Model Dependent and Independent Methods of Cosmological Parameter Estimation

Parth Kapoor
Roll No: MS16053

A dissertation submitted for the partial fulfilment
of BS-MS dual degree in Science

Under the guidence of
Dr. Harvinder Kaur Jassal

April 2020

Indian Institute of Science Education and Research Mohali
Sector - 81, SAS Nagar, Mohali 140306, Punjab, India

Certificate of Examination

This is to certify that the dissertation titled **"Model Dependent and Independent Cosmological Parameter Estimation"** submitted by **Parth Kapoor** (Reg. No. MS16053) for the partial fulfillment of BS-MS dual degree programme of the Institute, has been examined by the thesis committee duly appointed by the Institute. The committee finds the work done by the candidate satisfactory and recommends that the report be accepted.

Prof. Jasjeet S Bagla Prof. Sudeshna Sinha Dr. Harvinder K Jassal (Supervisor)

Dated: 9.04.2020

Declaration

The work presented in this dissertation has been carried out by me under the guidance of Dr. Harvinder Kaur Jassal at the Indian Institute of Science Education and Research Mohali.

This work has not been submitted in part or in full for a degree, a diploma, or a fellowship to any other university or institute. Whenever contributions of others are involved, every effort is made to indicate this clearly, with due acknowledgement of collaborative research and discussions. This thesis is a bonafide record of original work done by me and all sources listed within have been detailed in the bibliography.

Parth Kapoor
(Candidate)

Dated: April 9, 2020

In my capacity as the supervisor of the candidate's project work, I certify that the above statements by the candidate are true to the best of my knowledge.

Dr. Harvinder Kaur Jassal
(Supervisor)

Acknowledgement

First and foremost, I would like to thank Dr. Harvinder Kaur Jassal for her guidance throughout the project and on the work contained in this thesis.

I would like to thank my parents and my sister for their continued support, without which none of this would have been imaginable.

Lastly, I would like to thank Deesha Divecha for always being there to lend an ear, provide valuable insight, and thought provoking conversation.

Parth Kapoor

MS16053

IISER Mohali

List of Figures

Contents

Abstract

The universe is known to be expanding, the cause of which is said to be a mysterious, *dark energy*, named so, as it is of unknown origin. There are several candidate models, each of which bring with them their own set of free parameters that can be tuned to fit available data. In fact, too many such models exist; that *is* the problem — not enough of them have been falsified.

In this thesis, we will look at two ways to recreate dark energy — a parametric and a non-parametric, model independent approach. Along the way, we will learn about Bayesian statistics, multi-variate probability distributions, Markov Chain Monte Carlo methods, and Gaussian processes — tools we will use to analyze data from observations in an attempt to nail down the origin of dark energy.

Chapter 1

Introduction

The universe is expanding, this much we know. The origin of this expansion however is an entirely separate matter, and its investigation is what is contained in this thesis.
In this thesis, we will encounter a historical account of events that have led to the postulation of dark energy, we will learn the theory behind Cosmology - the study of the universe, its origins, and its dynamics; we will also get familiar with some of the statistical techniques needed to make use of the wealth of data that has been made available thanks to recent developments in observational techniques, and finally we will learn to make use of computational methods that are needed to implement those statistical techniques.

1.1 A Historical Perspective

Einstein's theory of General Relativity gave us a description of spacetime as a dynamic, ever-changing backdrop against which events took place, while also giving us a prescription for exactly how those changes were brought on by the presence of matter and energy. At the time however, common sense guided him to assume a universe that was static; changing in shape and curvature due to the presence of matter, but fixed in size. This was not a universe his own theory allowed for, and to enforce it, he made what in his own words was the "greatest blunder of his life" - the introduction of a cosmological constant into his field equations - something to counter the attractive force of gravity, and keep the universe from shrinking or expanding.
This cosmological constant, Λ, is a free parameter, and its value was fine tuned to achieve

equilibrium against the force of gravity. This equilibrium solution to the Field Equations was valid, but unstable. No matter what, a stable, static universe was not something his own theory allowed for [Wei89].

This problem however, was not a problem for long, as observations confirmed that the universe was far from static - it was expanding, and accelerating in fact. The problem was then turned on its head. The cosmological constant that was earlier being used to force a static solution was now required to better fit the accelerated, expanding model of the universe, no longer a cohesive but a repulsive factor. This mysterious new mathematical addition to the energy densities that made up the universe was dubbed Dark Energy.

In chapter 2, we will learn about the Cosmology required to understand the problem more rigorously. We will talk about how the distribution of matter and energy affect the evolution of the universe as a whole, and how we can measure that effect in the form of two cosmological observables, the Hubble parameter and the Luminosity Distance.

Chapters 3 and 4 will be spent learning about the methods we will be employing to tackle our phenomenological problem - Markov Chain Monte Carlo as an alternative to a brute force grid search in order to reconstruct dark energy parametrically, and Gaussian Processes to do the same independent of any particular dark energy model.

Finally, in chapter 5 we will encounter a restatement of our problem in terms of the methods described above, and the results that follow.

Chapter 2

Cosmology

Cosmology is the study of the universe itself - it's past, present, and future. On these unfathomable scales, a few assumptions go a long way, and reduce the complexity of this study to just a couple of simple equations, what we call the Friedman Equations, and that is where we will begin too. We will take our knowledge of general relativity, and add on to it what we learn about the distribution of matter and energy in the universe, which will lead us to the Friedmann Equations. This discussion of distributions, and the final equations would finally pave the way to really appreciating the problem of the origin of dark energy as we begin to learn about some of the observables of cosmology.
For more details on the mathematics contained in this chapter, refer to [Car19] [Pad02] [Wei72].

2.1 The Friedmann Equations

Einstein's theory of general relativity gave us a system of equations, called the Einstein Field Equations,

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} = 8\pi G T_{\mu\nu}. \tag{2.1}$$

The left and right hand sides (LHS & RHS) of this equation can be thought of to represent two distinct things. On the LHS, the presence of the Ricci tensor, $R_{\mu\nu}$, the Ricci scalar, R, and the metric tensor, $g_{\mu\nu}$, translate into that part of the equation holding the geometrical properties of space-time. On the RHS, we have the stress energy tensor, $T_{\mu\nu}$, which holds the information about the distribution of the various matter or energy densities, the presence of which affect the geometrical properties of space-time, and that relation is what is held in

this equation.

The **Ricci tensor** can be obtained as a contraction of the **Riemannian tensor**, $R^{\rho}_{\sigma\mu\nu}$, which is defined by,

$$R^{\rho}_{\sigma\mu\nu} = \partial_{\mu}\Gamma^{\rho}_{\nu\sigma} - \partial_{\nu}\Gamma^{\rho}_{\mu\sigma} + \Gamma^{\rho}_{\mu\lambda}\Gamma^{\lambda}_{\nu\sigma} - \Gamma^{\rho}_{\nu\lambda}\Gamma^{\lambda}_{\mu\sigma}, \tag{2.2}$$

where the Γ terms are the **Christoffel Symbols**. Contracting the indices using tensor calculus rules,

$$R_{\mu\nu} = \partial_{\lambda}\Gamma^{\lambda}_{\mu\nu} - \partial_{\mu}\Gamma^{\lambda}_{\lambda\nu} + \Gamma^{\lambda}_{\lambda\delta}\Gamma^{\delta}_{\mu\nu} - \Gamma^{\lambda}_{\mu\delta}\Gamma^{\delta}_{\lambda\nu}. \tag{2.3}$$

The final term of interest on the left hand side is the familiar metric tensor. As an note, when talking about Lorentzian space-time, we will take the form of the tensor to be, $\eta_{\mu\nu} = diag\ (-1,\ 1,\ 1,\ 1)$, instead of the general $g_{\mu\nu}$.
On the right hand side, we have G, Newton's gravitational constant, and the stress energy tensor. The particular form of the stress energy tensor depends on the form of the matter and energy distributions we choose to work with. This choice is guided by what is known as **The Cosmological Principle**.

2.1.1 The Cosmological Principle

This is one of the most fundamental assumption that cosmology works under, that no place or direction in the universe is special. In more concrete terms, the universe is

- homogeneous, so there are no special points in space where the geometric properties of space would be different from any other point.
- isotropic, meaning that there are no special direction, where measurements would yield results different from any other direction.

2.1.2 The Distribution of Matter and Energy

Given the homogeneous, isotropic universe, we can go ahead and assume the matter and energy distributions to behave as a perfect fluid at cosmic scales. This simplifies our job greatly, as the form for the stress energy tensor simply becomes,

$$T = diag\ (-\rho,\ p,\ p,\ p), \tag{2.4}$$

where ρ is the energy density, and p is the flux associated with the momentum of that particular energy density.

