GLLiM-ABC method for the analysis of the Cosmic Microwave Background

Gabriel Ducrocq

October 2022

1 Introduction

Today's cosmologists estimate the age of the universe to be 13 billions years, the birth of which is called the Big Bang. Some 380,000 years after that Big Bang, the universe cooled down, allowing the electrons to bond with protons to form hydrogen, making the universe transparent, and allowing light to travel through the universe. Looking far away in our universe today, we can still observe this light (in the form of a temperature signal) and this signal is called the Cosmic Microwave Background (CMB), see Figure 1. The anisotropies of the temperature signal (of the order of the micro Kelvin) tell us about our universe and its birth.

More precisely, cosmologist have devised the Λ-CDM model: 6 cosmological parameters $\theta \in \mathbb{R}^6$ (e.g the age of the universe) are used to describe the universe. Thanks to a statistical model provided by the cosmology theory, we are able to relate these parameters to the observed map of temperature:

$$\begin{array}{rcl} s|\theta & \sim & \mathcal{N}(0, C(\theta)) \\ d & = & As + n \end{array} \tag{1}$$

where s is the signal expressed in the spherical harmonics basis, $C(\theta)$ is the covariance of the signal, depending in a deterministic fashion on the cosmological parameters, A is the matrix transforming the spherical harmonics based signal into a pixel map, n is a Gaussian noise and d is the observed temperature skymap (see Fig 1), of dimension 10^5 to 10^7 for today's missions.

We are interested in recovering the cosmological parameters from the observed skymap. Unfortunately, even though the likelihood is Gaussian, the dimensionality of the problem is so high that it rules out a maximum likelihood method. Instead, in a Bayesian framework, we can set a prior on the cosmological parameters and try to sample from the corresponding posterior distribution. Even though MCMC approaches have been useful in the past, the dimensionality of the current observations and the high computational cost associated tend to make these methods unefficient.

The recent years have seen a development of the likelihood free approaches (e.g Approximate Bayesian Computation) in the statistical literature. Attempts have been made to apply these methods to the described project. Unfortunately, they require to find low dimensional statistics summarizing the data, which is difficult. The project aims to:

1. Implement the GLLiM-ABC [1] approach in the context of CMB data analysis, allowing for a low dimensional statistic.
2. Compare it with a recent deep learning based likelihood-free approach [2] in terms of the sample size required and ease of implementation.
3. If we have time, compare with other likelihood-free approaches.

Only a generic statistical background is required, but a taste for computational statistics is a plus, especially if you already have some experience with likelihood-free methods. No knowledge of cosmology required.

Contact: Gabriel Ducrocq
Email: gabriel.ducrocq@liu.se

References

[1] F. Forbes, H. D. Nguyen, T. Nguyen, and J. Arbel, "Summary statistics and discrepancy measures for approximate Bayesian computation via surrogate posteriors," *Statistics and Computing*, vol. 32, p. 85, Oct. 2022.

[2] A. Cole, B. K. Miller, S. J. Witte, M. X. Cai, M. W. Grootes, F. Nattino, and C. Weniger, "Fast and credible likelihood-free cosmology with truncated marginal neural ratio estimation," *Journal of Cosmology and Astroparticle Physics*, vol. 2022, p. 004, sep 2022.

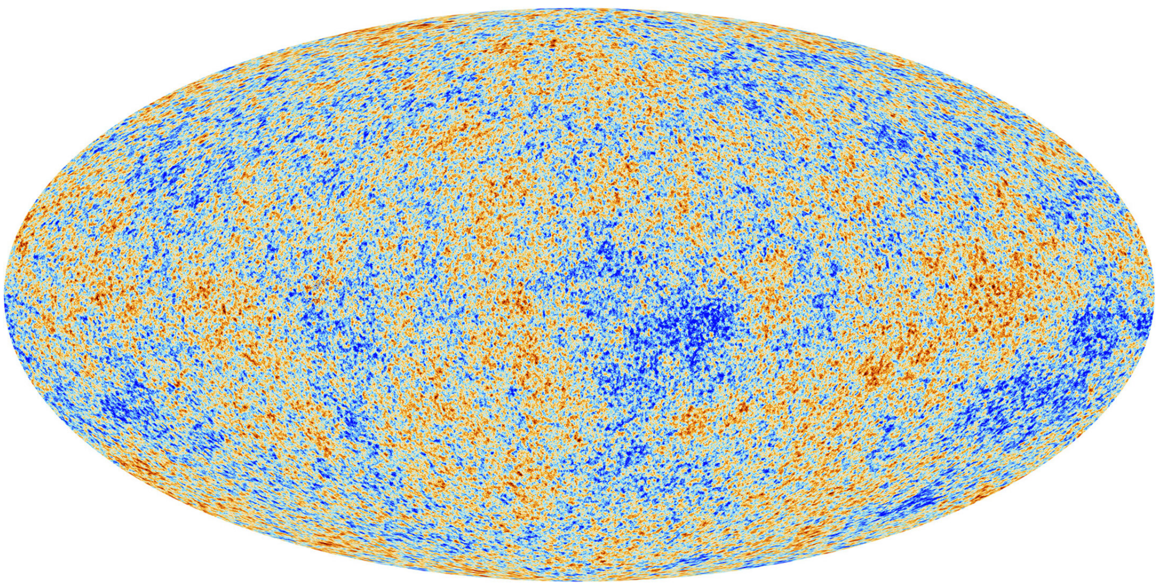

Figure 1: At each time step, the Planck satellite points toward a direction in the sky, measuring the temperature coming from the universe (the Cosmic Microwave Background). We are thus observing a sphere around us, which explains the shape of the map. We are interested in the anisotropies of this signal, which are of the order of the micro-kelvin.