Advanced statistical methods for astrophysical probes of cosmology

Marisa Cristina March
Astrophysics Group
Blackett Laboratory, Department of Physics
Imperial College London

Thesis submitted for the Degree of Doctor of Philosophy to
Imperial College London

· 2012·

Abstract

The work presented in my thesis develops advanced Bayesian statistical methods for using astrophysical data to probe our understanding of the Universe, I cover three main areas:

Should we doubt the cosmological constant?

While Bayesian model selection is a useful tool to discriminate between competing cosmological models, it only gives a relative rather than an absolute measure of how good a model is. Bayesian doubt introduces an unknown benchmark model against which the known models are compared, thereby obtaining an absolute measure of model performance in a Bayesian framework. I apply this new methodology to the problem of the dark energy equation of state, comparing an absolute upper bound on the Bayesian evidence for a presently unknown dark energy model against a collection of known models including a flat Lambda cold dark matter (ΛCDM) scenario. I find a strong absolute upper bound to the Bayes factor between the unknown model and ΛCDM. The posterior probability for doubt is found to be less than 13 per cent (with a 1 per cent prior doubt) while the probability for ΛCDM rises from an initial 25 per cent to almost 70 per cent in light of the data. I conclude that ΛCDM remains a sufficient phenomenological description of currently available observations and that there is little statistical room for model improvement

Improved constraints on cosmological parameters from supernovae type Ia data

I present a new method based on a Bayesian hierarchical model to extract constraints on cosmological parameters from SNIa data obtained with the SALT-II lightcurve fitter. I demonstrate with simulated data sets that our method delivers considerably tighter statistical constraints on the cosmological parameters and that it outperforms the usual chi-square approach 2/3 of the times. As a further benefit, a full posterior probability distribution for the dispersion of the intrinsic magnitude of SNe is obtained. I apply this method to recent SNIa data and find that it improves statistical constraints on cosmological parameters from SNIa data. From the combination of SNIa, CMB and BAO data I obtain $\Omega_m = 0.28 \pm 0.02, \Omega_\Lambda = 0.73 \pm 0.01$ (assuming $w = -1$) and $\Omega_m = 0.28 \pm 0.01$, $w = -0.90 \pm 0.05$ (assuming flatness; statistical uncertainties only). I constrain the intrinsic dispersion of the B-band magnitude of the SNIa population, obtaining $\sigma_\mu^{\text{int}} = 0.13 \pm 0.01[\text{mag}]$.

Robustness to systematics for future dark energy probes

I extend the Figure of Merit formalism usually adopted to quantify the statistical performance of future dark energy probes to assess the robustness of a future mission to plausible systematic bias. I introduce a new robustness Figure of Merit which can be computed in the Fisher Matrix formalism given arbitrary systematic biases in the observable quantities. I argue that robustness to systematics is an important new quantity that should be taken into account when optimizing future surveys. I illustrate our formalism with toy examples, and apply it to future type Ia supernova (SNIa) and baryonic acoustic oscillation (BAO) surveys. For the simplified systematic biases that I consider, I find that SNIa are a somewhat more robust probe of dark energy parameters than the BAO. I trace this back to a geometrical alignment of systematic bias direction with statistical degeneracy directions in the dark energy parameter space.

Dedicated to:

‘STELLA MARIS’

Both the Vessel Singular
and the vessel nautical
and all those who sail with her.

Contents

Declaration and Copyright

The research presented in this thesis has previously been published in a different form in the three papers listed on page 14. The work presented here has been carried out in collaboration with my colleagues who are listed as co-authors on these papers, and as such we have shared ownership of the research presented. I list below the specifics of my contribution and acknowledge the specific contributions of my collaborators:

(i) Chapter 6: Bayesian Doubt

   (a) Initial concept of Bayesian doubt is credited to the other listed co-authors and is not my own.

   (b) Implementation of the *wbins* model is by Pascal Vaudrevange.

   (c) All other numerical inplementation and analysis is my own.

(ii) Chapter 7: Bayesian parameter inference for SNeIa data:

   (a) All of the numerical implementation and analysis are my own.

   (b) Concept and calculations carried out in equal collaboration with listed co-authors.

(iii) Chapter 8

   (a) Initial concept of Robustness is credited to the other listed co-authors and is not my own.

   (b) The BAO data set and discussion of that data set in section 8.4.2 are by Luca Amendola and are not my own.

   (c) The WL data set and discussion of that data set in section 8.4.1 are by Dragan Huterer and are not my own.

(d) All of the numerical implementation of Robustness to the toy model and the two data sets listed above are my own.

Marisa Cristina March

September 2011

Acknowledgments

First and foremost I would like to thank my supervisor Dr. Roberto Trotta for his help and guidance over the past three years, for identifying interesting and relevant areas of research, for giving me the opportunity to work in various collaborations and for assistance in securing funding. I would also especially like to thank my collaborators who have added a great depth and breadth to my research experience, in chronological order of collaboration they are Prof. Glenn Starkman, Dr. Pascal Vaudrevange, Dr. Pietro Berkes, Prof. Luca Amendola and Prof. Dragan Huterer. In particular I would like to thank Prof. Glenn Starkman and Case Western Reserve University, and Prof. Luca Amendola and the Institute for Theoretical Physics, Heidelberg for hospitality.

I would like to thank all those in the Astrophysics Group and Theoretical Physics Group at my home institution, Imperial College London. I thank especially my academic advisor Prof. Andrew Jaffe for advice on many occasions; our Deparmental Head Prof. Joanna Haigh; our Head of Group Prof. Stephen Warren; our Post Graduate Tutor Dr. Yvonne Unruh and also Dr. Dave Clements.

I am grateful to the many people who have given me feedback and advice on my research either directly or indirectly in conversations and in reviews of first drafts of the various papers presented here. For this I would like to extend my thanks to Bruce Bassett, Alex Conley Ariel, Goobar, Josh Frieman, Andrew Jaffe, Rick Kessler, Marek Kowalski, Tom Loredo, Louis Lyons, Kaisey Mandel, Mathew Smith, Mark Sullivan, David van Dyke and Ben Wandelt.

I would like to acknowledge the use of the Imperial College High Performance Computing (HPC) facility, I thank Matt Harvey and Simon Burbridge for their care and efficiency in maintaining the facility and assisting me with many enquiries. I am also grateful for the assistance of our administrative and support staff whose untiring help with essential paperwork and many other details is much appreciated: Sandie Bernor, Louise Haywood, Dilly Osbahr and Loli Sanchez.

I thank the Science and Technology Facilities Council (STFC) for providing the bulk of my funding during the course of my post graduate research. I am very

grateful to the Royal Astronomical Society (RAS) for providing additional financial support and travel funding. I would like to thank the Imperial College Trust also for contributing to my travel funding.

I of course thank my family, friends and colleagues, at home, abroad, at Imperial and More House for friendship, advice and moral support throughout these years.

Publications

- “Should we doubt the cosmological constant?”
  March M. C., Starkmann G. D., Trotta R., Vaudrevange P.M., 2010, MNRAS, 410,2488-2496(2011)
- “Robustness to systematics for future dark energy probes”
  March M. C., Trotta R., Amendola L., Huterer D. 2011, MNRAS, 415,143-152(2011)
- “Improved constraints on cosmological parameters from SNIa data”
  March M. C., Trotta R., Berkes P., Starkmann G. D., Trotta R., Vaudrevange P.M., arXiv:1102.3237

Foreword

This thesis is based on work carried out at Imperial College London between October 2008, and September 2011.

Chapter 1

Introduction

1.1 Epistemology of the Cosmos

This thesis is about using advanced statistical methods, namely Bayesian, to make best use of astrophysical probes of cosmology. The particular aspect of cosmology on which it focusses on is the apparent late time acceleration of our Universe. As observers, our immediate experience of the Universe is limited to a very small sphere of sensory perception that has been extended by a few space probes which have ventured no further than the edge of our Solar System. Even then, the information gathered by our space-borne instruments is limited to what lies within our past light cone.

Our *a posteriori* knowledge of the wider Universe is based on our local observations of distant regions of space. Many of these observations are of astrophysical objects for example stars, galaxies and supernovae. The question is how to make best use of these locally collected photons from distant astrophysical objects, to tell us about the Universe in which we live.

The fundamental cosmological questions I am interested in answering are: How do we translate our observations of astrophysical phenomena into probabilistic statements about the parameters of the underlying physical theories such as dark energy? How do we quantify our degree of belief about those fundamental models? How do we discriminate between different models such as a cosmological constant vs. modified gravity? My research presented in this thesis focuses on using Bayesian methods to investigate the apparent late time acceleration of the Universe and how this affects our choice of cosmological model and determination of the cosmological parameters.

The evidence for the apparent late time acceleration of the Universe which came in the form of observations of supernovae type Ia (Riess et al., 1998; Perlmutter et al., 1999b) resulted in a paradigm shift away from the Cold Dark Matter

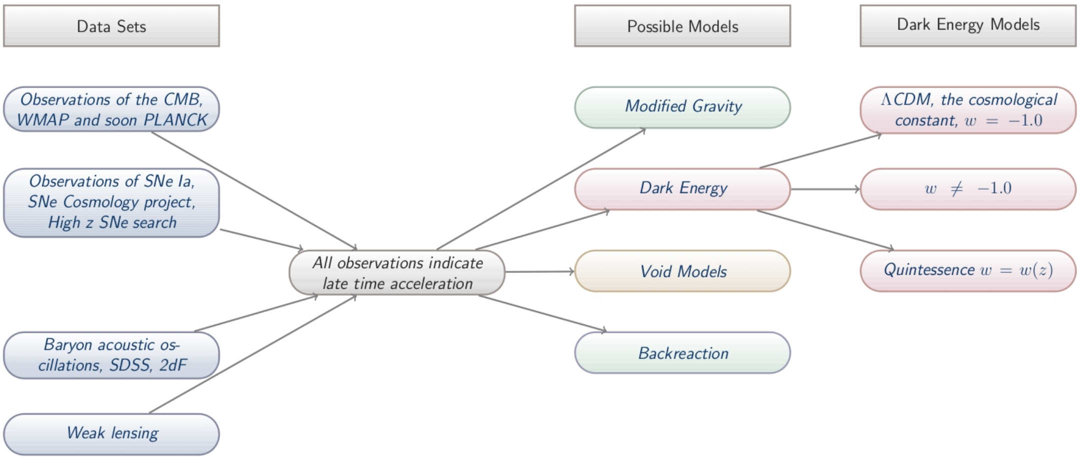


Figure 1.1: Schematic showing that various observations point towards the apparent late time acceleration of the Universe, and some of the possible different types of models which could explain this. The goal is to be able to use astrophysical observations to discriminate between the different cosmological models and infer the values of the parameters within those models. The research presented in this thesis develops statistical tools with which to do this.

(CDM) model to the current reigning model: Lambda Cold Dark Matter (ΛCDM). Will new observations gathered from the next generation of surveys such as LSST, DES, EUCLID and WFIRST confirm the status of ΛCDM or will they point beyond to a new model? Figure 1.1 shows how various observations all point towards the apparent late time acceleration of the Universe, and some of the possible different cosmological models offered in explanation. The current most widely accepted explanation for the apparent late time acceleration is the ΛCDM model which belongs to a sub set of the dark energy class of models. The most popular alternatives are other types of dark energy models, or modified gravity models. Cosmological models in which the Universe is not really accelerating but only *appears* to be accelerating include void models and differing descriptions of the backreaction. My research is about developing advanced statistical techniques to (1) Discriminate between different cosmological models; (2) Infer the parameters within those models; and (3) plan future observing strategies in order to be able to better constrain the various models and their parameters.

The question as to what is causing the apparent late time acceleration of our Universe has been identified by the Decadal Survey NRC (2010) as being one of the ‘most fundamental questions in astrophysics’. Answering this question has far reaching implications not just for astrophysics, but also for our understanding about the fundamental laws of nature: either there is a new and previously unknown energy field with exotic properties, or the laws of gravity need to be re-thought. The findings of our astrophysical observations will change our understanding of physics of the most fundamental level. To quote from the Decadal Survey NRC (2010): 'this survey recommends supplementing our current ability to use the universe as a giant cosmic laboratory to study dark energy, inflation, and black holes. Success in this endeavour would provide critical constraints on the laws of physics and the behaviour of the universe that would inform efforts to realize a unification of gravity and quantum mechanics through string theory or other approaches'. Well planned and well executed observational surveys will shed light on some of the most fundamental questions in physics. Given the large financial outlay required to initiate and run these large scale surveys, it is crucial that the best possible statistical methods are used to analyse the data and ensure that no information is wasted.

These surveys utilize several different astrophysical probes: supernovae Ia, baryon acoustic oscillations, weak lensing and galaxy clustering, and involve huge data sets. We need to give careful consideration to how this data is analysed, and how the different astrophysical probes are best combined. What are the best and most interesting questions we can ask about the Universe with this data? How can we extract the maximum information about the Universe from our observations?

A Bayesian statistical approach offers a holistic world view in which to answer these questions. Bayesian statistical methods are becoming increasingly popular in cosmology [e.g. Trotta (2008); Hobson, M. P., Jaffe, A. H., Liddle, A. R., Mukeherjee, P., & Parkinson, D. (2010) and refs. therein] because this Bayesian perspective gives a universal strategy for deciding how to ask the right questions, how to do model selection, parameter inference and how to quantify the unknown. My research presented in this thesis looks at applying these methods to problems at the interface between astrophysical probes and cosmological theory. The three types of problems I look at are: Firstly: What is the best way to constrain the parameters of a given model using a given data set? (i.e. problems of parameter inference); Secondly: How do we use the observational data to distinguish between competing theoretical models? (i.e. problems of model selection); And thirdly: How do we design and optimise future surveys to obtain the most useful information?

Aspects of these three types of question are addressed in the three work chapters of this thesis, which are based on three papers written with my collaborators during the course of my PhD. The three areas of investigation can be categorized as model building, parameter inference and future surveys. In the following paragraphs I give a brief overview of these three areas of investigation covered in this thesis.

1.2 Areas of research presented in this thesis

Cosmological Model Building: Bayesian Doubt

Several different types of explanation are possible for the apparent late time acceleration of the Universe, including different classes of dark energy model such as ΛCDM, wCDM or quintessence models; theories of modified gravity; void models or the back reaction. Through comparison of the model evidences, Bayes theory offers a way of comparing the relative degree of goodness of the different models on offer.

However, Bayesian model selection only offers a relative not an absolute measure of goodness of fit - it tells us which is the best model in a given set, but it does not tell us whether an unknown and potentially better model may exist. The research presented in chapter 6 is about extending the concept of Bayesian model selection to the concept of Bayesian Doubt (Starkman et al., 2008; March et al., 2011a) which gives an absolute measure of the degree of goodness of a model. I have applied the methodology of Bayesian Doubt to the question of whether we should doubt ΛCDM given the current data, and found the answer to be negative.

Parameter Inference: Cosmology from the supernovae type Ia data

I have been developing a new fully Bayesian method for constraining cosmological parameters from SALT lightcurve fits of SNIa data, based on hierarchical modelling, this work is presented in chapter 7. My new method has several advantages over the standard χ^2 method:

- In trials with simulated data it outperforms the χ^2 method in the recovery of the cosmological parameters 2/3 of the time.
- My new method is fully Bayesian and as such can be used with the full suite of other Bayesian techniques such as model selection.
- It can be used to investigate supernovae evolution, contamination by other supernovae populations and how this affects recovery of the cosmological parameters.
- My new method will allow me to develop a new tool for the analysis of possible systematic error sources in the SNIa data.
- My new fully Bayesian approach can be used with existing, fast sampling techniques which are expected to vastly increase the efficiency of current SNIa parameter determination analysis pipelines.

Preliminary results for this work were presented at the 217th AAS meeting, January 2011, and a preprint of the associated paper is available (March et al., 2011c). This chapter and associated initial paper expounds the basic implementation of the new Bayesian methodology and gives preliminary results using currently available supernova data.

Future Surveys: Robustness to systematic errors

Future proposed surveys and survey techniques are frequently assessed by a figure of merit which often takes the form of the inverse of the area of the Fisher matrix of the proposed new probe, marginalized down to the parameter space of interest. The traditional inverse area figure of merit for the joint posterior of several probes is independent of the location of the probes in parameter space - i.e. that figure of merit remains the same regardless of the potential size or direction of any bias that may affect the location of the proposed probe.

- How to we optimize future surveys which may be subject to unforeseen bias?
- How do we quantify the degree of complementarity between the different probes of dark energy?

Chapter 8 describes work I have been contributing to in developing an additional statistic, ‘Robustness’, (March et al., 2011b) based on the Bayesian evidence, which gives a measure of how compatible any two probes are. Probes which have a high degree of overlap in parameter space have a higher numerical value for Robustness, and probes are said to be more compatible, probes which have a lower degree of overlap in parameter space have a lower value for Robustness. As a test case I evaluated Robustness for a EUCLID type BAO survey and a SNAP type supernovae Ia survey.

1.3 Structure of this thesis

After this introduction there are three chapters describing the background pertaining to the work chapters following. Chapter 2 briefly describes the cosmological background to this work, outlining the Friedman Robertson Walker and ΛCDM model of the Universe. Chapter 4 describes in some detail the supernovae type Ia which are the major astrophysical probe treated in this work, in particular in chapter 7, whilst chapter 5 describes the statistical methods (primarily Bayesian) utilised. These background chapters are not intended to give an exhaustive description of current cosmology and statistics, but serve to place in context and give sufficient technical background for the understanding of the work chapters.

Following the work chapters, chapter 9 summarises the main conclusions of the work presented in this thesis and gives an overview of the future directions of this work.

Chapter 2

Cosmology background

The current standard model of the Universe is one which is homogeneous and isotropic (cosmological principle) in which we as observers occupy no privileged vantage point (Copernican principle). This model is described as the Friedmann-Lemaître-Robertson-Walker model (FLRW) and the particular matter-energy densities of roughly 70% dark energy (Λ), 25% dark matter (CDM) and 5% baryonic matter are described as the ΛCDM model. As a sign that the FLRW ΛCDM model has become the current accepted concordance model of cosmology, we see that there are many excellent and comprehensive standard texts describing this model. The content of this background chapter is largely drawn from the following standard texts: (Dodelson, 2003; Mukhanov, 2005; Weinberg, 2008; Peacock, 1999; Schneider, 2006).

This chapter serves to give a brief overview of some of the cosmological background necessary for the research presented in later chapters. The bulk of the research in this thesis is concerned with the extraction of cosmological parameters from the supernovae data, and in that context, it is sufficient to consider the smooth expanding Universe and its distance measures described in sections 2.2 and 2.3. However, this thesis also makes use of CMB data sets, and baryon acoustic oscillation (BAO) data sets, which depend on our understanding of the perturbed Universe, hence a brief overview of the inhomogeneous Universe is also included.

2.1 Prelude: An historical overview

Since much of the work in this thesis is focussed on investigating the apparent late time acceleration of the expansion of our Universe it seems fitting to include a brief historical note about the first observations which suggested that our Universe is indeed expanding, Einstein's independent suggestion of the cosmological constant, and an early use of standard candles.

In 1929 Edwin Hubble published a paper describing the relationship between the radial velocity and distance of certain extragalactic ‘nebulae’. (Hubble, 1929). Hubble’s methodology for determining distance depended on the knowledge that certain types of astrophysical objects within the extragalactic ‘nebulae’ had a known absolute magnitude, or at the very least had an absolute magnitude which lay within a small, well known range. By observing the apparent magnitude of known types of objects and comparing this to their supposed absolute magnitude, their distances could thus be calculated. Hubble essentially used objects of known brightness (such as Cepheid variables, novae and blue stars) as ‘standard candles’, that would appear dimmer at greater distances and brighter at closer distances. The velocities of the extra galactic nebulae were obtained from observations of their redshift.

Hubble concluded that a ‘roughly linear’ relationship existed between the velocities and distances of these extragalactic nebulae, as shown in fig. 2.1. The velocity, $v_{\rm obs}$, as observed from Earth, was proportional to the distance, $d(t_0)$, for low redshift $z \lesssim 0.1$, extragalactic objects:

$$v_{\rm obs} = H_0 d(t_0) + v_{pec} \tag{2.1}$$

Where H_0 is the value of the Hubble parameter today and $v_{\rm pec}$ is the peculiar velocity of the object due to local gravitational interactions. The recessional velocity, $v(t_0)$, is defined as:

$$v(t_0) = H_0 d(t_0) \tag{2.2}$$

where subscript 0 denotes values today. On scales larger than 100Mpc, the universe appears to be both isotropic and homogeneous and the Hubble parameter has no spatial dependence. For scales greater than the scale of homogeneity, the peculiar velocity is small with respect to the recessional velocity and can be neglected. Hubble’s observations that the further away an object is, the more rapidly it is receding, confirmed the theoretical predictions of Friedmann (1924) that an expanding universe was indeed a valid solution to Einstein’s theory of General Relativity.

Hubble’s tentative 1929 result for extragalactic objects at low red shift, has been subsequently confirmed by modern observations. Furthermore, it has been shown that the Hubble parameter varies with higher redshift (or earlier times), such that $H = H(t)$ and:

$$\dot{r}(t) = H(t) r(t) \tag{2.3}$$

where $r(t)$ is the distance in the radial direction and overdots denote derivatives with respect to time, t. The relationship between the physical distance, $r(t)$ and the comoving distance, χ is given by:

$$r(t) = \chi a(t) \tag{2.4}$$

Where $a(t)$ is known as the scale factor and the Hubble parameter is defined as:

$$H(t) \equiv \frac{\dot{a}(t)}{a(t)} \tag{2.5}$$

Having defined the Hubble parameter which describes the expansion rate of the universe, it is of interest to know what physical processes control this parameter. As shall be discussed in some detail later, it is Friedmann's equation which describes the physics governing the Hubble Rate:

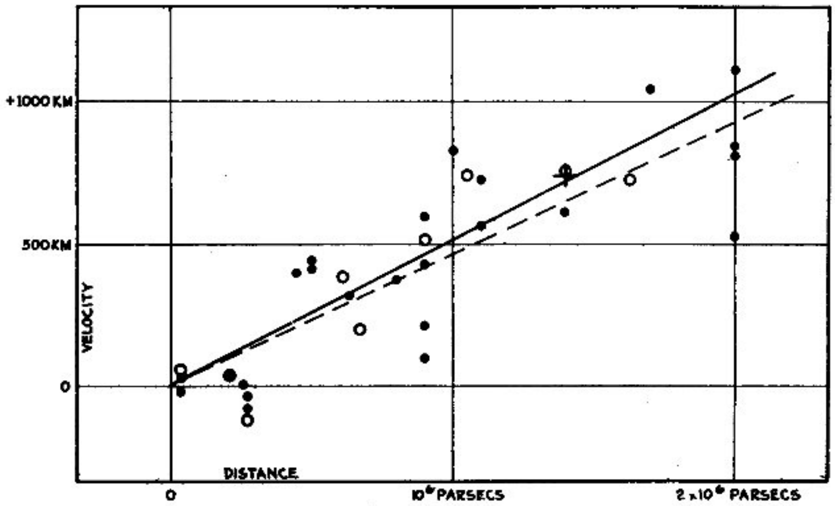


FIGURE 1

Velocity-Distance Relation among Extra-Galactic Nebulae.

Figure 2.1: Plot from Hubble (1929) showing that more distant objects recede more rapidly than nearby objects, a result that challenged the conventional wisdom of the time, suggesting that the universe was not static, but dynamic and expanding.

$$H^2(t) = \left(\frac{\dot{a}(t)}{a(t)}\right) = H_0^2(\rho_m(t) + \rho_r(t) + \rho_\Lambda(t) + \rho_K(t)) \tag{2.6}$$

H_0, is the value of the Hubble Rate today and is a constant, meaning that the time dependent Hubble rate, $H(t)$ is governed entirely by the sum of the evolving energy densities, (i.e. $\rho(t)s$) of the principal constituents of the universe. The equations describing the energy densities due to matter, $\rho_m(t)$ and radiation, $\rho_r(t)$ are well understood and shall be discussed in due course. $\rho_K(t)$ represents the contribution to the energy density due to the curvature of the universe, a term which vanishes in the case of a flat universe. However, it is $\rho_\Lambda(t)$ which is the term of special interest for the research presented in this thesis.

When $\rho_\Lambda(t)$ is considered to be constant, such that $\rho_\Lambda(t) = \rho_{\Lambda_0}$, then it is traditionally termed the 'Cosmological Constant', but when it is allowed to vary with time, it is known as 'Dark Energy'. The form which $\rho_\Lambda(t)$ takes is by no means certain and there are many unanswered questions regarding its nature and origin. Some of the initial questions which this project seeks to answer include identifying whether $\rho_\Lambda(t)$ has evolved with time, whether it is still evolving with time, what values it may take and whether it is the same in all directions.

In order to constrain the form of the dark energy term, $\rho_\Lambda(t)$, it is necessary to map out how $H(t)$ varies in time and discern how this can be affected by the $\rho_\Lambda(t)$ term in Eq. (2.6). Edwin Hubble mapped out the evolution of $H(t)$ using a 'standard candle' technique and a similar methodology is used today with the supernovae type Ia - this is the particular focus of chapter 7. As well as 'standard candles', 'standard rulers' can be used to map out the expansion history of the Universe. The idea of a standard ruler is to take some object of fixed co-moving length, χ and measure its physical length, $r(t)$ at various known times (i.e. at different redshifts).

No astrophysical objects have been identified as being suitable candidates for standard rulers, but a cosmological phenomenon has been identified as being appropriate: Baryon Acoustic Oscillations. Although in accordance with the cosmological principle, the universe is isotropic and homogeneous on large scales, it is lumpy on small scales ($<$ 100Mpc). On small scales, matter is clumped into over and under densities which are distributed according to preferred length scales, preferred scales which were established by the physical processes occurring around the time of recombination. The idea is to use these preferred scales in the matter distribution as standard rulers to probe the history of the universe and the evolution or otherwise of the dark energy. Astrophysical observations of galaxies and their spatial distribution reveal the imprint of the baryon acoustic oscillations. Astrophysical observations of this type are used in chapters 6 and 8 as probes of cosmology and in particular dark energy.

2.2 The Friedmann Lemaître Robertson Walker Universe

Assuming that the Universe is isotropic and homogeneous, the simplest description of this type of expanding spacetime is give by the Friedman-Lemaître-Robertson-Walker (FLRW) metric; where ds is the spacetime interval. In Cartesian coordinates this is given by:

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu \tag{2.7}$$

where for an unperturbed universe the FLRW metric is:

$$g_{\mu\nu} = \mathrm{diag}(-1, a^2(t), a^2(t), a^2(t)) \tag{2.8}$$

Transforming into spherical polar coordinates gives another useful form of the FLRW metric, in which the universe can be considered as a 3 dimensional hypersurface embedded in 4-space:

$$ds^2 = -dt^2 + a^2(t)\left[\frac{dr^2}{1-Kr^2} + r^2 d\Omega^2\right] \tag{2.9}$$

where:

$$d\Omega^2 = d\theta^2 + \sin^2\theta d\phi^2 \tag{2.10}$$

K describes the curvature of the universe:

$$K = \frac{|a|^2}{a^2} \tag{2.11}$$

$$K = \begin{cases} +1 & \text{bounded, positively curved} \\ 0 & \text{unbounded, flat} \\ -1 & \text{unbounded, negatively curved} \end{cases} \tag{2.12}$$

In the case of the curved universe, $K \neq 0$, then the magnitude of the scale factor has geometrical meaning as the radius of curvature of the universe. A further coordinate transformation allows the FLRW metric to be re written in terms of co-moving distance, χ:

$$ds^2 = -dt^2 + a(t)^2(d\chi^2 + S_K^2(\chi) d\Omega^2) \tag{2.13}$$

where $S_K(\chi)$ is given by:

$$S_K(\chi) = \begin{cases} \sin(\chi) & \text{positively curved} \\ \chi & \text{flat} \\ \sinh(\chi) & \text{negatively curved} \end{cases} \tag{2.14}$$

The relationship between the matter and the gravitational forces (i.e. geometry) of the Universe, are codified in the Einstein equations; which are stated here without proof. For a detailed description see for example (Misner et al., 1973).

$$G_{\mu\nu} = 8\pi G T_{\mu\nu} \tag{2.15}$$

where the Einstein Tensor is:

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} \mathcal{R} \tag{2.16}$$

with and $R_{\mu\nu}$ and $\mathcal{R}$ being the Ricci tensor and scalar respectively:

$$R_{\mu\nu} = \Gamma^{\alpha}_{\mu\nu,\alpha} - \Gamma^{\alpha}_{\mu\alpha,\nu} + \Gamma^{\alpha}_{\beta\alpha}\Gamma^{\beta}_{\mu\nu} - \Gamma^{\alpha}_{\beta\nu}\Gamma^{\beta}_{\mu\alpha} \tag{2.17}$$

$$\mathcal{R} = g^{\mu\nu} R_{\mu\nu} \tag{2.18}$$

The comma notation indicates a partial derivative. $T_{\mu\nu}$ is the stress-energy tensor, which describes the characteristics of the 4-momentum flowing through spacetime at any point. The stress-energy tensor for a perfect fluid of pressure, p and density, ρ, flowing with 4-velocity u^{μ} may be written as:

$$T_{\mu\nu} = (\rho + p) u^{\mu} u^{\nu} + p g^{\mu\nu} \tag{2.19}$$

The right hand side of Einstein's equation quantifies the amount of matter and the left hand side is a statement about the geometry of the spacetime. The Einstein tensor is constructed from the Ricci tensor and scalar. Einstein realized that the initial equation he proposed allowed for a dynamic, expanding universe, which at the time was counter intuitive to him, as a static universe was the favoured model of the day. In order to compensate for this and force the universe described in his equation to be static, Einstein added an additional term, the cosmological constant, Λ:

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} \mathcal{R} - g_{\mu\nu} \Lambda \tag{2.20}$$

Einstein originally added his additional term $g_{\mu\nu}\Lambda$ to the left hand side of Eq. (2.20), signifying a modification to gravity, but in modern discussions of the cosmological constant the term is generally subtracted from the right hand side of Eq. (2.20) indicating a modification to the matter-energy content of the Universe. The effect of introducing Λ was to alter the dynamics to prevent the possible expansion or collapse of the universe, but once it became established through Hubble's observations that the universe was indeed expanding, the additional term could be removed. However as modern cosmological observations seem to indicate that the universe is accelerating (Riess et al., 1998; Perlmutter et al., 1999b), a fact which cannot be accounted for by ordinary matter and energy densities, recent years have seen a reinstatement of Λ as a means of explaining this acceleration. Whereas Einstein sought to halt the universe with Λ, modern cosmologists are seeking to accelerate the universe with Λ.

For this background description of the ΛCDM Universe, we shall consider the cosmological constant to be in the form of an additional energy density, in the stress energy tensor. Alternative explanations for the apparent late time acceleration of the Universe are discussed in chapter 3. In order to obtain solutions to the Einstein equation, it is necessary first of all to obtain the metric connections for the Ricci tensor, which may be calculated directly from the metric. The non zero metric connections obtained are as follows; where overdots denote derivatives with respect to time:

$$\Gamma^i_{0j} = \Gamma^i_{j0} = \delta_{ij}\frac{\dot{a}(t)}{a(t)} \tag{2.21}$$

$$\Gamma^0_{ij} = \delta_{ij}a(t)\dot{a}(t) \tag{2.22}$$

Using the metric connections in Eq. (2.21) and Eq. (2.22), with the FLRW metric Eq. (2.8)), the Ricci scalar can be calculated via the Ricci tensor to give:

$$\mathcal{R} = 6\left[\frac{\ddot{a}(t)}{a(t)} + \left(\frac{\dot{a}(t)}{a(t)}\right)^2\right] \tag{2.23}$$

Using this to calculate the time, time component of the Einstein equation gives the first Friedmann equation for spatially flat spacetime:

$$H(t)^2 = \left(\frac{\dot{a}(t)}{a(t)}\right)^2 = \frac{8}{3}\pi G\rho(t) \tag{2.24}$$

Calculation of the space, space components of the Einstein equation leads to the

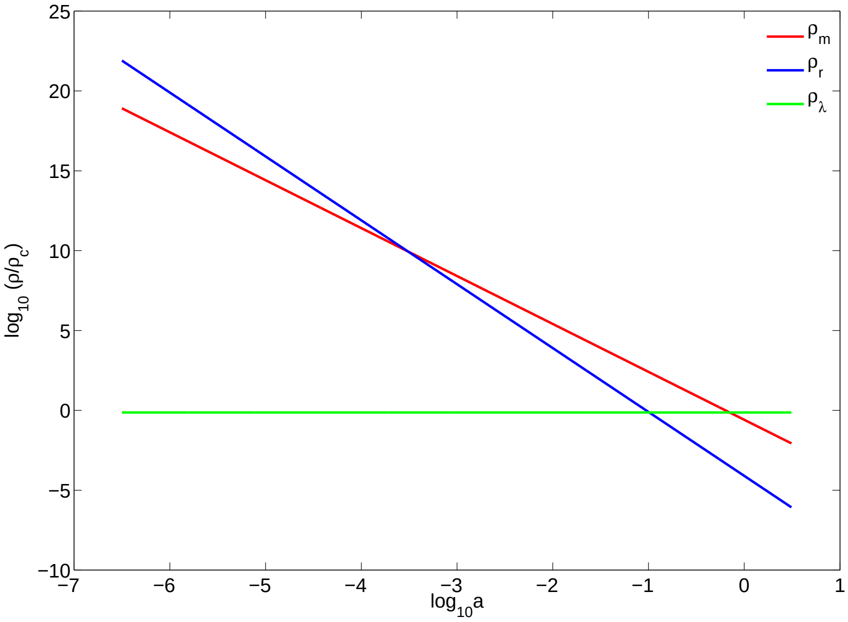


Figure 2.2: The evolution of energy density with scale factor, shown here for the WMAP best fit parameters (table 2.1). After the radiation era, the universe passed through matter-radiation equality, and entered the matter dominated era. Matter and radiation densities are diluted as the universe expands, but the relative amount of dark energy or cosmological constant has increased, allowing its density to remain constant or near constant. It is only in the present era that dark energy has begun to dominate the universe.

second Friedmann equation:

$$\frac{\ddot{a}(t)}{a(t)} = -\frac{4}{3}\pi G(\rho(t) + 3p(t)) \tag{2.25}$$

These two Friedmann equations are of fundamental importance, as they summarise the history of the universe. For a curved universe, the curvature term appears in the Friedmann equation as:

$$H(t)^2 + \frac{K}{a^2(t)} = \left(\frac{\dot{a}(t)}{a(t)}\right)^2 = \frac{8}{3}\pi G\rho(t) \tag{2.26}$$

The pressure $p(t)$ and density $\rho(t)$ terms in Eq. (2.24) and Eq. (2.25) can be expressed as sums of the $p(t)s$ and $\rho(t)s$ of the various fluids that make up the universe, provided that the fluids are non-interacting. The cosmological constant and curvature term may be included as fluids with contributing densities, such that:

$$\rho(t) = \rho_m(t) + \rho_\gamma(t) + \rho_\Lambda(t) + \rho_\kappa(t) \tag{2.27}$$

Description	Symbol	Value
Hubble constant	H_0	71.9 ± 2.7 km/s/Mpc
Baryon density	Ω_b	0.0441 ± 0.0030
Dark matter density	Ω_c	0.214 ± 0.0027
Dark energy density	Ω_Λ	0.742 ± 0.030
Curvature fluctuation amplitude, $k_0 = 0.002$ Mpc^{-1}	$\Delta^2_{\mathcal{R}}$	$(2.41 \pm 0.11)\text{x}10^{-9}$
Scalar spectral index	n_s	0.963 ± 0.015
Reionization optical depth	τ	0.087 ± 0.017

Table 2.1: Summary for WMAP 5-year update best fit parameters for flat ΛCDM model Dunkley et al. (2009)

where:

$$\begin{array}{lll} \rho_m(t) & \text{non relativistic pressureless matter (dust)} & w = 0 \\ \rho_\gamma(t) & \text{relativistic matter (photons)} & w = 1/3 \\ \rho_\Lambda(t) & \text{cosmological constant} & w = -1 \\ \rho_\kappa(t) & \text{curvature} & w = -1/3 \end{array} \tag{2.28}$$

The energy densities can be calculated, using the relativistic continuity equation (Dodelson, 2003), where w is the equation of state parameter $w = p/\rho$:

$$\frac{\partial \rho}{\partial t} + 3H(t)(\rho + p) = 0 \tag{2.29}$$

which in terms of a is:

$$\frac{\partial \rho}{\partial t} + 3\frac{\dot{a}(t)}{a(t)}(1 + w)\rho = 0 \tag{2.30}$$

which in the case of constant w has solutions of the form below:

$$\rho(t) = \rho_0 a(t)^{-3(1+w)} \tag{2.31}$$

So the Friedmann equation can be written as:

$$H(t)^2 = \left(\frac{\dot{a}(t)}{a(t)}\right)^2 = \frac{8}{3}\pi G(\rho^0_m a(t)^{-3} + \rho^0_\gamma a(t)^{-4} + \rho^0_\Lambda + \rho^0_\kappa a(t)^{-2}) \tag{2.32}$$

Typically, the densities of the various components of the universe are written in terms of the critical density as cosmological parameters, Ω:

$$\Omega(t) = \frac{\rho(t)}{\rho_c(t)} \tag{2.33}$$

where the critical energy density $\rho_c(t)$ is defined as:

$$\rho_c(t) = \frac{3H^2(t)}{8\pi G} \tag{2.34}$$

In terms of the critical energy density, the Friedmann equation becomes:

$$H(t)^2 = H_0^2(\Omega_m^0 a(t)^{-3} + \Omega_\gamma^0 a(t)^{-4} + \Omega_\Lambda^0 + \Omega_\kappa^0 a(t)^{-2}) \tag{2.35}$$

which also leads to:

$$\Omega_\kappa = 1 - \Omega_{\text{total}} \tag{2.36}$$

where:

$$\Omega_{\text{total}} = \Omega_m + \Omega_\Lambda + \Omega_\gamma \tag{2.37}$$

and

$$\Omega_\kappa = \frac{-K}{a_0^2 H_0^2} \tag{2.38}$$

giving a value for the scale factor today, in a non flat, $(K \neq 0)$, universe:

$$a_0^{-1} = \sqrt{|\Omega_{\text{total}} - 1|}H_0 \tag{2.39}$$

The energy density of the dust scales as the inverse volume of the universe, the energy density of the photons also decreases with increase in scale factor, as $\sim a^{-4}$. However, in the model currently under discussion, with $w_\Lambda = -1$ the energy density of the cosmological constant does not vary with time, it is not conserved in the same way as the other quantities - it is not diluted as the universe expands, which means that the relative dominance of the cosmological constant must increase as the universe expands and $\Omega_\Lambda \to 1$.