When we attempt to conserve energy, and apply the condition that,

$$\nabla_\nu T^{\mu\nu} = 0, \tag{2.5}$$

we find that its zeroth, time component tells us,

$$\nabla_\mu T^{0\mu} = -\partial_0 \rho - 3\frac{\dot{a}}{a}(p + \rho) = 0. \tag{2.6}$$

We can simplify that relation by assuming an equation of state that relates ρ and p,

$$p = \omega\rho, \tag{2.7}$$

which allows us to conclude that,

$$\frac{\dot{\rho}}{\rho} = -3(1 + \omega)\frac{\dot{a}}{a}, \tag{2.8}$$

which can be integrated to give a simple relation,

$$\rho \propto a^{-3(1+\omega)}. \tag{2.9}$$

Here, a is the scale factor of the universe. What this equation tells us is how the energy density changes as the universe changes in scale. We know for a fact that the universe is expanding, but so far in this thesis we have not yet introduced that, and so it suffices to treat this equation in general.

The actual value of the term to which the scale factor is raised depends on what particular energy density is being considered. We will now look at two examples of the same - matter, and radiation.

Matter in the universe is modelled as a pressure-less dust, and so for it, $p = 0$. So, for the relation, $\omega\ \rho_M = 0$, to make sense, we must have $\omega = 0$. Which allows us to conclude that,

$$\rho_M \propto a^{-3}. \tag{2.10}$$

Radiation follows Maxwell's laws. When written in their field theoretic, tensor forms, we know that the trace of perfect fluid energy momentum tensor would be equal to that of the trace of the tensor for electromagnetic radiation,

$$T^\mu_\mu = -\rho + 3p = F^{\mu\lambda}F_{\mu\lambda} - \frac{1}{4}g^{\mu\mu}F^{\lambda\sigma}F_{\lambda\sigma} = 0. \tag{2.11}$$

Simplifying, we can get, $p_R = \frac{1}{3}\rho_R$, which gives us,

$$\rho_R \propto a^{-4}. \tag{2.12}$$

Both of these relations make intuitive sense too; if we imagine the universe expanding, any cube in space would have it's volume increase as a power of three, but the matter in it remains the same, and so the density of that matter would fall of as an inverse cube. For radiation, a similar dilation of photon density would take place, but on top of that, individual photons would also lose energy as they get red-shifted because of the expansion of the universe and the apparent motion of the source of said photon away from the observers.

2.1.3 The Equations

Assuming several different energy densities, overlaid on top of each other, all obeying the perfect fluid condition, we can take the combination of them, and plug the resulting energy momentum tensor in Einstein's Field Equations (insert label and link here) that we encountered earlier. Doing this and simplifying gives us two equations, namely the **Friedmann Equations**.

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{\kappa}{a^2}, \tag{2.13}$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3p). \tag{2.14}$$

Here, p and ρ have their usual meanings, as does a, and the κ term represents the curvature of space-time. It takes values $-1,\ 0,\ 1$ depending on whether the universe is closed, flat, or open respectively.

Individually, these equations are creatively called the first and second Friedmann equations. For our purposes, we will mainly be focusing on the former, as it will bring us to our very first observable, the Hubble Parameter.

2.2 The Hubble Parameter

The equations we obtained in the previous section, can be written down in a simpler way by introducing a few new parameters. We define the Hubble Parameter, H, as,

$$H = \frac{\dot{a}}{a}. \tag{2.15}$$

This allows us to write down the first Friedmann equation as,

$$H^2 = \frac{8\pi G}{3}\rho - \frac{\kappa}{a^2}.$$

We also introduce the density parameter, Ω,

$$\Omega = \frac{8\pi G}{3H^2}\rho, \tag{2.16}$$

using which we can introduce a energy density for the curvature term in the Friedmann equations. So we say that,

$$\Omega_c = -\frac{\kappa}{H^2 a^2}, \tag{2.17}$$

which we can either work with as is, and write down the first Friedmann equation as,

$$\sum \Omega_i = 1, \tag{2.18}$$

or we can "reverse engineer" the definition for Ω_c, to get,

$$H^2 = \frac{8\pi G}{3}\sum \rho_i, \tag{2.19}$$

where in both cases, the summation over index i represents the different energy densities we would like to consider in our model (including curvature).

This parameter, H, measures the rate at which the scale of the universe is changing, whether it is expanding or contracting, accelerating or decelerating. The Hubble parameter actually comes in two "varieties", H_0 and $H(z)$, depending on whether we fix a frame of reference or not, as explained below and visualised in figure 2.1 [Col18]. To understand the difference between the two, let us assume that the universe is expanding. What this means is that every other point in the universe is moving away from every other point as the space between them expands, irregardless of any other velocities they might have had. If there existed some kind of lattice in space with a separation of one unit, after expansion, the separation between each point on the lattice would be two units. No particular point on the lattice moved away from another point specifically. But, if we were to focus on only one of such points, and overlay our standard lattice with the expanded lattice, such that the point we had chosen was aligned to itself in both lattices, then we would see that the expansion *appears* to be radially away from that one particular point on the lattice.

The Hubble Parameter at current epoch, H_0, is a measure of the actual change in scale, at the current point in the universe's age; while $H(z)$ measures the apparent rate of other

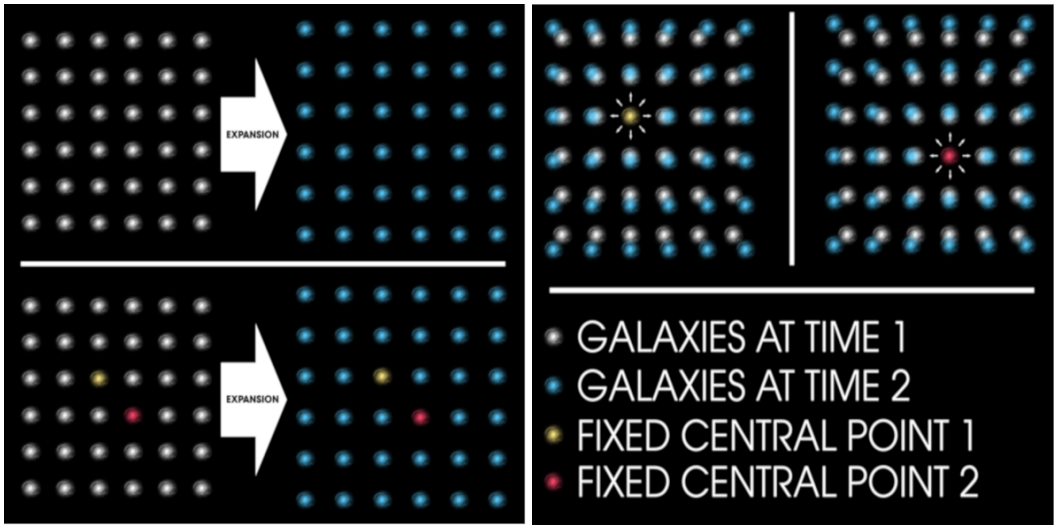

Figure 2.1: Once we choose a special point in space, it appears that the rest of the universe is expanding away from that point.

points in the universe moving farther or closer with respect to a particular point - like the Earth, as a function of z, the red-shift, which in cosmology is used as a measure of distance.

2.3 Luminosity Distance

This will serve as the second cosmological observable we will be making use of. Consider a bright object, like a light-bulb in one corner of a room. We know that the intensity of the light will be falling off as an inverse square law as the light-bulb would serve as a point source of light from which photons would be radiating radially. Having that knowledge in hand, and at least approximate knowledge of what the brightness of the light-bulb source might be, we could deduce the distance that the light-bulb would be from us by measuring its apparent brightness as the light reaches us. Now, imagine the same on a cosmic scale, with light-bulbs replaced by supernovae, and the room replaced by the expanding universe itself. The distance so measured, is called the luminosity distance and is denoted by d_L.
For the former light-bulb, existing in Euclidean space, we define the luminosity distance as,

$$d_L^2 = \frac{L}{4\pi F}, \tag{2.20}$$

where L is the brightness or luminosity at the source, and F is the flux measured at the receiver.
Before we go ahead and define the luminosity distance for an FRW universe, we should

take a moment and loop back to the density parameters we encountered before. We know the relation between ρ and a from equation (2.9); using that and the definition of Ω, we can write,

$$\Omega_i = \Omega_{0_i}(1+z)^{n_i}, \tag{2.21}$$

where $1+z=a$, and $n_i = -3(1+\omega_i)$.

