A Final Warning

Assignment- 6

Suprio Dubey
2013036

23/12/2022

Question

READ THE FOLLOWING TWO PAPERS

Arxiv: https://arxiv.org/abs/2203.16285
Title: A fast test to assess the impact of marginalization in Monte Carlo analyses, and its application to cosmology
Author: Adrià Gómez-Valent
P1Q1) What is the main message of paper 1?
P1Q2) What are the "volume effects" and why is it a problem for Bayesian analysis?
P1Q3) What is the profile likelihood approach and why it does not suffer from volume effects?

2) Arxiv: https://arxiv.org/abs/2112.12140
Title: New constraint on Early Dark Energy from Planck and BOSS data using the profile likelihood Authors: Laura Herold, Elisa G. M. Ferreira, Eiichiro Komatsu
P2Q1) What is the main message of paper 2?
P2Q2) Discuss why, in your opinion, volume effects appear in this paper, and if could be possible to avoid them changing something in the analysis.

ANSWER 1

1.1

The main message of the paper was to introduce a method called profile distribution(PD) that helps with the "volume effects " coming from the marginalization procedure which introduces a non-negligible bias into our conclusions. The PD allows to detection of marginalization biases directly from the Markov Chains.

1.2 Volume Effects

Markov Chain methods are used to sample multivariate distribution in high-parameter spaces. The posterior distributions of the subsets are obtained through marginalization which helps in the visualization and interpretation of results in Bayesian Analyses.
When the non-marginalized probability density has non-Gaussian features the marginalized posteriors derived from it do not show us how the parameters are distributed according to their ability to explain the data, but according to their integrated probability weight there can be a big mismatch between these two distributions, and in some cases, the posterior can hide points in parameter space that is able to fit very well the data. This is called the "Volume Effect"
The one-dimensional marginalized distribution of a particular parameter does not tell us directly what values of that parameter fit better the data, but the probability that a particular value of that parameter explains the data. The latter is computed by "summing" the contribution of all points in parameter space, given a fixed value of the parameter of interest, obtaining in this way an integrated probability. This introduces "volume effects". The "volume effect" causes some parameters to be underrated in the marginalized distribution, even though they may fit the data better or equally well. This is because a smaller volume of the parameter space is occupied by the points that result in a good description of the data for a given value of the parameter of interest. The marginalized distribution will typically hide values of that parameter that can fit the data extremely well but are located in too small regions of the parameter space.

1.3 Profile likelihood Approach

Given Profile likelihood $\mathcal{L}(\theta_1, \theta_2)$ is the normalized distribution for the parameter sets θ_1 and θ_2. We compute the Profile likelihood for θ_1, called $\tilde{\mathcal{L}}(\theta_1)$, by searching for the maximum of $\mathcal{L}$ for each θ_1 along the directions of θ_2, i.e.

$$\tilde{\mathcal{L}}(\theta_1) = \max_{\theta_2} \mathcal{L}(\theta_1, \theta_2).$$

$$R(\theta_1) = \frac{\tilde{\mathcal{L}}(\theta_1)}{\max_{\theta_1} \tilde{\mathcal{L}}(\theta_1)} = \frac{\tilde{\mathcal{L}}(\theta_1)}{\max_{\theta_1,\theta_2} \mathcal{L}(\theta_1, \theta_2)}$$

which is interpreted as a probability weight for every point in the space of θ_1 and, hence, we can employ this quantity to build the distribution of this parameter set. The main advantage here is that we don't need to marginalize hence avoiding the "Volume Effect" by avoiding integration.

$\mathcal{L}(\theta_1) = \int \mathcal{L}(\theta_1, \theta_2)\, d\theta_2$,

Hence, the ratio is not subject to volume effects, since we are not performing any integration. Its maximum is located at the very same value of θ_1 that maximizes $\mathcal{L}$, and the Profile likelihood shows explicitly what values of θ_1 lead to a better description of the data, taking also into account the information from the prior. The analogous quantities for θ_2 or any subset built from θ_1 and/or θ_2 can be also computed.

