Cosmology Final Proyect

Laura del Carmen Cabal Paramo

June 16, 2022

Abstract

In this project we explore how the Bayesian inference is used in cosmology and astrophysics. To view how this works, we will implement the Metropolis Hastings algorithm in python, we will try probing this code as well so that we can verify that this works.

1 Introduction

To begin, we need to understand what Bayes theorem is, Bayes theorem is a concept of probability theory viewed in statistics in which the main objective is determining the probability of an event occurring based on prior knowledge of conditions that might be related to the event. The theorem is stated as the following:

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)} \tag{1}$$

Where:

- $P(A|B)$ is the probability of event A occurring given that B is true. This is called the "posterior"
- $P(B|A)$ is the probability of event B occurring given that A is true. This is called the "likelihood"
- $P(A)$ is the probability of event A ocurring. This is also called the "prior"
- $P(B)$ is the probability of event B ocurring. This is also called the "evidence"

Before we continue with bayesian interference, we must know first what statistical interference is. Statistical interference is a process where we use data analysis to infer properties of a probability distribution, meaning it can be used to test hypotheses and obtaining or deriving some estimates.
Bayesian interference is a type of statistical interference where we use Bayes' theorem 1 to obtain the probability of an event or hypothesis of occurring or being true.

Now, referring to the likelihood we are going to use, the next equation 2 is what we are using in Bayesian inference.

$$L = \prod_i \frac{e^{-(yi - f(xi, parametros\ libres))^2} / 2\sigma_i^2}{\sqrt{2\pi}\sigma_i} \tag{2}$$

This form of likelihood is as we can observe, Gaussian, so with the Bayesian Inference we are following a normal statistics. This is because we are assuming or forcing (This is something I don't recall much) it to be this way and we are ignoring P(B) as we mention next since its a normalizing constant.
Bayesian methods allows greater emphasis to be given to scientific interest and less to mathematical convenience. [1] Bayesian interference is about maximizing the posterior and since the evidence P(B) is a normalizing constant that doesn't dependent on A, we can ignore it, also, the prior distribution is something we define with the information that we already know about the event.

Bayesian inference is represented as equation 3, which is very similar to Bayes theorem, as the name obviously suggests.

$$P(\theta|Data) = \frac{L(Data|\theta) \cdot Pr(\theta)}{E(Data)} \tag{3}$$

Where:

- θ is the free parameters we are varying.
- $Data$ is the data obtained to use bayesian inference in a model

In cosmology, bayesian interference or markov chain montecarlo (MCMC) algorithms are used because as time passed, cosmologists started to notice that observational data was not consistent with theory, so in recent years, as more cosmological observations were made, with a large variety of data, it became necessary to adapt the cosmological models with the obtained data. We cannot use frequentist statistics to interpret the data, since we only have one unique universe, that's why bayesian inference is the most appropriate in astronomy and cosmology. In other words, when observations are made, we get a lot of data since making this observations are difficult so, we need to use statistics that don't require frequent observations, hence Bayesian inference.

It is used more precisely to vary the parameters of a model we propose so that we find those that adapt to the data we observe, much like what we tried replicating in this project with the Distance Module and the data obtained from supernova. [2]

In this project, we are presenting the process described previously, using the model that is used to obtain the Module Distance, we are varying the parameters Ω_{lambda} and Ω_m to adapt these to the observed data that was obtained in [3] using Supernova explosions. Now that we know what Bayesian inference and its use in cosmology is, it is clear that this is related to Project 2 in the way that we varied the cosmological parameters to see what can adapt to the data already present in CLASS. This way it is much easier since there is a lot of data to work with, hence the many default files in output that we got.

2 Methods

For this project, I chose option 1 and I implemented the Metropolis Hastings algorithm so that I could infer the free parameters for a model. I started with

the model of a straight line, from which the parameters I infered were the m (slope of the line) and the b (ordinate to the origin of the line).
Since you had to define your prior based on the information you have before, I started varying the way I defined it, first I had a prior of the form of equation 4.

$$Prior = P(A) = \frac{1}{\theta_1}\frac{1}{\theta_2} \tag{4}$$

In this case, I was just varying or playing with how I could manipulate. I finally settled for 5

$$Prior = P(A) = \frac{\theta_1}{\theta_2} \tag{5}$$

When I finished the algorithm (for the second try since first attempt didn't work for me) I noticed that everytime I ran the algorithm, I got a diffrent configuration for my graph of θ_1 vs θ_2, so I ran it till I saw the points converging or grouping in a region. So the line test worked.