With this new notation, our first Friedmann equation can be written down in a form that we will be making use of most,

$$H(z) = H_0\Big\{\sum_i \Omega_{0_i}(1+z)^{n_i}\Big\}^{\frac{1}{2}}. \tag{2.22}$$

We can now also introduce the luminosity distance for an FRW universe,

$$d_L(z) = \frac{1+z}{H_0}\int_0^z \Big\{\sum_i \Omega_{0_i}(1+z')^{n_i}\Big\}^{-\frac{1}{2}} dz'. \tag{2.23}$$

This quantity is not directly used in calculations however, as logarithmic quantities are preferred, and so we define the **Distance Modulus**,

$$\mu = 5\ log(d_L) - 5. \tag{2.24}$$

2.4 The Problem

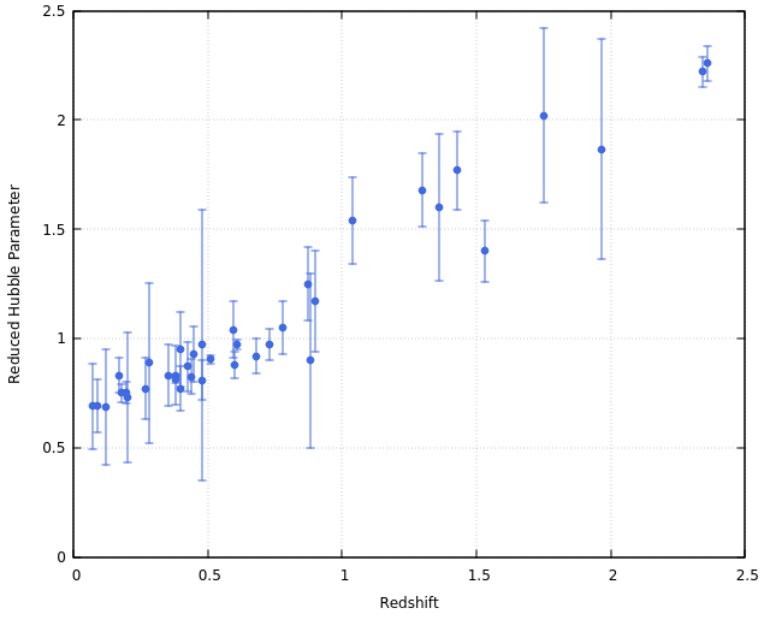

Figure 2.2: Observed data for the reduced Hubble Parameter.

Figure 2.2 shows us the observed values for the Hubble Parameter. As we can see, the size of the universe is not a constant. It is in fact, expanding at an accelerated rate. To account for this accelerated expansion, we add in another energy density to our model universe along with matter and radiation, and that new ingredient is what is called **Dark Energy**. Several models of this newly introduced energy density exists, each characterized by their own forms of the equation of state parameter for dark energy, $\omega(z)$. To name a few, we have Λ-Cold Dark Matter (CDM), ωCDM, Quintessence, etc. There are a number of polynomial, exponential, and logarithmic functions that serve the purpose as well. However, the existence of such a large number of possible candidates *is* the problem, as they have not yet been excluded by existing observations, and so the origin of dark energy is a mystery.

Chapter 3

Markov Chain Monte Carlo Methods

In this work, the main computational problem encountered is one of finding values of free parameters to get our models to fit the data best. To that end, we can follow two approaches, the first being a brute force method called a grid search, where we methodically run through the entire parameter space, calculating likelihoods for each point in the space to find the one most suited to data; the second approach, which is much more powerful, is to use Markov Chain Monte Carlo methods to do the same [CM01] [LB02] [Ver07].

3.1 The Likelihood

Let us assume that we have some observable $\mathcal{F}$, which can be theoretically modelled as $\mathcal{F}(\mathbf{\Theta},\ \mathbf{X})$, where $\mathbf{\Theta}$ is the set of free parameters in the model, and $\mathbf{X}$ is the set of variables that $\mathcal{F}$ depends on. We need to be able to gauge at any of the given independent variables of $\mathcal{F}$, how far away the model with a given set of values for $\mathbf{\Theta}$ is, and to do that, we define an error function,

$$\chi^2(\mathbf{\Theta}) = \sum \Big(\frac{\mathcal{F}_{calc}(\mathbf{\Theta},\ \mathbf{X_i}) - \mathcal{F}_{obs}(\mathbf{X_i})}{\sigma_i} \Big)^2, \tag{3.1}$$

where σ is the actual error in measurement, and the index i runs over all possible values for the independent variables for which we have measurements.

This chi-squared function, can be thought of as a summation over the indices of ratios of theoretical and experimental error. If the point in parameter space is close to its true value, then this ratio would ideally be close to one. That is why it is easy to judge whether a parameter is a good fit, as the value for chi-squared at that point would be close to the number of data points considered.

Now, depending on what we wish to do with this error function, we can either use it to define simple functions like,

$$\mathcal{L} = \frac{1}{\chi^2},$$

such that the higher the error, the lesser the likelihood of that particular point in the parameter space being the one closest to the true value.

If our goal was simply to find the most likely point in the parameter space, the above definition of likelihood, along with any other that preserved the shape of that distribution would work. However, our goals are not so simple, as we would also like to be able to gauge how confident we are in our estimation of the most likely parameter set, which means that we would like to be able to marginalize and normalize our resultant probability distributions with ease.

To that end, we reasonably assume that our errors are distributed normally, and define an exponential likelihood function,

$$\mathcal{L}(\boldsymbol{\Theta}) = \exp\Big(-\frac{\chi^2(\boldsymbol{\Theta})}{2}\Big), \tag{3.2}$$

to make our integrals simpler.

3.2 Grid Search

As the name suggests, in this method we divide up our m-dimensional parameter space into a grid, be it coarse or fine, and traverse the entire space, calculating the exponential likelihood value and assigning it to each point on the grid. What this ends up giving us is a distribution function, which we can turn into a probability distribution by integrating over the entire space and normalizing our distribution.

While this method is easy to implement, for larger dimension parameter spaces it is prohibitively slow. In practice it is only used as a preliminary tool to get a rough idea of what the likelihood surface looks like, traversed by a very coarse grid. The integration over the entire space is a very computationally expensive process. Further, if we have a multi-dimensional distribution, and wish to only look at a few important parameters after marginalization, that is another computationally expensive series of integrals that need to be performed.

All in all, it is best to not try to use this for actually getting results, as we have something much more powerful at our disposal - Markov Chain Monte Carlo methods.

3.3 Markov Chain Monte Carlo

Monte Carlo methods are a wide class of computational methods that rely on randomly sampling extremely large number of points, be it to integrate a function and calculate areas or in optimization problems.

Markov Chains refer to a sequence of events where the future path is not affected by the path taken previously, but only on the current position. More technically, the probability of taking a particular step in a chain is only affected by the current state of the system. Markovian systems are therefore also known as memory-less systems.

Put together, they give us ways to traverse only the interesting parts of a probability distribution, foregoing the need to sample the entire space to find the hills or valleys on our likelihood surface. One of those methods is called the **Metropolis-Hastings Algorithm**.

3.3.1 Metropolis-Hastings Algorithm

Starting at any random point in the parameter space, the algorithm tells us the fastest way to reach any peaks that might exist on our likelihood surface. Much like a mountain climber, the Markov chain in this algorithm generally prefers to climb towards the peak, but oftentimes will move to a point that takes it away from the peak, to keep checking for easier paths to the top in the vicinity. More concretely, the algorithm is as follows,

1. Define a likelihood function, $\mathcal{L}$.
2. Choose a random point in the sample space, θ_1, and calculate the likelihood, $\mathcal{L}_1$.
3. Propose to move to a new point, θ_2, in the parameter space chosen at random, and calculate its likelihood, $\mathcal{L}_2$.
4. (a) If the point is more likely, that is, if $\mathcal{L}_2 > \mathcal{L}_1$, accept that point as a point in the chain.

   (b) If on the other hand, $\mathcal{L}_2 < \mathcal{L}_1$, draw a random number r, and accept that point if $r < \frac{\mathcal{L}_2}{\mathcal{L}_1}$.

(c) If the point is rejected, stay at the same position in the chain.

5. Repeat steps 2 - 4 till convergence.

It is important to note the role that step $4.\ b)$ plays in this algorithm. If we were to only take the steps that bring us closer to the peak, we would reach the peak and get stuck. We would have the maximum likelihood point we were seeking, but would have no idea about the distribution around it to be able to infer any confidence estimate about our newly found most likely point. Since we occasionally step away from the peak, we get the chance to sample the space around it.

For our purpose, we will use the definition of L given in equation (3.2). Since we will only be dealing with ratios of likelihoods, we don't need to worry about any prefactors that would serve the purpose of normalizing an otherwise unnormalized probability distribution. A possible challenge that can come up is if we are dealing with very large data sets, then the value of χ^2 can be extremely large, in which case $\exp(-\chi^2)$ would be a ridiculously small value, and the ratio of the two likelihoods would be indeterminate. In that case, we can make use of a reduced χ^2,

$$\chi^2_{red} = \frac{\chi^2}{N}, \tag{3.3}$$

where N is the number of data points we have. Since this transformation is not a linear operation considering the likelihood function - as it's being performed in the exponent - it would alter the shape of the likelihood distribution, and so needs to be reverted back to obtain the actual ratio,

$$\frac{\mathcal{L}_2}{\mathcal{L}_1} = \exp\ \big(-(\chi^2_2 - \chi^2_1)\big). \tag{3.4}$$

So, after calculating the ratio with the χ^2_{red}, we can raise it to a power of N, to obtain the correct ratio.

3.3.2 Convergence

Finally, we discuss convergence; once we know our chains have reached convergence, the statistical information we obtain from it doesn't change the longer the chains keep going. In other words, the chains have spent enough time to sample the distribution, and running them longer will not change the resultant distribution any further. To check for convergence, various algorithms exist, but we'll be focusing on one used most commonly, and that is the

Gelman-Rubin Convergence Criteria [GR+92]. This is much easier to implement in multi-threaded approaches, as we can have one thread run its own individual chain through the parameter space.

