmi, A. Quintana, F. A. (2021), *The semi-hierarchical Dirichlet Process and its application to clustering homogeneous distributions*. [Bayesian Analysis \[arXiv:2005.10287\]](#)
3. Pegoraro*, M, **Beraha***, M. (2021), *Fast PCA in Wasserstein Spaces via B-splines Representation and Metric Projection*. [35th AAAI Conference on Artificial Intelligence](#)
2. **Beraha**, M., Guglielmi, A. (2019), *Invited discussion on “Latent Nested Nonparametric Priors” by F. Camerlenghi, D. B. Dunson, A. Lijoi, I. Prünster and A. Rodriguez*. [Bayesian Analysis](#)
1. **Beraha***, M., Metelli*, A. M., Papini*, M., Tirinzoni*, A., Restelli, M. (2019), *Feature Selection via Mutual Information: New Theoretical Insights*. [International Joint Conference on Neural Networks \(IJCNN\) \[arXiv:1907.07384\]](#)

TECHNICAL REPORTS AND ONGOING PROJECTS

9. **Beraha**, M., Favaro, S. (2023+), *Transform-Scaled Priors for BNP Trait Allocations*. In preparation
8. **Beraha**, M., Favaro, S. (2023+), *Random Measure Priors in Bayesian Frequency Recovery from Sketches*. Submitted [\[arXiv:2303.15029\]](#)
7. **Beraha**, M.*, Pegoraro, M.* (2023+), *A Wasserstein-PCA for Probability Measures on the Unit Circle*. In preparation
6. **Beraha**, M., Argiento, R., Camerlenghi, F., Guglielmi, A. (2023+), *Normalized Random Measures with Interacting Atoms for Bayesian Nonparametric Mixtures*. Submitted [\[arXiv:2302.09034\]](#)
5. Ghilotti, L., **Beraha**, M., Guglielmi, A. (2023+), *Bayesian clustering of high-dimensional data via latent repulsive mixtures*. [\[arXiv:2303.02438\]](#)
4. **Beraha**, M., Griffin, J. E. (2023+), *Normalized Latent Measure Factor Models*. Under Review [\[arXiv:2205.15654\]](#)
3. **Beraha**, M., Guindani, B., Gianella, M., Guglielmi, A. (2023+), *BayesMix: Bayesian Mixture Models in C++*. Under Review [\[arXiv:2205.08144\]](#)
2. **Beraha**, M., Corradin, R. (2023+), *Bayesian nonparametric model based clustering with intractable distributions: an ABC approach*. Submitted [\[arXiv:2112.10393\]](#)
1. **Beraha**, M., Falco, D., Guglielmi, A. (2021), *JAGS, NIMBLE, Stan: a detailed comparison among Bayesian MCMC software*. Technical Report [\[arXiv:2107.09357\]](#)

NATIONAL CONFERENCES PROCEEDINGS AND DISCUSSIONS

6. Ghilotti, M, **Beraha**, M., Guglielmi, A. (2021), *Anisotropic determinantal point processes and their application in Bayesian mixtures*. Book of Short Papers - SIS 2021 (pp. 1226-1231)

5. Gianella, M, **Beraha, M.**, Guglielmi, A. (2021), *Spatially dependent mixture models with a random number of components*. Book of Short Papers - SIS 2021 (pp. 936-942)
4. **Beraha, M.**, Corradin, R. (2020), *An ABC algorithm for random partitions arising from the Dirichlet process*. Book of Short Papers - SIS 2020 (pp. 632-637)
3. **Beraha, M.**, Pegoraro, M., Peli, R., Guglielmi, A. (2020), *A Bayesian model to induce dependence between mixtures*. Book of Short Papers - SIS 2020 (pp. 608-613)
2. **Beraha, M.**, Gualtieri, G., Villa, E., Vitali, R., Guglielmi, A. (2020), *Choosing the right tool for the job: a systematic analysis of general purpose MCMC software*. Book of Short Papers - SIS 2020 (pp. 661-666)
1. Bissoli, G., Principi, C., Rinaldi, G. M., **Beraha, M.**, Guglielmi, A. (2019), *A Bayesian model for network flow data: an application to BikeMi trips*. Book of Short Papers - SIS 2019 (pp. 673-679)

INVITED TALKS
AND SEMINARS

School of Mathematical Sciences, University of Nottingham (Mar. 2023). Random Measure Priors in Bayesian Frequency Recovery from Sketches (invited seminar)

MOX, Politecnico di Milano (Feb. 2023). A Bayesian nonparametric approach to data compression (department seminar)

CMStatistics (London, Dec. 2022). Random Measure Priors in Bayesian Frequency Recovery from Sketches

Department of Statistical Science, Università Cattolica di Milano (July 2022). Normalized Latent Measure Factor Models (invited seminar)

2022 ISBA World Meeting (Montreal, June 2022). Computational and modelling aspects of repulsive mixture models

Department of Statistical Science, University College London (May 2022). Distributional data analysis with the Wasserstein distance (invited seminar)

BNP Networking (Nicosia, Apr. 2022). NRMI Factor Models: scaling BNP models for partially exchangeable data to a large number of populations

CMStatistics (Online, Dec. 2021). Repulsive mixture models: modelling and computations, with applications to high-dimensional data

Department of Economics – Università di Bergamo (Nov. 2021). Some recent advances on nonparametric Bayesian analysis of grouped data (invited seminar)

MOX, Department of Mathematics – Politecnico di Milano (Jan. 2021). Projected statistical methods in the Wasserstein space (invited seminar)

CMStatistics 2020 (Online, Dec. 2020). MCMC computations for Bayesian mixture models using repulsive point processes

BAYSM:O 2020 (Online, Sep. 2020). MCMC computations for Bayesian mixture models using repulsive point processes

Inria Grenoble (Nov. 2019). Projected statistical methods for distributional data with the Wasserstein metric (invited seminar)

ISBA Webinar (January 2020). Discussant for "Latent Nested Nonparametric Priors"

ARS 2019 (Vietri sul Mare, Sep. 2019). Zero-inflated Poisson regression models for the analysis of network flow data

WORKING
EXPERIENCE

Data Scientist
Indigo A.I.

May. 2018 – Oct. 2018, Milan

Researched on: Neural Networks (LSTM, CNN) for text classification, Language Modeling for spell checking, GANs for text generation, Sentence level embedding for text clustering

Implemented the Python backend for a REST API using: Async & distributed programming, NoSQL, MongoDB, Redis, Protocol Buffers.

Intern in Big Data and Machine Learning
Brother Tongue

May 2017 – Nov. 2017, Paris

Analyzed and developed machine learning models, in the field of Natural Language Processing (NLP), over large datasets of text