Our main cosmological interest in the work presented in this thesis is in verifying whether or not this ‘cosmological constant’ is truly constant with time and whether w_Λ is well constrained to -1. We shall generally use the term ‘dark energy’ to refer to $\rho_\Lambda(t)$ and reserve the term ‘cosmological constant’ for models in which $\rho_\Lambda(t) = \rho_{\Lambda_0}$ and $w_\Lambda = -1$. For more details on dark energy models for which $w_\Lambda = -1$ see chapter 3.

2.3 Cosmological Distance Measures

For our later discussion on the use of standard candles and standard rulers, it is first useful to define a few cosmological distances and measures; for which we generally follow the methodologies of Mukhanov (2005) and Dodelson (2003).

2.3.1 Conformal Time

Thus far we have distinguished between physical size of an extended object, $r(t)$ and its comoving size, χ and their relationship via the scale factor:

$$r(t) = \chi a(t)$$

The comoving time, $\eta(t)$, is the maximum comoving distance a photon could have traveled in a given time. Supposing a photon travels a physical distance, dr during a time interval dt, at a velocity, c, in units where $c = 1$, from the metric:

$$dr = dt$$

in comoving coordinates:

$$\frac{d\eta}{dt} = a(t) \tag{2.40}$$

which may be integrated to give the conformal time, $\eta(t)$:

$$\eta(t) \equiv \int_{t_1}^{t_2} \frac{1}{a(t)} dt \tag{2.41}$$

A comoving distance of particular interest is the maximum comoving distance, $\eta_p(t)$ a photon could have traveled since the begining of the universe:

$$\eta_p(t) = \int_0^t \frac{1}{a(t)} dt \tag{2.42}$$

The comoving distance, $\eta_p(t)$ acts as a boundary which demarcates regions which are causally connected from those which are not causally connected. Any regions which are separated by a comoving distance greater than $\eta_p(t)$ are not causally connected. $\eta(t)$ is known as the 'particle horizon' or 'comoving horizon' and its size increases with the age of the universe.

2.3.2 Redshift and Comoving Distance

The expansion of the universe affects the wavelength of photons in such a way as to be shifted from shorter wavelengths to longer wavelengths, by a measurable amount, known as 'redshift', $z(t)$, which is also a useful measure of distance or time in the universe. This increase in wavelength is indicative of the energy loss of the photon

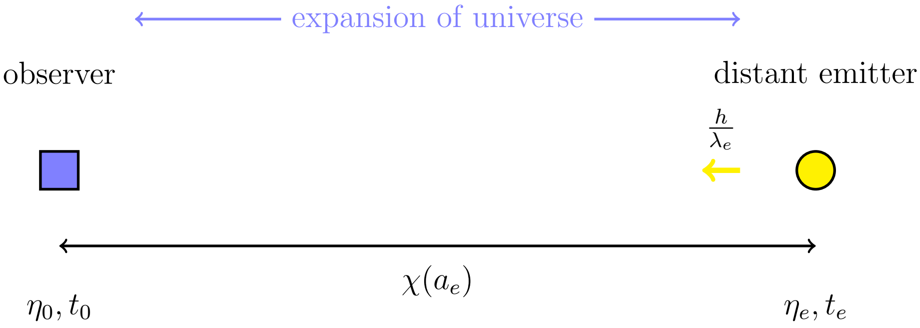


Figure 2.3: A pulse of light, of wavelength, λ_e, physical duration, Δt_e, conformal duration $\Delta\eta$ is emitted from a distant source, at conformal time, η_e and is received by an observer at conformal time, η_0. χ is the conformal distance, between the observer and the distant emitter.

as it does work against the expansion of the universe. Suppose a source at η_e emits a pulse of light of physical duration Δt_e and conformal duration $\Delta\eta$, as shown in fig.2.3. The conformal duration is fixed, but the physical duration will vary as the light pulse travels through the expanding universe and will change between the point of emission and the point of observation. At the point of emission, the duration of the pulse is:

$$\Delta t_e = \Delta\eta a(t_e)$$

at the point of observation, the duration of the pulse is:

$$\Delta t_0 = \Delta\eta a(t_0)$$

and as the conformal period is invariant:

$$\frac{\Delta t_e}{a(t_e)} = \frac{\Delta t_0}{a(t_0)}$$

The time interval of the pulse, Δt can be considered to be the time period of the photon, such that $\Delta t = c\lambda$, with $c = 1$ to give :

$$\frac{\lambda_0}{\lambda_e} = \frac{a(t_0)}{a(t_e} \tag{2.43}$$

As can be seen from Eq. (2.43), the wavelength of the photon on arrival at the observer today, λ_0 is longer its wavelength on emission from the source, λ_e, some time in the past. The photon's wavelength has been shifted towards the red end of the spectrum. The 'cosmological redshift', z is defined as the fractional shift in

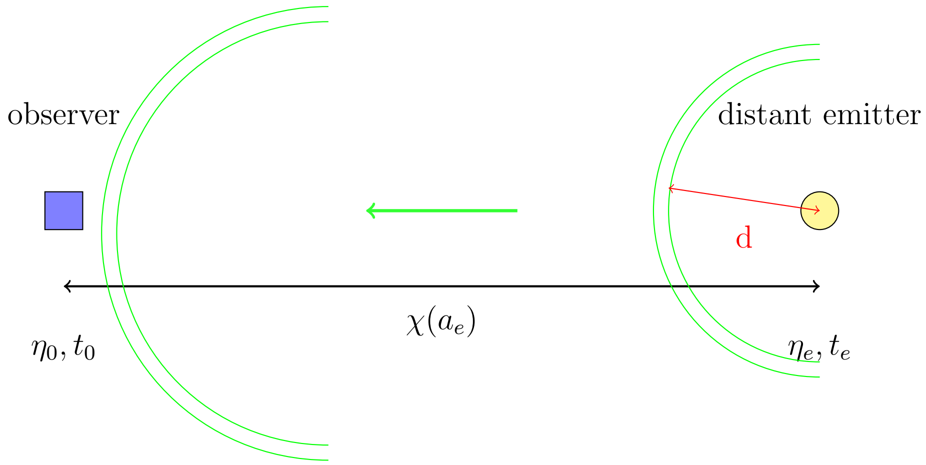


Figure 2.4: A distant source, of luminosity, L, emits photons in a thin spherical shell, of comoving thickness $\Delta\chi \equiv \Delta\eta$ and physical radius d, which travels out towards the observer.

wavelength:

$$z \equiv \frac{\lambda_0 - \lambda_e}{\lambda_e} \tag{2.44}$$

such that:

$$\frac{a_e}{a_0} = \frac{1}{(1+z)} \tag{2.45}$$

It is also useful to have an expression for the comoving distance from the observer out to the distant emitter, labeled $\chi(a_e)$ in fig. 2.3. This comoving distance can be obtained fromEq. (2.41), but instead of integrating from the begining of time to the present day, the limits need to be changed to find the the conformal time between the time of emission of the light pulse, t_e, z and time of its observation $t_0, z = 0$. It is also convenient to change to the coordinate of integration to z, such that the comoving distance out to an object at redshift z is:

$$\chi(z) = \frac{1}{a_0} \int_0^z \frac{1}{H(z)} dz' \tag{2.46}$$

2.3.3 Luminosity Distance

In a generic non-expanding spacetime, the observed flux, F, a distance, d from a far away object of intrinsic luminosity, L is given by:

$$F = \frac{L}{4\pi d^2}$$

Where the luminosity is the total energy, E per unit time, t. To calculate flux in an expanding universe, it is necessary to switch to comoving coordinates, as shown in fig. 2.4 and consider the path of a thin shell of photons, of comoving thickness $\Delta\chi \equiv \Delta\eta$ and physical duration Δt, emited from the distant source at η_e. The energy emited by the source is given by:

$$\Delta E(t_e) = L\Delta t_e \tag{2.47}$$

which in comoving coordinates is:

$$\Delta E(\eta_e) = La(\eta_e)\Delta\eta \tag{2.48}$$

The energy of a single photon is $E = 1/\lambda$, as the photons in the thin spherical shell move through the expanding universe, they are redshifted and loose energy, as described in Eq. (2.43), such that the energy of each observed photon is $E_0 = E_e\frac{a_e}{a_0}$. This means that there is an overall decrease in the energy contained in the thin shell by the time it reaches the observer:

$$\Delta E(\eta_0) = L\frac{a(\eta_e)^2}{a(\eta_0)}\Delta\eta \tag{2.49}$$

On reaching the observer, the physical surface area, $A(\eta_0)$ of the shell is:

$$A(\eta_0) = 4\pi\chi(a_e)^2 a(\eta_0)^2 \tag{2.50}$$

and its physical duration is:

$$\Delta t(\eta_0) = a(\eta_0)\Delta\eta \tag{2.51}$$

Thus in a flat expanding universe, the flux, i.e the energy per unit area per unit time, at the observer is given by:

$$F(\eta_0) = \frac{L}{a_0^2 4\pi\chi_e^2(1+z)^2} \tag{2.52}$$

where the comoving distance out to the object, $\chi_e \equiv \chi(a_e)$ is given by Eq. (2.46) in the general case of a curved universe, the flux is:

$$F(\eta_0) = \frac{L}{a_0^2 4\pi S_\kappa(\chi_e)^2(1+z)^2} \tag{2.53}$$

Where $S_\kappa(\chi_e)$ is given byEq. (2.14) and a_0 is given byEq. (2.39). The flux received by the observer can be measured, which is usually recorded in terms of the astronomical units of apparent bolometric magnitude, m_{bol}:

$$m_{bol}(z) = -2.5\log_{10} F(z) \tag{2.54}$$

which for an expanding universe is:

$$m_{bol}(z) = 5\log_{10}(S_\kappa(\chi_e)(1+z)) + \text{constant} \tag{2.55}$$

The distance modulus, μ_{bol} is the difference between the intrinsic magnitude of an object, M_{bol} and the observed magnitude of an object:

$$\mu_{bol} = m_{bol} - M_{bol} \tag{2.56}$$

The luminosity distance, d_L, is formally defined as:

$$d_L \equiv \left(\frac{L}{4\pi F}\right)^{1/2} \tag{2.57}$$

substituting Eq. (2.52) into Eq. (2.57), gives, for a flat universe:

$$d_L = (1+z)\int_0^z \frac{1}{H(z)}dz \tag{2.58}$$

The distance modulus can be written in terms of the luminosity distance:

$$\mu = 5\log\left[\frac{c}{H_0}\frac{d_L}{1\mathrm{Mpc}}\right] + 25 \tag{2.59}$$

This is the key equation for probing the expansion history of the universe using so-called standard candles. For a class of objects of fixed intrinsic magnitude, distributed over various redshifts, the distance modulus can be predicted for a given cosmology and compared with observed measurements to identify which cosmological model gives the best description of the observed universe.

2.3.4 Angular Diameter Distance

If a distant object at redshift, z, of known physical length, l, is subtended by an angle, θ, as in fig. 2.5, then the distance, d_A out to that object is the angular diameter distance; which in physical coordinates is:

$$\theta = \frac{l}{d_A} \tag{2.60}$$

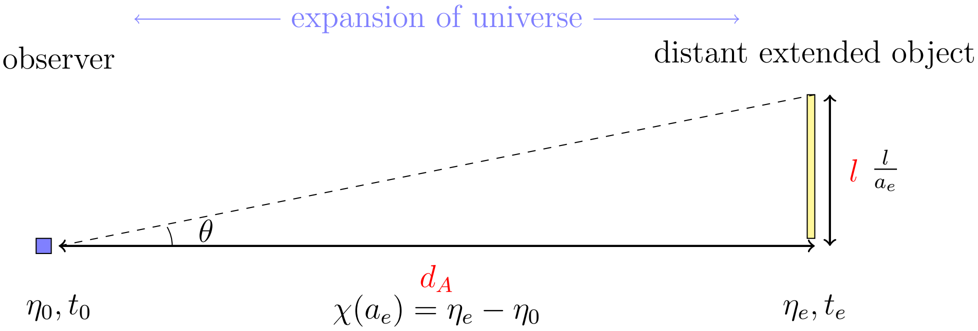


Figure 2.5: Angular Diameter Distance, d_A: A distant extended object of physical length, l, subtends an angle θ on the sky. Physical dimensions are shown in red; comoving distances and coordinates are shown in black; θ is the same in both comoving and physical dimensions.

or in comoving coordiantes:

$$\theta = \frac{l}{a_e}\frac{1}{\chi(a_e)} \tag{2.61}$$

Equating Eq. (2.60) andEq. (2.61) gives an expression for the angular diameter distance in a flat universe:

$$d_A = a_e\chi(a_e) = \frac{\chi(z)}{(1+z)} \tag{2.62}$$

where the comoving distance out to the object, $\chi(a_e)$ is given byEq. (2.46). It can also be useful to write the luminosity distance in terms of the angular diameter distance in this way:

$$d_L = (1+z)^2 d_A \tag{2.63}$$

where the scale factor today is unity $a_0 = 1$.

2.4 Inflation

The theory of inflation was devised primarily by Guth (1981) and Linde (1984) in the early 1980s as a means of explaining several cosmological problems, which were yet unexplained by the standard big bang model. A very brief overview of inflation is included here, partly because of its role in the description of the initial inhomogeneities and partly because there are some similarities between the mechanisms which drove inflation and some of the ideas used to describe dark energy. The similarities arise because both inflation and dark energy deal with the case of an accelerating universe; the acceleration during inflation was much more extreme than the slow acceleration seen today, however, the two scenarios share some common ground.

The main problems which inflation sought to solve were the so-called flatness and horizon problems. The distribution of the photons released at recombination and observed today initially by Penzias and Wilson and later by space borne instruments such as COBE and WMAP, was seen to be inhomogeneous on small scales, but statistically homogeneous on large scales, $\theta \gtrsim 1^\circ$. The problem is why do regions that would not have been not causally connected at recombination (when the CMB was released) appear to be statistically statistically homogeneous today? According to inflation, this problem can be solved if the universe undergoes a brief period of rapid expansion at early times.

The expression for the comoving horizon, Eq. (2.41) may be re-written as:

$$\eta(a) = \int_0^a \frac{1}{aH(a)} \frac{da}{a} \tag{2.64}$$

where the comoving Hubble radius is defined as:

$$\frac{1}{aH(a)}$$

The comoving horizon separates regions which are currently causally separated and have been causally separated since all time. The comoving Hubble radius separates regions which are currently causally separated, but may have been causally connected in the past (Dodelson, 2003). If the Hubble radius is smaller than the comoving horizon, then regions which are today separated by the Hubble radius, but lie within the comoving horizon, could have been causally connected in the past, even though they are causally disconnected.

In order to have a decreasing Hubble radius, $aH(a)$ must be increasing:

$$\frac{d}{dt}\left(\frac{a}{a}\frac{da}{dt}\right) > 0$$

such that we have an accelerating universe:

$$\frac{d^2a}{dt^2} > 0 \tag{2.65}$$

Our main interest here is to discern what drives this acceleration. Substituting Eq. (2.65) into the second Friedmann equation, Eq. (2.25) tells us something about the energy density involved:

$$0 < -\frac{4}{3}\pi G \underbrace{(\rho(t) + 3P(t))}_{\text{must be negative}} \tag{2.66}$$

$$P(t) < -\frac{\rho(t)}{3} \tag{2.67}$$

which demands that the pressure, $P(t)$ is negative, as the energy density, $\rho(t)$ is positive. This describes something completely different from ordinary matter and radiation, which always has a positive pressure. Guth identified this negative pressure as the driving force behind the exponential expansion Guth (1981) and this negative pressure could, for example, be supplied by a scalar field with a Lagrangian:

$$\mathcal{L} = \frac{1}{2}g^{\mu\nu}\partial_\mu\phi\partial_{nu}\phi - V(\phi) \tag{2.68}$$

which has a stress-energy tensor:

$$T^{\mu\nu} = \frac{1}{2}\dot{\phi}^2 - \frac{1}{2}(\nabla\phi)^2 - g^{\mu\nu}V(\phi) \tag{2.69}$$

whose time-time component gives the energy density:

$$\rho = \frac{1}{2}\dot{\phi}^2 + V(\phi) \tag{2.70}$$

and whose spatial components give the pressure:

$$P = \underbrace{\frac{1}{2}\left(\frac{\partial\phi}{\partial t}\right)^2}_{\text{kinetic term}} - \underbrace{V(\phi)}_{\text{potential term}} \tag{2.71}$$

In order for P to be negative, then the potential energy must be greater than the kinetic energy. Guth's original idea for such a scenario was to have a field trapped in a false minimum; because the field was trapped, it would have little kinetic energy and because it was in a false minimum, the potential energy could be a suitable non-zero value. The problem with Guth's original idea was how a 'graceful exit' could be made from entrapment in the false minimum, bringing the epoch of inflation to an end.

Linde (1984) reviewed some of the various modifications to Guth's 'old' inflationary theory and developed a 'new' inflationary model in which the field was not trapped in a false minimum, but began at some positive potential from which it could freely and slowly 'roll' towards the true vacuum potential. As well as solving the problem of escape from the false minimum, Linde's new model also had the added and important benefit of predicting that the density perturbations produced during the slow roll inflation are inversely proportional to $\dot{\phi}$ (Linde, 2005), such that the over-densities, $\delta\rho$ for different momenta, k are given by Linde (1984):

$$\frac{\delta\rho(k)}{\rho} = \frac{1}{\sqrt{2}\pi^3} \frac{H^2}{|\dot{\phi}|}\bigg|_{\phi=\phi^*} \tag{2.72}$$

where ϕ^* is the value of the field at the moment at which the momentum k of this perturbation was equal to $k_* = H$. The end of inflation occurs when the field finally reaches the true minimum and oscillates about that minimum for some time before decaying and producing elementary particles as it decays. Thus ends our brief and highly simplified description of inflation. The so-called 'new' inflation that we have outlined serves to illustrate the main principles of inflation, however there are other inflationary theories offering refinements and developments. Some of these developed theories include 'chaotic inflation', which does not assume thermal equilibrium for its initial conditions and 'hybrid inflation' which uses multiple scalar fields.

2.5 The Evolution of the Inhomogeneities

One of the useful outcomes of the inflation model is that it predicts the initial distribution of density inhomogeneities, in Eq. (2.72), which eventually give rise to the large scale structure of the present day universe and the temperature anisotropies in the CMB. The power spectrum of these initial over and under densities is usually parametrized in the following way, (e.g. Bridle et al. (2003)) and is termed the 'primordial power spectrum':

$$P_{\chi}(k) = A_s \left(\frac{k}{k_{s0}}\right)^{n_s-1} \tag{2.73}$$

The power spectrum amplitude, A_s, determines the variance of the fluctuations. The potential of the inflation field determines the spectral index, n_s, with $n_s = 1$ corresponding to the scale invariant power spectrum or Harrison-Zel'dovich-Peebles spectrum. In the slow roll models, where the field is rolling towards the true vacuum, n_s varies very slowly such that $|n_s - 1| << 1$.

Having now defined the primordial power spectrum, we wish to know what changes the matter density distribution undergoes in order to produce the present day matter power spectrum and the CMB power spectrum.

The principal constituents of the early universe were photons, baryons and dark matter. Protons and electrons were tightly coupled due to Coulomb scattering and since the mass of the protons is much greater than the mass of the electrons, electrons are included under the title of ‘baryons’. Neutrinos were also present, but we shall omit them from our current discussion.

At early times, the photons and baryons were tightly coupled by Compton scattering into a photon-baryon fluid. Oscillations were set up in this fluid, driven by perturbations in the density distribution of dark matter: the mass of the baryons meant that the fluid tended to fall into the gravitational potential wells in dark matter over dense regions and the radiation pressure of the photons would tend to force the fluid out of the gravitational potential wells. Hence acoustic waves were set up within the photon-baryon fluid with areas of compression having an over density of baryons and areas of rarefaction having an under density of baryons; as long as the photons and baryons were coupled, the fluid continued to oscillate, at characteristic wavelengths related to the speed of sound in the photon-baryon fluid.

Near the time of recombination, the photon-baryon fluid decoupled and the oscillations ceased; baryon over-densities and under-densities were ‘frozen in’ on a characteristic scale determined by the wavelength of the acoustic oscillations. Decoupled from the baryons, the photons are then free to travel or ‘free stream’ from the place which they were last scattered, through the universe to be observed today. The surface which connects all the points from which a photon last scattered is known as the last scattering surface or LSS, these photons which were released at the LSS and are observed by us today make up the cosmic microwave background radiation, or CMB. After recombination, the baryons and dark matter continued to interact gravitationally, giving rise to the large scale matter inhomogeneities we see today, and the characteristic matter power spectrum

The physics which governs growth of the inhomogeneities and the behaviour of the photon-baryon fluid is given by perturbation theory, of which a brief overview shall be given. We will give a brief description of the evolution of perturbations in some of the principal constituents of the universe, specifically, we shall look at how photons, baryons and dark matter are distributed and how these distributions change with time. We will also take note of how the perturbations in these affect the metric and how the perturbations in the metric in turn affect the distribution of the photons, baryons and dark matter. We shall not reproduce the full derivation of these perturbations here, (for the full derivation see Dodelson (2003)) however, we shall highlight some of the main ideas involved.

The general strategy is to perturb the metric at first order and then look at how this affects the distribution functions of the various quantities of interest, by looking at how the perturbations to the metric affect the paths of the particles as they move through the universe.

Here, we are considering the growth of perturbations in the early universe, up until the time of recombination. These are the perturbations which are manifested as small inhomogeneities in the CMB, hence we can make linear approximations in this calculation.

For a flat, smooth, expanding universe, spacetime is described by the Friedmann-Robertson-Walker metric; as in Eq. (2.8). In a perturbed universe, small changes in the Newtonian potential, $\Phi(\vec{x}, t)$ and spatial curvature, $\Psi(\vec{x}, t)$, can be introduced to perturb the smooth FLRW metric of Eq. (2.8) in this way:

$$g_{\mu\nu} = \begin{pmatrix} -1 - 2\Psi(\vec{x}, t) & 0 & 0 & 0 \\ 0 & a^2(t)(1 + \Phi(\vec{x}, t)) & 0 & 0 \\ 0 & 0 & a^2(t)(1 + \Phi(\vec{x}, t)) & 0 \\ 0 & 0 & 0 & a^2(t)(1 + \Phi(\vec{x}, t)) \end{pmatrix} \tag{2.74}$$

Where the sign convention is such that $\Phi(\vec{x}, t) < 0$ and $\Psi(\vec{x}, t) > 0$ describe overdense regions, whilst $\Phi(\vec{x}, t) > 0$ and $\Psi(\vec{x}, t) < 0$ describe underdense regions. One of the key points here is that $\Phi(\vec{x}, t)$ and $\Psi(\vec{x}, t)$ are small and therefore terms which are quadratic and above in these terms may be dropped. In Eq. (2.74), an explicit choice of gauge has been made, that of the conformal Newtonian gauge. Only scalar perturbations are being considered here.

The distribution function, $f(\vec{x}, \vec{p})$ describes the phase space distribution of the various particles;

$$f(\vec{x}, \vec{p}) = \frac{1}{e^{(E(p)-\mu)/T} \pm 1} \tag{2.75}$$

(+ for fermions, - for bosons, where $E(p)^2 = m^2 + p^2$, in units where $c = k_B = 1$). The Boltzmann relationship describes how the rate of change in the phase space distribution is related to the collision term $C[f(t)]$ for a given species, it can be expressed as:

$$\frac{df}{dt} = C[f(t)] \tag{2.76}$$

The derivation of the Boltzmann equations looks at how the distribution function for each of the species changes with respect to time in the presence of the perturbed metric. If the species under consideration were non interacting, then the collision term on the right of Eq. (2.76) would be zero; if the species is interacting, then a suitable expression for the interactions must be given, describing the scattering of particles in to and out of the phase space element.

The first order Boltzmann equations we present here are written in Fourier space, in terms of the wave vector, $\bar{k}$ of magnitude k. Primes, ($'$), represent differentials with respect to conformal time, η.

2.5.0.1 Photon and Baryon Perturbations

Photons, protons and electrons are coupled to the metric via gravity and are also coupled to each other. Photons are coupled to electrons through Compton scattering and protons are coupled to electrons through Coulomb scattering; the tight coupling between the the protons and electrons means they share a common over-density, δ_b and a common velocity, v_b

As the photons travel through the perturbed universe, they gain energy falling into potential wells, lose energy climbing out of potential wells and lose energy overcoming the expansion of the universe. The change in photon energy can be characterized by a change in the photon temperature, δT and we define $\Theta = \frac{\delta T}{T}$, where T is the average photon temperature.

A calculation of the amplitude for Compton scattering provides the collision terms for photon - electron interactions and this scattering affects the direction of the scattered photons. The direction of the photons is given by the momentum direction of the photons, $\hat{p}$ and the relationship between the direction of the photons and the direction of the wave vector is characterized by μ, where $\mu = \frac{\bar{k}.\hat{p}}{k}$. To deal with the directional importance of the photons, the photon perturbations can be described by the $\Theta_l(k, \mu)$ which are expanded in terms of Legendre polynomials, $\mathcal{P}_l$, such that:

$$\Theta_l \equiv \frac{1}{(-i)^l} \int_{-1}^{l} \frac{d\mu}{2} \mathcal{P}_l(\mu) \Theta(\mu) \tag{2.77}$$

$l = 0$ is the monopole $\Theta_0(k)$ which corresponds to the difference between the temperature perturbation at a specific point, and the the average temperature perturbation over all space; $l = 1$ is the dipole $\Theta_1(k)$ which is related to the gradient of the energy density of the photons. Omiting contributions from polarisation, the Boltzmann equation (Dodelson, 2003) describing photon temperature perturbations is:

$$\Theta' + ik\mu\Theta = -i\Phi' - ik\mu\Psi - \tau' \left[\Theta_0 - \Theta + \mu v_b - \frac{1}{2}\mathcal{P}_2(\mu)\Theta_2 \right] \tag{2.78}$$

which may be expanded using Eq. (2.77). $\tau(\eta)$ is the optical depth and is the distance which a photon may travel on average before being scattered:

$$\tau(\eta) = \int_{\eta}^{\eta_0} a(\eta) n_e \sigma_T \, d\eta' \tag{2.79}$$

where n_e is the number density of the electrons and σ_T is the Thompson scattering cross section. The Boltzmann equations for the baryon perturbations are:

$$\delta_b' + ikv_b = -3\Phi' \tag{2.80}$$

$$v_b' + \frac{a'}{a}v_b = -ik\Psi + \frac{\tau'}{R}[v_b + 3i\Theta_1] \tag{2.81}$$

where the photon to baryon ratio, R, is defined as:

$$\frac{1}{R} \equiv \frac{4\rho_\gamma^{(0)}}{3\rho_b^{(0)}} \tag{2.82}$$

The above equations describe the oscillations of the photon-baryon fluid. Although they seem not to contain an explicit dependence on the dark matter distribution, the dependence on the dark matter is implicit through the dependence on the perturbations to the metric, which are in turn coupled to the perturbations in the dark matter density.

2.5.0.2 Cold Dark Matter Perturbations

Cold dark matter accounts for some $\sim 25\%$ of the energy density of the universe and dominates the matter distribution; it is non-baryonic and interacts only weakly with other particles, but does interact gravitationally. The precise identity of cold dark matter is unknown, but favoured candidates include WIMPs such as the neutralino. The first order Boltzmann equations (Dodelson, 2003) for cold dark matter with over density, δ and velocity, v are as follows:

$$\delta' + ikv = -3\Phi' \tag{2.83}$$

$$v' + \frac{a'}{a}v = -ik\Psi \tag{2.84}$$

The first order Boltzmann equations, Eq. (2.78), Eq. (2.80) and Eq. (2.83) described in this section govern the development of the perturbations in the distributions of photons, baryons and dark matter. In the following sections we shall look at how these equations can be used to explain the inhomogeneities in the large scale distribution of matter today and the anisotropies in the photon distribution of the CMB.

2.6 CMB Angular Power Spectrum

The anisotropies we observe in the CMB today are due to four main factors: Firstly the distribution of photon under densities and over densities at the time of recombination, described by the monopole, $\Theta_0(\eta_\star)$ (where $\eta_\star$ is the conformal time at recombination). Secondly, there are anisotropies due to the perturbations in the gravitational potential at the time of recombination $\Phi(\eta_\star)$; photons released at recombination which were in over dense regions at the time of recombination had to do work to climb out of the gravitational well; photons observed today from over dense regions are less energetic than photons from under dense regions. Thirdly, there is the 'Doppler' effect due to the peculiar velocity of the photons at recombination, as described by the dipole term, $\Theta_1(\eta_\star)$. Fourthly, the effect of a photon falling in to a potential well which is deepening with time causes an overall redshift of that photon.

The effects which are principally responsible for the multiple peak structure in the CMB angular power spectrum are the photon density distribution at recombination, $\Theta_0(\eta_\star)$, and the Doppler effect, $\Theta_1(\eta_\star)$. Since our principal interest in the CMB is the prospect of using the spacing of the acoustic peaks as a standard ruler to probe the expansion history of the universe, we will briefly highlight a few of the necessary steps for obtaining the monopole and dipole terms at recombination, and hence the spacing of the CMB peaks.

We need to obtain suitable expressions for the monopole and dipole from the Boltzmann equation for photons Eq. (2.78). Because we are dealing with the time before recombination, in the tight coupling limit, we can make the approximation that there is no anisotropic stress, such that:

$$\Psi = -\Phi \tag{2.85}$$

and the photon distribution can be completely described by the monopole and dipole terms. Expanding Eq. (2.78) and retaining only the monopole and dipole terms gives:

$$\Theta_0'(k) + k\Theta_1(k) = -\Phi' \qquad \text{monopole equation} \tag{2.86}$$

$$\Theta_1'(k) + \frac{k}{3}\Theta_0(k) = \frac{k}{3}\Psi + \tau'\left[\Theta_1 - iv_b\right] \qquad \text{dipole equation} \tag{2.87}$$

Along with the dark matter perturbation equations Eq. (2.83) and baryon perturbation equations Eq. (2.80) there is one other equation we need to describe the evolution of the inhomogeneities: an equation to describe the perturbations in the gravitational potential - for this, we quote without proof (Dodelson, 2003) the perturbed time-time component of the Einstein equation:

$$k^2\Phi + 3\frac{a'}{a}\left(\Phi' + \frac{a'}{a}\Phi\right) = 4\pi G a^2[\rho_c\delta + 4\rho_r\Theta_0] \tag{2.88}$$

where ρ_c and ρ_r are the dark matter and radiation energy densities respectively.

In theory, there are two main types of initial conditions possible, as set by inflation: adiabatic modes and isocurvature modes. In practice however, observations of the CMB rule out pure isocurvature modes (Väliviita et al., 2001), although mixtures of adiabatic and isocurvature modes are possible, (Trotta et al., 2001). Specific possibilities for isocurvature initial conditions include the baryon isocurvature mode, cold dark matter isocurvature mode, neutrino isocurvature density mode and the neutrino isocurvature velocity mode (Bucher et al., 2000). Adiabatic initial conditions are those for which there were spatial fluctuations in the total energy density, but not in the entropy, giving initial temperature fluctuations. Isocurvature initial conditions are those for which there were fluctuations in the entropy, but not in the total energy density, leading to near 'isothermal' initial conditions. Here we choose to use adiabatic initial conditions, which for initial time η_i are:

$$\Phi(\eta_i) = 2\Theta_0(\eta_i) \tag{2.89}$$

$$\delta(\eta_i) = 3\Theta_0(\eta_i) \tag{2.90}$$

$$\delta_b(\eta_i) = 3\Theta_0(\eta_i) \tag{2.91}$$

where the initial velocities and monopole are given by:

$$\Theta_1(\eta_i) = \frac{iv_b(\eta_i)}{3} = \frac{iv(\eta_i)}{3} = -\frac{k\Phi(\eta_i)}{6aH} \tag{2.92}$$

In principle, the differential equations for the photon perturbations Eq. (2.86), baryon perturbations Eq. (2.80), dark matter perturbations Eq. (2.83), and gravitational potential Eq. (2.88), can be solved numerically using the adiabatic initial conditions Eq. (2.89), in order to obtain the photon distribution at the time of recombination $\Theta(\eta_\star)$. However, there is also an analytic approach, aspects of which we will briefly highlight here as it gives a physical insight into the physics of the early universe, and the resulting spacing of the CMB peaks.

