




# FIGARO: hierarchical non-parametric inference for population studies

Stefano Rinaldi <sup>1,2</sup> and Walter Del Pozzo <sup>3,4</sup>

1 Institut für Theoretische Astrophysik, ZAH, Universität Heidelberg, Albert-Ueberle-Str. 2, 69120 Heidelberg, Germany 2 Dipartimento di Fisica e Astronomia “G. Galilei”, Università di Padova, Via F. Marzolo 8, 35121 Padova, Italy 3 Dipartimento di Fisica “E. Fermi”, Università di Pisa, Largo Bruno Pontecorvo 3, 56127 Pisa, Italy 4 INFN, Sezione di Pisa, Largo Bruno Pontecorvo 3, 56127 Pisa, Italy

DOI: [10.21105/joss.06589](https://doi.org/10.21105/joss.06589)

## Software

- [Review](#) 
- [Repository](#) 
- [Archive](#) 

Editor: [Dan Foreman-Mackey](#) 

## Reviewers:

- [@dgerosa](#)
- [@Uddiptaatwork](#)

Submitted: 15 February 2024

Published: 25 May 2024

## License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)).

## Summary

The astrophysical graveyard is populated by black holes (BHs) and neutron stars (NSs). These are the remains of the most massive stars, and studying them will teach us about the physics that governs the stars’ lives and deaths. Gravitational wave (GW) observations are now revealing the population of BHs via the detection of binary black hole mergers, and the numbers are set to grow rapidly in the coming years. The astrophysical distribution of BHs is inferred making use of the probability distribution of the parameters of detected BHs, combined together in a hierarchical population study. Currently, the characterisation and interpretation of the available observations is mainly guided by astrophysical models which depend on the not-well-understood physics of massive binary evolution (see [Abbott et al., 2023](#) and references therein): this approach is, however, intrinsically prone to potential biases induced by inaccurate modelling of the underlying population.

Bayesian non-parametric methods are models with a countably infinite number of parameters to model arbitrary probability densities. These parameters do not have any connection with the modelled distribution, making these a convenient and agnostic way of describing some unknown population. These are key tools to reconstruct probability densities without being committal to a specific functional form: the basic idea is to let the data speak for themselves, retrieving the distribution that is the most likely to have generated the observed data. In a certain sense, this is the most phenomenological approach possible: unlike the standard parametric approach, where we specify a functional form inspired by what we expect to find in the data, with non-parametric methods all the information comes from the data, thus avoiding the risk of biasing the inference with inaccurate models. Features in the inferred distribution will arise naturally without the need of including them in the model, leaving astrophysicists tasked with explaining them in terms of formation channels and astrophysical processes.

The GW community is currently exploring this direction ([Callister & Farr, 2024](#); [Edelman et al., 2023](#); [Tiwari, 2021](#); [Toubiana et al., 2023](#)): FIGARO fits in this framework, being a non-parametric inference scheme designed to reconstruct arbitrary probability densities in a hierarchical fashion under the requirement of minimal mathematical assumptions.

## Statement of need

FIGARO (Fast Inference for GW Astronomy, Research and Observations) is a python package that implements a variation of the Gibbs sampling scheme with a *Hierarchy of Dirichlet Process Gaussian Mixture Models* ([Rinaldi & Del Pozzo, 2022a](#)), or (H)DPGMM for short,

as non-parametric model<sup>1</sup> to reconstruct arbitrary probability densities given a set of observations. These observations can either be samples from an unknown distribution<sup>2</sup> or a set of posterior probability densities represented by samples from these posteriors. Differently from other publicly-available DPGMM implementations such as the one included in `scikit-learn` (Pedregosa et al., 2011), FIGARO performs a stochastic sampling over the (potentially infinite) parameter space of the DPGMM, thus allowing for an efficient marginalisation over such parameters.

Despite being originally developed in the context of GW physics and in particular to work with the data released by the LIGO-Virgo-KAGRA (LVK) collaboration, FIGARO can take as input generic data and therefore it can be applied to a variety of problems beyond GWs (see *Publications*). FIGARO output objects are modelled after `scipy.stats` (Virtanen et al., 2020) classes to make them intuitive for users that are already familiar with the `scipy` package. The flexibility of (H)DPGMM in reconstructing arbitrary probability densities united with the speed provided by the Gibbs sampling variation we implemented in this package makes FIGARO an ideal tool for population studies.

## Availability and usage

FIGARO is available via [PyPI](#) and is compatible with `python<3.12`. The code is hosted on [GitHub](#) and the documentation is hosted at [figaro.readthedocs.io](#). FIGARO comes with two CLIs to perform both the reconstruction of a probability density given a set of samples (`figaro-density`) and the hierarchical inference (`figaro-hierarchical`). The documentation also includes a guide on how to use FIGARO in a custom python script.

## Publications

This is a list of the publications that made use of FIGARO so far:

- Rinaldi & Del Pozzo (2022b) – online sky localisation of the potential electromagnetic (EM) counterpart of a GW signal to maximise the likelihood of a joint GW-EM detection;
- Rinaldi et al. (2023) – non-parametric analysis of systematic errors in the determination of Newton’s constant of gravitation  $G$ ;
- Rinaldi et al. (2024) – non-parametric inference of the joint BH primary mass, mass ratio and redshift distribution. First evidence for the evolution of the BH mass function with redshift;
- Cheung et al. (2023) – use of non-parametric methods to minimise the impact of population bias in the search for lensed GW events;
- Sgalletta et al. (2023) – detailed study of the NS population in the Milky Way. FIGARO is used to approximate the predicted observed population of NSs;
- Morton et al. (2023) – investigation of the potential origin of GW190521 in an AGN. FIGARO is used to approximate the multivariate posterior probability density for GW190521 parameters;
- Rallapalli & Desai (2023) – applies the framework described in Rinaldi et al. (2023) to the inference of the W boson mass.

