FIGARO: hierarchical non-parametric inference for population studies

Stefano Rinaldi [1,2] **and Walter Del Pozzo** [3,4]

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

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: 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.

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 (**F**ast **I**nference for **G**W **A**stronomy, **R**esearch and **O**bservations) 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 to reconstruct arbitrary probability densities given a set of observations. These observations can either be samples from an unknown distribution[1] or a set of posterior probability densities represented by samples from these posteriors. Differently from other publicly-available

---

[1]In this case, the model used will be a DPGMM.

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 can be found at [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 code repository also
includes a [jupyter notebook](#) with a tutorial 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, Middleton, et al. ([2023](#)) – non-parametric analysis of systematic errors in the
  determination of Newton's constant of gravitation G;
- Rinaldi, Del Pozzo, et al. ([2023](#)) – 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, Middleton, et al.
  ([2023](#)) to the inference of the W boson mass.

Acknowledgements

References

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

84 Morton, S. L., Rinaldi, S., Torres-Orjuela, A., Derdzinski, A., Vaccaro, M. P., & Del Pozzo, W.
85   (2023). GW190521: A binary black hole merger inside an active galactic nucleus? *Physical*
86   *Review D*, *108*(12), 123039. https://doi.org/10.1103/PhysRevD.108.123039

87 Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M.,
88   Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D.,
89   Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python.
90   *Journal of Machine Learning Research*, *12*, 2825–2830.

91 Rallapalli, A., & Desai, S. (2023). Bayesian inference of W-boson mass. *European Physical*
92   *Journal C*, *83*(7), 580. https://doi.org/10.1140/epjc/s10052-023-11754-x

93 Rinaldi, S., & Del Pozzo, W. (2022a). (H)DPGMM: a hierarchy of Dirichlet process Gaussian
94   mixture models for the inference of the black hole mass function. *Monthly Notices of the*
95   *Royal Astronomical Society*, *509*(4), 5454–5466. https://doi.org/10.1093/mnras/stab3224

96 Rinaldi, S., & Del Pozzo, W. (2022b). Rapid localization of gravitational wave hosts with
97   FIGARO. *Monthly Notices of the Royal Astronomical Society*, *517*(1), L5–L10. https:
98   //doi.org/10.1093/mnrasl/slac101

99 Rinaldi, S., Del Pozzo, W., Mapelli, M., Lorenzo Medina, A., & Dent, T. (2023). Evidence for
100   the evolution of black hole mass function with redshift. *arXiv e-Prints*, arXiv:2310.03074.
101   https://arxiv.org/abs/2310.03074

102 Rinaldi, S., Middleton, H., Del Pozzo, W., & Gair, J. (2023). Bayesian analysis of systematic
103   errors in the determination of the constant of gravitation. *European Physical Journal C*,
104   *83*(10), 891. https://doi.org/10.1140/epjc/s10052-023-12078-6

105 Sgalletta, C., Iorio, G., Mapelli, M., Artale, M. C., Boco, L., Chattopadhyay, D., Lapi, A.,
106   Possenti, A., Rinaldi, S., & Spera, M. (2023). Binary neutron star populations in the
107   Milky Way. *Monthly Notices of the Royal Astronomical Society*, *526*(2), 2210–2229.
108   https://doi.org/10.1093/mnras/stad2768

109 Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D.,
110   Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson,
111   J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., … SciPy
112   1.0 Contributors. (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in
113   Python. *Nature Methods*, *17*, 261–272. https://doi.org/10.1038/s41592-019-0686-2