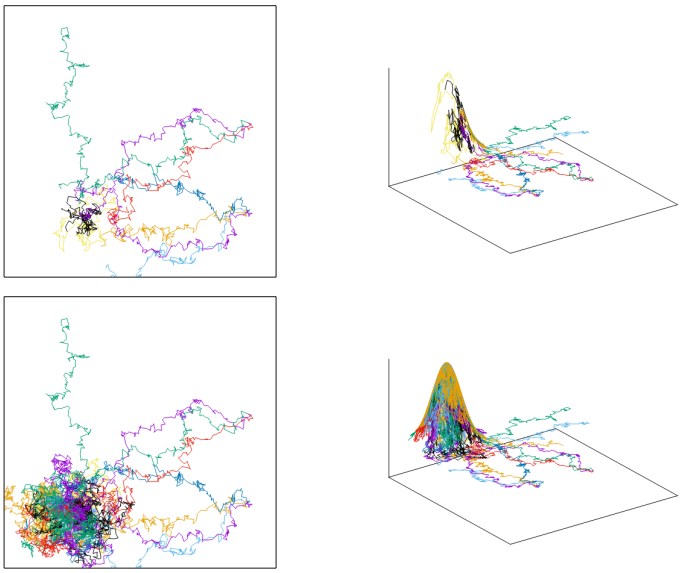

Figure 3.1: A 2 dimensional and a 3 dimensional rendition of several chains run on a likelihood surface that is a Gaussian near the bottom left hand corner, which is where the chains eventually converge. The top row shows chains that have not yet converged to the distribution, and the bottom row is when they're allowed to run for slightly longer.

For the algorithm, we assume that there are M individual chains in total, each containing $2N$ points. Of these, we will only take the last N for our calculation. We define the mean of the chain, $\bar{y}^j$, and the mean of the entire distribution, $\bar{y}$, as,

$$\bar{y}^j = \frac{1}{N}\sum_{i=1}^{N} y_i^j, \quad \bar{y} = \frac{1}{NM}\sum_{ij=1}^{NM} y_i^j, \tag{3.5}$$

using which we can define two variances,

$$B = \frac{1}{M-1}\sum_{j=1}^{M}(\bar{y}^j - \bar{y})^2, \tag{3.6}$$

and,

$$W = \frac{1}{M(N-1)}\sum_{ij}(y_i^j - \bar{y}^j)^2, \tag{3.7}$$

where B is the variance between chains, and W is the variance within a chain. Finally, we define,

$$R = \frac{\frac{N-1}{N}W + \frac{1+M}{M}B}{W}. \tag{3.8}$$

When the value of $R < 1.03$, we say that the chains have sufficiently converged [Ver07].

Chapter 4

Gaussian Processes

While Markov Chain Monte Carlo methods can give us constraints on the free parameters in different parameterizations of dark energy, they cannot really give us information about which one is correct. To take the model out of the picture therefore, we employ non-parametric methods. Where previously we were concerned with constraining free parameters, we will now look at a method to reconstruct our observables independent of any underlying assumptions of dark energy models [QCR05] [Wan20].

4.1 Gaussian Distributions

A one dimensional, normally distributed random variable follows what we call the Gaussian Distribution,

$$P(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\Big(- \frac{(x-\mu)^2}{2\sigma^2}\Big). \tag{4.1}$$

Similarly, if we wanted to model a multi-variate Gaussian distribution of dimension D, we say,

$$\mathcal{N}(\mathbf{x} \mid \boldsymbol{\mu},\, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{D/2}|\boldsymbol{\Sigma}|^{1/2}} exp\Big(- \frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}(\mathbf{x}-\boldsymbol{\mu})\Big), \tag{4.2}$$

where $\mathbf{x}$ and $\boldsymbol{\mu}$ are D dimensional vectors, and $\boldsymbol{\Sigma}$ is a $D \times D$ matrix. $\boldsymbol{\Sigma}$ is called the correlation matrix, and encodes information about how some of the variables affect the others.

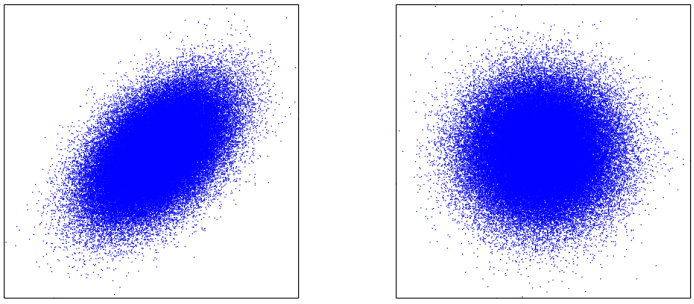

Figure 4.1: Samples drawn from two different bi-variate distributions, one with a non-zero correlation and one where both dimensions are independent of each other.

4.2 Generating Random Functions

At the most fundamental level, Gaussian Processes is a method to generate random functions. Like how a Gaussian distribution is a distribution of a random variable, a Gaussian process is a distribution of a random function; it is analogously defined by a mean function with the associated error.

More formally, a Gaussian process is nothing but a multi-variate Gaussian distribution with a label associated to each of its dimensions such that it can be ”unfolded” on the x axis.

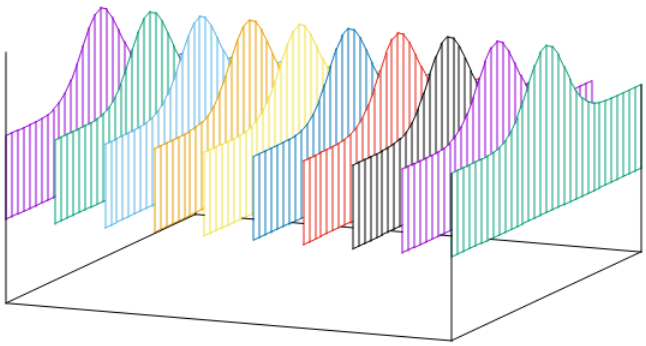

Figure 4.2: Imagine cutting the x axis into slices, and there being a Gaussian distribution along the y axis for every slice in x. An uncorrelated multi-variate Gaussian distribution ”unfolded”.

For each slice of the x axis as given in figure 4.2, we can draw a sample, and connect each of those points. Such a sample is said to be drawn from a Gaussian process.If we make the

slices fine enough, we can make this sample smooth enough to model analytically smooth functions.

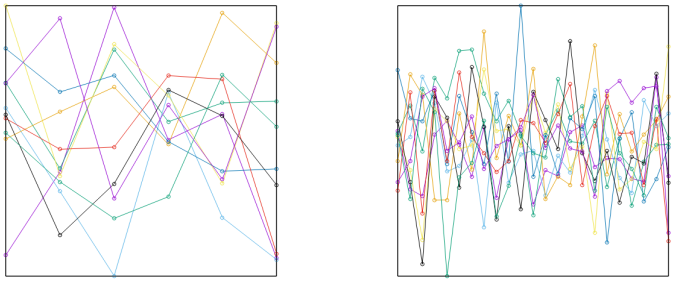

Figure 4.3: Samples drawn from an uncorrelated Gaussian process. On the left, we have one where the x axis was coarsely sliced, and on the right, it was done more finely.

As can be seen from figure 4.3, when samples are drawn from an uncorrelated Gaussian process, the resultant functions can be very jagged. This makes sense, as no point on x is affected by any of its nearby points, and therefore each point is sampled independently.

To obtain smoother samples, we need to introduce correlations between neighbouring points on the x axis, and to do that, we use what we call Kernel functions. Several such functions exist, the most commonly used being the Radial Basis Function,

$$k(x,\ x') = \exp(-(x - x')^2), \tag{4.3}$$

which is calculated for each point on the x axis, and corresponding values are inserted into the correlation matrix.

When this particular Kernel function is applied and samples are drawn from this new Gaussian process, they turn out to be much smoother, as the probability for picking up a points that are close to each other both in x and in y is increased, as we can see in figure 4.4.

So far, we have only been generating random functions, with no input about the data they're meant to model. Functions obtained in such a way are called priors. What we need however, is samples from a distribution that is conditional on the data present.

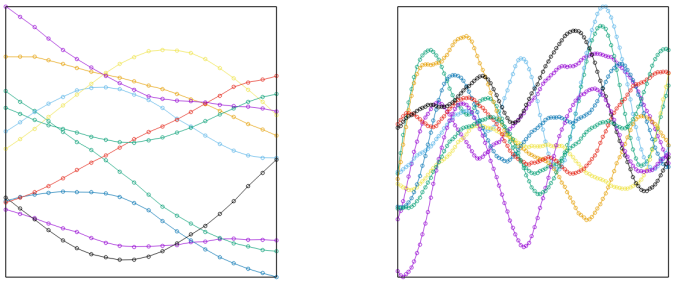

Figure 4.4: Samples drawn from a smoothed Gaussian process where correlations were given by the Radial Basis Function. On the left, we have one where the x axis was coarsely sliced, and on the right, it was done more finely.