Hu & Sugiyama (1995) developed an analytical method for obtaining the photon density and velocity distribution: From Eq. (2.86), along with the equation for the baryon velocity Eq. (2.80), a second order differential equation can be obtained for

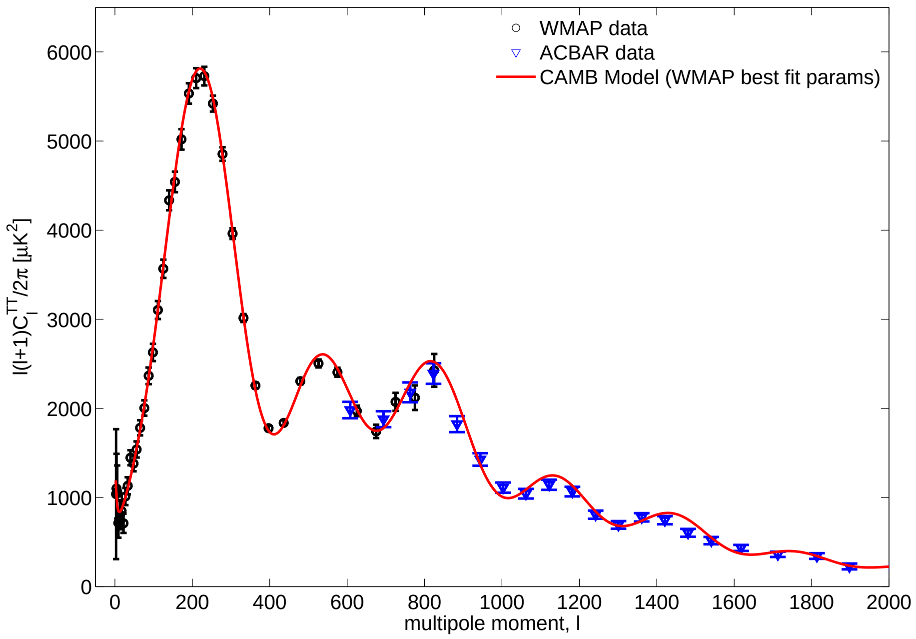


Figure 2.6: Solid line shows model CMB angular power spectrum produced using CAMB (Lewis et al., 2000) for a ΛCDM model using WMAP best fit parameters (table 2.1). Data points are from WMAP 5 (Nolta et al., 2009) and ACBAR (Reichardt et al., 2008). The 'acoustic peaks' due to oscillations in the photon-baryon fluid can be seen very clearly from $l = 200$ onwards; these acoustic peaks are also manifest in the galaxy power spectrum. The idea is to use the spacing of these acoustic peaks as a 'standard ruler' for probing the expansion history of the universe.

the monopole:

$$\Theta_0'' + \frac{a'}{a}\frac{R}{1+R}\Theta_0' + k^2 c_s^2 \Theta_0 = -\frac{k^2}{3}\Psi - \frac{a'}{a}\frac{R}{1+R}\Phi' - \Phi'' \tag{2.93}$$

where the sound speed of the photon-baryon fluid is defined as:

$$c_s \equiv \sqrt{\frac{1}{3(1+R)}} \tag{2.94}$$

Hu and Sugiyama obtained the following monopole solution to Eq. (2.93), for the case of adiabatic initial conditions:

$$\Theta_0(\eta)+\Phi(\eta) = [\Theta_0(0)+\Phi(0)]\cos(kr_s)+\frac{k}{\sqrt{3}}\int_0^{\eta} d\eta'[\Phi(\eta')-\Psi(\eta')]\sin[k(r_s(\eta)-r_s(\eta'))] \tag{2.95}$$

where $r_s(\eta)$ is the sound horizon, i.e. the co-moving distance traveled by a sound wave in conformal time η, and is defined as:

$$r_s(\eta) \equiv \int_0^\eta d\eta' c_s(\eta') \tag{2.96}$$

Differentiating Eq. (2.95) gives an expression for the dipole:

$$\Theta_1(\eta) = \frac{1}{\sqrt{3}}[\Theta_0(0) + \Phi(0)]\sin(kr_s) - \frac{k}{\sqrt{3}}\int_0^\eta d\eta'[\Phi(\eta') - \Psi(\eta')]\cos[k(r_s(\eta) - r_s(\eta'))] \tag{2.97}$$

The physical insight we gain here is that at recombination the locations of the peaks in the monopole are related to the cosine of the sound horizon, and the locations of the peaks in the dipole are related to the sine of the sound horizon - hence the terminology ‘acoustic peaks’. The monopole and dipole terms are out of phase, with the effect that the peaks in the dipole serve to ‘fill in’ the troughs of the monopole to some extent.

Eq. (2.95) and Eq. (2.97) go most of the way to describing the photon perturbation at the time of recombination, but to completely describe the photon distribution at the time of recombination, an additional term must be added to describe diffusion damping - i.e to account for the fact that even before recombination, the photons did have some freedom to free stream short distances. However, Eq. (2.95) and Eq. (2.97) do accurately predict the location of the ‘acoustic peaks’ in the CMB angular power spectrum.

Today, we observe the CMB photons coming from all directions on the sky, with temperature:

$$T(\vec{x}, \hat{p}, \eta) = T(\eta)[1 + \Theta(\vec{x}, \hat{p}, \eta)] \tag{2.98}$$

Since we are observing the projection of the CMB on to a spherical sky, we generally expand $\Theta(\vec{x}, \hat{p}, \eta)$ in terms of spherical harmonics:

$$\Theta(\vec{x}, \hat{p}, \eta) = \sum_{l=1}^{\infty}\sum_{m=-l}^{l} a_{lm}(\vec{x}, \eta) Y_{lm}(\hat{p}) \tag{2.99}$$

where $Y_{lm}(\hat{p})$ are the spherical harmonics and $a_{lm}(\vec{x}, \eta)$ are the observables given by:

$$a_{lm}(\vec{x}, \eta) = \int \frac{d^3k}{(2\pi)^3} e^{i\vec{k}.\vec{x}} \int d\Omega Y^*_{lm}(\hat{p})\Theta(\vec{x}, \hat{p}, \eta) \tag{2.100}$$

The quantity of interest is the angular power spectrum, C_l which is the variance of the $a_{lm}(\vec{x}, \eta)$:

$$\langle a_{lm} a^{*}_{l'm'} \rangle = \delta_{ll'} \delta_{mm'} C_l \tag{2.101}$$

The observed angular power spectrum for the CMB, as measured by WMAP (Nolta et al., 2009) and ACBAR (Reichardt et al., 2008) is shown in fig. 2.6, along with the ΛCDM model for the WMAP best fit parameters (table 2.1). The 'acoustic peaks' due to the oscillations in the photon-baryon fluid can be clearly seen from $l = 200$ onwards - these same peaks show up as 'baryonic wiggles' in the matter power spectrum. The idea is to use the spacing of the peaks manifest in the CMB angular power spectrum and the galaxy power spectrum as a 'standard ruler' to probe the expansion history of the universe and constrain dark energy.

The CMB data alone do not constrain dark energy, since dark energy only began to dominate at late times (see for example fig. 2.2), however the CMB data can be used in conjunction with other astrophysical probes of cosmology to place tighter constraints on the cosmological parameters. The full CMB data set and likelihood is used directly in our work on Bayesian doubt in chapter 6, section 6.3.2. In the work on cosmological parameter inference from the SNe Ia data, the cosmological constraints are combined with the effective constraints from the CMB. Wang & Mukherjee (2007) have shown that the information from the CMB observations relevant to dark energy constraints in the case of a redshift independent dark energy equation of state can be summarised by the CMB shift parameters:

$$R \equiv \sqrt{\Omega_m H_0^2} r(z_{CMB}) \tag{2.102}$$

$$l_A \equiv \pi \frac{r(z_{CMB})}{r_s(z_{CMB})} \tag{2.103}$$

where $r_s(z_{CMB})$ is the comoving sound horizon at recombination as in Eq.(2.96), and $r(z_{CMB})$ is the comoving distance to the recombination surface, where the comoving distance to redshift surface z is given by:

$$r(z) = \frac{c}{H_0} |\Omega_\kappa|^{-1/2} \text{sinn} \left[|\Omega_\kappa|^{1/2} \int_0^z \frac{H_0}{H(z)} \, \mathrm{dz} \right] \tag{2.104}$$

Following Kessler et al. (2009a) we use the R shift parameter to include the constraints from the CMB from the WMAP-5 data (Komatsu et al., 2009) in this way:

$$\chi^2_{CMB} = \left[\frac{R(z_{CMB}; w, \Omega_m, \Omega_\Lambda) - 1.710}{0.019} \right]^2 \tag{2.105}$$

Inclusion of the CMB constraints in this way are described in section 7.8 and the confidence intervals produced can be seen in fig. 7.17.

2.7 Matter Power Spectrum and Baryon Acoustic Oscillations

The distribution of dark matter may be inferred by making observations of galaxies which trace the underlying distribution of dark matter. From these observations, the galaxy (or dark matter) power spectrum can be obtained, which is the Fourier transform of the correlation function and is the matter analogue of the CMB angular power spectrum. The dark matter power spectrum is defined as:

$$P(k) = \langle \delta(\vec{k}) \delta^{\star}(\vec{k}') \rangle \tag{2.106}$$

The acoustic oscillations described in the previous section should show up in the matter power spectrum in a similar way, albeit less pronounced, to the acoustic peaks in the CMB angular power spectrum. The matter power spectrum at late times is given by:

$$P(k,a) = P_{\chi}(k) T(k)^2 \left(\frac{D_1(a)}{D_1(a=1)} \right)^2 \tag{2.107}$$

where $P_{\chi}(k)$ is the primordial power spectrum described by Eq. (2.73) and where the transfer and growth functions, $T(k)$ and $D_1(a)$ are defined as:

$$T(k) \equiv \frac{\Phi(k, a_{late})}{\Phi_{\text{large scale}}(k, a_{late})} \tag{2.108}$$

(for $a > a_{late}$)

$$\frac{D_1(a)}{a} \equiv \frac{\Phi(a)}{\Phi(a_{late})} \tag{2.109}$$

In the regime where the linear approximations are valid, a function for the gravitational potential today $\Phi(k,a)$ may be obtained by numerically solving the differential equations for the photon perturbations Eq. (2.86), baryon perturbations Eq. (2.80), dark matter perturbations Eq. (2.83), and gravitational potential Eq. (2.88), using the adiabatic initial conditions Eq. (2.89) described in the previous section.

The description we have outlined above, is based on the the Boltzmann equations and perturbed Einstein equations for which a number of linear approximations were made. Hence following this methodology leads to the linear matter power spectrum. Although the linear power spectrum may give a good approximation on large

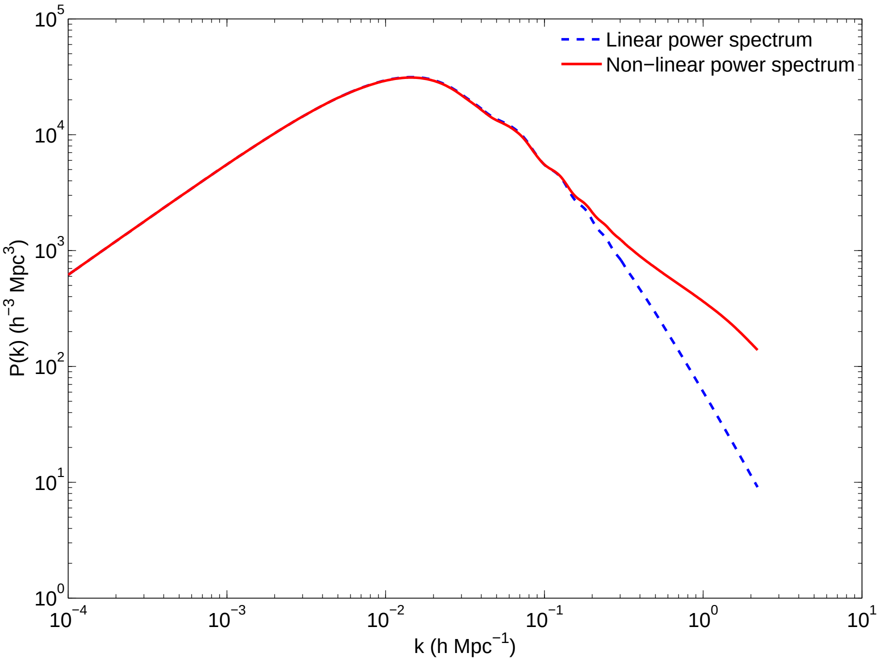


Figure 2.7: Linear and non-linear matter power spectra produced using CAMB (Lewis et al., 2000) for a ΛCDM model using WMAP best fit parameters (table 2.1). Calculation of non-linear power spectrum makes use of the HALO fit model (Smith et al., 2003). Linear and non-linear power spectra diverge at small scales, where the linear approximations are no longer valid. Baryon acoustic oscillations show up as 'wiggles' in the power spectrum, to the right of the turnover from the radiation dominated to matter dominated era.

scales, it is not valid on small scales. Fig. 2.7 shows that the linear and non-linear matter power spectra diverge on small scales (large k). To obtain an accurate non-linear matter power spectrum, an alternative method must be used, such as large N-body simulations (Smith et al., 2003).

The 'baryon acoustic oscillations', steming from the same origin as the acoustic peaks in the CMB angular power spectrum can be seen in fig. 2.7, they appear as small wiggles to the right of the turnover point. These baryon acoustic oscillations are washed out by non-linear structure growth at high k. However, baryon acoustic oscillations can still be seen quite distinctly at lower k, well within the regime which is well described by the linear power spectrum. At higher redshifts, the validity of the linear power spectrum extends to higher k.

The location of the peaks of the baryon acoustic oscillations in the matter power spectrum are directly related to the size of the sound horizon at recombination. The empirical relationship between the oscillations in the matter power spectrum, the acoustic peaks in the CMB power spectrum and the sound horizon at recombination is given by Blake & Glazebrook (2003):

$$\frac{P(k)}{P_{\mathrm{ref}}} = 1 + Ak \exp\left[-\left(\frac{k}{0.1h\mathrm{Mpc}^{-1}} \right)^{1.4} \right] \sin\left(\frac{2\pi k}{k_A} \right) \tag{2.110}$$

where A is an amplitude fitting parameter and P_{ref} is a smooth reference power spectrum with no baryons. Crucially, k_A is given by $k_A = \frac{2\pi}{r_s}$ where r_s is the sound horizon at recombination given by Eq. (2.96). The idea is that the co-moving separation of the acoustic peaks or acoustic oscillations in Fourier space act as a cosmological standard ruler. The values for k_A were fixed by the physics of the early universe, which is well understood.

In practice, in the work presented in chapter 7 the BAO constraint is included using the method suggested by Kessler et al. (2009a) based on Eisenstein et al. (2005) where the value A which is independent of the dark energy model is defined as:

$$A(z_1; w, \Omega_m, \Omega_\Lambda) = \frac{\sqrt{\Omega_m} H_0^{1/3}}{H(z_1)^{1/3}} \left[\frac{1}{z_1 \sqrt{|\Omega_\kappa|}} \mathrm{sinn} \left(|\Omega_\kappa|^{1/2} \int_0^{z_1} \frac{H_0}{H(z)} \mathrm{dz} \right) \right]^{2/3} \tag{2.111}$$

The SDSS LRG constraint can be included in the parameter inference analysis in the following way:

$$\chi^2_{BAO} = \left[\frac{(A(z_1; w, \Omega_m, \Omega_\Lambda) - 0.469)}{0.017} \right]^2 \tag{2.112}$$

The inclusion of this constraint from the BAO data is discussed in section 7.8 and the confidence intervals produced can be seen in fig. 7.17.

Chapter 3

Dark energy and apparent late time acceleration

Astrophysical observations suggest that the expansion of the Universe is accelerating. This cannot be explained by matter, either ordinary or dark since matter is gravitationally attractive and would lead to deceleration not acceleration. Instead, another explanation must be sought. There are four main categories of cosmological model offered in explanation as indicated in figure 1.1 and as listed below:

(i) Dark energy: An additional energy density which has a repulsive gravitational force, this category includes the cosmological constant model. Models in this category modify the right hand side of Einstein's equation Eq. (2.16).

(ii) Modified gravity: the laws of gravity need to be adjusted for different scales to match our observations. Models in this category modify the left hand side of Einstein's equation Eq. (2.16).

(iii) Void models: we as observers are at the centre of a large under density - this has the effect of making it appear as if the Universe is accelerating.

(iv) Backreaction: the appearance of acceleration is an artefact of the way in which we smooth and average the matter content of the Universe.

The distinction between (i) and (ii) above is a fairly arbitrary classification depending on which side of the Einstein equation is modified, dark energy or 'modified matter' models modify the right hand side and modified gravity models modify the left hand side (Amendola & Tsujikawa, 2010). In this chapter we shall outline briefly some of the theoretical ideas behind these different cosmological models. The research in this thesis focusses primarily on the ΛCDM model and its close cousin the wCDM model. Do we live in a ΛCDM Universe or not? However we also include a few brief comments on the alternative explanations for completeness. Two recent standard texts which give an excellent and comprehensive overview of dark energy are Amendola & Tsujikawa (2010); Ruiz-Lapuente, P. (2010), for review papers see for example Peebles & Ratra (2003); Tsujikawa (2010); Frieman et al. (2008).

3.1 Dark energy models

In dark energy models an additional energy density component is added to the total matter-energy content of the Universe. Phenomenologically, this additional energy density appears in the Friedmann equation in the following way:

$$H(z)^2 = H_0^2 \left(\Omega_m (1+z)^3 + \Omega_\gamma (1+z)^4 + \Omega_\kappa^2 (1+z)^2 + \Omega_{DE} \exp \left[\int_0^z \frac{1+w(z)}{1+z} \mathrm{dz} \right] \right) \tag{3.1}$$

where Ω_{DE} is the dark energy density today and $w(z)$ is the dark energy equation of state. In this thesis we describe dark energy models with $w(z) \equiv -1$ as ΛCDM or cosmological constant models; models with $w(z) \equiv w$ where w is some fixed number as wCDM models. More exotic dark energy models allow $w(z)$ to assume a range of different functional forms.

3.1.1 The cosmological constant and the energy density of the vacuum

Although convincing astrophysical evidence for the presence of a cosmological constant or ΛCDM model has only arisen since the late 1990s with the advent of the supernovae Ia data (Riess et al., 1998; Perlmutter et al., 1999b), theoretical ideas about the existence of a cosmological constant have been around for longer, in the form of ideas about the significance of the energy of the vacuum. Einstein's equation (2.16) states that the curvature of spacetime is governed by the total energy density, as expressed in the stress energy tensor. At first glance it would seem that in empty space, that is, in a vacuum, there would be no energy densities contributing to the stress energy tensor. However, ideas in quantum field theory suggest that the energy density of the vacuum is not in fact zero.

Even before the development of modern quantum field theory, as early as 1916, Nernst and later Pauli in 1920, postulated that the vacuum had an energy density due to the zero point energy of the electromagnetic field, which would have observable physical manifestations (Peebles & Ratra, 2003). In 1948 Casimir made a classical calculation to predict the change in the electromagnetic zero point energy (Casimir, 1948), this prediction was proven experimentally by Sparnaay in 1957 and has been subsequently confirmed by many modern experiments Bordag et al. (2001).

In quantum field theory, everything (except gravity) is described by all pervasive fields which are made up of a series of harmonic oscillators, which give rise to particles and interactions, or forces between particles. When the field is in its lowest energy state, all of the harmonic oscillators are in the ground state, however, it is not necessarily the case that the ground state energy of a harmonic oscillator should be zero. For example, as an oscillating particle is cooled, its oscillations will decrease in amplitude, but it can never be cooled to a point where it is completely stationary; Hesienberg's Uncertainty Principle forbids it. This residual energy which remains and allows the particle to move even in its lowest energy state is the zero point energy.

The energy of the vacuum, or the zero point energy, may be calculated by constructing the energy operator, or Hamiltonian of a field and applying it to the lowest energy state, the ground state. Here we will obtain an order of magnitude estimate for the zero point energy, by considering the energy of the ground state of a free scalar field.

Consider a scalar field, ϕ with Lagrangian, $\mathcal{L}$:

$$\mathcal{L}(\phi, \dot{\phi}) = \frac{1}{2} g^{\mu\nu} \partial_\mu \phi \partial_\nu \phi - m\phi\phi \tag{3.2}$$

the field can be expressed in Fourier space as:

$$\phi(x) = \int \frac{d^3k}{(2\pi)^3} \frac{1}{\sqrt{2\omega_k}} (a_{\bar{k}} e^{i\bar{k}\bar{x}} + a^\dagger_{\bar{k}} e^{-i\bar{k}\bar{x}}) \tag{3.3}$$

and the conjugate momentum, π, is:

$$\pi(x) = \frac{\partial \mathcal{L}}{\partial \dot{\phi}(x)} = \dot{\phi}(x) \tag{3.4}$$

The Hamiltonian, H is defined as:

$$H = \int d^3x (\pi(x) \dot{\phi}(x) - \mathcal{L}) \tag{3.5}$$

The Hamiltonian for the scalar field can be constructed by substituting (3.2), (3.3) and (3.4) into (3.5), to give:

$$H = \int \frac{d^3k}{(2\pi)^3} \omega_{\bar{k}} (a^{\dagger}_{\bar{k}} a_{\bar{k}} + \frac{1}{2}(2\pi)^3 \delta^{(3)}(0)) \tag{3.6}$$

when this operates on the ground state, the first expression in (3.6) vanishes, as by definition, $a_{\bar{k}}|0\rangle = 0$, leaving only:

$$H|0\rangle = \int \frac{d^3k}{(2\pi)^3} \omega_{\bar{k}} \frac{1}{2}(2\pi)^3 \delta^{(3)}(0) \tag{3.7}$$

which is the energy of the vacuum.

Most problems in quantum field theory are concerned with changes in energy, rather than absolute values of energy, hence this energy of the vacuum is generally set to zero and omitted from most calculations Peskin et al. (1996). However, in our investigation of the cosmological constant, we are dealing with a problem of gravity and as gravity 'sees' all energy, the absolute value of vacuum energy density becomes important, so it becomes necessary to evaluate (3.7).

The first problem with (3.7) is that the delta function evaluated at zero is infinite, which arises because the total energy of infinite space is being calculated. We proceed by dividing by the total volume, V, to get the energy density, ε_0:

$$\varepsilon_0 = \frac{H_{\text{vacuum}}}{V} = \int \frac{d^3k}{(2\pi)^3} \frac{\omega_{\bar{k}}}{2} \tag{3.8}$$

To obtain the contributions to the vacuum density from all of the possible modes, $\bar{k}$ it is necessary to integrate over all possible values of wavenumber, k:

$$\varepsilon_0 = \int_{-\infty}^{\infty} \frac{d^3k}{(2\pi)^3} \frac{1}{2} \sqrt{k^2 + m^2} \tag{3.9}$$

where $\omega^2_{\bar{k}} = \bar{k}^2 + m^2$. It is easiest to evaluate this integral by switching to spherical polar coordinates:

$$\begin{aligned} \varepsilon_0 &= \int_0^{\infty} \frac{dk}{(2\pi)^3} \frac{k^2}{2} \sqrt{k^2 + m^2} \int_0^{\pi} \int_0^{2\pi} \sin\theta d\theta d\phi & (3.10) \\ &= \int_0^{\infty} \frac{d^3k}{(2\pi)^3} \frac{k^2}{2} \sqrt{k^2 + m^2}\ 4\pi & (3.11) \end{aligned}$$

as in Weinberg (1989). This integral will diverge as k approaches ∞, i.e. at higher energies and short wavelengths we seemingly have the problem of ultraviolet divergence. However, it is not the case that quantum field theory is valid at arbitrarily high energies, indeed, there is an upper energy limit, beyond which ordinary quantum field theory is expected to break down. The actual upper limit where quantum field theory breaks down is unknown, but we shall label this upper limit as Λ. In the limit where $\Lambda >> m$, we may write:

$$\varepsilon_0 = \int_0^\Lambda \frac{dk}{(2\pi)^3} \frac{k^3}{2} \, 4\pi \tag{3.12}$$

$$= \frac{\Lambda^4}{16\pi^2} \tag{3.13}$$

Our main question now concerns assigning a suitable value for Λ. Weinberg (1989) suggests that the Planck scale would be an appropriate limit, such that $\Lambda = G^{-1/2}$, as both General Relativity and Quantum Field Theory are expected to break down at these energies.

Can this energy of the vacuum be identified as the 'dark energy' of cosmological observations? It happens that the energy density of the vacuum which is obtained by evaluating (3.12) at the Planck scale is $\sim$ 120 orders of magnitude greater than the current observed value of the dark energy density ρ_Λ. This discrepancy between the observed dark energy density obtained from cosmological observations and the vacuum energy density predicted by quantum field theory was discussed in some detail by Weinberg in 1989, who described it as the 'Cosmological Constant Problem' (Weinberg, 1989).

The main so-called 'problem' here really stems from our incomplete understanding of the relationship between gravity and quantum field theory. Quantum field theory has been derived with no reference to or description of gravity and yet we have attempted to insert it directly into Einstein's description of gravity, but it is not clear it is entirely correct to do so. The solution to this problem may well lie in a better description of the link between gravity and quantum field theory; the vacuum energy of a quantum field in the presence of gravity may turn out to be vanishingly small.

3.1.2 wCDM and more general dark energy models

Current observations are consistent with $w(z) = -1$ but only constrain w to within 10%, meaning that there is still room for dark energy models for which $w(z) \neq 1$. Here we give a brief mention of physical theories which give rise to dark energy equations of state for which $w(z) \neq 1$. If $w(z) \neq -1$, then what other values might it take, and what type of dark energy models could give rise to $w(z) \neq -1$? The purpose of introducing the dark energy fluid is to cause acceleration of the Universe, similar to the way in which inflation caused acceleration in the early Universe, albeit on a much shorter time scale. From our discussions on inflation in chapter 2, section 2.4, we know from Eq. (2.67) that $w(z) < -1/3$ in order to cause acceleration, hence $w < -1/3$ places an upper bound on the value w may take, and forms the upper bound for the prior on w in our work on Bayesian Doubt (chapter 6) and cosmological parameter estimation (chapter 7).

Scalar field or 'Quintessence' models of dark energy can be constructed using the same Lagrangian as given for the basic inflation model described by Eq. (2.68). The dark energy equation of state is given by the ratio of Eq. (2.70) and Eq. (2.71)

$$\begin{aligned} w &= \frac{p}{\rho} && (3.14)\\ &= \frac{\frac{1}{2}\dot{\phi}^2 - V(\phi)}{\frac{1}{2}\dot{\phi}^2 + V(\phi)} && (3.15)\\ &= \frac{\dot{\phi}^2/2V(\phi) - 1}{\dot{\phi}^2/V(\phi) + 1}. && (3.16) \end{aligned}$$

In this scenario, a slowly evolving field would give $\dot{\phi}^2/2V(\phi) << 1$, hence $w \approx -1$ and a very rapidly evolving field would give $\dot{\phi}^2/2V(\phi) >> 1$ hence $w \approx 1$ (Frieman et al., 2008). Of course models for which $w > -1/3$ are ruled out as they do not cause acceleration. This scalar field model has a lower bound on $w > -1$.

Although the scalar field model places a lower bound on $w > -1$, values for w below this are not ruled out observationally. Models which allow $w < -1$ are known as Phantom dark energy models, and they require more unorthodox Lagrangians, see for example Caldwell (2002); Kunz & Sapone (2006).

Once we have relaxed the condition that $w(z) = -1$, there is no special reason to constrain $w(z)$ to be a fixed number, but it is also not clear what form $w(z)$ should take. Several simple empirical parametrizations have been suggested for investigating a redshift dependent dark energy equation of state, the most common of these is the Chevalier-Polarski-Linder (CPL) formalism (Chevallier & Polarski, 2001; Linder, 2003). The CPL formalism allows for a constant component w_0 and a component which scales with redshift w_a.

$$w(z) = w_0 + w_a \frac{1}{1+z} \qquad (3.17)$$

This CPL parametrization is used in our work on Robustness described in chapter 8. Although our work on Bayesian Doubt and cosmological parameter inference is currently restricted to the investigation of fixed w models, the intention is to extend that work to investigate CPL models at a later stage.

In addition to quintessence and phantom models, there are a whole host of other dark energy models including k-essence models, Chameleon models, Holographic models, and unified dark energy dark matter models to name a few (for a comprehensive review see Amendola & Tsujikawa (2010)). In theory these models will all have a predictable different dark energy equation of state, however as we shall discuss in chapter 6 it is very difficult to constrain $w(z)$ using current observational data as even large variations in $w(z)$ only show up as small variations in the observables.

3.2 A note on Modified Gravity

The most closely related alternative explanations to dark energy models for explaining apparent late time acceleration are modified gravity models. Modified gravity models are not explicitly addressed in this thesis, but it is important to highlight their existence as real alternative to dark energy or 'modified matter' models. Modified gravity models give the appearance of late time acceleration by weakening gravity on large scales (Ruiz-Lapuente, P., 2010). As for dark energy, there are many different approaches to modifying gravity, including scalar-tensor theories such as the Brans-Dicke theory, $f(R)$ models and Brane world models. For reviews see for example Copeland et al. (2006); Clifton et al. (2011) in addition to the standard texts listed at the start of this chapter.

Distinguishing between dark energy and modified gravity models presents a particular observational and statistical challenge, the data sets and methods discussed in this thesis cannot make that distinction, and there are theoretical reasons why the supernovae data alone will never be able to discriminate. Differentiation between modified gravity and dark energy models requires the particular combination of baryon acoustic oscillation data and weak lensing data, a particular combination of astrophysical probes which will be investigated by the proposed EUCLID (Laureijs et al., 2009) and wFIRST (Green et al., 2011) missions.

3.3 Alternatives: Voids and the Backreaction

Completely different alternative explanations which do not require modifications to the Einstein equation, and require neither new physics nor new theories of gravity have also been put forward. Two of these theories are Void models, and explanations concerned with the backreaction.

Void models are based on the idea of an inhomogeneous Universe, which is less dense at the centre (the underdense 'void') and more dense at the edges. Typically these void models are based on a spherically symmetric Lemaître-Tolman-Bondi (LTB) Universe (Bondi, 1947; Tolman, 1934) with the observers placed close to the centre of the void, for a discussion of how void models produce the appearance of late time acceleration see for example Enqvist (2008). One of the main criticisms of void models are that they violate the Copernican principle, however this criticism may be side stepped to some degree by constructing a model in which there are many voids and that the void in which the observers find themselves is no different to any of the other voids. Void models of this type are known by their descriptive title as 'Swiss Cheese models', see for example Marra et al. (2007). Future observations of BAO data could rule out large scale inhomogeneities and void models, see Clarkson (2009) for one method.

In section 2.2, when describing the background cosmology, we presented the FRLW Universe as a smooth Universe - the Friedman equation and the Hubble rate were presented with the assumption that all the matter and energy in the Universe could be treated as if it were smoothed out and equally distributed throughout the whole Universe; in calculating the metric and expansion of the background Universe, an assumption is made that there is a uniform matter and density throughout the Universe. This is true for the early Universe at the end of inflation, but it is not true today. Today the Universe is only homogeneous on large scales, on smaller scales this is manifestly untrue: matter, ordinary and dark, forms clusters around galaxies, galaxy clusters and galaxy superclusters (Clarkson, 2009).

It has been argued that this process of averaging over all the matter and energy and then solving the Einstein equations for the homogeneous Universe is an incorrect approach that could cause us to draw wrong conclusions about the Universe we observe. Instead, it has been argued that a correct approach would be to solve the Einstein equations for the inhomogeneous Universe and then take the average. The discrepancy which arises between the two methods could explain why we see an apparent late time acceleration, without the need to introduce ark energy or modified gravity. Explanations concerning the correct methodology for averaging over the matter-energy content of the Universe are described as 'Backreaction' explanations, for a review of the Backreaction, see for example Rasanen (2010); Kolb et al. (2005)

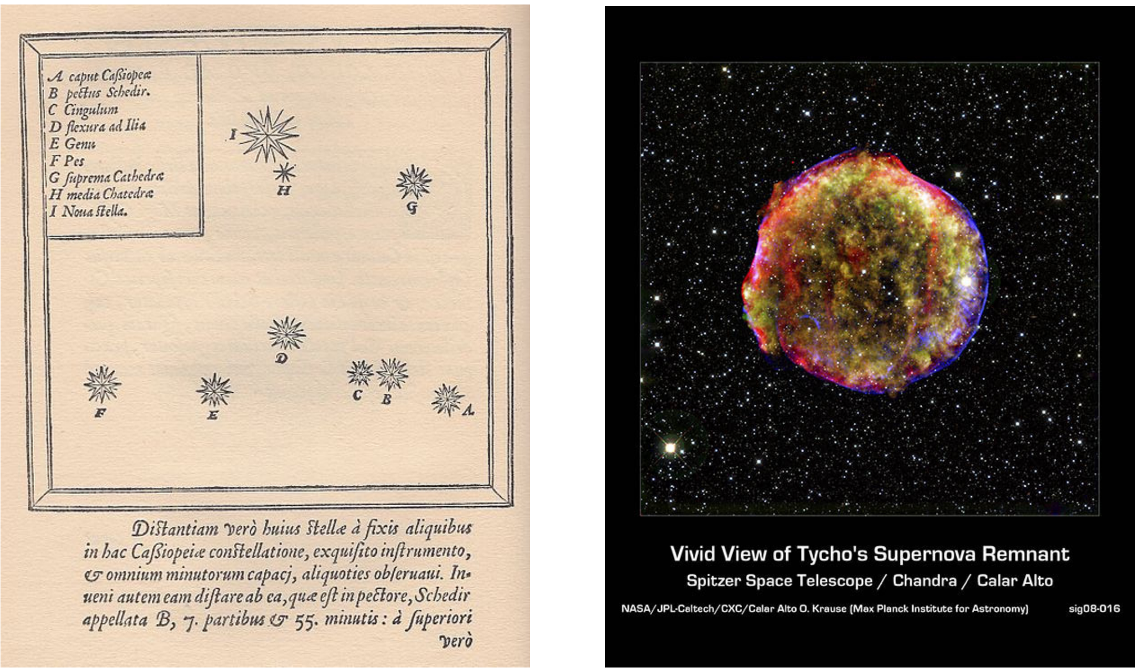
Chapter 4

Supernovae Ia

The purpose of this chapter is to give a brief introduction and background to the supernovae type Ia sufficient to put my work on cosmological parameter inference from SNe Ia into context.

Observations of supernovae have been recorded since ancient times, transient objects which flare brightly and appear as 'new stars' against the unchanging backdrop of fixed stars, before fading after a month or so. The term 'nova' was first used by the Danish astronomer Tycho Brahe (1546-1601) to describe the 'new star' which appeared in Cassiopeia on November 11th 1572, which he observed from Herrevad Abbey, Sweden. Brahe's sketch of the location of this nova is shown in the left hand panel of fig.4.1. Much later, Swiss and German astronomers Fritz Zwicky and Walter Baade differentiated between 'two well-defined types of new stars or novae which might be distinguished as common novae and super-novae'. Baade & Zwicky (1934); Zwicky (1940).

It is a special class of supernovae, supernovae type Ia (SNe Ia) which are useful to us for investigating dark energy, as they all have similar absolute magnitudes and can be made into standard candles by applying small empirical corrections based on their colour and the shape of their light curves. This correction process is described in section 4.5. Historically, SNe Ia were classified by their spectra and it was only later that physical models were developed to explain the different classes of observed spectra. Understanding the physics of the SNe Ia explosions is important as it will ultimately enable us to refine the methodology which we use to make the SNe Ia into standard candles. Currently this standardization process is essentially empirical, and lack of physical understanding limits our ability to improve cosmological parameter inference from SNe Ia. We shall briefly outline the spectral classification in section 4.1 and introduce some of the physics of SNe Ia models in sections 4.2 and 4.3.