## Acknowledgements

We acknowledge contributions from Daniele Sanfratello and Vera Delfavero. SR acknowledges financial support from the European Research Council for the ERC Consolidator grant DEMOBLACK, under contract no. 770017.

<sup>1</sup>(H)DPGMM is based on the Dirichlet Process Gaussian Mixture Model, introduced in Escobar & West (1995). Mathematical details about the Dirichlet Process can be found in Teh (2010).

<sup>2</sup>In this case, the model used will be a DPGMM.

## References

- Abbott, R., Abbott, T. D., Acernese, F., Ackley, K., Adams, C., Adhikari, N., Adhikari, R. X., Adya, V. B., Affeldt, C., Agarwal, D., Agathos, M., Agatsuma, K., Aggarwal, N., Aguiar, O. D., Aiello, L., Ain, A., Ajith, P., Akutsu, T., de Alarcón, P. F., ... others. (2023). Population of Merging Compact Binaries Inferred Using Gravitational Waves through GWTC-3. *Physical Review X*, 13(1), 011048. <https://doi.org/10.1103/PhysRevX.13.011048>
- Callister, T. A., & Farr, W. M. (2024). Parameter-Free Tour of the Binary Black Hole Population. *Physical Review X*, 14(2), 021005. <https://doi.org/10.1103/PhysRevX.14.021005>
- Cheung, D. H. T., Rinaldi, S., Toscani, M., & Hannuksela, O. A. (2023). Mitigating the effect of population model uncertainty on strong lensing Bayes factor using nonparametric methods. *arXiv e-Prints*, arXiv:2308.12182. <https://doi.org/10.48550/arXiv.2308.12182>
- Edelman, B., Farr, B., & Doctor, Z. (2023). Cover Your Basis: Comprehensive Data-driven Characterization of the Binary Black Hole Population. *Astrophysical Journal*, 946(1), 16. <https://doi.org/10.3847/1538-4357/acb5ed>
- Escobar, M. D., & West, M. (1995). Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association*, 90(430), 577–588. <https://doi.org/10.1080/01621459.1995.10476550>
- Morton, S. L., Rinaldi, S., Torres-Orjuela, A., Derdzinski, A., Vaccaro, M. P., & Del Pozzo, W. (2023). GW190521: A binary black hole merger inside an active galactic nucleus? *Physical Review D*, 108(12), 123039. <https://doi.org/10.1103/PhysRevD.108.123039>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Müller, A., Nothman, J., Louppe, G., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830. <https://doi.org/10.48550/arXiv.1201.0490>
- Rallapalli, A., & Desai, S. (2023). Bayesian inference of W-boson mass. *European Physical Journal C*, 83(7), 580. <https://doi.org/10.1140/epjc/s10052-023-11754-x>
- Rinaldi, S., & Del Pozzo, W. (2022a). (H)DPGMM: a hierarchy of Dirichlet process Gaussian mixture models for the inference of the black hole mass function. *Monthly Notices of the Royal Astronomical Society*, 509(4), 5454–5466. <https://doi.org/10.1093/mnras/stab3224>
- Rinaldi, S., & Del Pozzo, W. (2022b). Rapid localization of gravitational wave hosts with FIGARO. *Monthly Notices of the Royal Astronomical Society*, 517(1), L5–L10. <https://doi.org/10.1093/mnras/slac101>
- Rinaldi, S., Del Pozzo, W., Mapelli, M., Lorenzo-Medina, A., & Dent, T. (2024). Evidence of evolution of the black hole mass function with redshift. *Astronomy & Astrophysics*, 684, A204. <https://doi.org/10.1051/0004-6361/202348161>
- Rinaldi, S., Middleton, H., Del Pozzo, W., & Gair, J. (2023). Bayesian analysis of systematic errors in the determination of the constant of gravitation. *European Physical Journal C*, 83(10), 891. <https://doi.org/10.1140/epjc/s10052-023-12078-6>
- Sgalletta, C., Iorio, G., Mapelli, M., Artale, M. C., Boco, L., Chattopadhyay, D., Lapi, A., Possenti, A., Rinaldi, S., & Spera, M. (2023). Binary neutron star populations in the Milky Way. *Monthly Notices of the Royal Astronomical Society*, 526(2), 2210–2229. <https://doi.org/10.1093/mnras/stad2768>
- Teh, Y. W. (2010). Dirichlet process. In *Encyclopedia of machine learning* (pp. 280–287). Springer US. [https://doi.org/10.1007/978-0-387-30164-8\\_219](https://doi.org/10.1007/978-0-387-30164-8_219)

- Tiwari, V. (2021). VAMANA: modeling binary black hole population with minimal assumptions. *Classical and Quantum Gravity*, 38(15), 155007. <https://doi.org/10.1088/1361-6382/ac0b54>
- Toubiana, A., Katz, M. L., & Gair, J. R. (2023). Is there an excess of black holes around 20  $M_{\odot}$ ? Optimizing the complexity of population models with the use of reversible jump MCMC. *Monthly Notices of the Royal Astronomical Society*, 524(4), 5844–5853. <https://doi.org/10.1093/mnras/stad2215>
- Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., ... SciPy 1.0 Contributors. (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17, 261–272. <https://doi.org/10.1038/s41592-019-0686-2>