M2 internship project

Inferring galaxy properties from their observations

M2 internship 2024

Jenny SORCE, Pierre CHAINAIS,

Centre de Recherche en Informatique, Signal et Automatique de Lille

CRIStAL UMR 9189 - Lille, France

Dates: 4 to 6 months, starting date around March-April 2024.

Contacts:
Jenny SORCE (jenny.sorce@univ-lille.fr),
https://jennygsorce.appspot.com
Pierre CHAINAIS (pierre.chainais@centralelille.fr),
http://pierrechainais.ec-lille.fr/

Laboratory: Centre de Recherche en Informatique, Signal et Automatique de Lille (UMR 9189 CRIStAL), Villeneuve d'Ascq, France.

Keywords: Inverse problem, galaxy properties, Bayesian inference.

---

1 Project overview

According to the standard cosmological model, dark matter and dark energy make up about 95% of the Universe. Recent analyzes of large galaxy surveys reveal discrepancies with this model. For example, the local measurement of the expansion rate of the Universe (Freedman 2021) differs by more than three standard deviations from that deduced from the first light of the Universe (Planck Collaboration et al. 2016). The debate consists in determining whether this apparent mismatch is a signature of new physics or of systematic effects in the observation processing pipeline. The data processing pipeline is indeed complex. Part of this pipeline consists in determining galaxy properties using methods most often calibrated independently of the environment, of the redshift (linked to the Universe age) and of the survey specificities. This can lead to critical biases (Puech et al. 2019; Sorce and Guo 2016).

Consequently, the main objective of this internship is to formalize the inverse problem of the estimation of galaxy intrinsic luminosity from their observations in a Bayesian framework while taking into account measurement uncertainties. In practice, the student will implement the corresponding algorithm and, depending on the time budget, consider either applications to observational data or addition of redshift effects to the problem. More precisely, current methods are most often relations linking one or several galaxy observables to their properties. The observables of galaxies include, for example, their relative luminosity, their observational redshift and their rotation on themselves. The Tully-Fisher relation (Tully et al. 1977) is the relation between their rotation rate and their 'intrinsic luminosity'. This property combined with their relative luminosity and their observational redshift make it possible to estimate the expansion rate of the Universe. There are other relationships that permit to infer the intrinsic luminosity of galaxies such as the fundamental plane (Colless et al. 2001) or the surface brightness (Tonry et al. 2001). All of these relationships have the common denominator of having large variances and systematic biases due, for example, to the environment or redshift. All this has an impact on the measurement of the intrinsic luminosity of galaxies and, consequently, on the estimate of the expansion rate of the Universe (Wojtak et al. 2014).

So-called end-to-end machine learning techniques are increasingly used to replace the usual expensive techniques for analyzing galaxies in order to infer their properties. However, they are rarely successfully applied to observations after being developed and calibrated on simulations (de Santi et al. 2023). When they do so, the interpretability and uncertainty quantification of the results are generally missing (Hong et al. 2021).

This project aims at determining the best combinations of observables to infer the intrinsic luminosity of galaxies. Inspired by the work of Schafer et al. 2023, the study will be decomposed into the following steps:

1. explore the provided synthetic dataset, obtained from a cosmological simulation (Sorce, Dubois, et al. 2021). It will contain different observables of galaxies as well as their intrinsic luminosity;
2. identify appropriate techniques to look for relationships with minimum intrinsic variance between observables and the intrinsic luminosity to be inferred. Intrinsic luminosities and observables are independently available in the simulated data to determine and calibrate these relationships;
3. formalize the inverse problem in a Bayesian framework to infer the intrinsic luminosity from the observables taking into account measurement uncertainties; then, implement the algorithm to infer the intrinsic luminosity from the observables;
4. time permitting, either applications to observational data or addition of redshift effects will be considered.

Note that Stone et al. 2021 showed that the current difficulty is in disentangling the intrinsic variance of the relationships from those due to measurement uncertainties of observables and to the limits of theoretical models. This will be the challenge of this project which will have to build a technique capable of inferring intrinsic luminosity with the minimum uncertainty and ideally applicable to all galaxy types.

Important note: This project may be followed by an interdisciplinary PhD thesis in data science and astrophysics.