Diſtantiam verò huius ſtellæ à fixis aliquibus in hac Caſsiopeiæ conſtellatione, exquiſito inſtrumento, & omnium minutorum capacj, aliquoties obſeruaui. Inueni autem eam diſtare ab ea, quæ eſt in pectore, Schedir appellata B, 7. partibus & 55. minutis: à ſuperiori verò

Figure 4.1: SN 1572: Tycho's SNe. The left hand panel show's Tycho Brahe's sketch included in *De nova stella* of the 'new star' he observed in Cassiopeia on 11th November 1572. The right hand panel shows a modern composite image of Tycho's supernovae remnant using optical and X-ray data. Tycho's SN 1572 has been spectroscopically confirmed as a type Ia supernova

4.1 Spectral classification of SNe Ia

Once spectral data became available for supernovae, these objects were subdivided into two types: supernovae type I which have no hydrogen H_α lines in their spectra, and supernovae type II which do have hydrogen H_α lines. The type I supernovae were further subdivided in to those with strong silicon Si II absorption lines ($\lambda = 6150\mathring{A}$) in their spectra, and those which do not, see fig. 4.2 for an illustration of the spectra from the different classes of supernovae. As cosmological distance indicators, it is the supernovae type Ia which command our attention. For a modern review see for example Hillebrandt & Niemeyer (2000); Howell (2011). Modern observations of the remnants of Tycho's supernova, SN 1572 have confirmed spectroscopically that SN 1572 was a type Ia supernova. (Krause et al., 2008).

The classification of supernovae has historically been made based on empirical observations of their spectra and the presence or absence of particular features in those spectra. However, we are also interested in the physical processes which give rise to these particular objects, as by better understanding the physics of these supernovae, we are better able to model their observed properties and make better

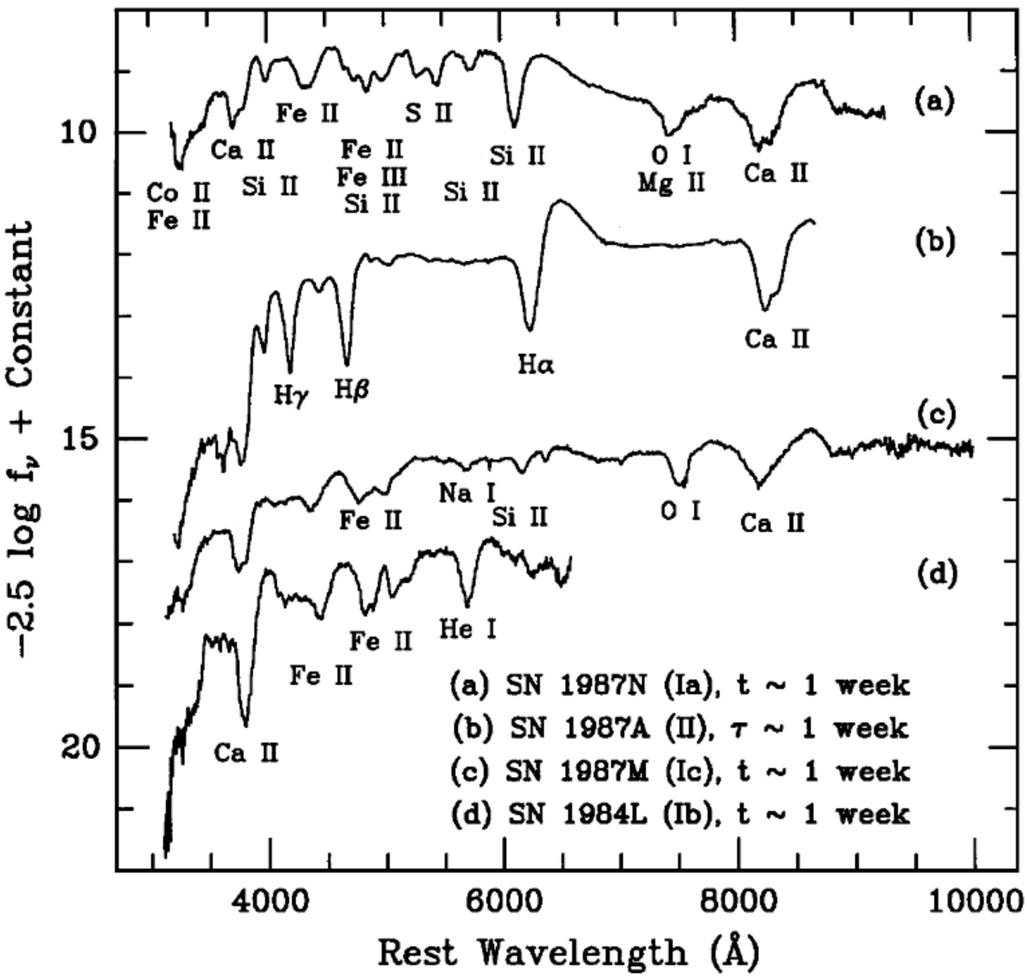


Figure 4.2: Figure taken from Filippenko (1997) showing differentiation of supernovae by their spectra. The supernovae type Ia spectrum is labelled (a), it is distinguished by an absence of hydrogen H_α features and the presence of a strong silicon Si II absorption line.

use of them as standardizable candles. Supernovae are not ‘new stars’, but rather they are explosive events that occur at the end of stellar life cycles when a star undergoes a catastrophic disruption. There are two principle categories of supernovae progenitors, which cut across the classification by spectral type:

(i) Core collapse supernovae (which produce Type II and Type Ib,c supernovae spectra) which occur when a massive ($M \gtrsim 8M_\odot$) star dies.

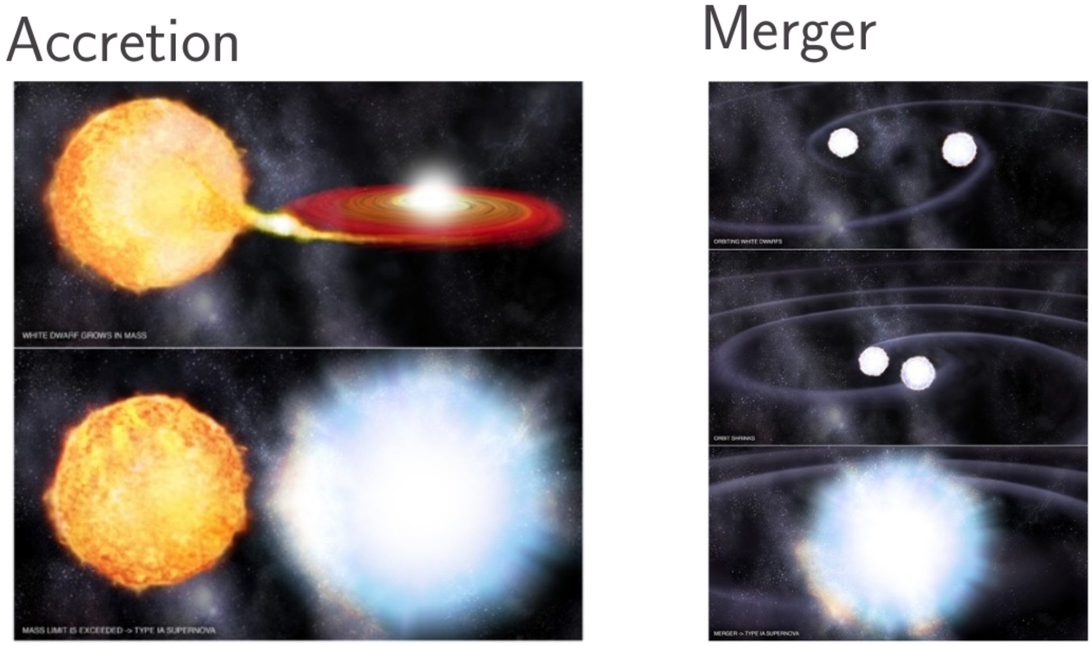
(ii) Supernovae produced from a degenerate star, which produce Type Ia supernovae spectra.

The useful property of supernovae type Ia is that they all have a similar absolute magnitude, making them ideal candidates for use as standardise-able candles. The absence of hydrogen features in their spectra suggests that the original progenitor star has already expended all or almost all of its hydrogen supply before exploding, and the presence of silicon is an indication that carbon burning has occurred. The similar luminosities of these objects suggest they had a similar mass when they exploded. A number of different progenitor systems have been proposed for the type Ia, all of which involve mass transfer in a binary system, in which a white dwarf gradually accumulates enough matter to just push it to the brink of the Chandrasekar limit of 1.4 Solar masses, ($1.4M_{\odot}$).

4.2 Progenitor scenarios for SNe Ia

The most popular scenario was originally the single degenerate accretion system, which consists of a binary pairing of a white dwarf and a red giant. The white dwarf is a star made up of carbon, nitrogen and oxygen and is prevented from collapsing under its own mass by the electron degeneracy pressure. This electron degeneracy pressure can prevent gravitational collapse for stars up to a mass of the Chandrasekar limit ($1.4M_{\odot}$), above this, the degeneracy pressure is insufficient, and the star will begin to collapse. In the single degenerate system, matter from the binary partner, in this case the red giant, gradually accretes on to the white dwarf, the mass of the white dwarf steadily increases until it approaches $1.4M_{\odot}$. At this point, the white dwarf begins to collapse under gravity and this collapse triggers carbon burning.

The principle alternative progenitor scenario is the double degenerate merger system, which begins with a binary pairing of two white dwarves which eventually merge and coalesce. The main difference between this and the single degenerate system is that the mass of the star at the point of explosion is much more variable. In the single degenerate system all of the progenitor stars go supernova at around $1.4M_{\odot}$ whereas in the double degenerate system the final mass may be considerably greater if two more massive white dwarves merge. The greater variation in final mass in the double degenerate scenario means a greater variation in fuel available and a greater variation in luminosity of the ensuing supernova explosion, which has consequences for the ability of SNe Ia produced in this way to act as standardizable candles. Recent work by Gilfanov & Bogdán (2010) based on X-ray observations suggests that the double degenerate merger scenario may be responsible for a much greater fraction of the SNe Ia than we first thought, which has implications for the way in which we use SNe Ia to infer the cosmological parameters.



Images:NASA/CXC/M Weiss.

Figure 4.3: Schematic showing the two most popular SNe Ia progenitor scenarios - both involve mass transfer in a binary system. The left hand panel shows the single degenerate accretion scenario in which a red giant or other main sequence star gradually accretes on to the white dwarf. The right hand panel shows the double degenerate merger scenario in which two white dwarfs merge.

4.3 Detonation and deflagration models

Early simulations of SNe Ia explosions focused on detonation (supersonic explosion) models, which produced the iron peak elements (Fe,Co,Ni) but were unable to reproduce other important characteristics of observed SNe Ia light curves. As an alternative, the Carbon deflagration model was put forward Nomoto et al. (1976), which describes a subsonic explosion. The most important of these early deflagration models was the W7 model of Nomoto et al. (1984). In the deflagration model the thermonuclear flame spreads more slowly through the white dwarf so more of the intermediate mass elements are produced (Si,S,Ca,Mg,O), which characterise SNe Ia spectra (see fig.4.2).

Neither detonation nor deflagration models alone can fully reproduce observed SNe Ia light curves, but combined models which begin with deflagration and then move to detonation have been more successful and are the current best physical explanation for SNeIa. Khokhlov (1991); Gamezo et al. (2005). The initial set up of the physical model can affect the subsequent SNe Ia light curve and spectrum in a number of different ways: The SNe Ia light curve is driven by the radioactive decay of ^{56}Ni into ^{56}Co then ^{56}Fe. The amount of ^{56}Ni produced in the explosion controls the luminosity of the SNe Ia Howell et al. (2009). The ^{56}Ni yield depends in turn on the initial mass of the progenitor (which can vary depending on whether it is a double degenerate or single degenerate system) and the metallicity of the progenitor. Deflagration produces less ^{56}Ni than detonation and so produces dimmer SNe Ia. Howell (2011). Other factors such as the location of the original deflagration flame, whether it is centrally located or off centre also affect the resulting SNe Ia light curve. Kasen et al. (2009).

This discussion on supernovae progenitors and physical models for deflagration and detonation serves to illustrate the point that there are many different physical factors which can affect the luminosity of SNe Ia and the properties of their observed light curves, but the light curve fitting techniques use very simple empirical approach which do not distinguish between the physical causes of the different light curve properties.

4.4 SNe Ia as standardizable candles

Standard candles are any class of object for which all members of the class have the same absolute magnitude M. The distance modulus $\mu(z)$ is the difference between the absolute magnitude M and the apparent magnitude $m_B(z)$ in the B-band

$$\mu(z) = m_B(z) - M \tag{4.1}$$

The distance modulus can be calculated from the luminosity distance D_L as follows

$$\mu = 5\log_{10}\left(\frac{D_L(z)}{1Mpc}\right) + 25 \tag{4.2}$$

where the luminosity distance is given by

$$D_L(z) = \frac{c(1+z)}{H_0\sqrt{-\Omega^0_\kappa}}\sin\left(\sqrt{-\Omega^0_\kappa}\int_0^z \frac{H_0}{H(z)}\mathrm{dz}\right) \tag{4.3}$$

and the Friedmann equation relates the cosmological parameters $\{H_0, \Omega^0_m, \Omega^0_\kappa, \Omega^0_\Lambda, w\}$ to Hubble rate $H(z)$

$$H(z)^2 = H_0^2\left(\Omega^0_m(1+z)^3 + \Omega^0_r(1+z)^4 + \Omega^0_\kappa(1+z)^2 + \Omega^0_{DE}\exp\left[3\int_0^z \frac{1+w(z)}{1+z}dz\right]\right) \tag{4.4}$$

The idea of standard candles is that since they all have the same absolute magnitude, they appear brighter or dimmer depending on their redshift and the cosmological parameters. If their redshifts are known, then the relative brightness or dimness of the standard candles can be used to derive the cosmological parameters.

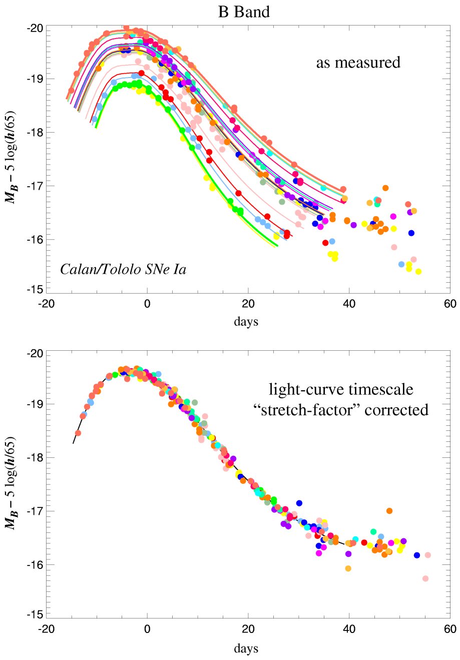
SNe Ia are not themselves perfect standard candles, as their absolute magnitudes vary somewhat (by ±0.8 mag in the B band) due to the physical reasons mentioned previously, and also due to photon absorption by dust. However, by applying small corrections for absolute magnitude, this scatter can be reduced considerably (to around ±0.15 mag in the B band) hence SNe Ia may be made into standard candles. How these corrections are made is the subject to the light curve fitting method used, as will be discussed section 4.5

4.5 Light curve fitting

The aim of SNe Ia light curve fitting is to be able to standardise supernovae absolute magnitudes in order that the supernovae can be used as standard candles.

SNe Ia are found by repeatedly imaging large areas of the same patch of sky every few days. Image subtraction is then carried out and any bright spots remaining become SNe Ia candidates. These candidates are then followed up spectroscopically to determine whether they are indeed SNe Ia. The apparent magnitude of the supernova in all pass-bands is recorded over several weeks, as for example in fig. 4.5. The redshift z of the supernovae is usually taken by measuring the redshift of the host galaxy.

After the initial explosion, the luminosity of the SNe Ia increases before reaching a maximum after around 10 days then turning over and declining. Phillips (1993) noticed that there was a tight correlation between the absolute magnitude of the supernova and the initial decline rate of the light curve after the maximum, supernovae with broader light curves and slower decline rates are brighter than supernovae with narrower light curves and a fast decline rate. Phillips (1993) also noted that the fastest declining light curves corresponded to the reddest supernovae and the slower declining light curves corresponded to the bluer supernovae. The 'Phillips corrections' refer to the original method of correcting supernovae magnitudes using the decline rate Δm_{15}. From this original concept, several different phenomenological methods for light curve fitting and supernovae standardization have been developed, the most popular methodologies currently in use are SALT-II (Guy et al., 2007) and MLCS2k2 (Jha et al., 2007), both of which have are motivated by slightly different philosophies, other light curve fitters include CMAGIC (Wang et al., 2003). The different methods of light curve fitting do give slightly different cosmological results, e.g. (Kessler et al., 2009a) and the relative merits of the various light curve fitters is a topic of much debate. My work on cosmological parameter inference presented in this dissertation is based exclusively on the SALT-II light curve methodology.



Kim, *et al.* (1997)

Figure 4.4: The upper panel shows the light curves from nearby supernovae, for which it can be seen there is considerable scatter $\sim \pm 0.8$mag in the maximum magnitude. The lower panel shows the same light curves after the application of the stretch correction, here the scatter in maximum magnitude is greatly reduced but not eliminated. Understanding this residual scatter will be one of the main motivations for my work presented in chapter 7

4.5.1 SALT-II Light curve fitting

The second Spectrally Adaptive Lightcurve Template (SALT) method for light curve fitting and cosmological parameter inference was the method originally used by Perlmutter et al. (1999a), the current version is SALT-II. SALT-II is essentially a two step process: first, the data points for each supernova are fitted to a light curve template to fit a model to the observed data (see for example fig.4.5). For SALT-II, the light curve template is ‘trained’ using nearby and distant SNeIa. Step one derives the following parameters for each supernovae, see for example Guy et al. (2005); Astier & Guy (2006):

(i) m^*_{Bi} The rest frame B-band maximum magnitude of the supernovae at maximum luminosity.

(ii) x_{1i} The stretch or shape parameter, which is related to the width of the fitted light curve.

(iii) c_i The colour parameter, which is the colour excess in B-band at maximum luminosity. $c = (B - V)_{Bmax} + 0.057$

The information we have at the end of the first step are estimates for the best fit values for m^*_{Bi}, x_{1i}, c_i and the covariance matrix $\hat{C}_i$ for each supernovae.

$$\hat{C}_i = \begin{pmatrix} \sigma^2_{m^*_B i} & \sigma_{m^*_B i, x_1 i} & \sigma_{m^*_B i, ci} \\ \sigma_{m^*_B i, x_1 i} & \sigma^2_{x_1 i} & \sigma_{x_1 i, ci} \\ \sigma_{m^*_B i, ci} & \sigma_{x_1 i, ci} & \sigma^2_{ci} \end{pmatrix}. \tag{4.5}$$

where $\hat{C}_i$ describes the covariances between the three fitted parameters from the light curve fitting.

In the second step, the free parameters α, β and the absolute magnitude M_0, are fitted simultaneously with the cosmological parameters, where α, β and M_0 are global parameters common to all SNe Ia and control how the stretch and colour corrections should be applied to make SNe Ia into standard candles. The SALT-II corrected version of 4.1 becomes (see Guy et al. (2007)):

$$\mu_i = m^*_{Bi} - M_0 + \alpha x_{1i} - \beta c_i \tag{4.6}$$

where $i = 1 \ldots N$ and N is the total number of SNeIa in the sample. An important conceptual point is that in the standard SALT-II literature the absolute magnitude of the supernovae always appears as a global parameter M_0 in Eq. (4.6), whereas in our Bayesian Hierarchical Model we present this differently, see chapter 7. The lack of uniformity in the supernovae absolute magnitudes has been corrected for by the application of the small correction terms αx_{1i} and βc_i. Notice here that regardless of whether the colour excess is due to host galaxy reddening or intrinsic colour variation of the SNe Ia, the colour excess is parametrized by a single parameter c_i, it is a purely empirical parametrization. Chapter 7 presents a detailed analysis of how the cosmological parameter inference step should be carried out.

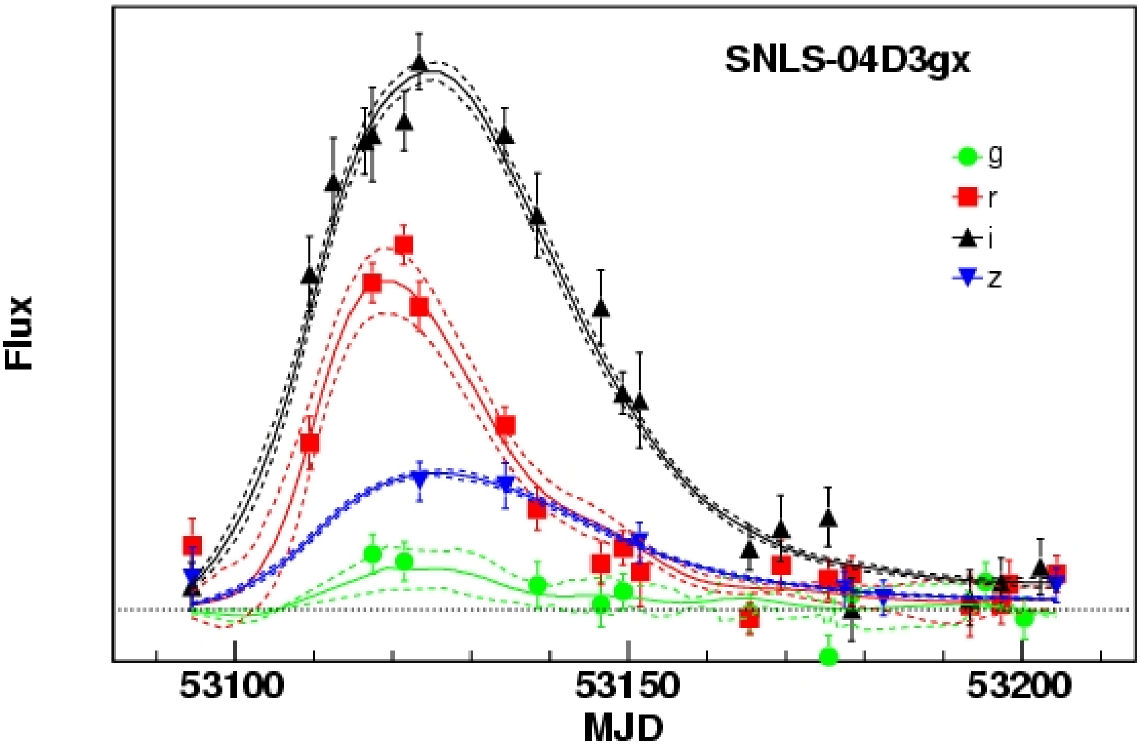


Figure 4.5: Example light curve from Guy et al. (2007). Solid points are the observed data and their error bars, solid lines are the best fit light curve model and the dashed lines are the 1σ uncertainties on that model

4.5.2 MLCS2k2 Light curve fitting

Multicolour Light Curve Shapes, MLCS, and the latest version, MLCS2k2 (Jha et al., 2007) is the major alternative to the SALT-II light curve fitting method, and MLCS was the method used by Riess et al. (1998) in one of the first papers to infer cosmological acceleration from the SNe Ia data. We do not go into detail discussing this method since it is not used in the work presented in this thesis, but we will highlight some of the differences with respect to the SALT-II method. The MLCS2k2 fitter is trained only on the nearby sample of SNE Ia, for which the distance moduli are known (unlike SALT-II which includes high-z SNe Ia in the training set). Whereas the first stage of the SALT-II light curve fitting process gives estimates for the stretch, colour and apparent B-band magnitude, MLCS2k2 gives a direct estimation of the distance modulus $\hat{\mu}$.

MLCS2k2 attributes all colour variation in the SNe Ia to host galaxy reddening, which is parametrized by A_v, and there is a strong prior which controls this parameter. Since dust from the host galaxy can only cause reddening of the SNe Ia, there is a sharp cut off on the A_v prior which forces $A_v > 0$ - this is in contrast to the SALT-II method which allows the colour parameter c to indicate a bluer SNe

Ia $c < 0$ (Kessler et al., 2009a). The shape of the light curve is parametrized by Δ (which is similar to the stretch parameter x_1 in SALT-II), and is the free parameter in the light curve fitting process.

Once the MLCs2k2 light curve fitting process has produced distance estimates, then the cosmological parameters are inferred by minimizing over the χ^2_μ which is given by:

$$\chi^2_\mu = \sum_i \frac{\left[\hat{\mu_i} - \mu(z_i; w, \Omega_m, \Omega_\Lambda, H_0)\right]^2}{\sigma^2_\mu} \tag{4.7}$$

where the theoretical μ is given by Eq. (2.59). There is a tension between the cosmological results produced from the MLCS2k2 method and the SALT-II method, Kessler et al. (2009a) trace discrepancy back to two causes: (i) A difference in the way the colour parameter is treated: SALT-II allows bluer SNe Ia, MLCS2k2 cuts them off with a strong prior. (ii) A discrepancy arising from problems with calibration of u-band data: because SALT-II uses both high and low redshift SNe Ia for calibration it is less sensitive to u-band calibration than MLCS2k2 which only uses the nearby sample. Currently it is not clear which method is more ‘correct’, it depends on a better understanding of the systematics and calibration techniques. One of the best ways of improving the standardization of SNe Ia is to include data from other wavelengths, most notably the near infra-red (NIR). For important work using NIR data to improve SNe Ia as distance indicators see Wood-Vasey et al. (2008); Mandel et al. (2011, 2009).

This chapter has given a brief overview and discussion of the current understanding of SNe Ia both in terms of their physical origins, and their observed properties. There is currently a gap between the physical understanding of SNe Ia and the way in which the observed SNe Ia light curves are fitted and standardized. Ultimately the ability of SNe Ia to be used for cosmological parameter inference will only be considerably improved when there is a better crossover between the physics of SNe Ia and the observed properties of their light curves. The work I shall present in chapter 7 describes a way of improving the cosmological parameter inference step of the SALT-II light curve fitter, and opens a way for investigating other properties of SNe Ia such as their evolution with redshift and correlation with their host galaxies.

Chapter 5

Statistical techniques

5.1 Elements of Bayesian Statistics

In statistics, there are two schools of thought about the exact nature of probability, one being the Frequentist school which defines probability as 'the ratio of the times the event occurs in a test series to the total number of trials in the series' (D'Agostini, 1995), or the 'frequencies of outcomes in random experiments' (Mackay, 2003) the other being the Bayesian school of thought which defines probability as 'a measure of the degree of belief that an event will occur' (D'Agostini, 1995). In my work in cosmological parameter inference and model selection I adopt a Bayesian approach, partly as it allows for a more holistic approach to solving problems involving probability, and partly as it allows us to explicitly include in the probability, information based on our prior experience and physical understanding of the situation. Rather than a collection of statistical tests, a Bayesian approach to probability provides a more flexible framework within which to construct solutions to problems of parameter inference and model selection. Standard texts describing Bayeisan statistical methods used in this research include Jaynes & Baierlein (2004); Gregory (2005); Sivia & Skilling (2006); Box & Tiao (1992) for Bayesian statistics in the cosmological context see especially Hobson, M. P., Jaffe, A. H., Liddle, A. R., Mukeherjee, P., & Parkinson, D. (2010); Trotta (2008); Loredo (1990)

Bayes theorem is named after the Rev. Thomas Bayes (1701?-1761) (Bellhouse, 2004), whose now famous theorem was published posthumously. In the simplest terms, Bayes theorem can be seen as a rearrangement of some of the basic rules in probability theory Riley (2004). Consider an event, A, whose probability of occurring is denoted by $p(A)$ and an event, B, whose probability of occurring is denoted by $p(B)$. The probability of both events A *and* B occurring is denoted by $p(A, B)$, whilst the probability of event A occurring *given that* event B has already occurred is shown by $p(A|B)$ and vice versa, $p(B|A)$. The product rule states that:

$$p(A, B) = p(B)p(A|B)$$

and

$$p(A, B) = p(A)p(B|A)$$

setting the two expressions equal gives:

$$p(B)p(A|B) = p(A)p(B|A)$$

which can be rearranged to give Bayes Theorem:

$$p(A|B) = p(B|A)\frac{p(A)}{p(B)} \tag{5.1}$$

5.1.1 Bayesian parameter inference

Eq. (5.1) is uncontroversial until event A is replaced by the unknown parameter of interest, θ and event B is replaced by the observed data, d and the question is then asked; 'What is the probability of the unknown set of parameters $\theta = \{\theta_1, \dots \theta_i \dots \theta_n\}$, given the data, d? The answer to this question is given by:

$$p(\theta|d) = p(d|\theta)\frac{p(\theta)}{p(d)} \tag{5.2}$$

where all we have done is re-labeled the quantities in Eq. (5.1).

5.1.1.1 Bayes Theorem

More usefully, we can write Eq. (5.1) being conditional on a particular model, $\mathcal{M}_j$, as follows:

$$p(\theta|d, \mathcal{M}_j) = \frac{p(d|\theta, \mathcal{M}_j)p(\theta|\mathcal{M}_j)}{p(d|\mathcal{M}_j)} \tag{5.3}$$

Eq. (5.3) is Bayes theorem as it is commonly written out for problems of parameter inference. $p(\theta|d, \mathcal{M}_j)$ is the joint posterior probability for the parameters, it is a function of θ and in Bayesian parameter inference is the quantity we are most interested in. $p(d|\theta, \mathcal{M}_j)$ is also a function of θ, it is the probability of the data given the parameters, it is known as the likelihood and often labelled $\mathcal{L}(\theta)$, it is the main quantity of interest in a Frequentist approach to parameter inference. $p(\theta|\mathcal{M}_j)$ is also a function of θ and is known as the prior. The prior is a function which describes our prior belief about where in parameter space we believe the true value of the parameter lies.

The slight controversy from the Frequentist point of view lies in the necessity of specifying the prior probability of the unknown parameter, $p(\theta)$, as it implies that some form of informed guess must be made, in advance of the collection of the observed data, about the plausible values of the unknown parameter. Priors can be a sensitive issue and must be carefully specified. However, the advantage of being able to explicitly specify priors and indeed one's prior belief about the parameters is that theoretical ideas, scepticism about the quality of the experiment or other considerations may be knowingly quantified. A well motivated prior means that we are making the best possible use of whatever prior information we have before gathering data. (See the following section for more discussion on the choice and role of priors)

The quantity $p(d|\mathcal{M}_j)$ is a single number obtained by integrating the product of the likelihood and prior over all of parameter space:

$$p(d|\mathcal{M}_j) = \int p(d|\theta, \mathcal{M}_j)p(\theta|\mathcal{M}_j)\ d\theta \tag{5.4}$$

In problems of parameter inference, this number $p(d|\mathcal{M}_j)$ can be thought of as a normalizing constant. $p(d|\mathcal{M}_j)$ goes by several names, it is most commonly known as the Bayesian evidence (or simply the evidence), or the model likelihood, as it gives the probability of the model given the data. For problems of parameter inference only, it is not necessary to compute the Bayesian evidence, as one may simply consider:

$$p(\theta|d, \mathcal{M}_j) \propto p(d|\theta, \mathcal{M}_j)p(\theta|\mathcal{M}_j) \tag{5.5}$$

but for problems of model selection, the Bayesian evidence plays a crucial role, as will be explained in following sections.

To obtain the posterior probability for a single parameter θ_k from the joint posterior probability distribution over all the parameters θ, one must integrate over all the other parameters $\theta_{\neq k}$ in the set:

$$p(\theta_k|d, \mathcal{M}_j) \propto \int p(d|\theta, \mathcal{M}_j)p(\theta|\mathcal{M}_j)\ \mathrm{d}\theta_{\neq \mathrm{k}} \tag{5.6}$$

this posterior probability for a single parameter is known as the marginalised posterior.

5.1.2 Priors and the Bayesian definition of a model

The prior quantifies our existing knowledge, and may be based on past experiments, or on theoretical predictions. There are three types of prior used in this thesis: flat, Gaussian and Jeffreys'. A flat (or uniform) prior represents a belief that all parts of parameter space within the prior range are equally probable, it is an appropriate choice of prior for location parameters. One (possible) disadvantage of a flat prior is that it definitively excludes some parts of parameter space - if that is not intended, one must ensure that the prior range is wide enough that the likelihood lies completely within the prior and is not truncated by the prior. We make frequent use of Gaussian priors in this work, and they have several advantages: They have long tails and so do not categorically exclude any part of parameter space; They can be easily adjusted to be broader or narrower, and they are easy to manipulate algebraically. A Jeffreys' prior is $1/\theta$ and equivalently is flat (uniform) in log space of the parameter. A Jeffreys' prior is used for range or scale parameters, representing indifference about the value of a range or scale parameter.

With less information, a broader prior should generally be used, although sometimes a prior may be specified with a sharp cut off to eliminate non-physical results such as negative masses. With more information, perhaps arising from a well tested theory, it may be advantageous to use a narrower prior which will prevail even in the case when poor data are used. The choice of prior can be controversial when the prior dominates over the likelihood, but this is usually a sign that either the data are insufficient to make a strong claim about the parameters of interest, or that an inappropriately strong prior has been chosen.

In the Bayesian sense, a specific model $\mathcal{M}_j$ is defined as being a particular function with a number of free parameters, and the priors on those parameters. The important point to note here is that within the Bayesian sense a model is not defined only by its free parameters, but also the priors on those parameters. If the priors are not specified, the model is not fully defined.

For an example in the cosmological context, consider that the standard model of cosmology today is the flat ΛCDM model, with its parameters. For ΛCDM to be fully defined as a model in the Bayesian sense, one would also need to specify the priors on those parameters, e.g. $p(\Omega_m) = 0.0 < \Omega_m < 2.0$. A different choice of prior on any of those parameters, either in form or in range would constitute a different model in the Bayesian sense, for example ΛCDM with a prior on Ω_m of $p(\Omega_m) = 0.0 < \Omega_m < 1.0$ would be a different model from the one just mentioned, and would be different again from a ΛCDM model with a Gaussian prior on Ω_m.

A model is said to be 'nested' within another model if the more complex model reduces to the simpler model when one of its free parameters is fixed. For example, curved ΛCDM with a curvature density parameter with a prior of $p(\Omega_\kappa) = -1.0 < \Omega_\kappa < 1.0$ reduced to flat ΛCDM when one of its free parameters Ω_κ is fixed to to zero $\Omega_\kappa \equiv 0$. Here the flat ΛCDM model is nested within the curved ΛCDM model.

5.1.3 Bayesian model selection

Thus far we have discussed Bayesian techniques for extracting model parameters for a specific model. However the question often arises as to which is the 'best' model to use; 'model selection' is a technique that can be used when we wish to discriminate between competing models and identify the best model in a set, $\{\mathcal{M}_1........\mathcal{M}_n\}$, given the data (see for example Trotta (2008)).

For a specific model, $\mathcal{M}_j$, with a specific choice of parameters and prior, Bayes theory states that:

$$p(\theta_j|\mathcal{M}_j, d) = \frac{p(d|\theta_j, \mathcal{M}_j)p(\theta_j|\mathcal{M}_j)}{\underbrace{p(d|\mathcal{M}_j)}_{\text{model evidence}}} \tag{5.7}$$

The denominator of (5.7) is known variously as the model likelihood, Bayesian evidence, model evidence or marginal likelihood and is given by:

$$p(d|\mathcal{M}_j) = \int p(d|\theta_j, \mathcal{M}_j)p(\theta_j|\mathcal{M}_j)d\theta \tag{5.8}$$

For a specific model, the model evidence acts as a normalisation constant, so for questions of parameter inference it was not necessary to compute it. However, it does become important in questions of model selection. Once the model likelihood (i.e. model evidence) has been calculated for each model in the set, then Bayes theorem can be applied to calculate the posterior probability $p(\mathcal{M}_j|d)$ of each model in the following way:

$$p(\mathcal{M}_j|d) = \frac{\overbrace{p(d|\mathcal{M}_j)}^{\text{model evidence}}\ p(\mathcal{M}_j)}{p(d)} \tag{5.9}$$

Where $p(\mathcal{M}_j)$ is the prior belief in the model and $p(d)$ is a normalisation constant given by:

$$p(d) = \sum_i p(d|\mathcal{M}_i)p(\mathcal{M}_i) \tag{5.10}$$

The relative 'goodness' of models is given by a comparison of their posterior probabilities, so to compare two models $\mathcal{M}_a$ and $\mathcal{M}_b$, we look at the ratio of the model

$\lvert \ln B_{ij} \rvert$	Odds	Strength of evidence
< 1.0	$\lesssim 3:1$	Inconclusive
1.0	$\sim 3:1$	Weak evidence
2.5	$\sim 12:1$	Moderate evidence
5.0	$\sim 150:1$	Strong evidence

Table 5.1: Empirical scale for evaluating the strength of evidence from the Bayes factor B_{ij} between two models (so–called 'Jeffreys' scale'). The right–most column gives our convention for denoting the different levels of evidence above these thresholds, following Gordon & Trotta (2007).

posterior probabilities:

$$\frac{p(\mathcal{M}_a|d)}{p(\mathcal{M}_b|d)} = \frac{p(d|\mathcal{M}_a)p(\mathcal{M}_a)}{p(d|\mathcal{M}_b)p(\mathcal{M}_b)} \tag{5.11}$$

The Bayes factor, B_{ab} is defined as the ratio of the model likelihoods:

$$B_{ab} = \frac{p(d|\mathcal{M}_a)}{p(d|\mathcal{M}_b)} \tag{5.12}$$

The Bayes factor gives a measure of the 'goodness' of a model, regardless of the prior belief about the model; the higher the Bayes number, the better the model. In many cases, the prior belief in each model in the set of proposed models will be equal, so the Bayes factor will be equivalent to the ratio of the posterior probabilities of the models. The relative strength of the Bayesian evidence between two competing models may be interpreted using the Jeffrey's scale, shown in table 5.1.

