Playing dice with the universe: Bayesian statistical analyses of cosmological models and new observables

Cañas Herrera, G.

Summary

The fascination with the night sky. A simple gaze at the wonderful landscape above us opens a window toward imagination. This fascination together with our innate human curiosity paved the passion for understanding the world we live in. From ancient cultures to nowadays, we were all concerned about the same questions: "*Where did we come from?*" and "*What is our place in the Universe?*".

Understanding and learning more about the cosmos is not only appealing to those who dedicate their lives to science but also to the lay population in general. A quick search within the most popular forums on the internet demonstrates that all have threads up-to-date dedicated to conversations about the universe[12] Moreover, every time there is an exciting astronomical discovery, it spreads within a few hours, as was the case for the first-ever image captured by the Event Horizon Telescope (EHT) of the supermassive black hole Sgr A*.

The composition of the universe, as well as its believed origin, and the main force dominating, gravity, are part of the popular culture of our civilization. But, how is (and was) the Universe studied? We rely on astronomical observations of, for instance, stars and other massive compact objects present in the Universe, to build our scientific knowledge based on the comparison of those observations with some physical assumptions, which form a cosmological model. For example, an average lay citizen is familiar with the concept of Big Bang.

Indeed, cosmologists believe that the universe was formed approximately 13.77 billion years ago from a very dense and hot state. The young Universe cooled down while it expanded, and it is still expanding today at a faster rate. Our Universe is composed of two main ingredients: an unknown matter substance that interacts gravitationally denominated dark matter and the agent responsible of the accelerated expansion of the universe called dark energy. The matter we are all made of, baryonic matter, accounts only for approximately 5% of the content of the universe. Therefore, it is not surprising that concepts such as dark energy and dark matter are surrounded by a halo of mystery, which attracts the interest of the public. Talking in scientific terms, most of the scientific knowledge about the composition of the universe was obtained from observations of the cosmic microwave background: the relic radiation

[12]See for instance `https://www.quora.com/Why-do-you-love-to-gaze-at-the-stars-of-the-night-sky`.

emitted in the early universe when the first neutral atoms were formed and photons could finally scatter away and move freely throughout the Universe. From these observations, we concluded that our Universe is flat.

As the Universe cooled down and expanded, gravity started to play a more significant role and the current structures observed in our Universe started to form: from the first stars to the first galaxies, and from the first galaxies to the first clusters of galaxies. These objects formed the so-called large scale structure of the universe, which looks like a web. However, gravity is not the only player in the formation of the large scale structure of the universe. We need the existence of an underlying Gaussian distribution of primordial density perturbations that can explain the origin of the current structure. So far, inflation, an exponential expansion phase in the primordial universe, is believed to be the mechanism of production of these primordial seeds. The simplest model of inflation is compatible with the assumption of a Gaussian distribution of the primordial density perturbations, apart from explaining why the cosmological observations indicate that the universe is causally connected and flat. The inflationary paradigm together with the assumption that our Universe is mostly composed of dark matter and dark energy are the basis of the standard cosmological model.

In the coming years, the cosmological observations available to study the Universe are expected to increase. In particular, we will exploit the information encoded in the large scale structure of the universe. For that, new missions and experiments are focusing on creating large galaxy surveys that will contain information about their positions in the sky, their shapes and also their redshifts. With the catalogues, cosmologists will study the composition of the universe using observables such as galaxy clustering (GC) and weak lensing (WL). Among these experiments, the European Space Agency medium-size *Euclid* mission, whose satellite is expected to be launched in the near future, aims to understand the physical origin of the accelerated expansion of the universe, as well as understand the nature of dark matter. *Euclid* will also study the initial conditions that seeded the early universe and were responsible for the formation of the cosmic-web structure. For that, *Euclid* will map the large scale structure of the universe by creating one of the largest galaxy catalogues ever known.

What is the methodology used to compare astrophysical observations with physical models to extract conclusions about the universe we live in? Cosmologists have used statistics, in particular, a Bayesian statistical approach, during the last 30 years to test different cosmological models against data. Bayesian statistics are based on Bayes' theorem. This theorem tells us that our degree of belief can be encoded in a probability distribution. Bayesian statistics assume that the probability distribution of some parameters of a model, given some observed data, is directly proportional to the product of two probability distributions: the distribution that gives the probability of observing the data given the model M and the parameter values, times the probability distribution that encodes some prior knowledge information known about the theory or the experiment. In the last decades, the success of Bayesian statis-

tics applied to cosmology is mostly due to the exponential increase in computational power that made massive numerical inference feasible for the first time.
Bayesian statistics can be used for a large set of scientific questions. For instance, it allows us to constrain, given a model and some data, the best fit values of the parameters of that model. On top of that, with this statistical approach we can also discern whether an alternative cosmological model is statistically favoured with respect to the standard cosmological model. Moreover, Bayesian statistics are also useful in forecasting analyses, where we aim to test how sensitive a future experiment will be to the possible detection of a new observable or even, given the experimental set-up of a future mission, if there is any chance that some extensions of the standard cosmological model can be ruled out statistically.

This thesis is dedicated to the Bayesian statistical analyses of extensions of the standard cosmological model using several astronomical data sets, and to the forecast of new observables or experiments. The use of this methodology is the common motto in all the chapters of the thesis. After **Chapter 1**, where we introduce all the main concepts of cosmology as well as the basics of Bayesian statistics, the thesis is divided into three different parts depending on which goal the Bayesian statistical methodology was used for:

- The first part focuses on data science and inflation, and it aims to constrain inflationary models using advanced inference techniques and forecasting tools. **Chapter 2** shows the first-ever results of the reconstruction of the speed of sound of the field responsible for driving inflation, using the latest cosmic microwave background (CMB) data from Planck 2018 and modern algorithms (Gaussian processes). **Chapter 3** is dedicated to the forecast of a particular class of single-field inflation models, known as α-attractors, for a future CMB stage-IV experiment using a model-dependent alternative approach for the sampling of the inflationary parameters based on current constraints obtained by cosmic microwave background and large scale structure data.

- The second part of the thesis is dedicated to the novel concept in cross-correlations of gravitational-wave (GW) physics and large scale structure observables; in particular, galaxy clustering. In two projects we study how we can exploit the information contained in these new observables by forecasting their behaviour and possible detection using future experimental set-ups and Bayesian statistics. **Chapter 4** studies unresolved GW events that form the astrophysical gravitational-wave background, and how we can use the cross-correlation of the anisotropies of that background with galaxy clustering to extract both astrophysical and cosmological information. On the other hand, in **Chapter 5** we investigate how machine learning techniques can be used to reconstruct the propagation of tensor perturbations by combining the spatial correlation between resolved GW mergers and galaxies.

- The third part of this thesis is dedicated to the *Euclid* mission and the tasks of the Euclid Consortium: goals, main survey features and the primary *Euclid* observational probes. In particular, **Chapter 6** focuses on a crucial data science analysis software for the mission: the code *Cosmological Likelihood for Observables in Euclid*, also known as `CLOE`. The Bayesian analysis tool `CLOE` is designed to be able to constrain the values of the parameters of a model given the future *Euclid* data through performing Bayesian inference. In this chapter, we describe the implemented cosmological recipe of the primary probes, as well as the description of the *Euclid* likelihood. The results concerning the structure of `CLOE` and its performance to constrain cosmological parameters are a preview of the current state of the *Euclid* mission, for which I have dedicated a vast percentage of my time.