4.3 Sampling From The Posterior Distribution

When we perform Bayesian analysis on a previously unconditioned distribution, the resultant distribution is called the posterior. Let us consider a multi-variate distribution,

$$P(\mathbf{f} \mid \mathbf{X}) = \mathcal{N}(\mathbf{f} \mid \boldsymbol{\mu},\ \mathbf{K}), \tag{4.4}$$

where $\mathbf{X} = [x_1,\ x_2, ...,\ x_n]$, $\mathbf{f} = [f(x_1),\ f(x_2), ...,\ f(x_3)]$, $K_{ij} = k(x_i,\ x_j)$, and $\boldsymbol{\mu} = [m(x_1),\ m(x_2),\ ...,\ m(x_n)]$. Here, $\mathbf{X}$ represents the independent variable, $\mathbf{f}$ is the function we are modelling, $\boldsymbol{\mu}$ is the mean function of the multi-variate distribution, and $\mathbf{K}$ is the correlation matrix, the elements of which are given by the kernel, $k(x_i,\ x_j)$.

We are now ready to introduce the data we have from observations, denoted by non-starred symbols. The starred symbols are the points at which we are going to be sampling our posterior function at. We can write down the joint probability distribution,

$$\begin{bmatrix} \mathbf{f} \\ \mathbf{f}_* \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mathbf{m}(\mathbf{X}) \\ \mathbf{m}(\mathbf{X}_*) \end{bmatrix},\ \begin{bmatrix} \mathbf{K} & \mathbf{K}_* \\ \mathbf{K}_*^T & \mathbf{K}_{**} \end{bmatrix} \right), \tag{4.5}$$

where $\mathbf{K}_* = K(\mathbf{X},\ \mathbf{X}_*)$, with the rest of the correlation matrix defined similarly. In most cases, the mean function for the prior is set to zero, especially when the function we are trying to model is oscillatory around it. From this joint probability distribution, $P(\mathbf{f}, \mathbf{f}_* \mid \mathbf{X},\ \mathbf{X}_*)$, we can calculate the conditional distribution $P(\mathbf{f}_* \mid \mathbf{f},\ \mathbf{X},\ \mathbf{X}_*)$ as,

$$\mathbf{f}_* \mid \mathbf{f},\ \mathbf{X},\ \mathbf{X}_* \sim \mathcal{N}(\mathbf{K}_*^T\mathbf{K}^{-1}\mathbf{f},\ \mathbf{K}_{**} - \mathbf{K}_*^T\mathbf{K}^{-1}\mathbf{K}_*). \tag{4.6}$$

The distribution on the RHS of this equation is the posterior distribution we need. However, this is only valid for the very rare case of when our observations have no error or noise

associated with them. Assuming the error is given by σ_i^2, we simply add it to the $\mathbf{K}$ quadrant of the correlation matrix,

$$\mathbf{K} \to \mathbf{K} + \sigma_i^2 \mathbf{I}.$$

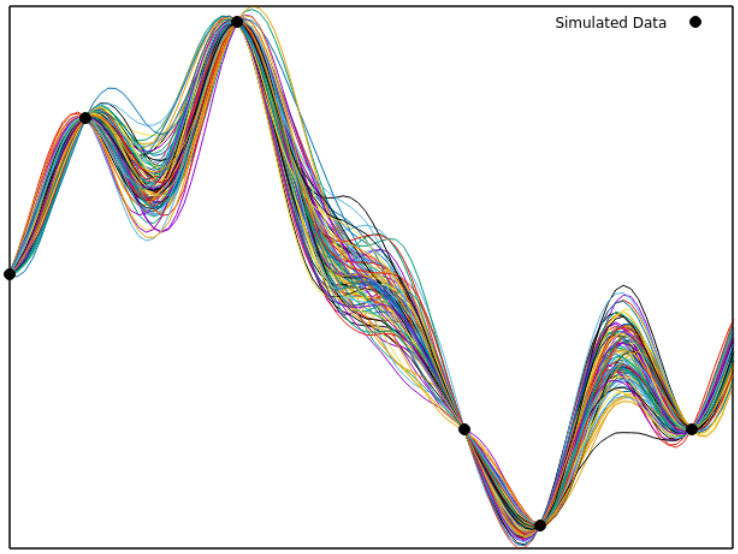

Figure 4.5: 100 samples drawn from the posterior distribution after performing Bayesian analysis on simulated data, which is simply randomly chosen points of a sine function.

So, Gaussian Processes can be used as a valuable tool to reconstruct underlying functions if we have enough data to model off of. As we can see in figure 4.6, at the points where there was no available data, the mean is far off from the true function, as it tends to stay close to the prior mean that was fed to it. Another point to note is that this was only for the case when the data we had was noiseless; in the case where our data has inherent errors associated with it, we get a much better fit, as seen in figure 4.7.

4.4 Hyperparameter Selection

The kernel function we had encountered in equation (4.3) was only a very simplified version of the real Radial Basis function,

$$k(x_i,\ x_j) = \sigma^2 \exp\Big(- \frac{|x_i - x_j|^2}{2l^2} \Big). \tag{4.7}$$

The additional parameters, σ and l, are known as **hyperparameters**, called so to differentiate them from parameters that may be present in whatever true function we are trying

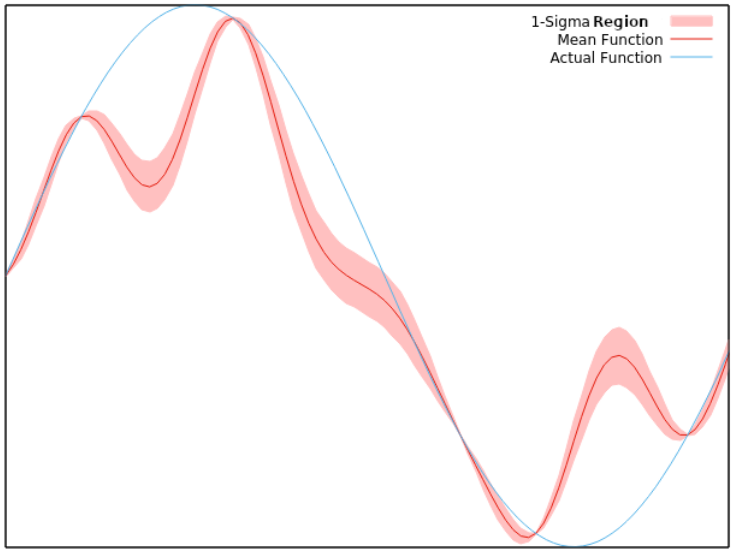

Figure 4.6: The mean function we obtain after averaging over all the randomly sampled functions from figure 4.5 along with the associated error, and the actual function from which our data points were derived.

to estimate. Finding the values for these hyperparameters is another regression problem in itself, for which we can either use a grid search or the Metropolis-Hastings algorithm.
For the same, we need to define an appropriate error function, and for that we use,

$$\xi(\mathbf{\Phi}) = \sum_i \Big\{ \sum_j |f_{data}(x_i) - f_{calc}(x_i,\ \mathbf{\Phi})| * \exp((x_i - x_j)^2) \Big\}, \tag{4.8}$$

where $\mathbf{\Phi}$ is the set of hyperparameters, and the indices i and j run over the observed data points and the entire length of the x axis respectively.

Once we have the correct hyperparameters, we are fully equipped to start sampling from the posterior and generate a mean function from it along with the relevant confidence intervals.

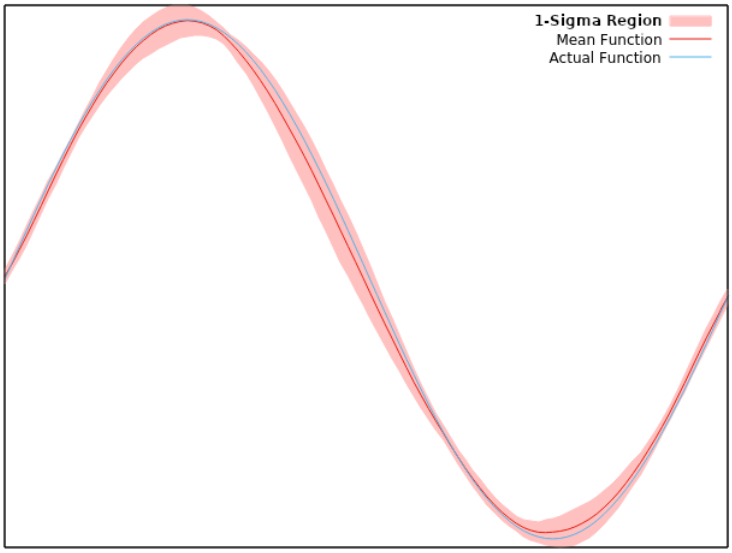

Figure 4.7: A much better estimate of the true function is obtained in a more realistic situation, with data that is slightly noisy.

Chapter 5

Results and Conclusions

Let us now turn to a restatement of the problem, and the results that follow.