The advantage of using the Bayes factor as a marker to discriminate between models is that it incorporates information about both the quality of fit of the model to the data and the number of free parameters in the model; models with fewer free parameters are favoured, which accords with Occam's maxim that the simplest model should be prefered. The 'best' model in the Bayesian sense is the one which gives the best fit to the data with the smallest parameter space.

5.2 Numerical Methods

We turn our attention now to the practical numerical methods used when investigating problems of parameter inference and model selection, which we illustrate with a simple Gaussian linear model, $\mathcal{M}$ which relates the true x_i and y_i thus:

$$y_i = \theta_0 + \theta_1 x_i \tag{5.13}$$

where the model parameters $\theta = \{\theta_0, \theta_1\}$ are the parameters of interest to be inferred, and the observed values from n observations $\hat{y} = \{\hat{y}_1 \ldots \hat{y}_i \ldots \hat{y}_n\}$ are subject to a small amount of Gaussian noise of variance σ^2 such that:

$$\hat{y}_i \sim \mathcal{N}(y_i, \sigma) \tag{5.14}$$

For simplicity in this example there are no errors in the x direction and we write $x = \{x_1 \ldots x_i \ldots x_n\}$. The data $d = \{x, \hat{y}\}$ are shown in the upper left panel of figure 5.1, in which the straight line represents our model $\mathcal{M}$ for one choice of θ_0 and θ_1. If we wished to investigate whether the data were better fitted to a different model, $\mathcal{M}_{\text{new}}$ such as a low order polynomial, then that would be a problem of model selection rather than parameter inference. Bayes theorem for this problem can be written as:

$$p(\theta|d, \mathcal{M}) = \frac{\mathcal{L}(\theta)p(\theta|\mathcal{M})}{p(d|\mathcal{M})} \tag{5.15}$$

Our goal is to infer the most probable region for the parameters θ, given the data; this probability distribution is the joint posterior. Before looking at the data, a prior must be specified, which can be based on past experiments or theoretical predictions. In this example, we specify broad Gaussian priors on both parameters. The likelihood function $\mathcal{L}(\theta)$ also needs to be specified, it describes the probability of obtaining the data for given parameters and it depends on the experimental technique used to gather the data. In this case, the data were gathered subject to a small amount of Gaussian noise, so a Gaussian likelihood is appropriate.

$$\mathcal{L}(\theta) \equiv p(d|\theta) \tag{5.16}$$

$$= (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2}\sum_{i=1}^{n} \frac{(\hat{y}_i - y_i(\theta))^2}{\sigma^2}\right) \tag{5.17}$$

where

$$y_i(\theta) = \theta_1 \, x_i + \theta_0 \tag{5.18}$$

In the case of parameter inference, it is not necessary to calculate the Bayesian evidence, which here acts as a normalization constant, the posterior distribution is sufficiently specified by the numerator of Eq. (5.15). An analytic solution exists for this simple Gaussian linear model, however, for illustration, we calculate the required quantities numerically. As can be seen in the lower left panel of figure 5.1, the likelihood is strongly peaked and when this is multiplied by the broad prior, to give the posterior, the posterior is dominated by the data rather than by the prior. If the data had been less conclusive, then the prior would have had a greater

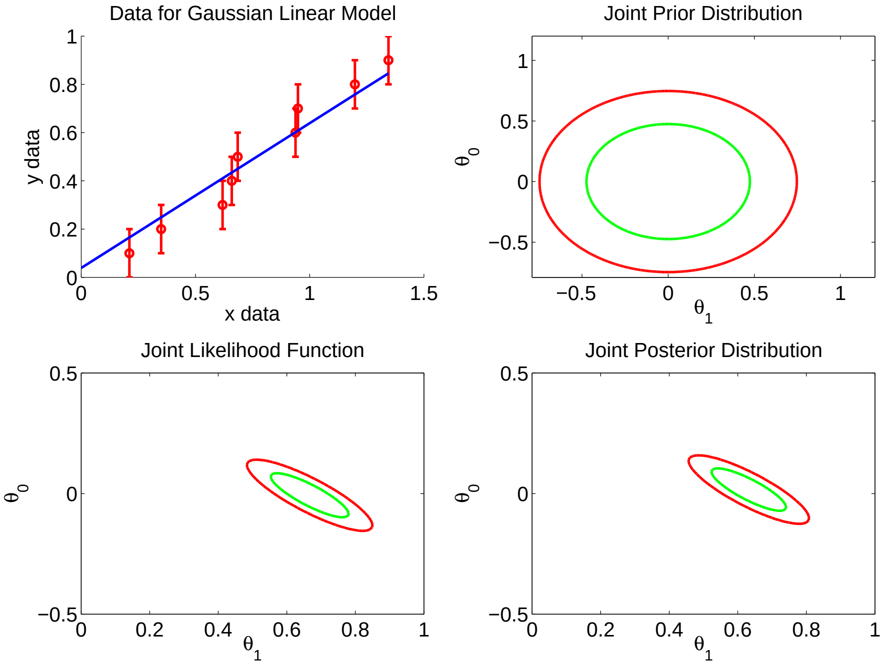


Figure 5.1: Upper left panel: points show experimental data, straight line shows model $\mathcal{M}$ for one choice of θ_0 and θ_1. We wish to infer the most probable values of θ_0 and θ_1, given the data - this is described by the posterior distribution, shown in the lower right panel. Red contours enclose 95% of the prior, likelihood and posterior probability; green contours enclose 68.3%.

influence in shifting the location of the posterior within parameter space. Likewise, if the prior had been more narrowly specified, it would have dominated over the data.

For this example, the posterior was calculated numerically for each point on a grid in parameter space. Although this is feasible for our simple model, it is not practical on grander scales for problems which have many more parameters and complicated likelihood functions. Parameter inference and model selection present two different numerical challenges: firstly, parameter inference requires the sampling of a potentially large parameter space in order to find the shape and peak of the joint posterior probability distribution; secondly, model selection requires the calculation of what can be a large multidimensional integral. For low dimensional problems with only one or two unknown parameters, these challenges become easy to solve, as the joint posterior may be evaluated over a one or two dimensional grid and the associated integral easily evaluated, as described in the preceding paragraphs. However, as the number of parameters increase, evaluating the posterior over a grid becomes increasingly unfeasible. Rather than calculating the posterior at every point in parameter space, it is more efficient to evaluate the posterior only at some points, concentrating on the regions around the peak of the posterior. One method for exploring the posterior in this way is to use a Markov Chain Monte Carlo (MCMC) method, for example the Metropolis-Hastings algorithm Metropolis et al. (1953); Hastings (1970), which is the MCMC algorithm used in this thesis; the other method used in this thesis to explore parameter space, which crucially also computes the Bayesian evidence, is the Nested Sampling technique (Skilling, 2004, 2006).

Metropolis-Hastings methods are useful for exploring parameter space and finding the peak whereas nested sampling techniques are primarily for calculating the Bayesian evidence but also explore parameter space and find the peak of the posterior. Alternative sampling techniques not used in this thesis include Gibbs Sampling, Importance Sampling, Rejection Sampling, Slice Sampling and Hamiltonian Monte Carlo techniques. Alternative techniques for calculating the Bayesian evidence include thermodynamic integration (simulated annealing) techniques, although these tend to be more computationally intensive than Nested Sampling. Here we shall give a brief overview of MCMC methods and Nested Sampling techniques which are used in the research presented in this thesis.

5.2.1 Markov Chain Monte Carlo (MCMC) techniques

In MCMC methods, a series of consecutive points (the ‘chain’) are produced in parameter space, with the density of the points being proportional to the posterior and converging to a stable distribution. We shall briefly describe here the Metropolis-Hastings algorithm and apply it to the Gaussian linear model described above. In the Metropolis-Hastings algorithm, a start point θ_s is chosen at random within a specified area of parameter space and the un-normalized posterior is evaluated at that point, P_s. A jump is then made in accordance with the proposal density (see later) to a candidate point, θ_c and the posterior at that point is evaluated, P_c. A decision is then made to either move to the candidate point and add it to the chain, or remain at the start point, again adding the start point to the chain and choose a different candidate point; the probability, α that the move is made to the candidate point is given by $\alpha = \min\left(\frac{P_c}{P_s}, 1\right)$ i.e. if $P_c > P_s$ then the jump is made, but if $P_c < P_s$ then the jump is less likely, but still possible. When the jump is made, the candidate point is added to the chain and becomes the new start point and the process is repeated. Allowing the jump to be made even to a point with a lower value of the posterior means that the tails of the distribution can be explored.

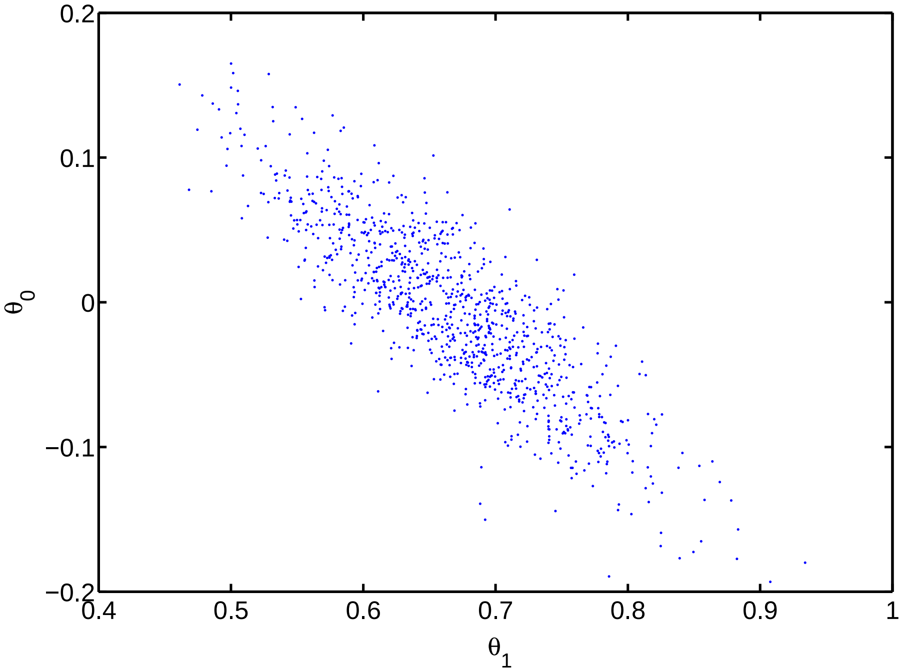


Figure 5.2: The application of MCMC methodology to the Gaussian linear model described by Eq. (5.13) and Eq. (5.14). The density of the points in parameter space corresponds to the posterior distribution. (Here the points are uniformly weighted) The first few points in the chain have been discarded as 'burn in' points and the chain has been 'thinned' by retaining only every 100th point.

How the candidate point is selected, is governed by the proposal density, which needs to be chosen with some care; in our example, we use a proposal density which is a Gaussian distribution centred on the location of the start point. This means that, depending on the width of the Gaussian, it is more likely that the candidate point will be quite near to the start point, but also allows the possibility of a candidate point being selected some distance away. If the width of the Gaussian is too small, the candidate will most likely be chosen very near to the start point, such that only a small region of the parameter space is investigated. If the width of the Gaussian is too broad, then the candidates will be selected far from the start point, in which case there will be little or no movement and it may take a long time for the chain to converge. The posterior distribution for a single parameter, i.e. the 'marginal distribution' may be obtained by integrating over the other parameters, for example, to obtain the un-normalised marginal posterior for θ_0, the numerator of (5.15) must

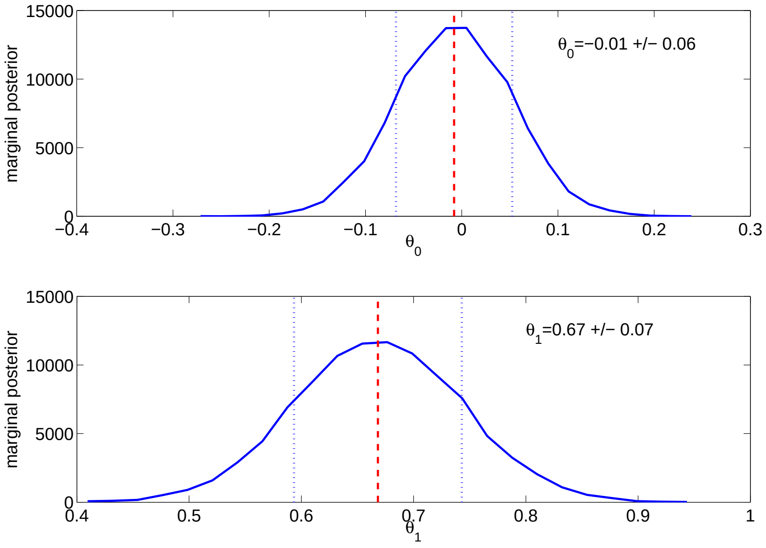


Figure 5.3: From the simple Gaussian linear model: The marginal posteriors for the two parameters are obtained by integrating out the unwanted parameter, which in practice means binning the points from the chain shown in figure 5.2. The dashed line indicates the mean of the distribution, the dotted lines indicate $\pm\sigma$

be integrated over with respect to θ_1:

$$p(\theta_0|d,\mathcal{M}) \propto \int \mathcal{L}(\theta)p(\theta|\mathcal{M})d\theta_1 \tag{5.19}$$

When the joint posterior distribution has been obtained by MCMC methods and is represented by the density of points in parameter space which form the 'chain', integrating over the unwanted parameter, θ_1 is simply a matter of dividing the range of the parameter of interest into a series of bins and counting the number of points that lie within each bin, since the density of samples is proportional to the posterior. Figure 5.3 shows the marginal posteriors of the two parameters.

5.2.1.1 MCMC Cosmological Parameter Inference in Practice: CosmoMC

Anthony Lewis and Sarah Bridle have developed an MCMC package, known as CosmoMC, for cosmological parameter inference (Lewis & Bridle, 2002), based on the Metropolis -Hastings algorithm described above. We include a brief description of CosmoMC as it is used for some of the numerical work in chapter 6. The MCMC sampler is a more sophisticated version of that described above using the example of the Gaussian linear model, and it has a cosmological interface for use in cosmological problems.

Within a specified cosmological model, e.g. ΛCDM, CosmoMC can be used to infer the various cosmological parameters, for a given data set such as WMAP (Dunkley et al., 2009). In the evaluation of the likelihood, CAMB (Lewis et al., 2000), itself based on CMBfast, (Seljak & Zaldarriaga, 1996), is used to calculate the theoretical model and flat priors are used throughout. CosmoMC produces chains which can be plotted and marginalised over in a similar way as the chains produced by our simple example code for our Gaussian linear model, to produce plots of the marginal posterior probabilities for the cosmological parameters, as shown in figures 5.4 and 5.5, using the supplied 'Getdist' package. (The slight difference in the implementation of the marginalization is that our basic example MCMC sampler produces unweighted chains, whereas CosmoMC produced weighted chains.) Thus CosmoMC provides a useful numerical method for Bayesian cosmological parameter inference, and for exploring the posterior. CosmoMC is used in our work on Bayesian Doubt in chapter 6 to explore a likelihood function and find its maximum.

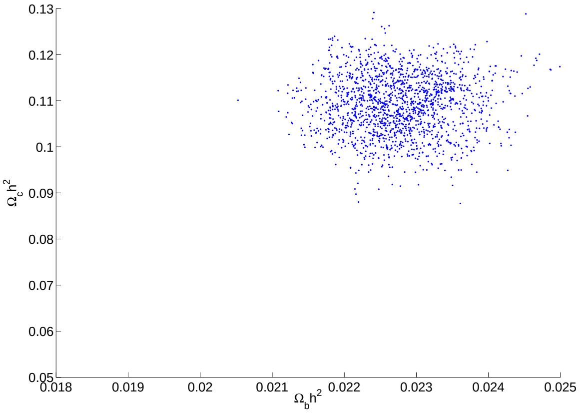


Figure 5.4: The application of MCMC methodology to WMAP data (Dunkley et al., 2009) using CosmoMC. As for figure 5.2, but here the samples are weighted. The weighted density of points in parameter space corresponds to the posterior distribution.

5.2.2 Nested Sampling

MCMC methods are useful for exploring uni-modal parameter spaces and doing parameter inference but less efficient when used to compute the Bayesian evidence.

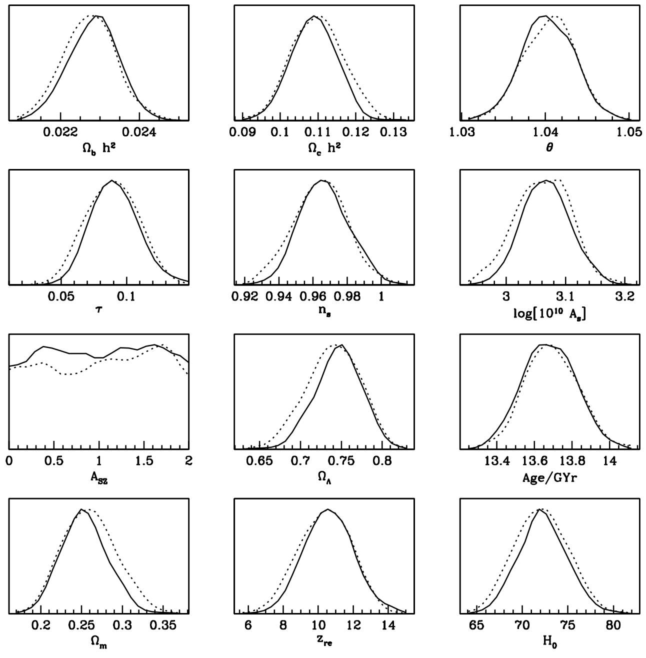


Figure 5.5: Cosmological parameter inference using WMAP data (Dunkley et al., 2009)and CosmoMC. 'Getdist' was used with the chains produced by CosmoMC shown in figure 5.4 to automatically produce plots of the marginal posteriors, using a similar technique to that demonstrated in figure 5.3, but taking into account the weight of the samples. Solid lines show marginal posterior, dotted lines show average likelihood.

Numerical calculations of Bayesian evidence in the cosmological context are often difficult and involve the calculation of a large multidimensional integral. Fortunately, several nested sampling codes have been developed based on the algorithm of Skilling (2004, 2006), such as MultiNest (Feroz & Hobson, 2008a; Feroz et al., 2009a) and CosmoNest (Mukherjee et al., 2006; Parkinson et al., 2006), these nested sampling algorithms compute the Bayesian evidence directly.

Rather than calculating the posterior first and then the evidence, Skilling's Nested Sampling algorithm aims for a direct calculation of the evidence and produces posterior samples as a subsidiary element. The methodology we briefly outline here is taken from Sivia & Skilling (2006) in which the the likelihood, $\mathcal{L}(\theta)$ and prior, $p(\theta)$ as are treated as inputs and the evidence, Z and posterior, $p(\theta|d)$ as outputs:

$$\mathcal{L}(\theta)p(\theta) = Zp(\theta|d) \tag{5.20}$$

and the evidence is defined in the usual way as:

$$Z = \int \mathcal{L}(\theta) p(\theta) d\theta \tag{5.21}$$

Skilling then asks the question: 'for a given value, λ of the likelihood, what proportion $\xi(\lambda)$ of the prior has a likelihood greater than λ?'. The calculation of this quantity is made easier if the values of the likelihood have already been sorted into order of decreasing magnitude. $\xi(\lambda)$ is defined formally as:

$$\xi(\lambda) = \int_{\mathcal{L}(\theta)>\lambda} p(\theta) d\theta \tag{5.22}$$

and we also define:

$$\mathcal{L}(\xi(\lambda)) \equiv \lambda \tag{5.23}$$

The evidence Z is then given by the area under the curve of a plot of $\mathcal{L}(\xi(\lambda))$ against $\xi(\lambda)$:

$$Z = \int_0^1 \mathcal{L}(\xi) d\xi \tag{5.24}$$

The trick here is in how to calculate the function $\xi(\lambda)$ given by Eq. (5.22) without having to calculate the likelihood and posterior over a grid of points in parameter space. The Nested Sampling method draws a number of random points with respect to the prior, initially across the whole range of $\xi(\lambda)$ (which can run from 0 to 1). The point within the set with the highest value of $\xi(\lambda)$ (corresponding to the lowest value of $\mathcal{L}(\xi)$) is then rejected and a new point is randomly selected on the condition that $\xi(\lambda)_{\text{new}} < \xi(\lambda)_{\text{rejected}}$ this new set of points covers a subset of the range of the old set of points; the rejection and selection process is repeated, each set of new points being a subset 'nested' within the previous set, gradually converging on the area of maximum likelihood. Mathematically, the effect has been to replace a potentially cumbersome multidimensional integral over a large parameter space Eq. (5.21) with a simpler one dimensional integral Eq. (5.24), which represents a great saving in computational power.

For solving cosmological problems, Multinest can be used as an alternative sampler within the CosmoMC package. As a natural by product, Multinest produces a set of samples from the posterior which can be processed by 'GetDist' to give the marginalised posterior distributions of the parameters in a similar fashion to that shown in figure 5.5.

The Multinest-CosmoMC configuration is used in our work on Bayesian Doubt in chapter 6 for computing Bayesian evidences. MultiNest (like the MCMC part of

CosmoMC) can also be used as an independent sampler, and it is in this configuration that MultiNest is used extensively in our work on parameter inference from the supernovae type Ia data described in chapter 7 - we chose to use MultiNest for this work as ultimately we wish to use the SNe I analysis to address problems of model selection.

5.3 Forecasting: Fisher matrix formalism

The background statistical ideas presented so far all concern the analysis and interpretation that takes place once the data have been gathered. A distinct but related question is how to plan the best data gathering strategy for future observation campaigns or astrophysical probes of cosmology, which includes the important question of how to choose the best metric for evaluating and comparing the expected returns for proposed strategies. A common metric for evaluating and comparing proposed probes is based on the Fisher matrix. Here we shall outline the basics of the Fisher matrix formalism, which is used in our work in chapter 8. we shall discuss some of the limitations of the Fisher matrix formalism and introduce a new statistic 'Robustness'.

The Fisher information matrix, named after R. A. Fisher (1890 - 1962) is a way of predicting the errorbars on parameters of interest given a proposed set of observations and expected errors on those observations. For a discussion of the Fisher matrix formalism in cosmology and a useful implementation, see for example (Hobson, M. P., Jaffe, A. H., Liddle, A. R., Mukeherjee, P., & Parkinson, D., 2010; Bassett et al., 2009a,b; Amendola & Tsujikawa, 2010). The Fisher matrix formalism approximates the Likelihood function by Taylor expanding the log likelihood $\ln \mathcal{L}(\theta)$ about the maximum likelihood:

$$\ln \mathcal{L}(\theta) \approx \ln \mathcal{L}_{max} - \frac{1}{2} \sum_{ij} \left. \frac{\partial^2 \ln \mathcal{L}(\theta)}{\partial \theta_i \partial \theta_j} \right|_{max} (\theta_i - \theta_i^{max})(\theta_j - \theta_j^{max}), \tag{5.25}$$

where $\mathcal{L}_{max}$ is the maximum likelihood and θ_i^{max} are the values of θ_i which maximize the likelihood. The Fisher matrix F_{ij} is then defined as the expectation value of the curvature of the log likelihood at the maximum likelihood, averaged over all the realizations of the data:

$$F_{ij} \equiv \left\langle \left. \frac{\partial^2 \ln \mathcal{L}(\theta)}{\partial \theta_i \partial \theta_j} \right|_{max} \right\rangle \tag{5.26}$$

The inverse of the Fisher matrix F_{ij} is the average expected covariance matrix. For Forecasting the Fisher matrix for future surveys, the Fisher matrix is evaluated at the point in parameter space corresponding to the fiducial model, which on average corresponds to the point in parameter space which gives the maximum likelihood.

2D contour plots can be made from the Fisher matrix showing the expected error ellipses on the parameters of interest, centred on some fiducial model. Figure 5.6 shows an example 95% error ellipse for the dark energy parameters of interest w_0, w_a. The figure of merit with which to compare future proposed astrophysical probes is usually related to the inverse area of the expected error ellipse derived from the Fisher matrix. The figure of merit chosen by the DETF is precisely 'the reciprocal of the area of the error ellipse enclosing the 95% confidence limit in the w_0, w_a plane.' (Albrecht et al., 2006). Large error ellipses give small figures of merit which indicate large uncertainties. Astrophysical probes with a large figure of merit are favoured as this indicates smaller expected errors on the parameters of interest. In chapter 8 we shall discuss some of the limitations of this inverse-area figure of merit and introduce a new statistic 'Robustness' as an extension to the inverse area figure of merit.

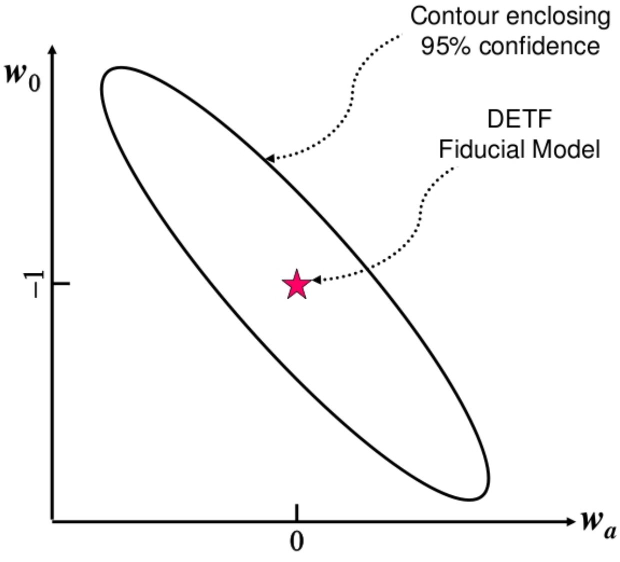


Figure 5.6: Example 95% error ellipse derived from the Fisher matrix for the dark energy parameters of interest, w_0, w_a taken from (Albrecht et al., 2006). The figure of merit is the inverse area of the 95% confidence level or error ellipse.

5.4 Some aspects of Frequentist statistics

Although the work presented in this thesis is based primarily on a Bayesian approach to probability, we make some reference to Frequentist methods for comparison especially in the work on parameter inference from supernovae type Ia discussed in chapter 7. In this section I will outline some of the ideas and concepts encountered in the Frequentist approach, and explain where appropriate how it relates to the Bayesian approach. Texts with sections on Frequentist approaches referenced for this section include Williams (2001); Frodesen et al. (1979); Young & Smith (2005). As mentioned in the introductory remarks to this chapter, one of the main differences between the Frequentist and Bayesian approaches are the differing concepts of probability. For the Bayesians, probability is a measure of their degree of belief about a particular statement, for the Frequentists, it concerns the number of expected outcomes in a series of repeated tests. The questions asked in both cases are different, and consequently the answers obtained differ in their meaning. Frequentists are primarily concerned with the likelihood $p(d|\theta)$ which is the probability of the data d given the parameters θ (or the *likelihood* of the parameters θ). In the Frequentist sense, the best estimator for a parameter, is the value of the parameter θ_{max} which maximises the likelihood. Along with the maximum likelihood, the 1-dimensional profile likelihood is often given, which is the maximum likelihood at every fixed value of the parameter of interest. Instead of the Likelihood, Bayesians are primarily interested in the posterior $p(\theta|d)$ which is the probability of the *parameters* θ given the data d, and this should be the quantity of interest in cosmology - we are interested in the probability of the cosmological parameters $\{\Omega_m, \Omega_\Lambda, \Omega_\kappa, H_0, w\}$ given the data, not vice versa.

The Frequentist quantity of interest $p(d|\theta)$ and the Bayesian quantity of interest $p(\theta|d)$ are related through the Bayes equation, as in Eq. (5.3). For problems of parameter inference, the normalizing Bayesian evidence may be neglected, and for the case where the priors are flat and fully enclose the likelihood, then the posterior is proportional to the likelihood

$$p(\theta|d) \propto p(d|\theta). \tag{5.27}$$

In this case, both the posterior and the likelihood peak in exactly the same place in parameter space, here both methods in some sense give the ‘same’ answer, but to two slightly different questions.

5.4.1 Maximum Likelihood and the Chi square statistic

Frequentists seek to identify the maximum likelihood estimator (MLE), which maximises the likelihood $\mathcal{L}(\theta)$. For the example of the Gaussian linear model described above, this means maximizing Eq. (5.17), which may be re-written as:

$$\begin{aligned}\mathcal{L}(\theta) &\equiv p(d|\theta) && (5.28)\\ &= (2\pi\sigma^2)^{-n/2}\exp\left(-\frac{1}{2}\sum_{i=1}^{n}\chi^2\right) && (5.29)\end{aligned}$$

where the random variable χ^2 has been defined as

$$\chi^2 \equiv \sum_{i=1}^{n}\frac{(\hat{y}_i - y_i(\theta))^2}{\sigma^2}. \tag{5.30}$$

Here, adjusting θ so as to minimize the χ^2 random variable will maximize the likelihood. We label the value of θ which gives the maximum likelihood $\mathcal{L}_{max}$ as θ_{max}. In this Gaussian linear example, maximization of the likelihood can be carried out neglecting the prefactor $(2\pi\sigma^2)^{-n/2}$ as it is constant and acts as a normalization factor, it does not change the value of θ_{max}. I emphasize here that in this example minimizing the χ^2 maximizes the likelihood, as the prefactor may be neglected - this becomes important for our work on the SNe Ia analysis in chapter 7 where the prefactor cannot be neglected.

When the data $\hat{y}_i$ are independent and identically distributed, and their errors are Gaussian, then the random variable χ^2 itself follows a chi-square distribution χ^2_ν where ν is the number of degrees of freedom given by $\nu = n - m$, where n is the number of data points and m is the number of free parameters, which is Wilks' theorem (Wilks, 1938). The mean of the χ^2_ν distribution is located at ν. When the theoretical model is a good fit for the data, the value for the χ^2 random variable is expected to fall close to the mean value of the chi-square distribution χ^2_ν, giving rise to the popular criteria that χ^2/d.o.f. ≈ 1 gives a good indication that the model is a good fit to the data, and that χ^2/d.o.f. $>> 1$ indicates that the model should be rejected. For a discussion of some of the problems with the chi-square goodness of fit test in specific scenarios, see Protassov et al. (2002).

In chapter 7 section 7.3 we shall look at an example where the χ^2_μ statistic as defined in Eq. (7.6) will not follow a chi-square distribution χ^2_ν because χ^2_μ in Eq. (7.6) is of a different form to the χ^2 statistic defined in Eq. (5.30). The chi-square statistic as defined in Eq. (5.30) only has unknown parameters in its numerator, whereas χ^2_μ statistic defined in Eq. (7.65) has a parameter dependence in both the numerator and denominator - hence we do not expect it to follow a chi-square distribution χ^2_ν, and cannot use the approximation that χ^2_μ/d.o.f. ≈ 1 is the mark of a good model.

Chapter 6

Bayesian Doubt: Should we doubt the Cosmological Constant?

6.1 Introduction

One of the most fundamental questions in cosmology today is to identify which cosmological model best describes the Universe in which we live, given the current observational data. The standard model of cosmology describes a hot big bang scenario followed by an expanding Universe. But the standard model of cosmology is not yet finalised, and there are several open questions, most notably, we desire to know which is the best explanation for the apparent late time acceleration which we seem to observe.

Observations of distance supernovae type Ia made in the late 1990s appear to show that the expansion of the Universe is accelerating. Is the Universe really accelerating or does it only appear to be doing so? Models which only give the appearance of acceleration, rather than actual acceleration include void models and alternative explanations of the backreaction. If the Universe is genuinely accelerating, then what is the cause of this acceleration? Ordinary matter, dark or otherwise cannot explain an acceleration in the expansion, being gravitationally attractive, matter may only explain a deceleration in the expansion rate. Instead, alternative theories must be postulated in explanation, most alternative theories fall into two categories: either they are modified gravity models, or they are dark energy models which posit an additional component which has a negative pressure. The cosmological constant model is a special case of a dark energy model for which the dark energy equation of state is fixed at precisely $w = -1.0$.

One way for discriminating between competing models of the Universe is to use

Bayesian model selection, the Bayesian evidences for each of the proposed models can be computed and the models can then be ranked by their Bayesian evidence. This is a good method for identifying which is the best model in a given set of models, but it gives no indication of the absolute goodness of the model. Bayesian model selection can identify the best model in a given set, but it says nothing about the overall quality of the set of models as a whole - the best model in the set may merely be the best of in a set of poor models.

Within the Bayesian model selection framework there is no mechanism for describing the absolute goodness of a model, however, it would be useful to have some method for quantifying just how good a model is. Knowing that the best model in the current set of models, while being better the known alternatives, is not particularly good model would provide motivation to search for a better model, and hence may lead to model discovery.