After this I tried to use the SN data to use the model of the module distance, using the library astropy cosmopy I defined the model with Ω_m and Ω_Λ as free parameters and did the thing I had done in Homework 4 (playing with this library and data so that we could compare to project 1). I started noticing I had complex number errors, so I used a try/except in python to ignore the values I couldn't use, much like the burn in (which are values or data form applying my algorithm that I simply can ignore, we can see this in our graph, this is the data that is not grouped). After this, I sometimes got an Integral error, but since I ignored this errors, I didn't pay much attention.
After a while of running the algorithm, I noticed that it took approximately 3 minutes each time I ran it, which is a lot compared to the line.
In the end, I obtained a graph of θ_1 vs θ_2, and a trace plot for each. I wasn't able to get a 1D and 2D plot since I couldnt get the library getDistPlot to work for me, maybe I needed to obtain more values than the ones I did.

3 Results and Discussion

As I mentioned, I was only able to obtain certain graphs, but I think we are able to get an idea at least of how this algorithm works. This is the plot of the straight line I was trying to apply the algorithm to, its simple and is a great view to see how it works. As we can see from graph 2, data starts to group in a certain region (2 for m or θ_1 and 1 for θ_2 or b), the points that are not in the grouping region are the burn in, or the data that isn't relevant and can be ignored. Finally for the straight line, we have the trace plot 3, which should look like a lot of lines of different size all grouped, like something fuzzy but in a straight line, showing that it is something Gaussian we are talking about, but something happened and there is a sudden well in the plot for both parameters.
Now lets discuss what happened with the Module distance model.
Graph 4 is the data we had to work with to vary the free parameters with Metropolis Hastings algorithm. The plot from 5 describes a certain iteration of the algorithm, as we can observe, nothing groups in a certain region nor converges, I ran this several times, hoping that something similar may be obtained like the straight line, instead every 3 minutes I ran this, I got worse

things. After obtaining absolutely nothing from this plot, I decided to do the trace plot in hopes that something could be visible there, instead I got graph 6. This graph seems weird, Ω_m keeps growing, while Ω_{lambda} seems like something reasonable, I think from this graph, we can see clearly that this doesn't converge.

4 Conclusions

This project can be found in https://github.com/l-cabal/CosmologyFinalProject_DCI_LCCP.
I don't know still how github works, so I couldn't delete my incomplete code from the repository. The final one is the one named Recent_MetropolisHastings-FinalProjectCOSMO_LCCP.ipynb.
In this project we were able to see and understand how statistics work in cosmology, particularly how the Metropolis Hastings algorithm worked. How we obtain the parameters value to get the theory of cosmology right is something I never asked myself and it really amazes me now that I know how this works. I never viewed the way how theory and observational data obtained worked, I always thought that what happened in theory, happened in the observations. Data analysis which is a bit of what we did in this project, is an area that can be so useful, but is often misunderstood, so I was glad to get to work a bit around this.
Even though I wasn't able to obtain a great graph or results from the Module Distance model, I feel somewhat relieved that I could at least obtain and see how it works in the straight line example. Now I know how we can use and apply algorithms and bayesian inference, as well as its relevance in cosmology and astrophysics.

5 Notes

- You will have to provide the code used to generate the plots, either in scripts, jupyter notebooks, or linking to a github page (preferable)
- Make a first submission on June 14 2022, or before, and final submission on June 17, 2022.

References

[1] George EP Box and George C Tiao. *Bayesian inference in statistical analysis*. John Wiley & Sons, 2011.

[2] Luis E Padilla, Luis O Tellez, Luis A Escamilla, and Jose Alberto Vazquez. Cosmological parameter inference with bayesian statistics. *Universe*, 7(7):213, 2021.

[3] N. Suzuki, D. Rubin, C. Lidman, G. Aldering, R. Amanullah, K. Barbary, L. F. Barrientos, J. Botyanszki, M. Brodwin, N. Connolly, K. S. Dawson, A. Dey, M. Doi, M. Donahue, S. Deustua, P. Eisenhardt, E. Ellingson, L. Faccioli, V. Fadeyev, H. K. Fakhouri, A. S. Fruchter, D. G. Gilbank,

M. D. Gladders, G. Goldhaber, A. H. Gonzalez, A. Goobar, A. Gude, T. Hattori, H. Hoekstra, E. Hsiao, X. Huang, Y. Ihara, M. J. Jee, D. Johnston, N. Kashikawa, B. Koester, K. Konishi, M. Kowalski, E. V. Linder, L. Lubin, J. Melbourne, J. Meyers, T. Morokuma, F. Munshi, C. Mullis, T. Oda, N. Panagia, S. Perlmutter, M. Postman, T. Pritchard, J. Rhodes, P. Ripoche, P. Rosati, D. J. Schlegel, A. Spadafora, S. A. Stanford, V. Stanishev, D. Stern, M. Strovink, N. Takanashi, K. Tokita, M. Wagner, L. Wang, N. Yasuda, H. K. C. Yee, and The Supernova Cosmology Project. The Hubble Space Telescope Cluster Supernova Survey. V. Improving the Dark-energy Constraints above z ¿ 1 and Building an Early-type-hosted Supernova Sample. , 746(1):85, February 2012.