ANSWER 2

2.1

The objective of the paper was to understand whether EDE is a possible solution to the Hubble tension. It is a dark energy-like component in the early universe. In this model, the λ_{CDM} cosmology is extended to include a dark energy-like component in the pre-recombination era, which reduces the size of the sound horizon and increases H_0. EDE is typically parametrized by three parameters: the initial value of the EDE field (θ_i), its maximum fractional energy density f_{EDE} and the critical redshift (z_c) at which this maximum fraction is reached.
A larger value of f_{EDE} leads to a higher H_0. To solve the Hubble tension, it was predicted that $f_{EDE} \approxeq 0.1$ would be necessary. However, there have been disagreements.
It was seen using two different analyses that the "Volume Effects" are present in the 3-parameter MCMC analysis as they prefer small f_{EDE}.
It is seen that higher values of f_{EDE} are more consistent in the data from grid sampling. Also, profile likelihood analyses confirmed that.
It was seen that profile likelihood is a suitable method to analyze EDE model and determine f_{EDE} as they don't suffer from "Volume Effects". The profile likelihood does not suffer from volume effects, since the minimum $\chi^2_{f_{EDE}}$ is the same as the maximum likelihood estimate. It was seen by both analyses that the f_{EDE} allowed by the data strongly depends on the particular choice of the other parameters of the model and that several choices of these parameters prefer larger values of f_{EDE}

than in the Markov Chain Monte Carlo analysis.

2.2

The main issue in the different values of f_{EDE} seemed to be in the Markov Chain Monte Carlo (MCMC) sampling of the three parameters of the EDE model. For $f_{EDE} = 0$, the EDE model is degenerate with λ_{CDM} for any choice of (θ_i) and z_c. This is a reason why the parameter volume for $f_{EDE} = 0$ is larger than for every $f_{EDE} > 0$. This can lead to a preference for $f_{EDE} = 0$ in the marginalized posterior, affecting the inferred amount of EDE allowed by the data. Also, was seen by Smith et al. (2021); that fixing some parameters of the model the result depends on the particular choice of the parameters. In the paper to check whether the effect on the 3-parameter model is due to the two unconstrained parameters (θ_i) and z_c or by volume effects they performed grid sampling. They fix the parameter space of $((\theta_i), z_c)$ to a wide range of values and explore the parameter space by performing a 1-parameter analysis.
It is found that f_{EDE} strongly depends on the choice of (θ_i) and z_c. It is seen that higher values of f_{EDE} are more consistent in the data from grid sampling hinting the "Volume Effects" are present in the 3-parameter MCMC analysis as they prefer small f_{EDE}.
It is possible to avoid the "volume effect".The Volume effect is avoided by the grid method since there is no larger prior volume at $f_{EDE} = 0$ compared to $f_{EDE} > 0$ when (θ_i) and z_c are fixed. Now, taking hints from the grid analysis a profile likelihood is constructed by fixing the parameter of interest i.e f_{EDE} to different values and maximizing the likelihood $\mathcal{L}$ or minimize $\chi^2 = -2ln(\mathcal{L})$ w.r,t to all other parameters as well as all nuisance parameter.
For every fixed value of f_{EDE}, a long MCMC (with at least 104 accepted steps) until the Gelman-Rubin criterion $(R1) < 0.25$ is reached. The profile likelihood shows that the best fit value obtained lay near the upper bound $f_{EDE} < 0.72(95\%CL)$ of Ivanov et al. (2020a) which was another strong indication that the MCMC analysis of the 3-parameter model is indeed plagued by volume effects and can be avoided by profile likelihood.
As the profile likelihood does not suffer from volume effects, since the minimum $\chi^2_{f_{EDE}}$ is the same as the maximum likelihood estimate.