One way of assigning some measure of the absolute goodness of a model is to use the concept of Bayesian doubt, first introduced by Starkman et al. (2008). Bayesian doubt works by comparing all the known models in a set with an idealized model, which acts as a benchmark model. In this chapter I review the concept of Bayesian doubt and apply it to cosmological model selection with regard to the dark energy equation of state. This chapter follows closely the work presented in the paper March et al. (2011a)

6.2 Bayesian doubt and model discovery

6.2.1 Bayesian doubt described

In standard Bayesian model selection, for a given set of N known models, $\{\mathcal{M}_1 \dots \mathcal{M}_i \dots \mathcal{M}_N\}$; the Bayesian evidences of each model may be computed and the models ranked accordingly:

$$p(\mathcal{M}_i|d) = \frac{p(d|\mathcal{M}_i)p(\mathcal{M}_i)}{p(d)}$$

where $p(\mathcal{M}_i)$ is the model prior, $p(d)$ is the probability of the data acts as a normalizing constant. $p(\mathcal{M}_i|d)$ is the posterior probability of the ith model and $p(d|\mathcal{M}_i)$ is the model likelihood or model evidence, given by the integral over all of parameter space, θ_i for the parameters within that model:

$$p(d|\mathcal{M}_i) = \int \mathrm{d}\theta_\mathrm{i} \mathrm{p}(\mathrm{d}|\theta_\mathrm{i}, \mathcal{M}_\mathrm{i})\mathrm{p}(\theta_\mathrm{i}|\mathcal{M}_\mathrm{i}) \tag{6.1}$$

$p(d|\theta_i, \mathcal{M}_i)$ is the likelihood of the parameters given the data for model $\mathcal{M}_i$ and $p(\theta_i|\mathcal{M}_i)$ is the prior on the parameters with in model $\mathcal{M}_i$. But what if the set of models is incomplete and does not contain the true model? $\{\mathcal{M}_1 \ldots \mathcal{M}_i \ldots ??? \ldots \mathcal{M}_N\}$. The methodology of Bayesian doubt dictates that an unknown idealized model X should be introduced against which the other models may be compared. Following Starkman et al. (2008) 'doubt' may be defined as the posterior probability of the unknown model:

$$\mathcal{D} \equiv p(X|d) = \frac{p(d|X)p(X)}{p(d)} \tag{6.2}$$

where $p(X)$ is the prior doubt, i.e. the prior on the unknown model, which represents the degree of belief that the list of known models does not contain the true model. The sum of all the model priors must be unity, so we can write:

$$\sum_{i=1}^{N} p(\mathcal{M}_i) + p(X) = 1. \tag{6.3}$$

If we assign equal prior probability to each of the known models within the set, then we may also write that the prior on each of the known models is:

$$p(\mathcal{M}_i) = \frac{1}{N}(1 - p(X)). \tag{6.4}$$

The probability of the data $p(d)$ is the product of the model likelihood and model prior, summed over all models:

$$p(d) = \sum_{i=1}^{N} p(d|\mathcal{M}_i)p(\mathcal{M}_i) + p(d|X)p(X) \tag{6.5}$$

Substituting for $p(d)$, doubt, $\mathcal{D}$ may be expressed as:

$$\begin{aligned} \mathcal{D} &= \frac{p(d|X)p(X)}{\sum_{i=1}^{N} p(d|\mathcal{M}_i)p(\mathcal{M}_i) + p(d|X)p(X)} && (6.6) \\ &= \left(1 + \frac{\sum_{i=1}^{N} p(d|\mathcal{M}_i)p(\mathcal{M}_i)}{p(d|X)p(X)}\right)^{-1} && (6.7) \end{aligned}$$

We can define the Bayes factor as:

$$B_{ij} \equiv \frac{p(d|\mathcal{M}_i)}{p(d|\mathcal{M}_j)} \tag{6.8}$$

The methodology of Bayesian doubt requires a baseline model, for which we will chose to use ΛCDM. The average Bayes factor between ΛCDM and each of the known models is given by:

$$\langle B_{i\Lambda} \rangle \equiv \frac{1}{N} \sum_{i=1}^{N} B_{i\Lambda}. \tag{6.9}$$

[N.B. the sum over the models $\mathcal{M}_i$ includes $i = \Lambda\mathrm{CDM}$ and as such $\langle B_{i\Lambda} \rangle \geq 1$] We may re-write Eq.(6.7) in terms of the ΛCDM model:

$$\mathcal{D} = \left(1 + \frac{\langle B_{i\Lambda} \rangle}{B_{X\Lambda}} \left(\frac{1 - p(X)}{p(X)} \right) \right)^{-1} \tag{6.10}$$

As well as looking at posterior doubt $\mathcal{D}$, we may also look at the ratio between the posterior doubt and prior doubt, which we call the relative change in doubt, $\mathcal{R}$.

$$\begin{aligned} \mathcal{R} &\equiv \frac{\mathcal{D}}{p(X)} && (6.11) \\ &= \left(p(X) + (1 - p(X)) \frac{\langle B_{i\Lambda} \rangle}{B_{X\Lambda}} \right)^{-1} && (6.12) \end{aligned}$$

For doubt to grow, i.e. the posterior doubt to be greater than the prior doubt ($\mathcal{R} \ll 1$), then the Bayes factor between the unknown model X and the baseline model, in this case ΛCDM, must be much greater than the average Bayes factor:

$$\frac{\langle B_{i\Lambda} \rangle}{B_{X\Lambda}} \ll 1, \tag{6.13}$$

To genuinely doubt the baseline model, ΛCDM, it is not sufficient that ($\mathcal{R} > 1$), but additional, the probability of ΛCDM must also decreace such that its posterior probability is greater than its prior probability, i.e. $p(\Lambda|d) < p(\Lambda)$. We can define:

$$\begin{aligned} \mathcal{R}_\Lambda &\equiv \frac{p(\Lambda|d)}{p(\Lambda)} && (6.14) \\ &= \left((1 - p(X)) \langle B_{i\Lambda} \rangle + p(X) B_{X\Lambda} \right)^{-1}. && (6.15) \end{aligned}$$

For ΛCDM to be doubted, the following two conditions must be fulfilled:

- $\mathcal{R} > 1$
- $\mathcal{R}_\Lambda < 1$

If these two conditions are fulfilled, then it suggests that the set of known models is incomplete, and gives motivation to search for a better model not yet included considered, which may lead to model discovery.

6.2.2 Bayesian evidence of the unknown model

One of the most important practical considerations in implementing the doubt formalism is how to estimate or select the Bayesian evidence (or model likelihood) $p(d|X)$ of the unknown model. By definition, the unknown model is not fully specified, hence $p(d|X)$ cannot be computed directly. Instead, a suitable value for $p(d|X)$ must be chosen, it should be representative of a ‘good’ Bayesian model, in as much as it gives a good fit to the data without being penalized by having a large number of free parameters. However, it must be based on a realistic but not perfect model, otherwise doubt will always grow. In this work, we choose to use an approach which uses the absolute upper bound of the model evidence $p(d|X)$. Using the absolute upper bound of the model evidence is a context specific approach, so I shall describe the context and data sets used before going on to describe how to compute the absolute upper bound of the evidence.

In this work, we are specifically investigating the dark energy component of the cosmological model. Our question is whether we should doubt the ΛCDM model, for which the dark energy equation of state is precisely -1, i.e for the ΛCDM model $w(z) \equiv -1.0$. Given the data from the standard astrophysical probes of dark energy, namely the cosmic microwave background (CMB), matter power spectrum (mpk) and supernovae type Ia (SNeIa), the ΛCDM model is currently the best model in the set of known models which is why we choose it to be the bench mark model.

In the context of exploring the dark energy sector, we are comparing ΛCDM with alternative models for $w(z)$, we can propose a number of different models, (e.g. the most basic would be $w(z) \neq -1.0$, but some other fixed number) which would constitute our set of known models. However, the point of Bayesian doubt is to investigate whether the set of known models is indeed complete, or whether another unknown model could exist. We seek a phenomenological description for the unknown model X for $w(z)$ which will provide a high degree of flexibility without incurring a high Occam’s Razor penalty for having a high number of free parameters.

In general, the maximum Bayesian evidence for a model is given when the priors on the parameters θ are delta functions co-located exactly at the peak of the likelihood function for those parameters, $\theta_{\max}$, such that the prior is $p(\theta = \delta(\theta - \theta_{max})$. These priors are not well motivated in the ordinary Bayesian sense, as they can only be applied aposteriori, however, they do delimit the absolute upper bound that may be placed on the Bayesian evidence. For two nested models, for which the simpler model $\mathcal{M}_0$ can be obtained from the more complex model $\mathcal{M}_1$ by fixing one of the parameters θ in the complex model, to a value θ^*, the maximum improvement in the Bayes factor $\bar{B}_{10}$ between the more complex and the simpler model is given by: (see Gordon & Trotta (2007))

$$\bar{B}_{10} = \frac{p(d|\theta_{max}, \mathcal{M}_1)}{p(d|\theta^*, \mathcal{M}_0)} \tag{6.16}$$

In the case we are considering here, the more complex model is the unknown model X which allows $w(z)$ to vary, and the simpler model nested within it is ΛCDM which has $w(z)$ fixed. The parameter likelihood is given by: $p(d|\theta, \mathcal{M}) = \exp -\frac{1}{2}\chi^2$, so we can write the maximum upper bound on the Bayes factor between the unknown model X and ΛCDM as:

$$B_{X\Lambda} < \bar{B}_{X\Lambda} \tag{6.17}$$

$$B_{X\Lambda} < \exp\left(-\frac{1}{2}\left(\chi^2_X - \chi^2_\Lambda\right)\right) \tag{6.18}$$

In order to evaluate the upper bound $B_{X\Lambda}$ it is sufficient to calculate the difference $\Delta\chi^2 = \chi^2_X - \chi^2_\Lambda$ from the best fit χ^2 for both the ΛCDM and X models. The question of how to obtain χ^2_X is context specific, and will be discussed in the next section.

6.2.3 Behaviour of doubt and the posterior probability for ΛCDM

In this subsection, we shall consider the behavior of doubt and the posterior probability of ΛCDM in three cases, which we shall describe in terms of the Bayes factors and the absolute upper bound on the Bayes factor between the unknown model and ΛCDM model. For a fixed choice of prior doubt $p(X)$ we can see that from Eqs. (6.10) and (6.14) the two factors which control the relative change in doubt $\mathcal{R}$ and the relative change in the probability of ΛCDM , $\mathcal{R}_\Lambda$ are the average Bayes factor $\langle B_{i\Lambda}\rangle$ and the Bayes factor between the unknown model and ΛCDM, $B_{X\Lambda}$. The three cases are as follows:

- CASE 1: $\overline{B}_{X\Lambda} \gg 1$ and $\langle B_{i\Lambda}\rangle \sim 1$ Here the unknown model has a much stronger Bayesian evidence than ΛCDM, and ΛCDM has a similar Bayesian evidence to the other known models. As $\overline{B}_{X\Lambda} \gg 1$, we expect that there will be a significant amount of doubt, and in this case, Eq. (6.7) will become:

$$\mathcal{D} \approx \left(1 + \frac{1}{p(X)\overline{B}_{X\Lambda}}\right)^{-1} \tag{6.19}$$

$$\approx 1 \tag{6.20}$$

if $p(X)\overline{B}_{X\Lambda} \gg 1$. Hence in this case doubt grows, as $\mathcal{R} > 1$ if $\overline{B}_{X\Lambda}$ is large enough to overcome the prior doubt $p(X)$. This fulfills the first necessary condition for ΛCDM to be doubted. In order to ascertain whether the second condition is fulfilled, we look at Eq. (6.14), which becomes:

$$\begin{aligned} \mathcal{R}_\Lambda &\approx \left(1 + p(X)\overline{B}_{X\Lambda}\right)^{-1} && (6.21) \\ &< 1 && (6.22) \end{aligned}$$

so in this set up, the change in probability for ΛCDM has decreased: both conditions have been fulfilled: $\mathcal{R} > 1$ and $\mathcal{R}_\Lambda < 1$, thus ΛCDM is genuinely doubted.

- CASE 2: $\overline{B}_{X\Lambda} \gg 1$ and $B_{i\Lambda} \ll 1 (i \neq \Lambda)$ This is similar to case 1, but here the difference is that ΛCDM has a much stronger Bayesian evidence than the other known models and is demonstrably the best model in the set of known models. Here the average evidence is $\langle B_{i\Lambda}\rangle \approx 1/N$. As in case 1, $\overline{B}_{X\Lambda}$ indicates that the unknown model is favored over ΛCDM as it has a larger Bayesian evidence. The doubt in this case is given by:

  $$\mathcal{D} \approx \left(1 + \frac{1}{Np(X)\overline{B}_{X\Lambda}}\right)^{-1} \approx 1. \tag{6.23}$$

  if $p(X)\overline{B}_{X\Lambda} \gg 1/N$. But now we check the relative change the probability of ΛCDM:

  $$\mathcal{R}_\Lambda \approx \left(\frac{1}{N} + p(X)\overline{B}_{X\Lambda}\right)^{-1}. \tag{6.24}$$

  In order for ΛCDM to be doubted in this case, then we require $p(X)\overline{B}_{X\Lambda} \gg 1$. If this condition is not met, then the probabilities of both ΛCDM and the unknown model will both grow and they will accrue probability from the other less favored models in the known set of models, whose probabilities will decrease.

- CASE 3: $\overline{B}_{X\Lambda} \sim 1$ In this case the absolute upper bound on the improvement between the unknown model and ΛCDM is of order unity. So the maximum value the Bayesian evidence of the unknown model can take is the same value as the Bayesian evidence of ΛCDM, hence the unknown model may have equal weight with ΛCDM but cannot be favored over ΛCDM. In this case, doubt is given by:

  $$\mathcal{D} \approx \left(1 + \frac{\langle B_{i\Lambda}\rangle}{p(X)}\right)^{-1} \tag{6.25}$$

> For doubt to approach unity, $\mathcal{D} \to 1$, we would require $p(X) \gg \langle B_{i\Lambda} \rangle$ which would mean that we would to have had to assign the unknown model a much higher prior probability than any of the other known models, to which we assigned an indifferent, equal prior.
>
> In the more usual case where we do not assign the unknown model an exceptionally high prior probability, then doubt would not grow, and we would have no reason to doubt that our set of known models was indeed complete and did contain the true model. The true model could be then be identified by ranking the known models in the usual way in accordance with standard Bayesian model selection.

To summarize, these three cases show that the baseline ΛCDM model will only be doubted when

$$p(X)\overline{B}_{X\Lambda} \gg 1 \tag{6.26}$$

in addition to the condition that $\mathcal{R}_\Lambda < 1$. If these necessary conditions are met, then the way is opened for the possibility of model discovery.

6.3 Application of doubt to the dark energy equation of state

We chose four models for our set of known models. For the baseline model we chose flat ΛCDM, with the following set of free parameters: $\theta = \{A_s, n_S, \omega_b, \omega_c, \Omega_\Lambda, H_0\}$, where A_s is the amplitude of scalar fluctuations, n_S is the spectral index, ω_b the physical baryon density, ω_c the cold dark matter density, Ω_Λ the density parameter for the cosmological constant and H_0 the Hubble constant today. Purely adiabatic fluctuations were assumed for this work.

The additional three models are one or two parameter extensions of the baseline model, such that the simpler baseline model is nested within the more complex extended models. The first extension is to allow non-zero curvature, which requires the addition of the curvature density parameter Ω_κ, the second extension is to allow the dark energy equation of state to vary such that $w \neq -1.0$. The most complex model in the set of known models allows both Ω_κ and w to vary.

For the priors on these additional parameters; we use the astronomers prior on the curvature density $-1.0 \leq \Omega_\kappa \leq 1.0$, as suggested by Vardanyan et al. (2009), and the prior on the dark energy equation of state is $-1.3 \leq w \leq -1/3$. The upper bound on w is fixed by physical constraints as $w \leq -1/3$ is a necessary condition for dark energy to be repulsive, the lower bound is somewhat arbitrary and is chosen so as to allow models which just slip into the phantom domain. Other choices of priors can of course also be well motivated, and these would constitute additional distinct models, since a model in a Bayesian sense is comprised of both its parameters and the choice of priors on those parameters.

It is also possible to deduce analytically the effect of changing those priors, since the models are nested so the Bayes factors between the various known models and the baseline flat ΛCDM model can be calculated analytically using the Savage-Dickey density ratio (SDDR) - see Trotta (2008) for details:

$$B_{\Lambda j} = \left. \frac{p(\theta|d, \mathcal{M}_j)}{p(\theta|\mathcal{M}_j)} \right|_{\theta=\theta_0} \tag{6.27}$$

Where θ represents the additional free parameter in the more complex model, in this case either w or Ω_κ, and θ_0 represents the fixed value of that parameter in the simpler model, which in this case is the flat ΛCDM model which has fixed values of $\Omega_\kappa = 0$ and $w = -1.0$.

In the case (such as the one considered in this work) where the additional parameters have flat priors, which fully enclose the likelihood, then from the SDDR we can see that the Bayes factor $B_{\Lambda j}$ is proportional to the width of the prior $\Delta\theta = \theta_{max} - \theta_{min}$

$$B_{\Lambda j} = \frac{p(\theta = \theta_0|d, \mathcal{M}_j)}{1/\Delta\theta} \tag{6.28}$$

$$= p(\theta = \theta_0|d, \mathcal{M}_j)\Delta\theta \tag{6.29}$$

Increasing the width of the prior decreases the evidence in favour of the more complex model - the additional wasted parameter space is penalised. In order to give a significant shift in the model odds in accordance with the Jeffreys' scale (see table 5.1) one would need to have a change in the log evidence of ~ 2.5, i.e. we would require $\Delta \ln B_{\Lambda j} \sim 2.5$, which in turn would require a change in width of the prior corresponding to a factor of $\exp(2.5) \sim 12$. A twelve fold increase in the width of the prior on either w or Ω_κ would give a much wider prior than any prior which could be physically well motivated. Hence we are confident that the results in this work are robust to different choices of reasonable priors on w and Ω_κ.

Other choices of known model could also be included in our set of known models, for example a redshift dependent parametrization of $w(z)$ such as the Chevalier-Polarski-Linder parametrization $w(z) = w_0 + (1 + z)w_a$. However it has already been shown that these models are penalized for their additional parameter space and have lower evidences than ΛCDM (see for example Liddle et al. (2006b)) and hence would make only a small contribution to $\langle B_{i\Lambda} \rangle$, resulting in only a small change to the posterior doubt $\mathcal{D}$.

Additional, more exotic models exist for the explanation of the apparent late time acceleration of the Universe which involve not simply the addition of extra parameters, but instead a completely different physical model of the underlying cosmology, such alternative models include void models, and a range of different modified gravity models. These models are not included in our current set of known models in this work, since our principle objective here is to showcase the concept and application of Bayesian doubt in a simple case study. Further investigations of Bayesian model selection and Bayesian doubt with these alternative models is left for future studies, but as long as it is feasible to compute the numerical Bayesian evidence for these models, they can be included in doubt calculations.

6.3.1 Parametrization of the unknown model

Thus far we have discussed every quantity necessary to compute doubt, except for the chi-square value for the unknown model χ^2_X. Up until now, our discussion has been completely general (bar our nomination of ΛCDM as the baseline model), but now we must make a specific choice as to how we calculate χ^2_X within the context of our investigations into the dark energy component of the standard model of cosmology. We need a value for χ^2_X which represents a well fitting model, but not a perfect model.

One idea may have been to use the common rule that a model with χ^2 per degree of freedom 1 represents a model which is a good fit to the data, but that rule only holds true when: (1) The number of data points approaches infinity; (2) The data points are independent and (3) The data points are Gaussian distributed. But in this case, conditions (2) and (3) are not true for all the data sets we are considering. For example, with the CMB data the C_ℓ's are neither independent nor Gaussian distributed. A further problem is that the SNe data as it is presented at the time of doing this work has its error bars constructed in such a way as to artificially give $\chi^2/d.o.f. = 1$ for the specific choice of a flat ΛCDM model. (The identification of this particular problem with the SNe data led to my work on the SNe Ia analysis which forms the major part of my dissertation). The consequence of all this is that an alternative method must be used to compute χ^2_X.

Instead, we choose a phenomenological model to use in the numerical computation of χ^2_X. Our phenomenological model should be one which has a high enough degree of flexibility to give a good fit to the data, without over fitting the model. In the context of dark energy, allowing $w(z)$ to be a function which is able to vary freely with time (i.e. redshift z) is the principle way of improving the fit of the model. One such suitable model is the parametrized-post Friedmann (PPF) prescription developed by Hu & Sawicki (2007); Hu (2008). The PPF was originally developed to describe the behaviour of theories of modified gravity in a metric framework that describes leading order deviations from General Relativity, but it is also useful for our purposes as it can be used to describe evolving dark energy models which cross the phantom divide of $w = -1.0$. For our purely phenomenological model we use the PPF to describe a model for $w(z)$ in which w may take a different value for each of N redshift bins. In this work we choose to use $N = 10$, so our model for computing χ^2_X is comprised of 16 parameters: the original six ΛCDM parameters and the additional ten w_i parameters from each of the ten redshift bins.

6.3.2 Numerical implementation and data sets

For the data sets, we used the 307 SNe Ia from the "Union" data set compiled by Kowalski et al. (2008). The CMB data and likelihood used was the WMAP five year data set (Dunkley et al., 2009). Tegmark & Eisenstein (2006) provided the data and likelihood code for the matter power spectrum using SDSS DR4.

The computation of χ^2_X was carried out numerically by modifying CosmoMC[1] to include the following additional parameters: $w_i \equiv w(z_i)$ where z_i are N uniformly spaced redshift bins at intervals of $z = 0, \dots, 1.5$. The PPF prescription was implemented by using CAMB[2] (Lewis et al., 2000) with the publicly available plugin developed by Fang et al. (2008a,b).

Eq. (6.17) requires the best fitting value for χ^2_X, which we obtain using CosmoMC to do an MCMC reconstruction of the posterior of the 16-parameter model. Each run yields 5×10^5 samples in each of eight parallel chains. We verified that the Gelman & Rubin mixing criterion (Gelman & Rubin, 1992) was satisfied (i.e., $R \ll 0.1$, where R is the inter-chain variance divided by the intra-chain variance). However, MCMC methods are not optimized for searching for the absolute best fit point in the probability distribution, and this is a particular problem when higher dimension spaces are being explored. As a result of this, we expect that the best fit χ^2_X values in the 16-dimensional models are going to be systematically under estimated for, and this must be corrected for. In order to determine what this

---

[1] Modification of CosmoMC was carried out by Pascal Vaudrevange

[2] `http://camb.info/ppf/`

systematic under estimate is, trials were conducted using a 16D Gaussian of unit variance and then comparing the MCMC recovered best fit χ^2 with the true location of the minimum χ^2, over 5000 trials. These trials showed that MCMC systematically overestimates the best fitting χ^2 value by 0.94 ± 0.14, hence this was quantity was subtracted from the MCMC recovered values for χ^2_X. The trials also showed that the systematic uncertainty on the MCMC recovered value for χ^2 for the lower dimensional ΛCDM models was negligible, and hence no correction had to be made to those values.

The total error on the best fitting chi-square value is given by adding in quadrature the above described systematic error and also the inter chain variance of the recovered χ^2 from each of the eight chains. Another source of uncertainty is the sensitivity of χ^2_X to the number of redshift bins used to parametrize $w(z)$. Trials conducted with half as many redshift bins and twice as many redshift bins shifts the value of χ^2_X by a maximum of 0.5 units, include this additional uncertainty as part of the systematic error.

The evidence for the known models is computed using the publicly available MultiNest code (Feroz & Hobson, 2008a; Feroz et al., 2009a; Trotta et al., 2008), which implements the nested sampling algorithm, employed as an add-in sampler to CosmoMC (Lewis & Bridle, 2002) and CAMB (Lewis et al., 2000).

6.4 Results and discussion

Tabel 6.1 shows the Bayes factors between each of the known models and ΛCDM (see columns 1-3) and also shows the upper bound on the Bayes factor between the unknown model and ΛCDM. In this table, a negative Bayes factor shows that ΛCDM is favoured, and the magnitude of the Bayes factor can be interpreted using the Jeffreys' scale to give an indication of by how much one model is favoured over another. Looking only at the Bayes factors between the known models, we can see that in all cases ΛCDM is favoured over the more complex alternatives. Evidence against flat wCDM is inconclusive when only the cmb data is used, but builds towards being weak evidence when the SNeI and matter power spectrum data sets are used. There is moderate evidence against spatially curved ΛCDM which agrees with the results presented by Vardanyan et al. (2009). Evidence against the most complex model which allows both spatial curvature $\Omega_\kappa \neq 0$ and a dark energy equation of state of $w \neq -1$ is moderate approaching strong, showing that the penalty incurred for the additional parameter space outweighs the advantage of whatever slight improvement in fit is achieved by using a more flexible model.

The results of the Bayes factors between the different known models are in good agreement with other published values e.g. Liddle et al. (2006a,b) and Li et al. (2009). Using ordinary Bayesian model selection, these results confirm that ΛCDM is indeed the best models in the set of known models considered here, and is a good choice for the baseline model.

The improvement in the best fitting χ^2 (or log-likelihood) of ΛCDM over the unknown model X, shown in column four is slight, even when all data sets are used. There are two explanations for this: First of all this confirms quantitatively the commonly held view of the cosmological community that ΛCDM is indeed a model that fits the data well with little room for improvement; Secondly, it shows numerically the difficulty in improving the fit of the model by increasing the flexibility of $w(z)$ - this is because moving from $w(z)$ to the observable $\mu(z)$ involves a double integral over $w(z)$, huge variations in $w(z)$ show up as only very tiny fluctuations in $\mu(z)$ - see for example Huterer & Turner (1999); Maor et al. (2001) and Clarkson (2009). Allowing a greater degree of freedom to $w(z)$ does not significantly improve the fit of the model to the data.

When the SNe Ia data are included in the analysis, then the improvement in the chi-square for the unknown model is much more pronounced and the upper bound on the Bayes factor between the two models shows that there is weak evidence in favour of the unknown model, against the baseline ΛCDM model. This would seem to suggest that the SNeIa which are distributed in redshift space from $z = 0.0 \ldots 0.5$ require the extra degrees of freedom which allow the dark energy equation of state to vary with redshift. However, other factors besides cosmology may be responsible for this apparent variation with redshift, for example, the evolution of the colour correction parameter with redshift, as highlighted by Kessler et al. (2009a)

Thus far we have shown that there is moderate to strong evidence that ΛCDM is the best model in the set of known models, and there is weak evidence in favour of the unknown model against ΛCDM. But to answer the question as to whether we should doubt ΛCDM and the cosmological constant, we need to look at whether the necessary condition for doubt to grow, $p(X)\overline{B}_{X\Lambda} \gg 1$, is fulfilled. From table 6.1, we can see that the upper limit on the Bayes factor $\langle B_{i\Lambda} \rangle$ is just around ~ 5. The most generous prior doubt would be to assign all models equal prior probability, such that $p(X) = 0.5$, hence even under these generous arrangements the condition that $p(X)\overline{B}_{X\Lambda} \gg 1$ is not fulfilled and ΛCDM is not genuinely doubted.

Table 6.2 presents the results in terms of doubt $\mathcal{D}$ and the posterior probability for ΛCDM, for two different choices of prior doubt: $p(X) = 1 \times 10^{-2}$ representing a prior belief that we have a 99% belief that the set of known models is complete,

	$-1.3 < w < -0.3$	$w = -1.0$	$-1.3 < w < -0.3$	"Unknown" model X		$\langle B_{i\Lambda} \rangle$	$\overline{B}_{X\Lambda}$
	$\Omega_\kappa = 0.0$	$-1.0 < \Omega_\kappa < 1.0$	$-1.0 < \Omega_\kappa < 1.0$				
	$\ln B_{j\Lambda}$	$\ln B_{j\Lambda}$	$\ln B_{j\Lambda}$	$\Delta\chi^2$	$\ln \overline{B}_{X\Lambda}$		
CMB only	-0.30 ± 0.09	-2.25 ± 0.09	-2.79 ± 0.09	-1.02 ± 0.2	0.51 ± 0.2	0.48 ± 0.02	1.67 ± 0.2
CMB + SN	-0.85 ± 0.09	-2.52 ± 0.09	-3.33 ± 0.09	-3.05 ± 0.2	1.53 ± 0.2	0.39 ± 0.01	4.60 ± 0.7
CMB + mpk	-0.98 ± 0.08	-3.79 ± 0.08	-4.39 ± 0.08	-0.84 ± 0.2	0.42 ± 0.1	0.35 ± 0.01	1.52 ± 0.2
CMB + SN + mpk	-0.96 ± 0.09	-3.73 ± 0.09	-4.43 ± 0.09	-3.38 ± 0.3	1.69 ± 0.2	0.35 ± 0.01	5.42 ± 1.0

Table 6.1: In the first three columns, we report the Bayes factors between the known models and ΛCDM for different combinations of data sets, where $\ln B_{j\Lambda} < 0$ favours ΛCDM. The fourth columns gives $\Delta\chi^2 = \chi^2_X - \chi^2_\Lambda$, the improvement in the best-fit log-likelihood obtained by using model X (specified in the text) over ΛCDM. The last column gives the corresponding absolute upper bound to the Bayes factor between model X and ΛCDM.

with a 1% chance that the true model is not included; $p(X) = 1 \times 10^{-6}$ represents a much higher degree of belief that the set of known models is complete, but still allows a one in an million chance that the true model is as yet undiscovered. We can see that for both choices of prior, doubt grows such that $\mathcal{D} > p(X)$, thus fulfilling one of the two necessary conditions for ΛCDM to be doubted, $\mathcal{R} > 1$. However, the second condition for doubt to grow, namely that $\mathcal{R}_\Lambda < 1$ is not fulfilled. Equal prior probability of around 25% was assigned to each of the four known models, so $p(\Lambda) \approx 0.25$. As we can see from column 3 of table 6.2, for all combinations of data sets, the degree of belief assigned to ΛCDM has increased, with the posterior probability of ΛCDM $p(\Lambda|d)$ accruing $50 - 70\%$ of the total posterior probability. Since the probabilities of both ΛCDM and the unknown model have increased, then the probability of some of the other known models must have decreased, as the total posterior probability must be unity.

The first two columns of table 6.2 show that doubt has increased by a factor of $3 - 15$ for the different data sets, irrespective of the prior. These results are shown in fig.6.1 which shows graphically the probabilities of the unknown model and ΛCDM have increased for both choices of prior.

By using the upper bound on the Bayes factor between the baseline model and the unknown model the doubt methodology gives a strong advantage to the unknown model, as by construction, any penalty that may have been incurred by the Occam's razor term has been removed. The idea of Bayesian doubt is not to compare the unknown model with the known model as if it had equal prior probability, but rather it is a way of measuring just how confident we are justified in being when we have found the best model in the set and are wondering if a better model could possibly exist. In this context, we do not wish to assign a high prior probability to the unknown model - the unknown model already has a huge advantage over the known models, as the Occam's razor term has been artificially removed.

	Doubt $\mathcal{D}$		Posterior for ΛCDM, $p(\Lambda\|d)$
	Prior doubt: $p(X) = 10^{-2}$	Prior doubt: $p(X) = 10^{-6}$	(with $p(X) = 10^{-2}$ and $p(\Lambda) \approx 0.25$)
CMB only	$(3.41 \pm 0.3) \times 10^{-2}$	$(3.49 \pm 0.3) \times 10^{-6}$	0.51 ± 0.02
CMB+SN	$(10.8 \pm 1.0) \times 10^{-2}$	$(11.9 \pm 1.2) \times 10^{-6}$	0.58 ± 0.02
CMB+mpk	$(4.18 \pm 0.4) \times 10^{-2}$	$(4.32 \pm 0.4) \times 10^{-6}$	0.68 ± 0.01
CMB+SN+mpk	$(13.4 \pm 1.6) \times 10^{-2}$	$(15.3 \pm 2.1) \times 10^{-6}$	0.61 ± 0.02

Table 6.2: First two columns: Posterior doubt for different data sets combinations and two prior doubt assumptions. Last column: posterior probability for the ΛCDM model when allowing for the possibility of a 1% prior doubt on the completeness of our list of known models.

	Doubt $\mathcal{D}$	Posterior for ΛCDM, $p(\Lambda\|d)$
	Prior doubt: $p(X) = 0.2$	(with $p(X) = 0.2$ and $p(\Lambda) = 0.2$)
CMB only	(0.46 ± 0.02)	0.28 ± 0.02
CMB+SN	(0.75 ± 0.02)	0.16 ± 0.02
CMB+mpk	(0.52 ± 0.03)	0.34 ± 0.03
CMB+SN+mpk	(0.79 ± 0.02)	0.15 ± 0.02

Table 6.3: Posterior doubt and posterior probability for ΛCDM for the case where doubt is given the same prior probability as the known models, i.e. $p(X) = 1/(N+1)$. Een in this extremely favourable case for doubt, the posterior probability for ΛCDM does not decrease, showing that ΛCDM cannot be genuinely doubted.

However, for the purposes of testing the robustness of the conclusions presented in this work, it is interesting to consider the case in which the unknown model is assigned the same prior as the known models, i.e.:

$$p(X) = p(\mathcal{M}_i) = \frac{1}{N+1} \tag{6.30}$$

The results for this choice of prior doubt $p(X) = p(\Lambda) = 0.2$ are presented in table 6.3. In this case, the posterior doubt grows, but by a smaller factor $\sim 2-4$, and the probability of ΛCDM either stays roughly the same or increases. Even in these extreme cases with a large prior doubt, ΛCDM cannot be genuinely doubted as the condition $\mathcal{R}_\Lambda < 1$ is not met.

6.5 Impact of the addition of further known models

We have carried out this work with a test case of a set of four known models, when in reality we know that many more models have also been proposed to describe dark energy and the apparent late time acceleration of the Universe. It would be

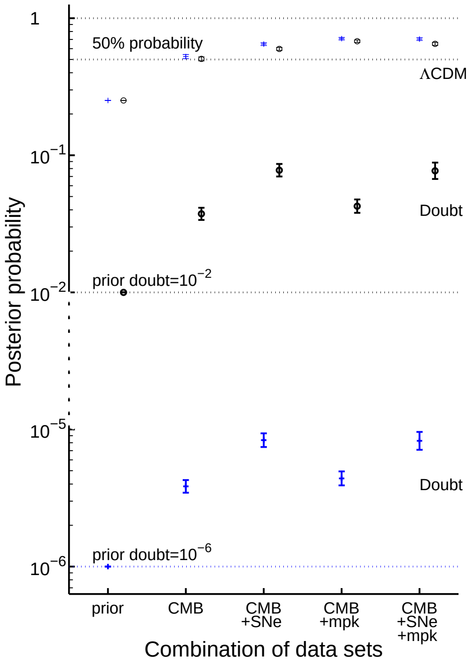


Figure 6.1: Posterior probability for doubt for the ΛCDM model as a function of different combinations of data sets. The probability of ΛCDM increases from the initial 25% to just about over 50%, while the probability of doubt increases from the initial 1% to just over 6%, mostly as a consequence of acquiring probability from the other 3 known models considered in the analysis. This signals that ΛCDM remains the most valid statistical description of the data.

reasonable to ask how using this limited set of known models impacts our final results, so we will now proceed to consider analytically what happens as additional models are included in the known set.

As we increase the number of models N we may continue to assume that the average value of the Bayes factors between ΛCDM and the other known models would continue to scale as $\langle B_{i\Lambda}\rangle \propto 1/N$, as we are expecting that ΛCDM will continue to be the best model in the set of known models. If ΛCDM was found not to be the best model in the set, then we would re-label whatever the best model in the set was as the new baseline model. We are interested in knowing what change in $\Delta\chi^2$ is required in order for the posterior doubt to approach the posterior probability of the baseline model ΛCDM as a function of the number of models N in the set of known models.

We begin by equating the posterior doubt (Eq. (6.10)) and the posterior probability of ΛCDM (Eq. (6.14)):

$$p(X|d) = p(\Lambda|d) \tag{6.31}$$

$$\left(1+\frac{\langle B_{i\Lambda}\rangle}{B_{X\Lambda}}\left(\frac{1-p(X)}{p(X)}\right)\right)^{-1} = \left(p(X)+(1-p(X))\frac{\langle B_{i\Lambda}\rangle}{B_{X\Lambda}}\right)^{-1} p(\Lambda) \tag{6.32}$$

Using the approximations $\langle B_{i\Lambda}\rangle \approx 1/N$ and $p(X) \ll 1$, we find that:

$$\Delta\chi^2 \approx 2\ln(Np(X)) \tag{6.33}$$

i.e. the change in $\Delta\chi^2$ required for posterior doubt to approach the posterior probability of ΛCDM scales with the logarithm on the number of models in the known set. For the case where $p(X) = 10^{-2}$, the required values of $\Delta\chi^2$ as a function of N are shown in the first column of table 6.4 - as more models are added to the set of known models it becomes easier to doubt ΛCDM. This effect is mainly due to the fact that when more models are added to the set, the prior on each known model becomes smaller, Eq.(6.4) whilst the prior on the unknown model remains unchanged. The prior on ΛCDM decreases while the prior doubt is constant, causing the posterior doubt to increase with N and the posterior of ΛCDM to decrease with N, as can be seen in fig.6.2.