6 Annexes

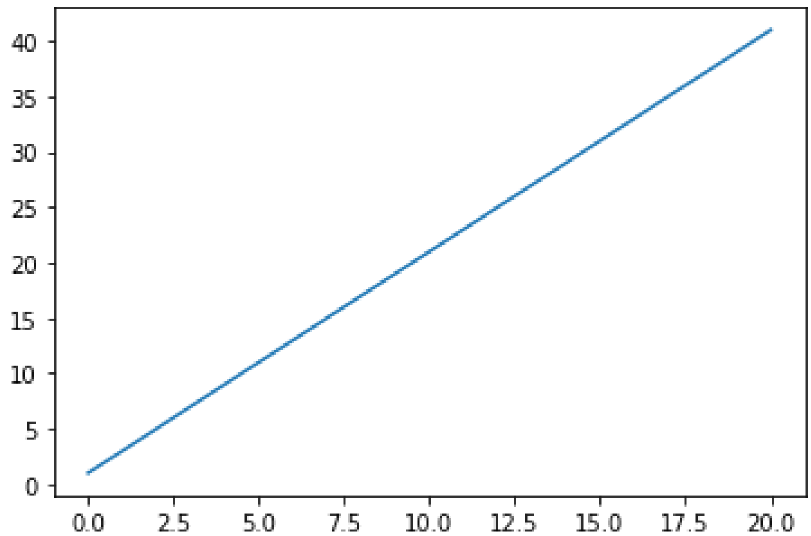

Figure 1: Plot of straight line y = 2x+1

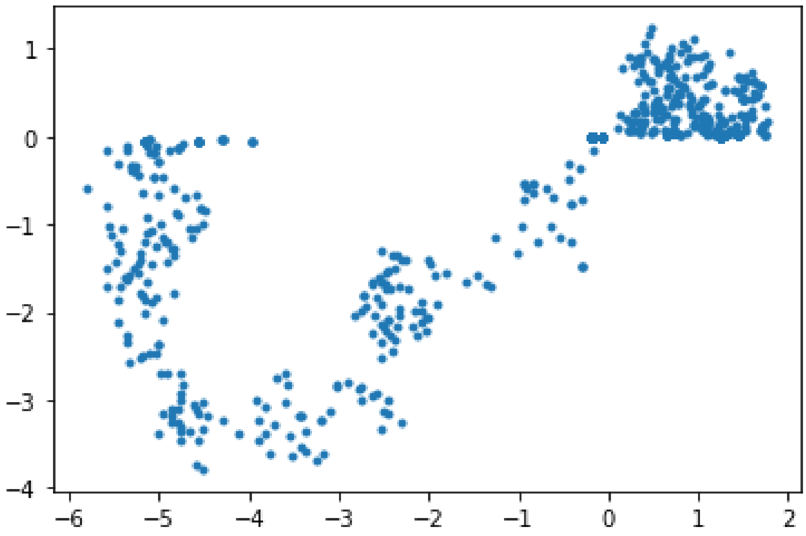

Figure 2: The plot of θ_1 vs θ_2 of the straight line y = 2x+1

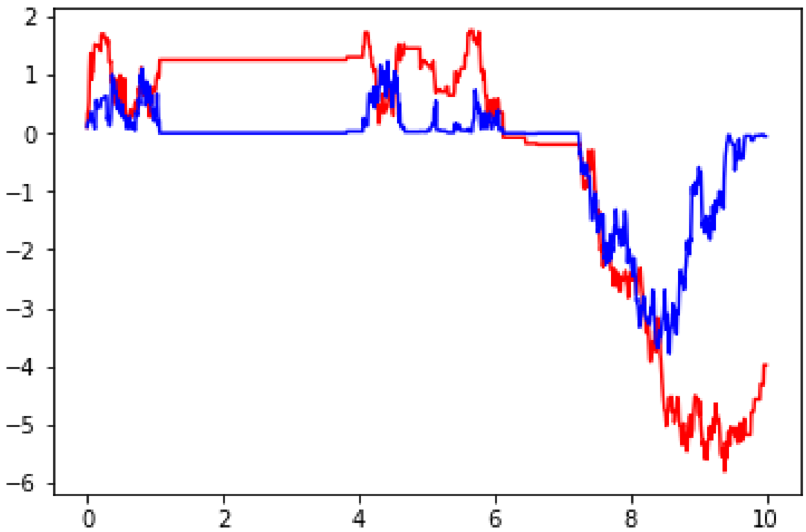

Figure 3: The trace plot of straight line y = 2x+1