5.1 Comparison of MCMC and Gaussian Process Methods

As we now understand the methods we will be employing, we are ready to restate the problem in terms that make it evident how we'll be tackling it. What we are attempting to tackle in this thesis is two-fold; an approach with defined parametric forms of the dark energy equation of state parameter where we will attempt to constraint the free parameters by using Markov Chain Monte Carlo methods to find the most likely set of values for them in order to reconstruct cosmological observables, and a non-parametric approach given by Gaussian processes, where we will use randomly drawn functions from a posterior distribution for the same cosmological observables, relying purely on observed data and not on any specific model for dark energy.

5.1.1 Parametric Reconstruction

For the parametric reconstruction, we will be following what was done in [TSJ17], with the two differences being using the Metropolis-Hastings algorithm, and that we will only be considering the simplest of dark energy models, the ωCDM model, where the ω is simply an undetermined, to-be-constrained constant, not dependent on the red-shift. For the same,

our observables can be defined as the Hubble parameter,

$$H(z) = H_0\left\{\Omega_m(1+z)^3 + \Omega_{DE}(1+z)^{3(1+\omega)}\right\}^{\frac{1}{2}}, \tag{5.1}$$

and the Luminosity Distance,

$$d_L(z) = \frac{c}{H_0}(1+z)\int_0^z \left\{\Omega_m(1+z')^3 + \Omega_{DE}(1+z')^{3(1+\omega)}\right\}^{\frac{1}{2}} dz', \tag{5.2}$$

which can be written in its more commonly used form, the Distance Modulus,

$$\mu = 5\log(d_L) - 5. \tag{5.3}$$

The density parameters follow the Friedmann equation, and give us another constraint on our equations,

$$\Omega_m + \Omega_{DE} = 1. \tag{5.4}$$

This means that we now have three free parameters to constrain, H_0, Ω_m, ω. The priors we input into the Metropolis-Hastings algorithm are taken to be uniform,

Minimum	Parameter	Maximum
0.01	Ω_m	0.60
-4.00	ω	0.00
65.00	H_0	75.00

Table 5.1: Priors for the ωCDM model.

The data we use for the distance modulus reconstruction is the Union data [ALR$^+$10], compiled by the Supernova Cosmology Project, and has 580 data points collected together from various observations over the years. For the Hubble parameter as well, we have a collection of 38 data points [FMCR17].

Given the priors and the data, we can obtain the posterior distribution for the three parameters, marginalize it as needed, and calculate confidence intervals.

5.1.2 Non-parametric Reconstruction

Observations do not currently rule out any model of dark energy, and hence it is reasonable to search for methods which do not a priori assume any Dark Energy properties. For this, we can use Gaussian processes to obtain a mean function for the Hubble parameter along with

the relevant confidence intervals. We can then compare the reconstructed Hubble parameter forms from the previous approach, using different models, and see if any of those models fall outside the confidence intervals, they would be falsified, but if they don't the method is flexible enough that we can wait for more observations in the future that would help narrow down our error margins. For this, we use the same data we did for the parametric reconstruction.

Since there are no models we depend on, there are no free parameters to constrain; instead, we have what we call hyperparameters - the parameters that help define the Gaussian Process from which we will be drawing samples. For the kernel we are using, as defined in equation (4.7), there are two hyperparameters to constrain, σ and l. σ represents how correlated the values are on the y axis, and l defines it for the x axis. For this, we can either use MCMC methods, or we can just use a grid search. Since in this case the dimension of our hyperparameter space is not that large, we can safely use a grid search algorithm. A note however, since there will matrix inversions involved, it would be advisable to employ the Gauss-Jordan elimination to calculate the inverse, as the dimension of the matrices involved will be the same as the number of data points, which in our case can go up to 580, and the recursive algorithm would take infinitely long to be completed for a matrix that large.

The priors we use for the hyperparameters are,

Minimum	Hyperparameter	Maximum
0.01	σ	1
0.01	l	1

Table 5.2: Hyperparameter priors.

Traversing the hyperparameter space, we can calculate our error as given in equation (4.8), and use the set of hyperparameters with the minimum error.

5.2 Results

Now that we understand the problem and our approach to it, we can attempt to finally reconstruct our observables, first taking the ωCDM model of dark energy, and then relying purely on observed data to obtain a non-parametric reconstruction.

5.2.1 Parametric Reconstruction

For the parametric reconstruction using the ωCDM model, we use the Metropolis-Hastings algorithm to constraint the values of w and Ω_M for both observables. The resultant confidence intervals can be seen in figures 5.1 and 5.2.

For the parametric reconstruction using the ωCDM model, we use the Metropolis-Hastings algorithm to constraint the values of w and Ω_M for both observables. The resultant confidence intervals can be seen in figures 5.1 and 5.2.

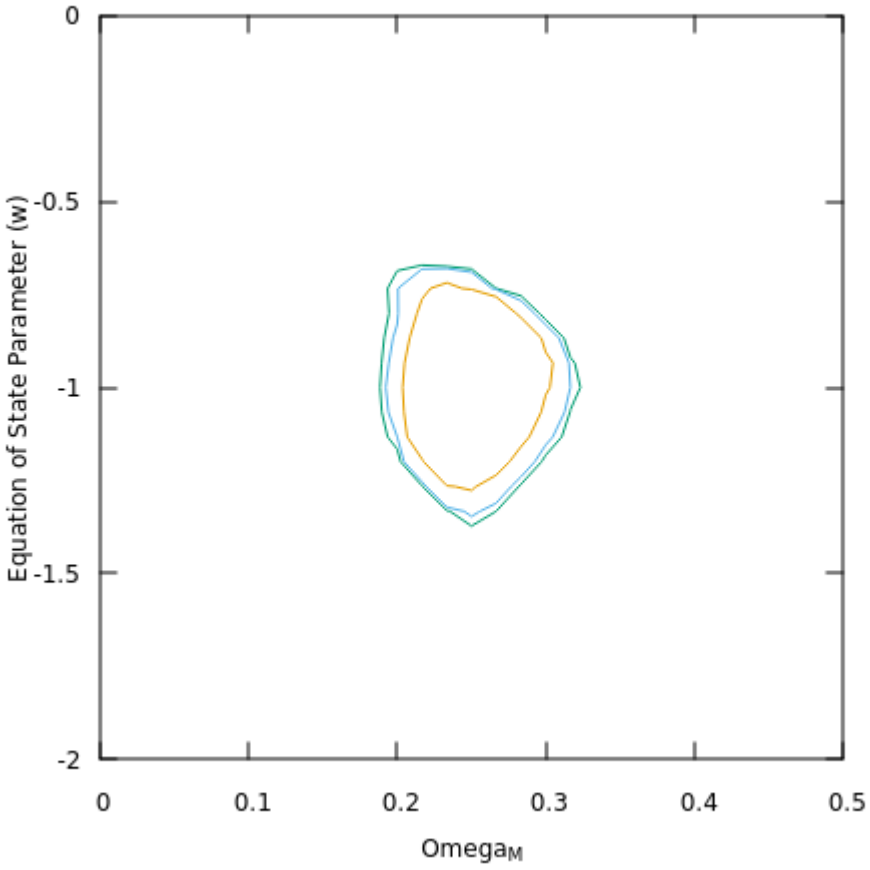

Figure 5.1: 1, 2, and 3-σ contours for the Hubble parameter data.

We can marginalize over each of the parameters to get a distribution for them separately as can be seen in figures 5.3 and 5.4. This would allow us to get a quantitative idea of how good our fit is with respect to the individual parameters.

The final values of the parameters and the corresponding confidence intervals can be sum-

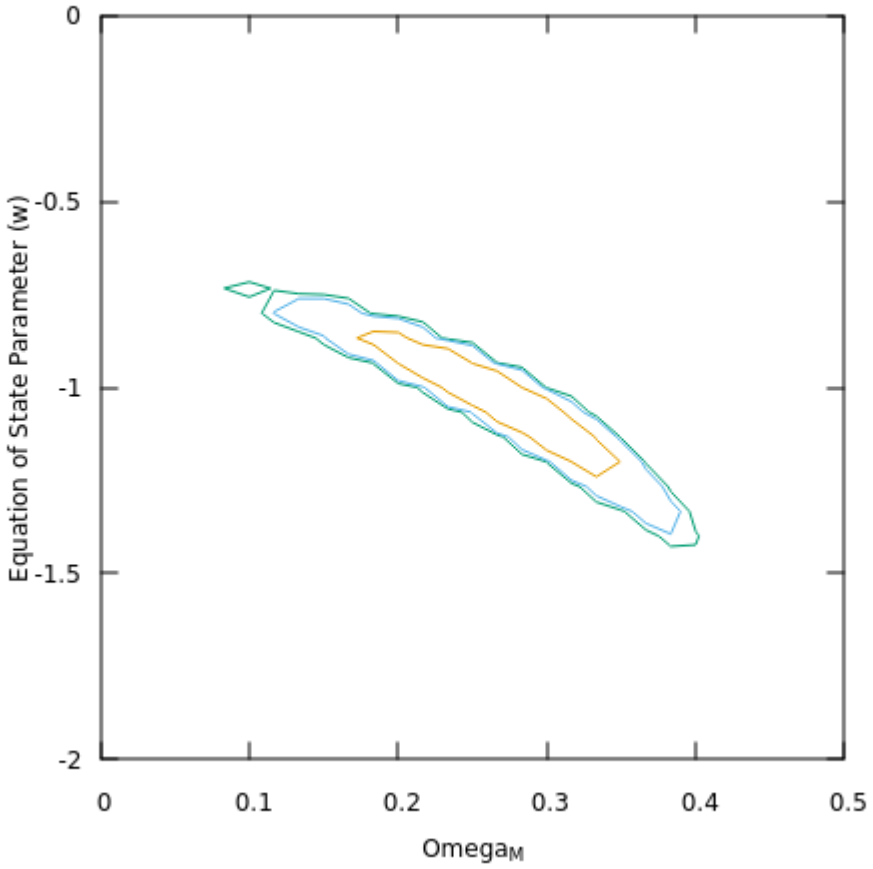

Figure 5.2: 1, 2, and 3-σ contours for the SN1a data.