In order to investigate the effect of increasing N independently of the change in relative sizes of the priors $p(\Lambda)$ and $p(X)$ we can change our assignment of priors such that the prior of X is always some fixed fraction f of the prior on Λ, replacing our original asignment of priors in Eq. (6.4) with the following:

$$p(\Lambda) = \frac{1}{N}(1-p(X)) \tag{6.34}$$

$$p(X) = \frac{p(\Lambda)}{f} \tag{6.35}$$

$$= (Nf+1)^{-1} \tag{6.36}$$

using the above formulation, the relative difference in the size of the priors on X and Λ remains the same as N increaces. As before, we can equate the posteriors of ΛCDM and the unknown model in order to investigate what change in $\Delta\chi^2$ is required in order for the $p(X|d)$ to approach $p(\Lambda|d)$, and we find that the requirement is:

$$\Delta\chi^2 = -4\ln f \tag{6.37}$$

Now the required change in $\Delta\chi^2$ is no longer a function of N and depends only on the relative difference between the priors of the unknown model and the baseline

Prior: $p(X) = 10^{-2}$		Prior: $p(X) = p(\Lambda)/f$	
N	Required $\Delta\chi^2$	f	Required $\Delta\chi^2$
4	-6.4	4	-5.5
10	-4.6	10	-9.2
20	-3.2	10^2	-18.4
50	-1.4	10^3	-27.6

Table 6.4: Improvement in the χ^2 of ΛCDM required for the unknown model X to have the same *a posteriori* probability as ΛCDM. First two columns: as a function of the number of known models, N, assuming a fixed prior doubt $p(X) = 10^{-2}$. Last two columns: assuming a fixed fractional prior doubt, $p(X) = p(\Lambda)/f$, and as a function of f. It is assumed that the evidence of the known models is much smaller than the evidence for ΛCDM.

ΛCDM model. The last two columns of table 6.4 show how the required $\Delta\chi^2$ changes with f, independently of N - even if the unknown model was just 4 times less likely than the baseline ΛCDM model a quite considerable of improvement of $\Delta\chi^2 = -5.5$ is required. For the more realistic case where the unknown model is assigned a much lower prior, such as $f = 20$ (which would correspond to our earlier case with $p(X) = 10^{-2}$ and $p(\Lambda) = 0.2$) then a larger still change in χ^2 is required, $\Delta\chi^2 = -12.0$. As f increases and the gap between the prior on X and the prior on ΛCDM widens, it becomes increasingly difficult to doubt ΛCDM.

Table 6.4 shows that for both choices of prior assignment, a minimum change in χ^2 of 5 units is required for ΛCDM to be doubted. Table 6.1 shows that for the data sets we used, the maximum change in χ^2 is less than the theoretical thresholds and is ~ 4. Analysis of the real data does not give a significant enough change in χ^2 for ΛCDM to be genuinely doubted.

6.6 Conclusions of Bayesian Doubt

In this chapter we have endeavoured to extend Bayesian model selection to give an absolute rather than relative measure of the goodness of the best model in the set of known models, by introducing an unknown idealized model to act as a benchmark against which the known models can be compared. The absolute upper bound on the Bayes factor between the current best model (i.e. the baseline model) and the unknown model was used to estimate the maximum evidence for the unknown model. The framework for Bayesian doubt is general up to the point at which a specific model must be chosen from which to compute a numerical value for χ^2_X.

In the work presented in this chapter, we chose specifically to investigate the

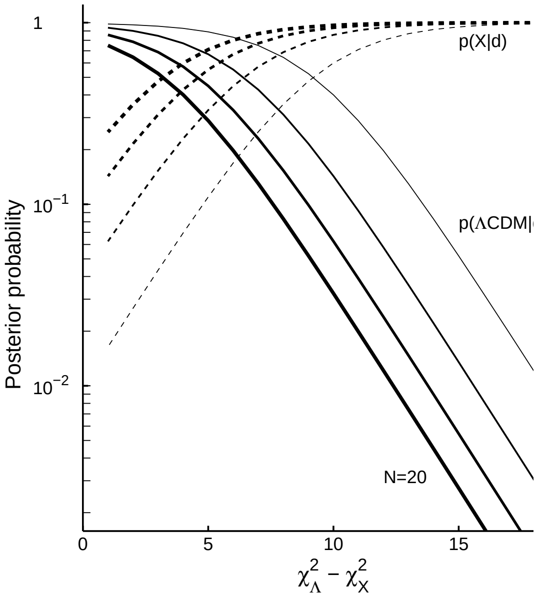


Figure 6.2: Posterior for doubt (dashed lines) and for ΛCDM (solid lines) as a function of $-\Delta\chi^2 = \chi^2_\Lambda - \chi^2_X$ assuming a fixed prior doubt $p(X) = 10^{-2}$. Different curves are for different numbers of known models, $N = 1, 4, 10, 20$ (from thin to thick), assuming that $\langle B_{i\Lambda}\rangle \approx 1/N$.

dark energy sector of the standard cosmological model and ask whether ΛCDM is indeed absolutely the best possible model, or whether there was statistical room for improvement to find a better fitting model. For this specific investigation we chose an empirical model to compute χ^2_X which allowed the dark energy equation of state to vary over a number of redshift bins. In this work, we found ΛCDM to be a good model for the data in the statistical sense with little room for improvement, and there is currently no reason to doubt ΛCDM.

Chapter 7

Bayesian parameter inference for SNeIa data

7.1 Motivations for this work

Whilst carrying out the work for Bayesian Doubt described in chapter 6, two problems became apparent with the supernovae type Ia data:

(i) The SNe Ia data is affected by an unknown intrinsic dispersion in the SNe Ia absolute magnitudes, increasing the number of SNe Ia observed will not reduce this error.

(ii) The common method for doing cosmological parameter inference from the SNe Ia data cannot be used for Bayesian model selection.

In this chapter I present a new method for Bayesian analysis of SNe Ia data fitted with the SALT2 light curve fitter. This method replaces the second step (parameter inference step) of the SALT2 method. The aims of this new methodology are twofold:

(i) To provide a rigorous statistical framework for assessing and understanding the unknown intrinsic dispersion.

(ii) To provide a fully Bayesian method for cosmological parameter inference from the SNe Ia data in order that the SNe Ia data can be used in Bayesian model selection, and exploited with the full suite of Bayesian methods.

This chapter describes how the new Bayesian method for cosmological parameter inference was developed and tested, and also describes the beginnings and plans for applications of this method to investigations of evolution of SNe Ia with redshift. Use of this new method in problems of Bayesian model selection is not covered in this work, but will be implemented in future papers.

7.2 Notation and a description of the problem

The objective of the parameter inference step of the SALT2 light curve fitting process is to constrain the cosmological parameters $\Omega_m, \Omega_\Lambda, \Omega_\kappa, w$ using the fitted parameters $\hat{m}^*_{Bi}, \hat{x_1}_i, \hat{c}_i$ resulting from the first step of the process. In this chapter, quantities which are measured are denoted by a circumflex, and we label the set of measured values as

$$D_i = \{\hat{z}_i, \hat{m}^*_{Bi}, \hat{x}_{1i}, \hat{c}_i, \hat{C}_i\}. \tag{7.1}$$

where index i labels each of the N SNe Ia in the dataset and $\hat{C}_i$ is the covariance matrix for the measured values

$$\hat{C}_i = \begin{pmatrix} \sigma^2_{m^*_B i} & \sigma_{m^*_B i, x_1 i} & \sigma_{m^*_B i, ci} \\ \sigma_{m^*_B i, x_1 i} & \sigma^2_{x_1 i} & \sigma_{x_1 i, ci} \\ \sigma_{m^*_B i, ci} & \sigma_{x_1 i, ci} & \sigma^2_{ci} \end{pmatrix}. \tag{7.2}$$

Since dark energy can mimic curvature and vice versa, we conduct this work either in a ΛCDM model of the Universe with w fixed $w \equiv -1$ but allowing non zero curvature, or alternatively in a flat wCDM model of the Universe with $\Omega_\kappa \equiv 0$ but allowing $w \neq -1$ but constant with redshift. We denote the complete set of cosmological parameters in the context of a ΛCDM model as:

$$\mathscr{C} = \{\Omega_m, \Omega_\Lambda, \Omega_\kappa, h\} \tag{7.3}$$

in considering a flat wCDM model of the Universe we use the following set of cosmological parameters:

$$\mathscr{C} = \{\Omega_m, \Omega_\Lambda, w, h\}. \tag{7.4}$$

h is defined as $H_0 = 100h$km/s/Mpc, where H_0 is the value of the Hubble rate today. H_0 is degenerate with the absolute magnitude of the supernovae M_0 and cannot be determined from the SNe Ia data alone, H_0 is effectively a nuisance parameter in this work.

In a Friedman-Robertson-Walker cosmology defined by the parameters $\mathscr{C}$, the theoretical distance modulus to a SN at redshift z_i is given by

$$\mu_i = \mu(z_i, \mathscr{C}) = 5 \log \left[\frac{D_L(z_i, \mathscr{C})}{\text{Mpc}} \right] + 25, \tag{7.5}$$

where D_L denotes the luminosity distance to the SN, and is related to cosmological parameters by Eq. (4.3) and the Friedman equation, Eq. (4.4). The problem is how to infer the cosmological parameters $\mathscr{C}$ given the measured values D. I shall describe the standard χ^2 method for solving this problem in section 7.3 before going on to describe our new Bayesian method in 7.5. Both methods describe the same set up, but there are a few conceptual differences in the way in which the set up is described. The real difference between the two methods is that in the Bayesian hirearchiacal model the full joint posterior probability is computed.

7.3 The standard χ^2 method

The most common method found in the literature for estimating the cosmological parameters from SNe Ia data fitted using SALT2 involves some variation of the χ^2 method outlined in this section (e.g. Astier & Guy (2006); Kowalski et al. (2008); Kessler et al. (2009a)). The exact χ^2 method varies between consortia, but the essential elements are common to all, and are outlined below.

The χ^2 statistic is defined as

$$\chi^2_\mu = \sum_{i=1}^{N} \frac{(\mu_i - \mu_i^{\text{obs}})^2}{\sigma^2_{\mu i}}. \tag{7.6}$$

where μ_i is the theoretical distance modulus given by Eq. (7.5) and is a function of redshift and the cosmological parameters $\mathscr{C}$. The 'observed' distance modulus μ_i^{obs} is given by Eq. (4.6) with the estimated values from step 1 of the light curve fitting process for $\hat{m}^*_{Bi}, \hat{x}_{1i}, \hat{c}_i$

$$\mu_i^{\text{obs}} = \hat{m}^*_{Bi} - M_0 + \alpha \cdot \hat{x}_{1i} - \beta \cdot \hat{c}_i \tag{7.7}$$

The total error on the distance modulus $\sigma^2_{\mu i}$ is the sum of several errors added in quadrature

$$\sigma^2_{\mu i} = (\sigma^{\text{fit}}_{\mu i})^2 + (\sigma^{z}_{\mu i})^2 + (\sigma^{\text{int}}_{\mu})^2, \tag{7.8}$$

The three components of the error are:

(i) Fitting error $\sigma^{\text{fit}}_{\mu i}$ which is given by

$$\left(\sigma^{\text{fit}}_{\mu i}\right)^2 = \underline{\Psi}^T \hat{C}_i \underline{\Psi} \tag{7.9}$$

where $\Psi = (1, \alpha, -\beta)$ and $\hat{C}_i$ is the covariance matrix given in Eq. (7.2). Here we see that the global fit parameters α, β enter into the denominator as well as the numerator of the χ^2_μ expression.

(ii) Redshift error $\sigma^z_{\mu i}$, error in the redshift measurement (given by host galaxy redshift) due to uncertainties in the peculiar velocity of the host galaxy and uncertainties in the spectroscopic measurements.

(iii) The intrinsic dispersion of the SNe Ia absolute magnitudes, σ_μ^{int} an unknown quantity which must be estimated from the data and parameter estimation process. This number describes the variation in the SNe Ia absolute magnitudes which remain after correction for stretch and colour, the variation may be due to physical differences in the SNe Ia population as described in chapter 4 or differences in the survey and data reduction technique.

Additional errors due to lensing, Milky Way dust extinction etc. can also be added in at this stage, but we do not consider those errors in this work.

The χ_μ^2 statistic of Eq. (7.6) is minimized by sampling over parameter space and simultaneously fitting for both the cosmological parameters $\Omega_m, \Omega_\Lambda, \Omega_\kappa, w$ and the SNeIa global fit parameters α, β, M_0. Variations on this include using an iterative process to update the α, β values in the denominator e.g. Astier & Guy (2006) who point out that Tripp (1998) realized that minimizing over α, β directly could result in artificially inflated values of α, β so as to reduce the χ_μ^2 value.

The problem with the unknown intrinsic dispersion σ_μ^{int} is dealt with by using an iterative process in which χ_μ^2 is minimized, then σ_μ^{int} is adjusted between minimizations in such a way as to give χ_μ^2 per degree of freedom to be unity i.e. $\chi_\mu^2/d.o.f \sim 1$.

Although the method described above has been fully tested by the consortia that use this method and has been found to give satisfactory results for cosmological parameter inference, several problems remain:

(i) The use of the χ_μ^2 expression is not well motivated statistically, but is based on a heuristic derivation.

(ii) The global fit parameters α, β act as both range and location parameters, appearing in both the numerator and denominator, hence the errors on these parameters are not Gaussian. The informal test which states that $\chi^2/d.o.f \sim 1$ for a good fit model only holds in the Gaussian case, and its use cannot be justified here.

(iii) The total error on the distance modulus $\sigma_{\mu i}^2$ is adjusted in a way which assumes that the model under consideration (generally either ΛCDM or flat wCDM) is a good fit for the data. This means that this method of parameter inference, or error bars produced using this method cannot be used to investigate problems of model selection, since the model used to derive σ_μ^{int} will by construction be the favoured model.

(iv) This method obtains a single value for σ_μ^{int}, giving no indication of the error on that value. A better approach would be to obtain a probability distribution function for σ_μ^{int} so that an indication can be given about the degree of belief about the value obtained for σ_μ^{int}.

(v) Because the different parameters are not treated in the same way (e.g. some are updated iteratively, sometimes marginalization is used, other times minimization is used), this method cannot be used with standard MCMC or nested sampling techniques.

The new method for Bayesian parameter inference from the SNe Ia data which I will present in this section seeks to address some of these problems with the χ_μ^2 method and provide a statistically well motivated framework for parameter inference which can also be used in problems of model selection. This method replaces the second step in the SALT2 light curve fitting process. I first describe the Bayesian Hierarchial Model for the system, and then present the details of the calculation.

7.4 In search of a Bayesian solution

7.4.1 Identifying the bias problem

One of the major features of our new Bayesian Hierarchical Model (BHM) method is the introduction of the latent variables and the priors on those variables. I will include here a short discussion of why this step is so crucial, and how we came to understand its necessity. The bias problem described here has to do with the measurement errors on $\{\hat{m}_{Bi}^*, \hat{c}_i, \hat{x}_{1i}\}$ and is independent of the unknown intrinsic dispersion problem σ_μ^{int}. Hence for simplicity in these examples we assume there is no intrinsic dispersion in the absolute magnitudes σ_μ^{int}. As mentioned in section 7.5, the problem described in this work is essentially an extension of the linear model. A simple linear model describing the relationship between the true variables x_i and y_i and characterized by a slope parameter a and intercept parameter b can be expressed as:

$$y_i = ax_i + b \tag{7.10}$$

The observed values for the dependent and independent variables are denoted by hats ($\hat{x}_i, \hat{y}_i$, $i = 1, \ldots, N$), and they are obtained from the latent values under the assumption of Gaussian noise (with known variances σ_x^2, σ_y^2):

$$\hat{x}_i \sim \mathcal{N}(x_i, \sigma_x^2) \text{ and } \hat{y}_i \sim \mathcal{N}(y_i, \sigma_y^2). \tag{7.11}$$

The Bayesian hierarchical network showing the dependencies of the different parameters in the linea toy model is shown in Fig. 7.1. As before, the solid lines show probabilistic connections and dashed lines show deterministic connections.

This linear toy model is very similar in form to the relationship between the latent parameters in the SNe Ia case described in Eq. 7.73 which is as follows:

$$\mu_i = m_{Bi}^* - M_i + \alpha \cdot x_{1i} - \beta \cdot c_i \tag{7.12}$$

The problem in both cases is how to use the true relationships above to infer the parameters (either a, b for the linear toy model or α, β and the cosmology parameters for the SNe Ia case) given that we only have noisy 'observed' values $\{\hat{x}_i, \hat{y}_i\}$ and $\{\hat{m}_{Bi}^*, \hat{c}_i, \hat{x}_{1i}\}$ respectively. We use the term 'observed' somewhat colloquially to signify that these quantities are subject to error, in reality $\{\hat{m}_{Bi}^*, \hat{c}_i, \hat{x}_{1i}\}$ are not observed directly, but are the result of the first stage of the SALT2 fitting process, and the errors on these quantities are fitting errors which depend indirectly on the observational errors.

The first special feature of this problem of parameter inference is that there are errors on all the axes - i.e. there are errors on both the x and y axes for the linear toy model and on the m_B^*, x_1 and c axes for the SNe Ia case. A proposed Bayesian solution for this type of problem for the linear toy model has been given by D'Agostini (2005) who gives the following expression for the posterior probability of a, b for the linear toy model with errors on both axes:

$$p(a, b|\underline{\hat{x}}, \underline{\hat{y}}) \propto \prod_i^n \frac{1}{\sqrt{\sigma_y^2 + a^2\sigma_x^2}} \exp\left[-\frac{1}{2}\frac{(\hat{y}_i - a\hat{x}_i - b)^2}{(\sigma_y^2 + a^2\sigma_x^2)}\right] p(a, b) \tag{7.13}$$

$$\propto \underbrace{(\sigma_y^2 + a^2\sigma_x^2)^{-\frac{n}{2}}}_{\text{prefactor}} \underbrace{\exp\left[-\frac{1}{2}\chi_x^2\right]}_{\text{exp. term}} p(a, b) \tag{7.14}$$

where $p(a, b) = p(a)p(b)$ are the uniform (and in this example separable) priors on the parameters a, b (see table 7.1 for priors used in the numerical trials). The chisquare χ_x^2 for the linear toy model is defined as:

$$\chi_x^2 = \sum_i^n \frac{(\hat{y}_i - a\hat{x}_i - b)^2}{(\sigma_y^2 + a^2\sigma_x^2)} \tag{7.15}$$

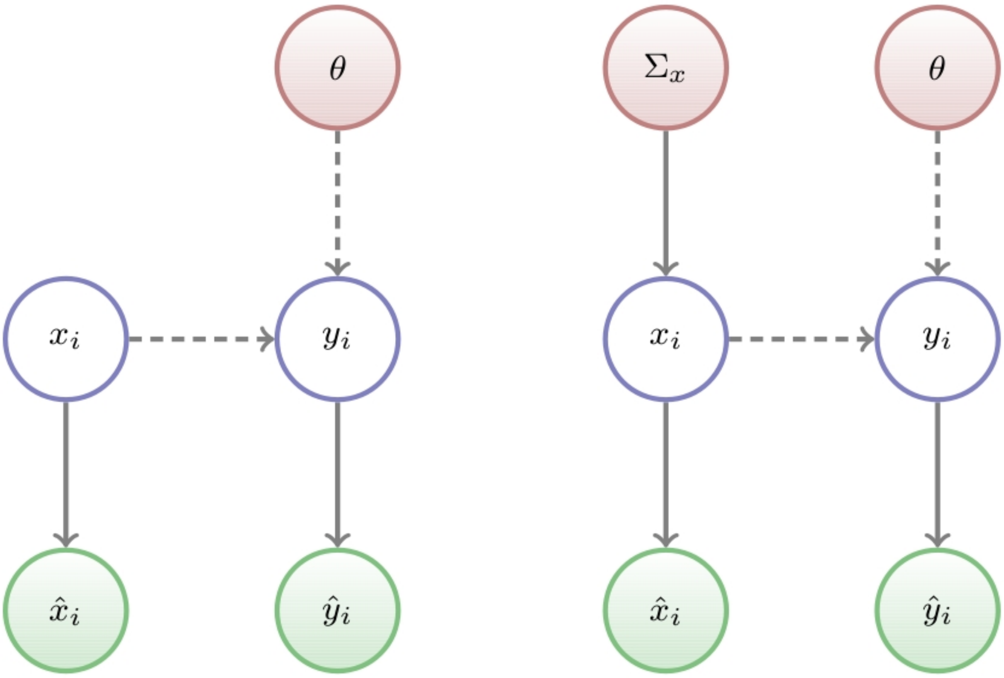


Figure 7.1: Bayesian networks showing the deterministic and probabilistic connections in the toy linear model. Solid lines indicate probabilistic connections, dashed lines represent deterministic connections. Parameters to be constrained are in red, latent variables in blue and data in green (denoted by hats). θ denotes the parameters a, b, i.e. the intercept and slope of the linear relation of Eq. (7.10). LH panel shows network as depicted by D'Agostini (2005); RH panel shows explicit description of parent distribution of latent x_i, necessary when using Gull (1989) methodology.

As we later show below, this expression for the posterior only works in a limited number of cases. An analogous expression can be derived for the SNe Ia case using the same principles. For clarity we consider the simplified case for which there is no intrinsic dispersion in the absolute magnitudes such that $M_i = M_0$ and we specify that there are no intra SNe Ia correlations. Based on the methodology of D'Agostini

(2005) the posterior would be:

$$
\begin{aligned}
p(\mathscr{C}, \alpha, \beta | \underline{m}_B^*, \underline{x}_1, \underline{c}) \propto & \prod_i^n \frac{1}{\sqrt{\sigma_m^2 + \alpha^2 \sigma_{x1}^2 + \beta^2 \sigma_c^2}} \\
& \times \exp\left[-\frac{1}{2} \frac{(\mu_i - \hat{m}_{Bi}^* + M_0 - \alpha \hat{x}_{1i} + \beta \hat{c}_i)^2}{(\sigma_m^2 + \alpha^2 \sigma_{x1}^2 + \beta^2 \sigma_c^2)} \right] \\
& \times p(\mathscr{C}, \alpha, \beta) \\
\propto & \overbrace{\underbrace{(\sigma_m^2 + \alpha^2 \sigma_{x1}^2 + \beta^2 \sigma_c^2)^{-\frac{n}{2}}}_{\text{prefactor}} \underbrace{\exp\left[-\frac{1}{2} \chi_\mu^2 \right]}_{\text{exp. term}}}^{\text{likelihood}} \times \overbrace{p(\mathscr{C}, \alpha, \beta)}^{\text{Prior}}
\end{aligned}
\tag{7.16}
$$

where χ_μ^2 is as defined in Eq. (7.6). Our initial trials for Bayesian cosmological parameter inference from the SNe Ia data used the expression for the posterior in Eq. (7.16). However, it became apparent that using this expression resulted in the biased recovery of the SNe Ia global parameters α and β. By bias, we mean that the recovered values for α and β are *always* systematically shifted several standard deviations away from their true values, in the same direction every time. The statistical definition of bias is:

$$
\text{bias} = \langle \hat{\theta} - \theta_{\text{true}} \rangle \tag{7.17}
$$

Data realizations are subject to noise, so one expects that there will be some discrepancy between the true value θ_{true} and the recovered value $\hat{\theta}$. However, averaged over many realizations, one expects this difference to be close to zero such that bias $= \langle \hat{\theta} - \theta_{\text{true}} \rangle \approx 0$ as sometimes the recovered value will be a little higher than the true value, and sometimes a little lower. However for the SNe Ia global parameters α and β in this specific case, because they are always systematically shifted in the same direction, the bias for these parameters is large and equivalent to several standard deviations. This bias can only be correctly identified as 'bias' rather than 'outlying realization' after looking at many realizations of the data.

Numerical trials with simulated SNe Ia data from a known cosmological model attempting parameter inference using Eq. (7.16) always resulted in systematically lower values for α and β being recovered, as can be seen in the left hand panel of fig.7.2. Many trials were conducted with different data sets, the example posteriors shown in the left hand panel of fig.7.2 are typical and are not merely the result of an outlying realization of the data. In the course of conducting the numerical trials and checking the methods of data simulation and parameter inference code, we realized

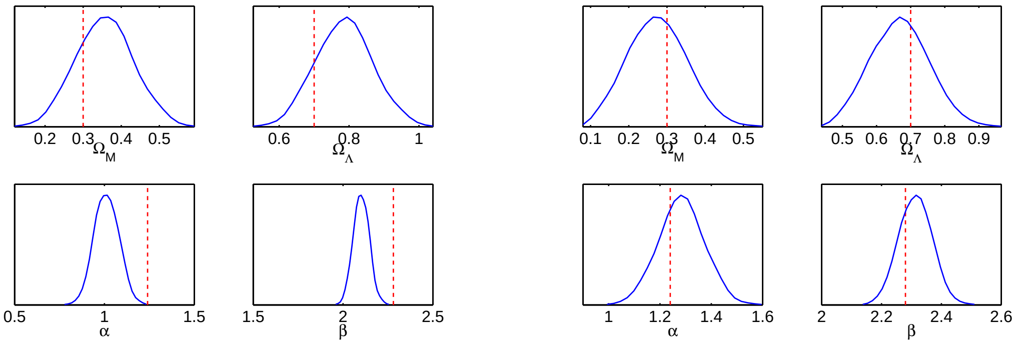


Figure 7.2: 1D Marginalized posterior plots of the cosmological parameters and SNe Ia parameters obtained using the naive Bayesian expression given by Eq. (7.16). In the LH panel the full expression is used and a clear bias can be seen in the recover of the α, β SNe Ia parameters. In the RH panel, the prefactor in Eq. (7.16) is dropped and α, β are recovered without bias - the small discrepancy between the true and recovered values is due to realization noise. Red dashed line indicates location of true parameter.

that if the prefactor of Eq. (7.16) was dropped, (effectively reducing it to a version of the χ^2) then α and β were no longer biased, as shown in the RH panel of fig.7.2. As will be shown later, this is because there is an approximate relationship between the true Bayesian solution (not yet presented) and the χ^2 method.

7.4.2 Locating the cause of the bias problem

Having ruled out coding error as being responsible for the apparent bias, we began numerical trials with the linear toy model described by Eq. (7.10) and the associated expression for the posterior Eq. (7.14) in order to have a better understanding of the general statistical problem without being concerned with the added complication of cosmological parameter inference. Initial trials with the linear model showed that for the simulated linear model data, a bias was seen in the recovery of the slope parameter a.

A lead in Sivia & Skilling (2006) regarding the particular problem of parameter inference in the presence of errors in both the x and y axes directions brought to light a short paper, Gull (1989), the abstract of which is reproduced below:

> ‘A Bayesian solution is presented to the problem of straight-line fitting when both variables x and y are subject to error. The solution, which is fully symmetric with respect to x and y, contains a very surprising feature: it requires an informative prior for the distribution of sample

Parameter	Prior
a	flat interval (0.1 3.0)
b	flat interval (0.1 3.0)

Table 7.1: Priors on linear model data parameters.

Series	Distribution of latent x_i	Σ_X	σ_x	σ_y	a	b
1	uniform grid $(-5, 5)$	-	0.05	0.15	2.0	1.0
2	$\sim \mathcal{N}(5, \Sigma_x)$	2.0	0.05	0.15	2.0	1.0
3	uniform grid $(-0.2, 0.2)$	-	0.05	0.15	2.0	1.0
4	uniform grid $(-0.2, 0.2)$	-	0.0	0.1	2.0	1.0
5	uniform grid $(-0.2, 0.2)$	-	0.05	0.1	2.0	1.0
6	uniform grid $(-0.2, 0.2)$	-	0.0	0.15	2.0	1.0

Table 7.2: Parameters used for the generation of the linear model data sets.

> positions. An uninformative prior leads to a bias in the estimated slope.' (Gull, 1989)

Further trials with the linear model confirmed that the bias problem described by Gull (1989) in the recovery of the slope parameter was exactly the same as the bias problem found in the linear toy model and the SNe Ia data. We generated 100 simulated data sets for $\hat{x}, \hat{y}$ for a series of six different noise models as described in table 7.2 and fig.7.3. We then used a numerical implementation of the D'Agostini expression, Eq. (7.16) coupled with the MultiNest sampler to recover the slope and intercept parameters a and b, using the mean value of the 1D marginalized posterior probability distribution for the parameter of interest. The results of the mean value of a recovered for each of the 100 trials for the series of six noise models are shown in histogram form in fig.7.4.

For noise models 3 and 5, a very clear bias can be seen in the recovery of the slope parameter a for every trial in noise models 3 and 5 the estimator for a is systematically and significantly lower than the true value of a. For data series 3 and 5, two conditions are fulfilled: (1) There are errors in both the x and y directions and (2) The error size is large compared with the range of the data points in that same direction. Data series 1 and 2 show that if the size of the error is small with respect to the range of the data points there is no bias in the estimator for a, and likewise there is no bias if the size of the error is large with respect to the range of the data points in one direction only.

Having established through numerical trials of the linear toy model that a biased value for the slope parameter was recovered when (1) there were errors on in both

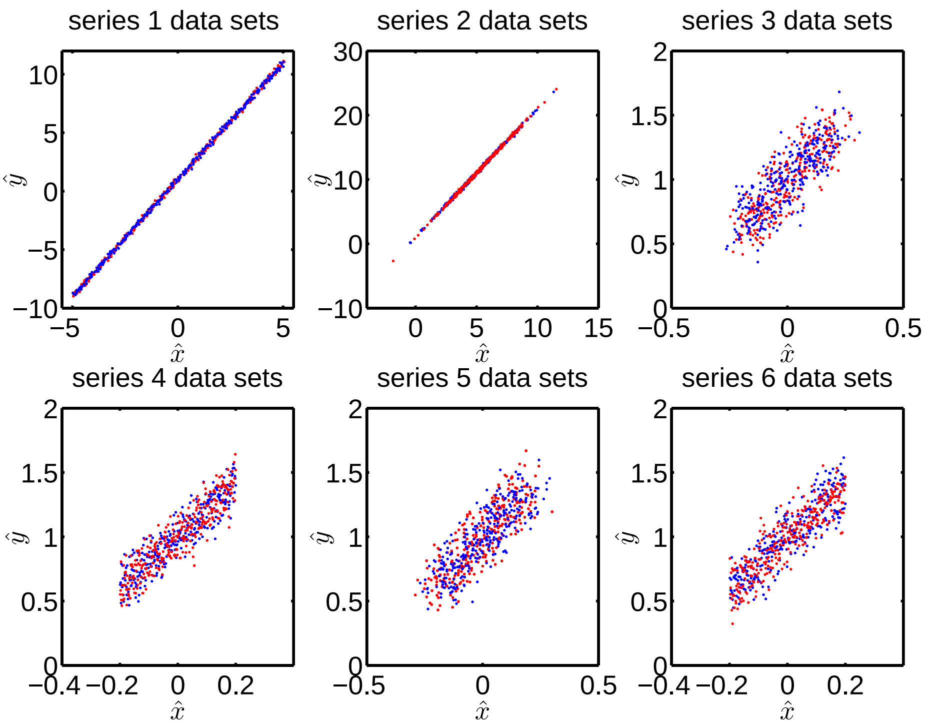


Figure 7.3: Examples of the simulated linear model data sets from series 1-6. For each series, 100 data sets were simulated. For the characteristics of each data set, see table 7.2. Two example data sets are shown for each series, one in red and one in blue. Each data set has 300 points.

axes directions and (2) the size of the errors was large compared with the range of the parameters we were able to conduct a quick check as to whether this was in fact the root of the bias problem in the SNe Ia case. A special test case realization of the SNe Ia data was created for which the ranges of the c_i and x_{1i} data points was increased ten fold whilst keeping the errors on those data points the same, such that in this special test case the size of the error bars was approximately ten times smaller than the range of those data points. (Ordinarily the size of the error bars on $\hat{x}_{1i}, \hat{c}_i$ are similar to the range of $\hat{x}_{1i}, \hat{c}_i$). The results of this simple test are shown in fig. 7.5 where it can be seen that the SNe Ia parameters α, β have been recovered apparently without bias. Although only trialled with a few such data sets (not all presented here), this test was sufficient to show that the cause of bias in when using the naive proposed posterior probability distribution given by Eq. (7.16) was due to the fact that (1) there were errors in all axes directions and (2) the size of the errorbars was large compared with the range of the data points. For the SNe Ia case, Kowalski et al. (2008) had already alluded to the fact that there was a known

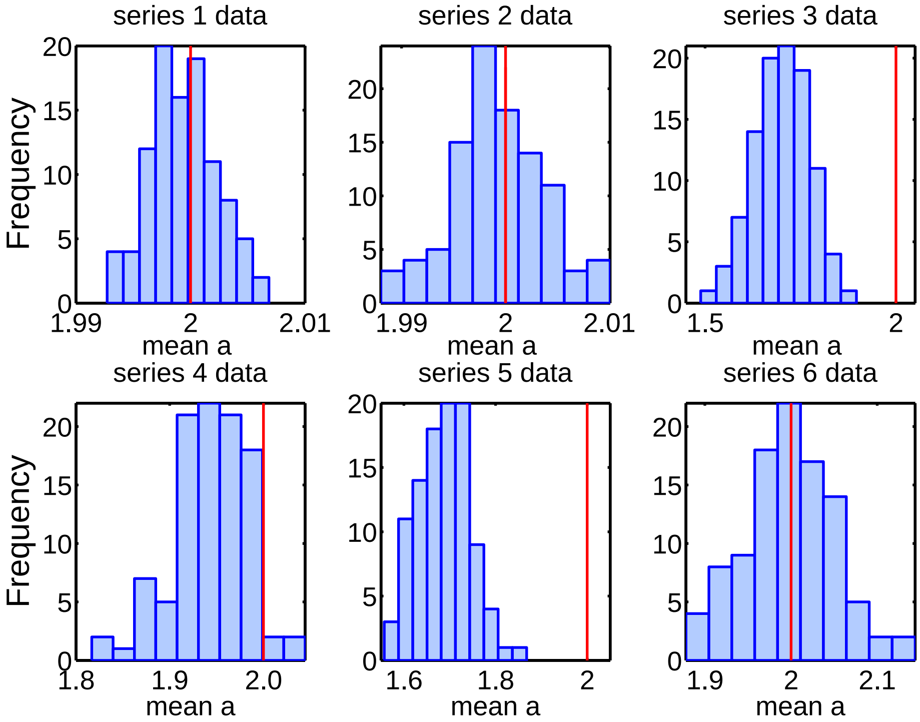


Figure 7.4: Linear model: $y_i = ax_i + b$. Histograms show mean values of a recovered from 1D posteriors from 100 trials for six different data series. For series 3 and 5 a clear bias can be seen and a systematically lower value for a is recovered.

problem with the recovery of the SNe Ia global parameters $\{\alpha, \beta\}$ which was related to the relative sizes of the error bars with respect to the range of the stretch and colour parameters, and stated that this was the cause of the bias seen in certain Frequentist methodologies for recovering $\{\alpha, \beta\}$.

The conclusions of these preliminary trials and preparatory survey of the literature were:

(i) The proposed Bayesian posterior of D'Agostini (2005) for the linear toy model with errors on both axes only holds when the size of the errors is small compared with the range of the data points in those respective directions.