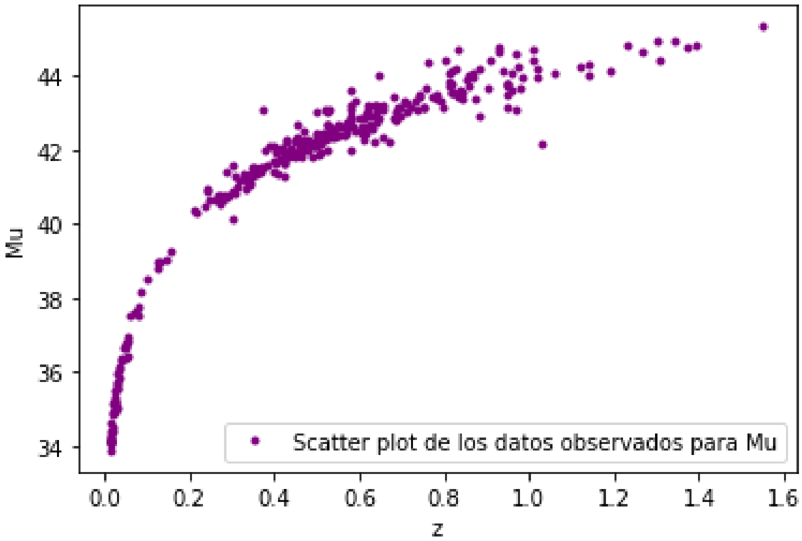

Figure 4: The data we have of Supernova, graphing the module distance.

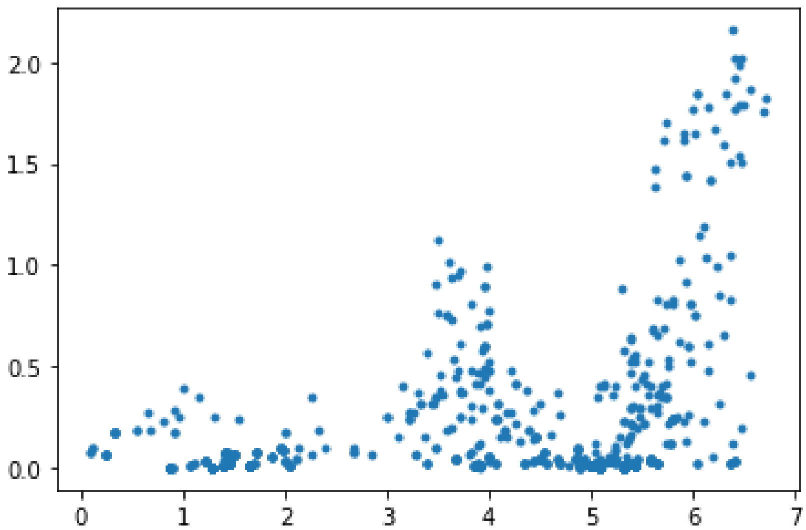

Figure 5: Plot of θ_1 vs θ_2 for the module distance model

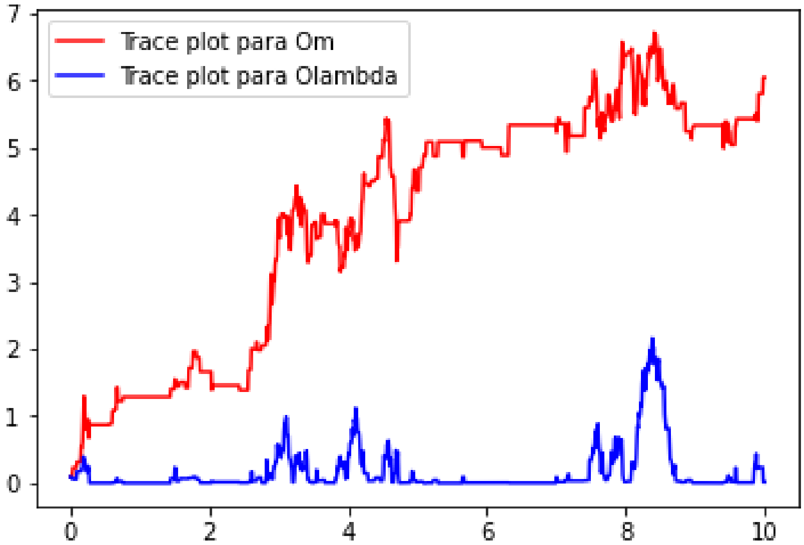

Figure 6: Trace Plot for the module distance model