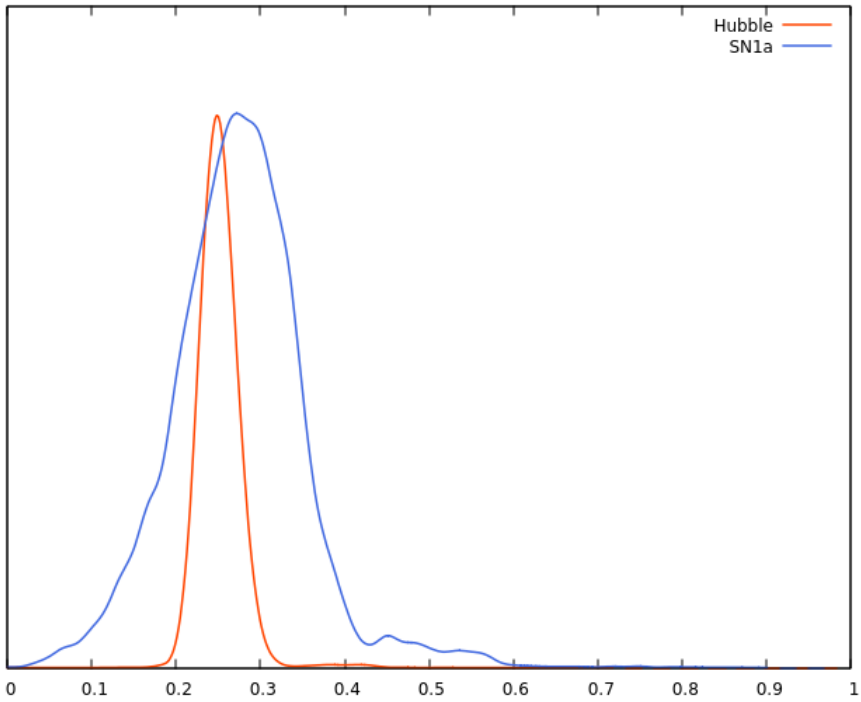

Figure 5.3: The final distribution we get for the parameter Ω_M after marginalizing over ω (and H_0 for the Hubble parameter data) and normalization.

marized as below,

	Data Set	Mean Value
Ω_M	H(z)	0.250 ± 0.017
	SN1a	0.267 ± 0.016
ω	H(z)	-1.000 ± 0.067
	SN1a	-1.000 ± 0.133

Table 5.3: Mean values of the parameters after marginalization along with error margins of 1 standard deviation, assuming a Gaussian fit.

5.2.2 Non-Parametric Reconstruction

Fitting the Gaussian Process to the data by finding the correct set of hyperparameters, we draw randomly sampled functions from it till the mean we recover does not change appreciably - in other words, till the sampled distribution converges to the underlying posterior we obtained from fitting.

As we can see from figures 5.5 and 5.6, the reconstruction fits the data well. Gaussian

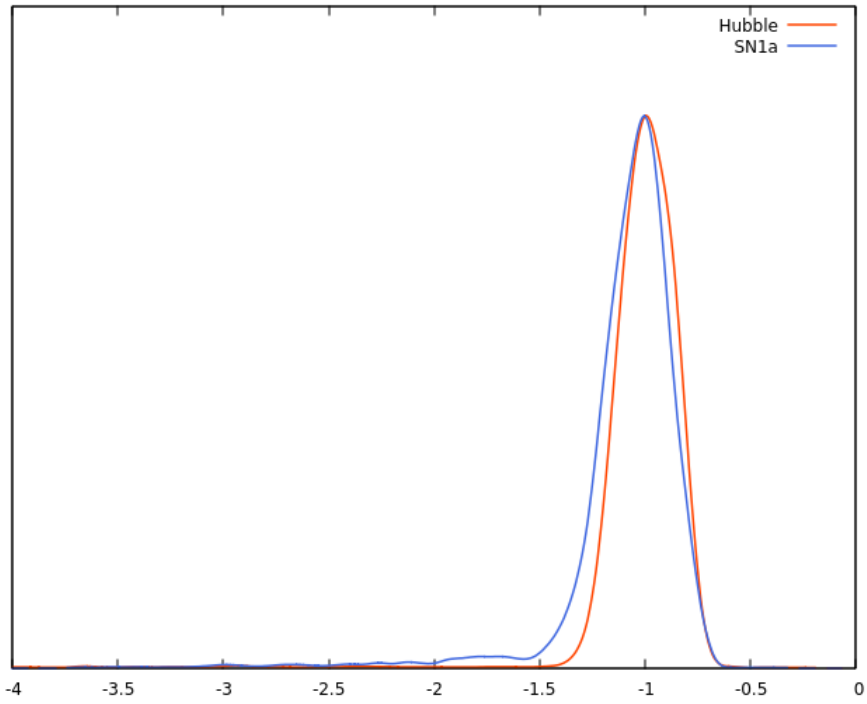

Figure 5.4: The final distribution for ω, the equation of state parameter in the ωCDM model for dark energy after marginalization and normalization.

Processes can therefore be used as a reliable tool to reconstruct cosmological observable independent of any dark energy model.

It is very clear that wherever we have more data points, the confidence interval is much tighter, and so as observational techniques improve in the future, we will be able to get ever smaller bounds on these reconstructed observables.

We can also compare these results with the reconstructed observables from the parametric approach we used earlier, with the goal in mind that if we can get the confidence intervals small enough with our Gaussian process approach, they can rule out various models of dark energy.

5.3 Conclusion

We have seen that with the parametric approach we can obtain a probability distribution for the parameters to be tuned, and found out that the distributions - their mean and variance - are in agreement with what is currently accepted in literature. The ωCDM model, tends to pick out a value of -1, which is equal to the equation of state parameter in the ΛCDM model.

Although the confidence intervals may not be an improvement over the current values, the computation time taken by MCMC methods is much, much smaller (by several orders of magnitude) than if we had used a grid search, and can also be scaled with dimension of our parameter space. It is therefore a viable substitute, as the models we wish to check and the data available with which to check them improve.

In the non-parametric approach as well, Gaussian processes have given us a reasonable reconstruction of our observables, along with a posterior distribution to gauge our confidence in said reconstruction.