2 Scientific context

The project is part of the *ANR Chaire IA SHERLOCK* (*Fast inference with controlled uncertainty: application to astrophysical observations*) led by Pierre Chainais (co-funded by Agence Nationale de la Recherche (ANR), I-SITE, Centrale Lille Institut and Région Haut-de-France). The successful candidate will be jointly supervised by Jenny Sorce (CNRS Researcher in cosmology) and Pierre Chainais (Professor, Centrale Lille) in the CRIStAL lab (UMR 9189), Lille, France.

3 Period

4 to 6-months internship, starting between March and April 2024, ending at the latest on September 30th. The precise starting and ending dates will be adjusted depending on the availability of the candidate. The intern will be granted the usual stipend of $\sim$ 615 euros/month (4,35 euros/hour).

4 Profile and requirements

Master or engineering school students with major in applied mathematics, computer science or electrical engineering. The project requires a good background in data science (statistics, optimization). Good Python coding skills are expected. Interest in physics, signal processing, C++ programming or parallel/distributed code implementation is a plus.

5 Application procedure

Applicants are invited to send the following documents in `.pdf` format to both co-advisors:

- a letter of motivation,
- a detailed curriculum,
- official transcripts (grade) from the institutions you have attended over the last 2 years.

For further information, please **contact both** co-advisors of the project:

- Jenny Sorce, jenny.sorce@univ-lille.fr
- Pierre Chainais, pierre.chainais@centralelille.fr

References

Colless, M. et al. (Feb. 2001). "The peculiar motions of early-type galaxies in two distant regions - VII. Peculiar velocities and bulk motions". In: *MNRAS* 321, pp. 277–305. eprint: arXiv:astro-ph/0008418.

de Santi, Natalí S. M. et al. (Oct. 2023). "Field-level simulation-based inference with galaxy catalogs: the impact of systematic effects". In: *arXiv e-prints*, arXiv:2310.15234, arXiv:2310.15234. arXiv: 2310.15234 [astro-ph.CO].

Freedman, Wendy L. (Sept. 2021). "Measurements of the Hubble Constant: Tensions in Perspective". In: *ApJ* 919.1, 16, p. 16. arXiv: 2106.15656 [astro-ph.CO].

Hong, Sungwook E. et al. (May 2021). "Revealing the Local Cosmic Web from Galaxies by Deep Learning". In: *ApJ* 913.1, 76, p. 76. arXiv: 2008.01738 [astro-ph.CO].

Planck Collaboration et al. (Sept. 2016). "Planck 2015 results. XIII. Cosmological parameters". In: *A&A* 594, A13, A13. arXiv: 1502.01589.

Puech, M. et al. (Sept. 2019). "Time-average properties of z $\sim$ 0.6 major mergers: mergers significantly scatter high-z scaling relations". In: *MNRAS* 488.1, pp. 876–893. arXiv: 1907.03785 [astro-ph.GA].

Schafer, Chad et al. (Oct. 2023). "Reinterpreting Fundamental Plane Correlations with Machine Learning". In: *arXiv e-prints*, arXiv:2310.18177, arXiv:2310.18177. arXiv: 2310.18177 [astro-ph.GA].

Sorce, Dubois, et al. (June 2021). "I - A hydrodynamical CLONE of the Virgo cluster of galaxies to confirm observationally driven formation scenarios". In: *MNRAS* 504.2, pp. 2998–3012. arXiv: 2104.13389 [astro-ph.CO].

Sorce and Guo (May 2016). "The baryonic Tully-Fisher relation cares about the galaxy sample". In: *MNRAS* 458, pp. 2667–2675.

Stone, Connor J. et al. (Dec. 2021). "AutoProf - I. An automated non-parametric light profile pipeline for modern galaxy surveys". In: *MNRAS* 508.2, pp. 1870–1887. arXiv: 2106.13809 [astro-ph.GA].

Tonry, J. L. et al. (Jan. 2001). "The SBF Survey of Galaxy Distances. IV. SBF Magnitudes, Colors, and Distances". In: *ApJ* 546, pp. 681–693. eprint: arXiv:astro-ph/0011223.

Tully et al. (Feb. 1977). "A new method of determining distances to galaxies". In: *A&A* 54, pp. 661–673.

Wojtak, R. et al. (Feb. 2014). "Cosmic variance of the local Hubble flow in large-scale cosmological simulations". In: *MNRAS* 438, pp. 1805–1812. arXiv: 1312.0276 [astro-ph.CO].