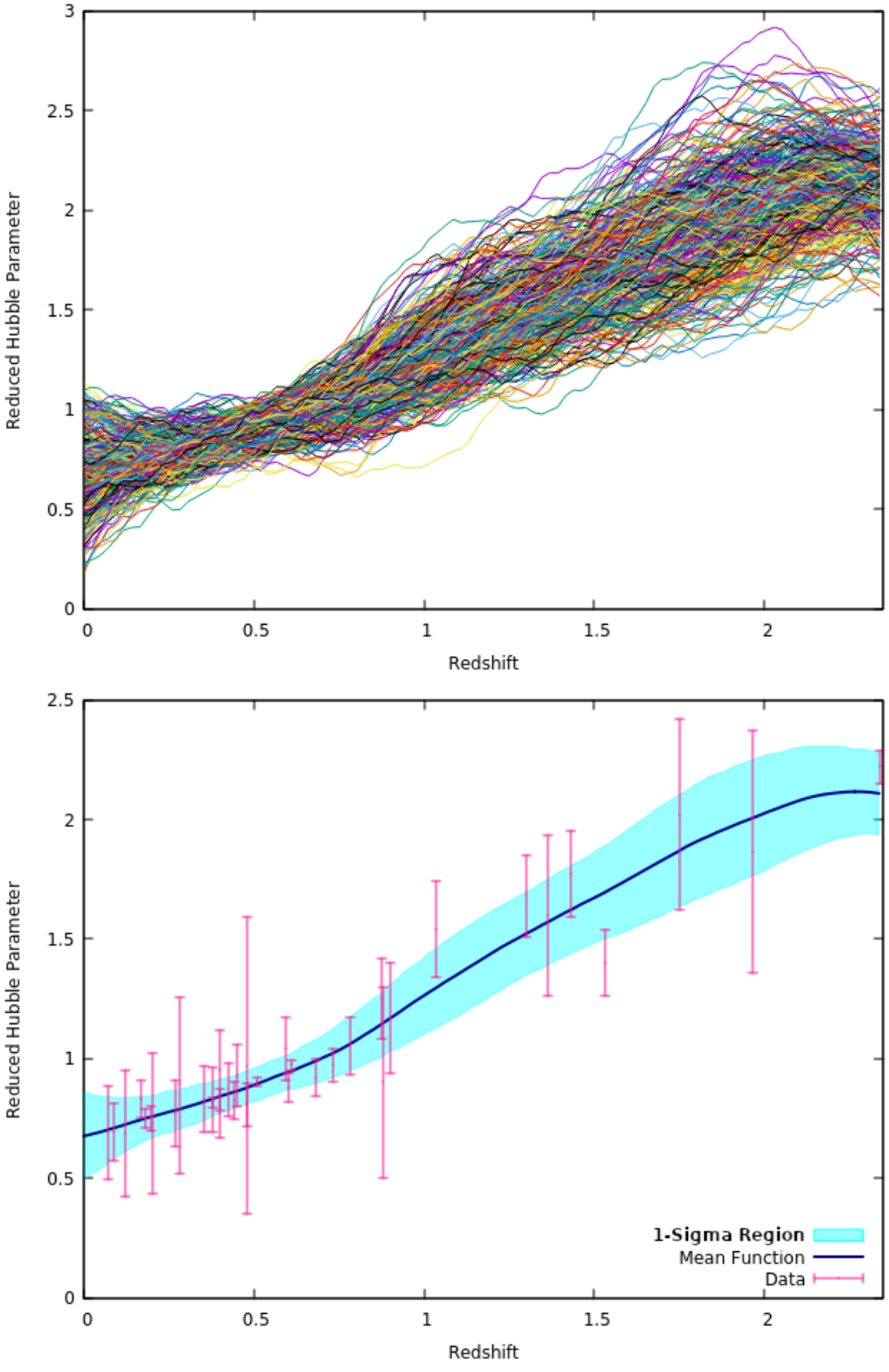

Figure 5.5: At the top, 400 samples that are randomly drawn from the posterior distribution fitted to the reduced Hubble parameter data, reduced by scaling down by a factor of 100. At the bottom, the mean of all of the functions, along with a standard deviation calculation, which gives us a 1-σ region.

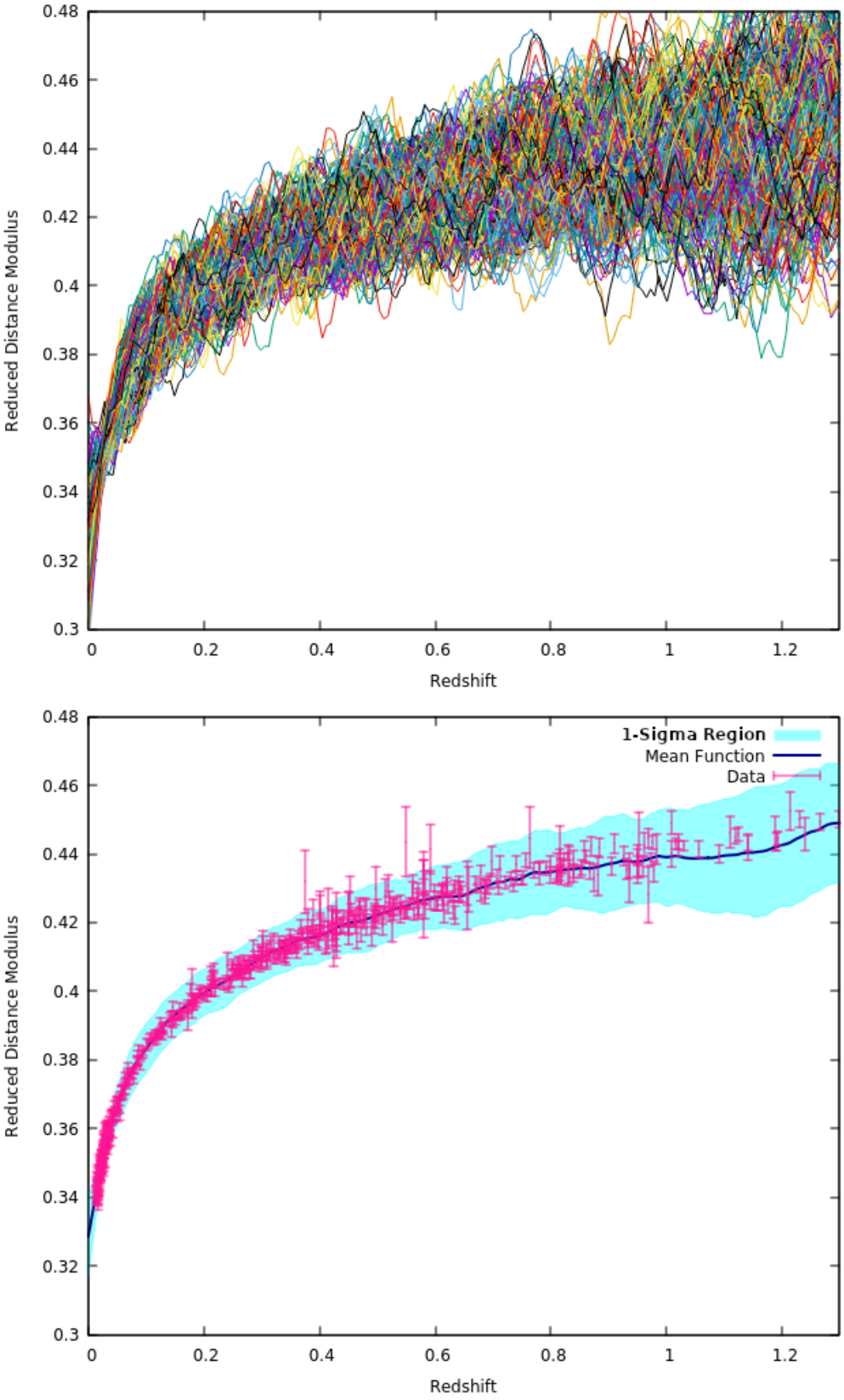

Figure 5.6: At the top, 400 samples that are randomly drawn from the posterior distribution fitted to the SN1a distance modulus data, reduced by scaling down by a factor of 100. At the bottom, the mean of all of the functions, along with a standard deviation calculation, which gives us a 1-σ region.

Bibliography

[ALR+10] Rahman Amanullah, Chris Lidman, D Rubin, G Aldering, P Astier, K Barbary, MS Burns, A Conley, KS Dawson, SE Deustua, et al., *Spectra and hubble space telescope light curves of six type ia supernovae at 0.511¡ z¡ 1.12 and the union2 compilation*, The Astrophysical Journal **716** (2010), no. 1, 712.

[Car19] Sean M Carroll, *Spacetime and geometry*, Cambridge University Press, 2019.

[CM01] Nelson Christensen and Renate Meyer, *Using markov chain monte carlo methods for estimating parameters with gravitational radiation data*, Physical Review D **64** (2001), no. 2, 022001.

[Col18] Andrew Z. Colvin, *Expansion*, 2018.

[FMCR17] Omer Farooq, Foram Ranjeet Madiyar, Sara Crandall, and Bharat Ratra, *Hubble parameter measurement constraints on the redshift of the deceleration–acceleration transition, dynamical dark energy, and space curvature*, The Astrophysical Journal **835** (2017), no. 1, 26.

[GR+92] Andrew Gelman, Donald B Rubin, et al., *Inference from iterative simulation using multiple sequences*, Statistical science **7** (1992), no. 4, 457–472.

[LB02] Antony Lewis and Sarah Bridle, *Cosmological parameters from cmb and other data: A monte carlo approach*, Physical Review D **66** (2002), no. 10, 103511.

[Pad02] Thanu Padmanabhan, *Theoretical astrophysics: Volume 3, galaxies and cosmology*, Cambridge University Press, 2002.

[QCR05] Joaquin Quinonero-Candela and Carl Edward Rasmussen, *A unifying view of sparse approximate gaussian process regression*, The Journal of Machine Learning Research **6** (2005), 1939–1959.

[TSJ17] Ashutosh Tripathi, Archana Sangwan, and HK Jassal, *Dark energy equation of state parameter and its evolution at low redshift*, Journal of Cosmology and Astroparticle Physics **2017** (2017), no. 06, 012.

[Ver07] Licia Verde, *A practical guide to basic statistical techniques for data analysis in cosmology*, arXiv preprint arXiv:0712.3028 (2007).

[Wan20] Jie Wang, *An intuitive tutorial to gaussian processes regression*, arXiv preprint arXiv:2009.10862 (2020).

[Wei72] Steven Weinberg, *Gravitation and cosmology: principles and applications of the general theory of relativity*, vol. 1, Wiley New York, 1972.

[Wei89] ———, *The cosmological constant problem*, Reviews of modern physics **61** (1989), no. 1, 1.