(ii) The SNe Ia are a special case of a data set for which the size of the errors on the stretch and colour parameters is large with respect to the range of the stretch and colour parameters, as noted by Kowalski et al. (2008)

(iii) The problem of bias in the recovery of the slope parameter in a linear toy model had already been identified and solved by Gull (1989)

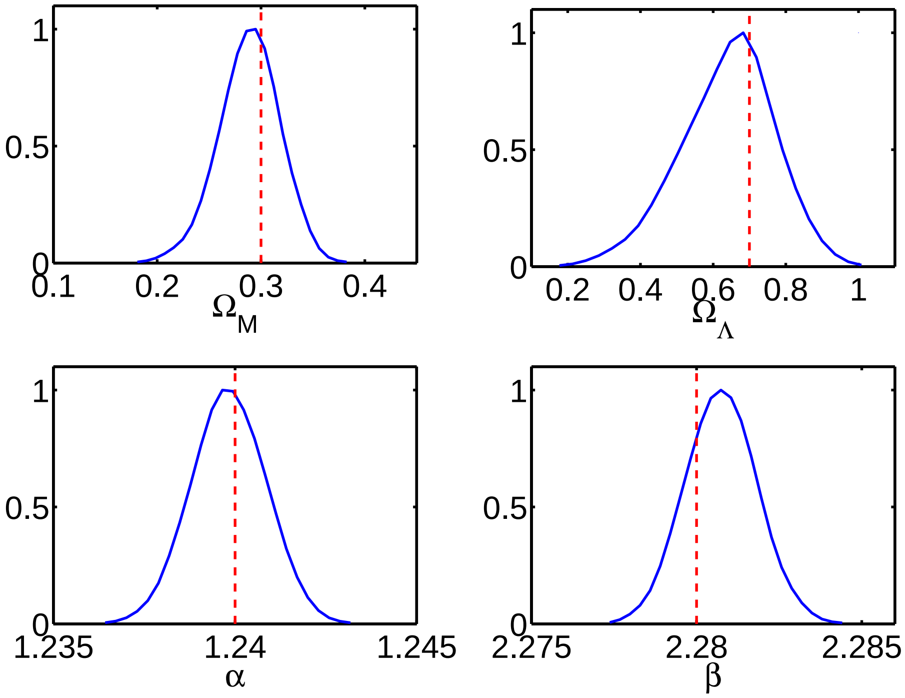


Figure 7.5: Multinest recovered parameters for data set with $\Sigma_x = 10.0$ and $\Sigma_c = 10.0$, all other parameters are as in table 7.3. Here there is no bias on α and β because the ranges of s and c are much larger than their associated observational noise. There is a small discrepancy between the true and recovered values of α and β, but this is due to realization noise rather than a systematic offset.

(iv) The structure of the relationship between the cosmological parameters, the SNe Ia global fit parameters and the SNe Ia data is essentially an extension of the linear toy model, hence whatever method was used by Gull (1989) to solve the bias problem in the linear toy model could in theory be adapted to solve the similar problem in the SNe Ia case.

Following this, the course of action chosen was to investigate the methodology of Gull (1989) more closely and adapt it to the SNe Ia case. I shall describe the methodology of Gull (1989) and explain how we adapted it to the SNe Ia problem. (N.B. The unknown $\sigma_\mu^{\rm int}$ is an additional problem not addressed by Gull (1989), but which we treat in the the full SNe Ia calculation described in section 7.5).

7.4.3 Bias solved in the linear toy model

Gull's insight into the bias problem was to recognise that it is necessary to place a prior on the distribution of the latent or true x_i. In implementing this solution, Gull also makes a coordinate transformation for convenience, rescaling all the observed $\hat{x}_i, \hat{y}_i$ into the interval $\{-1, 1\}$. Here we consider the linear toy model described by Eq. (7.10) and Eq. (7.11). The Bayesian network for this set up is shown in fig. 7.1. The observed values are indicated by circumflexes: $\hat{x}_i, \hat{y}_i$. Quantities $x, y, \hat{x}, \hat{y}$ without the subscript i denote N-dimensional vectors e.g. $x = \{x_1, \ldots, x_N\}$

There are four main components of Gull's derivation:

(i) Use of a prior for the range of the latent x_i. This is the key step which solves the bias problem.

(ii) Integration over the nuisance parameters x_i to remove them from the final expression for the joint posterior distribution.

(iii) A coordinate transformation to ease the algebra.

(iv) A further integration over the remaining nuisance parameters x_0 and y_0.

Here we shall outline a derivation based on these concepts which will lead to Gull's expression for the joint posterior distribution. Later we shall show how a similar idea may be applied to the case of the SNe Ia. For the SNe Ia we shall use the idea of a prior on the range of the latent variables, but we shall not use the coordinate transformation, because the coordinate transformation requires the intercept parameter b to be the same for each i, whereas for the supernovae case there is a different intercept μ_i for each supernova.

I will now present a derivation of Gull's expression for the joint posterior for the linear toy model. The coordinate transformation is as follows:

$$X = \frac{(\hat{x} - x_0)}{R_x} \tag{7.18}$$

$$Y = \frac{(\hat{y} - y_0)}{R_y} \tag{7.19}$$

$$a = \frac{R_y}{R_x} \tag{7.20}$$

$$b = y_0 - a x_0 \tag{7.21}$$

where x_0, y_0 are (unknown) offset parameters, and we have the following relationship:

$$X = \pm Y \tag{7.22}$$

The range parameter, R is defined as:

$$R = (R_x R_y)^{1/2} \tag{7.23}$$

which allows the useful relations:

$$R_x^2 = \frac{R^2}{a} \tag{7.24}$$

and

$$R_y^2 = aR^2 \tag{7.25}$$

The joint likelihood is:

$$\begin{aligned} p(\hat{x}, \hat{y}|x, y, a, b, \sigma_x, \sigma_y) &= (4\pi^2 \sigma_x^2 \sigma_y^2)^{-N/2} \\ &\times \exp(-\frac{1}{2}(\frac{\sum_i (\hat{x}_i - x_i)^2}{\sigma_x^2} + \frac{\sum_i (\hat{y}_i - y_i)^2}{\sigma_y^2})) \end{aligned} \tag{7.26}$$

Crucially, Gull assigns a Gaussian prior to the range of the latent x_i as follows:

$$p(x|x_0, y_0, R_x, R_y) = (4\pi^2 R_x^2)^{-1/2} \exp\left(-\frac{1}{2}\frac{\sum_i (x_i - x_0)^2}{R_x^2}\right) \tag{7.27}$$

Neglecting the normalizing Bayesian evidence term, the joint posterior can be written as:

$$\begin{aligned} p(x, y,&x_0, y_0, \log R_x, \log R_y|\hat{x}, \hat{y}, \sigma_x, \sigma_y,) \\ &\propto p(\hat{x}, \hat{y}|x, x_0, y_0, R_x, R_y) \\ &\times p(x|x_0, y_0, R_x, R_y) p(x_0, y_0, R_x, R_y) \end{aligned} \tag{7.28}$$

With Jeffries priors on the range parameters, uniform in log space $\log R_x, \log R_y$ and uniform priors on the location parameters x_0, y_0, this is:

$$\begin{aligned} p(x, y, x_0, y_0, \log a, \log R_x, \log R_y) &\propto (8\pi^3\sigma_x^2\sigma_y^2R_x^2)^{-N/2} \\ &\times \exp(-\frac{1}{2}(\frac{\sum_i(\hat{x}_i - x_i)^2}{\sigma_x^2} \\ &+ \frac{\sum_i(\hat{y}_i - y_i)^2}{\sigma_y^2} + \frac{\sum_i(x_i - x_0)^2}{R_x^2})) \end{aligned} \tag{7.29}$$

Where for convenience we have now dropped explicit reference to the conditionals $|\hat{x}, \hat{y}, \sigma_x, \sigma_y$ on the left hand side of the equation.

Eliminate y_i then b from Eq. (7.29) using Eq. (7.10) and Eq. (7.21) respectively to obtain:

$$\begin{aligned} p(x, x_0, y_0, \log a, \log R_x, \log R_y) &\propto (8\pi^3\sigma_x^2\sigma_y^2R_x^2)^{-N/2} \\ &\times \exp(-\frac{1}{2}(\frac{\sum_i(\hat{x}_i - x_i)^2}{\sigma_x^2} \\ &+ \frac{\sum_i(\hat{y}_i - ax_i - y_0 + ax_0)^2}{\sigma_y^2} \\ &+ \frac{\sum_i(x_i - x_0)^2}{R_x^2})) \end{aligned} \tag{7.30}$$

[For the SNe case we shall use the equivalent of Eq. (7.10) but we shall not use the coordinate transformation Eq. (7.21) to eliminate the analogue of b - this will be the main difference between the current application of Gull's ideas to the linear model and the future application of Gull's ideas to the SNe case.]

Now the aim is to rewrite the above expression into a form which can easily be integrated w.r.t. the x_i - this essentially means gathering together the x_i and writing the expression in a Gaussian form.

We make a temporary change of variables for ease of calculation:

$$\tau_i = x_i - x_0 \tag{7.31}$$

so the joint posterior becomes:

$$\begin{aligned} p(x, x_0, y_0, \log a, \log R_x, \log R_y) &\propto (8\pi^3 \sigma_x^2 \sigma_y^2 R_x^2)^{-N/2} \\ &\times \exp(-\frac{\sum_i}{2}(\frac{(\tau_i - (\hat{x}_i - x_0))^2}{\sigma_x^2} \\ &+ \frac{(a\tau_i - (y_0 - \hat{y}_i))^2}{\sigma_y^2} + \frac{\tau_i^2}{R_x^2})) \end{aligned} \tag{7.32}$$

Rearrange:

$$\begin{aligned} p(x,x_0, y_0, \log a, \log R_x, \log R_y) &\propto (8\pi^3 \sigma_x^2 \sigma_y^2 R_x^2)^{-N/2} \\ &\times \exp(-\frac{\sum_i}{2}(\frac{a^2\tau_i^2}{a^2 R_x^2} + \frac{(a\tau_i - a(\hat{x}_i - x_0))^2}{a^2\sigma_x^2} \\ &+ \frac{(a\tau_i - (y_0 - \hat{y}_i))^2}{\sigma_y^2})) \end{aligned} \tag{7.33}$$

Relabel:

$$t_i = a\tau_i \quad g_i = a(\hat{x}_i - x_0) \quad f_i = y_0 - \hat{y}_i \tag{7.34}$$

$$\sigma_1 = a^2 R_x^2 \quad \sigma_2 = a^2\sigma_x^2 \quad \sigma_3 = \sigma_y^2 \tag{7.35}$$

again, the joint posterior may be written as:

$$\begin{aligned} p(x,x_0, y_0, \log a, \log R_x, \log R_y) &\propto (8\pi^3 \sigma_x^2 \sigma_y^2 R_x^2)^{-N/2} \\ &\times \exp(-\frac{\sum_i}{2}(\frac{t_i^2}{\sigma_1^2} + \frac{(t_i - g_i)^2}{\sigma_2^2} + \frac{(t_i - f_i))^2}{\sigma_3^2})) \end{aligned} \tag{7.36}$$

Now we are able to integrate out the nuisance parameters x_i from Eq. (7.36)

$$p(x_0, y_0, \log a, \log R_x, \log R_y) = \int \mathrm{d}x \; p(x, \hat{x}, \hat{y}, x_0, y_0, R_x, R_y) \tag{7.37}$$

(Use $\mathrm{d}x_i = \frac{\mathrm{d}t_i}{a}$)

So the joint posterior marginalized over the latent x_i is:

$$\begin{aligned} p(x_0, y_0, \log a, \log R_x, \log R_y) &= (8\pi^3 \sigma_x^2 \sigma_y^2 R_x^2)^{-N/2} \frac{1}{a^N} \\ \times \int \mathrm{d}^N t_i \exp(-\frac{\sum_i}{2}&(\frac{t_i^2}{\sigma_1^2} + \frac{(t_i - g_i)^2}{\sigma_2^2} + \frac{(t_i - f_i))^2}{\sigma_3^2})) \end{aligned} \tag{7.38}$$

which is:

$$= (8\pi^3\sigma_x^2\sigma_y^2R_x^2)^{-N/2}\frac{1}{a^N} \times \int \mathrm{d}^N t_i \exp(-\frac{\sum_i}{2}(\frac{\sigma_2^2\sigma_3^2t_i^2 + \sigma_1^2\sigma_3^2(t_i - g_i)^2 + \sigma_1^2\sigma_2^2(t_i - f_i)^2}{\sigma_1^2\sigma_2^2\sigma_3^2})) \tag{7.39}$$

Complete the square and perform Gaussian integral to obtain:

$$= (8\pi^2\sigma_x^2\sigma_y^2R_x^2)^{-N/2}\frac{1}{a^N}(2\pi)^{N/2}\left(\frac{\sigma_1^2\sigma_2^2\sigma_3^2}{\sigma_1^2\sigma_2^2 + \sigma_1^2\sigma_3^2 + \sigma_2^2\sigma_3^2}\right)^{N/2} \times \exp\left(-\frac{1}{2}\sum_i \frac{\sigma_1^2\sigma_2^2\sigma_3^2((\sigma_1^2 + \sigma_3^2)g_i^2 + (\sigma_1^2 + \sigma_2^2)f_i^2 - 2\sigma_1^2 f_i g_i}{\sigma_1^2\sigma_2^2 + \sigma_1^2\sigma_3^2 + \sigma_2^2\sigma_3^2}\right) \tag{7.40}$$

Replace $\sigma_1^2, \sigma_2^2, \sigma_3^2, f_i, g_i$:

$$= (4\pi^2(\sigma_x^2\sigma_y^2 + \sigma_y^2R_x^2 + a^2\sigma_x^2R_x^2))^{-N/2} \times \exp\left(-\frac{1}{2}\sum_i \frac{\sigma_1^2\sigma_2^2\sigma_3^2((\sigma_1^2 + \sigma_3^2)g_i^2 + (\sigma_1^2 + \sigma_2^2)f_i^2 - 2\sigma_1^2 f_i g_i}{\sigma_1^2\sigma_2^2 + \sigma_1^2\sigma_3^2 + \sigma_2^2\sigma_3^2}\right) \tag{7.41}$$

This completes the first Gaussian integral - the nuisance parameters x_i have now been integrated out. The next step is to integrate out the nuisance parameters x_0, y_0. We begin by expanding $\sigma_1^2, \sigma_2^2, \sigma_3^2, f_i, g_i$ and rearranging the resulting expression into a simpler form.

Relabel:

$$\alpha^2 = \frac{a^2R_x^2 + \sigma_y^2}{a^2R_x^2\sigma_x^2 + R_x^2\sigma_y^2 + \sigma_x^2\sigma_y^2} \tag{7.42}$$

$$\beta^2 = \frac{R_x^2 + \sigma_x^2}{a^2R_x^2\sigma_x^2 + R_x^2\sigma_y^2 + \sigma_x^2\sigma_y^2} \tag{7.43}$$

$$\gamma = \frac{aR_x^2}{a^2R_x^2\sigma_x^2 + R_x^2\sigma_y^2 + \sigma_x^2\sigma_y^2} \tag{7.44}$$

so our expression for the joint posterior becomes:

$$\begin{aligned} p(x_0,&y_0, \log a, \log R_x, \log R_y) = \\ &(4\pi^2(a^2\sigma_x^2R_x^2 + \sigma_x^2\sigma_y^2 + \sigma_y^2R_x^2))^{-N/2} \\ &\times \exp(-\frac{1}{2}\sum_i \alpha^2(\hat{x}_i - x_0)^2 + \beta^2(\hat{y}_i - y_0)^2 \\ &- 2\gamma(\hat{x}_i - x_0)(\hat{y}_i - y_0)) \end{aligned} \tag{7.45}$$

The following relations may be derived:

$$\sum_i (\hat{x}_i - x_0)^2 = N[S_{xx}^2 + (\bar{x} - x_0)^2] \tag{7.46}$$

$$\sum_i (\hat{y}_i - y_0)^2 = N[S_{yy}^2 + (\bar{y} - y_0)^2] \tag{7.47}$$

$$\sum_i (\hat{x}_i - x_0)(\hat{y}_i - y_0) = N[S_{xy}^2 + (\bar{x} - x_0)(\bar{y} - y_0)] \tag{7.48}$$

where:

$$S_{xx}^2 = \frac{1}{N} \sum_i (\hat{x}_i - \bar{x})^2 \tag{7.49}$$

$$S_{xy}^2 = \frac{1}{N} \sum_i (\hat{x}_i - \bar{x})(\hat{y}_i - \bar{y}) \tag{7.50}$$

$$S_{yy}^2 = \frac{1}{N} \sum_i (\hat{y}_i - \bar{y})^2 \tag{7.51}$$

Which allows us to separate out the nuisance parameters x_0, y_0 from the other i dependent quantities such that Eq. (7.45) may be written as:

$$\begin{aligned} p(x_0, y_0, \log a, \log R_x, \log R_y) = \\ &= (4\pi^2(a^2\sigma_x^2 R_x^2 + \sigma_x^2\sigma_y^2 + \sigma_y^2 R_x^2))^{-N/2} \\ &\times \exp(-\frac{N}{2}(\alpha^2 S_{xx}^2 + \beta^2 S_{yy}^2 - 2\gamma_{xy}^2) \\ &\times \exp(-\frac{N}{2}(\alpha^2(x_0 - \bar{x})^2 + \beta^2(y_0 - \bar{y})^2 \\ &- 2\gamma(x_0 - \bar{x})(y_0 - \bar{y}))) \end{aligned} \tag{7.52}$$

Now we can write the required integral over the nuisance parameters x_0 and y_0 as:

$$\begin{aligned} p(\log a, \log R_x, \log R_y) = \\ &(4\pi^2(a^2\sigma_x^2 R_x^2 + \sigma_x^2\sigma_y^2 + \sigma_y^2 R_x^2))^{-N/2} \\ &\times \exp(-\frac{N}{2}(\alpha^2 S_{xx}^2 + \beta^2 S_{yy}^2 - 2\gamma S_{xy}^2) \\ &\times \int \exp(-\frac{N}{2}(\alpha^2(x_0 - \bar{x})^2 + \beta^2(y_0 - \bar{y})^2 \\ &- 2\gamma(x_0 - \bar{x})(y_0 - \bar{y}))) \, \mathrm{d}x_0 \, \mathrm{d}y_0 \end{aligned} \tag{7.53}$$

Looking only at the integral part, the exponential term may be rewritten in matrix form:

$$\begin{aligned}\int \exp(-\frac{N}{2}(\alpha^2(x_0-\bar{x})^2+\beta^2(y_0-\bar{y})^2 \\ &- 2\gamma(x_0-\bar{x})(y_0-\bar{y})))\ \mathrm{d}x_0\ \mathrm{d}y_0 \\ &= \int \exp\left(-\frac{1}{2}W^tFW\right)\mathrm{d}^2W\end{aligned} \tag{7.54}$$

Where:

$$W = \begin{bmatrix}(x_0-\bar{x})\\(y_0-\bar{y})\end{bmatrix} \tag{7.55}$$

and

$$F = \begin{bmatrix}N\alpha^2 & -2\gamma N\\ -2\gamma N & N\beta^2\end{bmatrix} \tag{7.56}$$

$$\begin{aligned}\int \exp\left(-\frac{1}{2}W^tFW\right)\mathrm{d}^2W &= \frac{(2\pi)^{2/2}}{\sqrt{\det F}}\\ &= \frac{2\pi}{N}(a^2\sigma_x^2R_x^2+\sigma_x^2\sigma_y^2+\sigma_y^2R_x^2)^{1/2}\end{aligned} \tag{7.57}$$

Substituting the result of the second Gaussian integral Eq. (7.57) back into Eq. (7.53) gives:

$$\begin{aligned}p(\log a,\log R_x,\log R_y) &= \frac{1}{N}(4\pi^2)^{-\frac{1}{2}(N-1)}\\ &\times (a^2\sigma_x^2R_x^2+\sigma_x^2\sigma_y^2+\sigma_y^2R_x^2)^{-\frac{1}{2}(N-1)}\\ &\times \exp(-\frac{N}{2}(\alpha^2S_{xx}^2+\beta^2S_{yy}^2-2\gamma S_{xy}^2)\end{aligned} \tag{7.58}$$

The terms in the exponential may be expanded to give Gull's expression for the marginalized joint posterior distribution:

$$\begin{aligned}&p(\log a,\log R) = \frac{1}{N}(4\pi^2)^{-\frac{1}{2}(N-1)}\\ &\times (a^2\sigma_x^2R_x^2+\sigma_x^2\sigma_y^2+\sigma_y^2R_x^2)^{-\frac{1}{2}(N-1)}\\ &\times \exp(-\frac{1}{2}(\frac{V_{xx}(a^2R_x^2+\sigma_y^2)-2V_{xy}aR_x^2+V_{yy}(R_x^2+\sigma_x^2)}{a^2\sigma_x^2R_x^2+\sigma_x^2\sigma_y^2+\sigma_y^2R_x^2})\end{aligned} \tag{7.59}$$

where: $V_{xx} = NS_{xx}^2$, $V_{xy} = NS_{xy}^2$ and $V_{yy} = NS_{yy}^2$

Substitute for R_x using $R_x^2 = \frac{R^2}{a}$:

$$
\begin{aligned}
p(\log a, \log R) &= \frac{1}{N}(4\pi^2)^{-\frac{1}{2}(N-1)} \\
&\times (a\sigma_x^2 R^2 + \sigma_x^2\sigma_y^2 + a^{-1}\sigma_y^2 R^2)^{-\frac{1}{2}(N-1)} \\
&\times \exp(-\frac{1}{2}(\frac{V_{xx}(aR^2+\sigma_y^2) - 2V_{xy}R^2 + V_{yy}(a^{-1}R^2+\sigma_x^2)}{a\sigma_x^2 R^2 + \sigma_x^2\sigma_y^2 + a^{-1}\sigma_y^2 R^2}) \qquad (7.60)
\end{aligned}
$$

Taking logs:

$$
\begin{aligned}
\log p(\log a, \log R) &= \text{constant} \\
&- \frac{1}{2}(N-1)\log(a\sigma_x^2 R^2 + \sigma_x^2\sigma_y^2 + a^{-1}\sigma_y^2 R^2) \\
&- \frac{1}{2}(\frac{V_{xx}(aR^2+\sigma_y^2) - 2V_{xy}R^2 + V_{yy}(a^{-1}R^2+\sigma_x^2)}{a\sigma_x^2 R^2 + \sigma_x^2\sigma_y^2 + a^{-1}\sigma_y^2 R^2}) \qquad (7.61)
\end{aligned}
$$

where the constant is a constant for a given N. Eq. (7.61) is Gull's expression for the log of the marginalized joint posterior distribution.

7.4.4 Gull's estimator for a

As an approximate estimator for a, for the case where $R \to \infty$, Gull recommends dropping the determinant term from Eq. (7.61) and using:

$$
= \min \frac{aV_{xx} - 2V_{xy} + a^{-1}V_{yy}}{a\sigma_x^2 + a^{-1}\sigma_y^2} \qquad (7.62)
$$

Gull's estimator for a Eq. 7.62 can be re written as

$$
= \min \sum_i \frac{(\hat{y}_i - a\hat{x}_i - \langle b \rangle)^2}{(\sigma_y^2 + a^2\sigma_x^2)} \qquad (7.63)
$$

This is essentially the χ_x^2 expression which looks similar to the log of the D'Agostini's expression in Eq. (7.14) with the prefactor dropped, but the important difference is that here b is replaced by an estimator $\langle b \rangle$ where

$$
\langle b \rangle = \bar{y} - a\bar{x} \qquad (7.64)
$$

where $\bar{x}$ and $\bar{y}$ are the means of the observed values $\hat{x}$ and $\hat{y}$ respectively. Although the estimator $\langle b \rangle$ does depend on a, $\langle b \rangle$ changes only slowly with a. Depending on the data set, a 100% change in a causes only a 10% or 1% change in $\langle b \rangle$. This means that even when a reasonable range for a is chosen, $\langle b \rangle$ will remain close to the true value of b. This is effectively like minimizing the simple chisquare χ_x^2 expression with b fixed at or near the true value of b.

7.4.5 Comparison between Gull's estimator for a and the marginalized simple χ^2 expression

For the linear model Eq. (7.10), the simple chisquare expression is:

$$p(a, b|\hat{x}, \hat{y}, \sigma_x, \sigma_y) = \exp\left(-\frac{1}{2}\sum_i \frac{(\hat{y}_i - a\hat{x}_i - b)^2}{\sigma_y^2 + a^2\sigma_x^2}\right) \tag{7.65}$$

In order to compare the simple chisquare expression Eq. (7.65) with Gull's estimator for a, we integrate b out of the simple chisquare expression in Eq. (7.65)

$$p(a|\hat{x}, \hat{y}) = \int \exp\left(-\frac{1}{2}\sum_i \frac{(\hat{y}_i - a\hat{x}_i - b)^2}{(\sigma_y^2 + a^2\sigma_x^2)}\right) \mathrm{d}b \tag{7.66}$$

relabel:

$$\hat{\gamma}_i = \hat{y}_i - a\hat{x}_i \tag{7.67}$$

$$S^2 = \sigma_y^2 + a^2\sigma_x^2 \tag{7.68}$$

separate b from i dependent quantities:

$$\begin{aligned} p(a|\hat{x}, \hat{y}) &= \int \mathrm{d}b \exp\left(-\frac{1}{2}\sum_i \frac{(\hat{\gamma}_i - b)^2}{S^2}\right) \\ &= \int \mathrm{d}b \exp\left(-\frac{1}{2}\left(\frac{N}{S^2}(b - \bar{\gamma})^2\right)\right) \\ &\quad \times \exp\left(-\frac{1}{2}\left(\sum_i \frac{\hat{\gamma}_i^2}{S^2} - N\frac{\bar{\gamma}^2}{S^2}\right)\right) \end{aligned} \tag{7.69}$$

where:

$$\bar{\gamma} = \frac{\sum_i \hat{\gamma}_i}{N} \tag{7.70}$$

perform Gaussian integral over b:

$$p(a|\hat{x}, \hat{y}) = \sqrt{\left(\frac{2\pi S^2}{N}\right)} \exp\left(-\frac{1}{2}\left(\frac{\sum_i \hat{\gamma}_i^2}{S^2} - \frac{N\bar{\gamma}^2}{S^2}\right)\right) \tag{7.71}$$

replace relabeled quantities and write in terms of estimator $\langle b \rangle$ as defined in Eq. (7.64):

$$p(a|\hat{x},\hat{y}) = \sqrt{\frac{2\pi}{N}(a^2\sigma_x^2 + \sigma_y^2)} \exp\left(-\frac{1}{2}\left(\frac{\sum_i(\hat{y}_i - a\hat{x}_i - \langle b \rangle)^2}{a^2\sigma_x^2 + \sigma_y^2}\right)\right) \tag{7.72}$$

which is similar to Gull's estimator for a. The exponential terms are identical, but the marginalized chisquare expression has an additional prefactor. This shows the relationship between Gull's joint posterior and D'Agostini's proposed joint posterior. The D'Agostini solution without the prefactor is similar to Gull's approximate case, and this is essentially analogous to the simple χ^2 method generally used in the SNe Ia analysis as described in section 7.3. D'Agostini's full posterior gives biased results when the errors are large compared to the range of the data points.

figs. 7.6 and 7.7 show a comparison between the results using the 1D marginalised posterior from the full Gull expression Eq. (7.60) and the profile likelihood from the simple χ_x^2 expression Eq.(7.15), for two different data sets. The two data sets and their respective errors are plotted in the upper left panels of figs. 7.6 and 7.7. Data set 1, fig. 7.6 has large errors in the x direction and small errors in the y direction whereas data set 2 fig. 7.7 is more extreme having large errors in both the x and y directions.

Dropping the prefactor from the D'Agostini posterior provides an apparently unbiased approximation in some cases, as shown in fig. 7.6 for the less extreme data set, where the χ^2 profile likelihood follows closely the 1D marginalized posterior from the full Gull method. However, this approximation breaks down with more extreme data sets e.g. as shown in fig. 7.7, here the profile likelihood from the χ^2 method diverges from the Gull 1D marginalized posterior: although the Gull 1D marginalized posterior is broader (i.e. puts weaker constraints on the a parameter), it provides an estimate for a which is closer to the true value of a. It is better to have weaker constraints and a more accurate estimator than to have tighter constraints con-straining a to the wrong value.

To summarize, there are three proposed expressions for the joint posterior for the linear toy model with errors on both the x and y axes, and only one of these, the full expression from Gull (1989) gives correct results for extreme data sets. To summarise, these three proposed expressions are:

(i) The D'Agostini (1995) proposed Bayesian solution, including prefactor, as given in Eq.(7.14). This gives biased results when there are errors in both directions and those errors are large compared with the range of the data points, as shown in panels 3 and 5 of fig. 7.4.

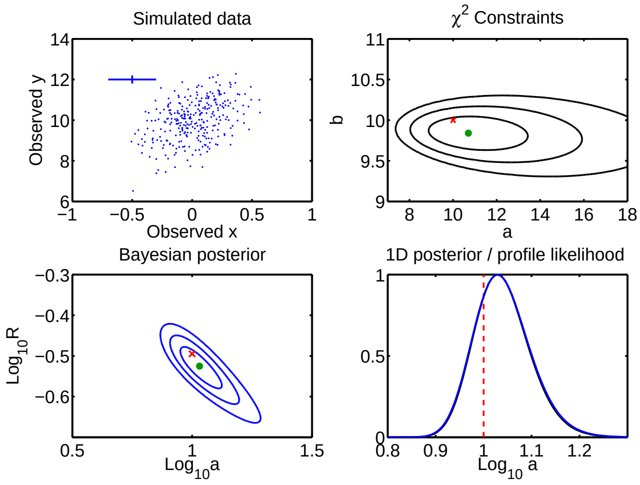


Figure 7.6: Numerical comparison between the χ^2 approach and the Bayesian method for linear fitting. Upper left panel: data set of $N = 300$ observations with errors both in the x and y directions, given by the error bar. Upper right panel: reconstruction of the slope a and intercept b using a χ^2 likelihood (red cross is the true value, green circle the maximum likelihood value). Lower left panel: Gull's Bayesian posterior (Eq. (7.60)) in the $\log a$, $\log R$ plane, with green circle showing posterior mean. In both panels, contours enclose 1σ, 2σ and 3σ regions. Lower right panel: marginalized Bayesian posterior (blue) and profiled χ^2 likelihood (black, lying exactly on top of the blue curve), with the dashed line showing the true value. The two methods give essentially identical results in this case.

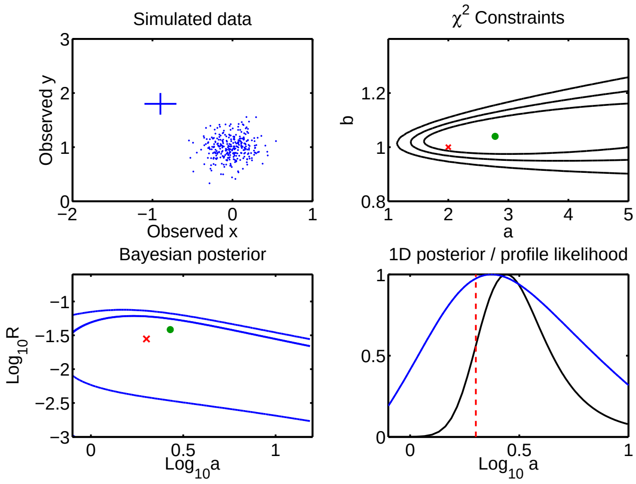


Figure 7.7: As in Fig. 7.7, but now with a larger statistical uncertainty in the data compared to their spread. The Bayesian marginal posterior for the slope (blue, lower right panel) peaks much closer to the true value than the χ^2 expression (black). In the lower left panel, the 3σ contour from the Bayesian method lies outside the range of the plot.

(ii) The Gull (1989) Bayesian solution, as given in Eq.(7.60). Initial trials show this method to give unbiased results even for more extreme data sets where the errors on both axes are large compared with the range of data points.

(iii) The approximate Gull (1989) solution, as given in Eq. 7.62 which is similar to the D'Agostini expression without the prefactor Eq.(7.14) or the simple χ^2 expression given in Eq. (7.15). This expression is valid in certain limits and does not suffer from the extreme bias of the full D'Agostini method, but does break down for more extreme data sets. The exact limits for the validity of this approximation have not been fully tested, but Gull (1989) suggests that it is valid for the case where $R \to \infty$. More extensive numerical trial described in the later sections of this work make a more detailed analysis as to whether the simple χ^2 methodology is valid for the levels of uncertainty found in SNe Ia data sets.

7.5 Bayesian Hierarchical Model (BHM): Description

The method which I present here takes the methodology of Gull (1989) and applies it to the SNe Ia case. We use a similar initial set up to the χ^2 approach described above, with one important difference. Whereas the in the χ^2 approach M_0 appears as a global fit parameter explicitly in Eq. (4.6), where it is the mean absolute magnitude of the SNe Ia population, we use M_i to represent the absolute magnitude of each individual SNe Ia. M_i is different for each SNe Ia because even after correction for colour and stretch some variation remains in the SNe Ia population. (M_0 also appears explicitly in our method as the mean absolute magnitude of the SNe Ia population in Eq. (7.74)). Our version of Eq. (4.6) is:

$$\mu_i = m^*_{Bi} - M_i + \alpha \cdot x_{1i} - \beta \cdot c_i \tag{7.73}$$

The graphical representation of the Bayesian hierachical network in fig. 7.8 shows the dependencies of the parameters within the problem. One can see that each SNe Ia has a true (unobserved) redshift z_i, and a true absolute magnitude M_i. The M_i are drawn from a Gaussian distribution with mean M_0 and standard deviation σ_μ^{int}.

$$M_i \sim \mathcal{N}(M_0, (\sigma_\mu^{\text{int}})^2). \tag{7.74}$$

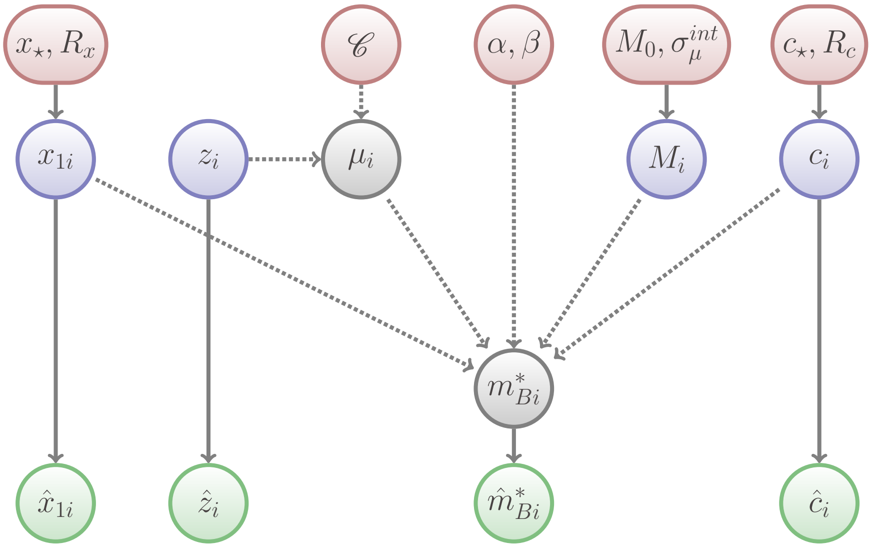


Figure 7.8: Graphical representation of the Bayesian hierarchical network showing the deterministic (dashed) and probabilistic (solid) connections between variables in our Bayesian hierarchical model (BHM). Variables of interest are in red, latent (unobserved) variables are in blue and observed data (denoted by hats) are in green.

This σ_μ^{int} is the intrinsic dispersion of the absolute magnitudes which remains in the SNe Ia absolute magnitudes even after correction for stretch and colour, σ_μ^{int} characterizes the scatter in absolute magnitudes remaining in the stretch corrected light curves shown in the lower panel of fig.4.4.

The cosmological parameters $\mathscr{C} = \{\Omega_m, \Omega_\Lambda \text{ or } w, h\}$ are unknown as are the SNe Ia global fit parameters α, β, these are the parameters we would like to infer. If $\mathscr{C}$ and z_i were known, then we could deterministically specify μ_i using Eq. (7.5). Each SNe Ia also has its own stretch parameter x_{1i} and colour parameter c_i, drawn from their parent distributions. In this work, we model the distributions of the true stretch and colour parameters as Gaussians parameterised each by a mean $(c_\star, x_\star)$ and a variance (R_c^2, R_x^2) as

$$c_i \sim \mathcal{N}(c_\star, R_c^2), \quad x_{1i} \sim \mathcal{N}(x_\star, R_x^2) \tag{7.75}$$

The choice of a Gaussian distribution for the latent variables $\underline{c}$ and $\underline{x}_1$ is justified by the fact that the observed distribution of $\underline{\hat{c}}$ and $\underline{\hat{x}}_1$, shown in Fig. 7.9 for the actual SNIa sample described in section 7.7 below, is fairly well described by a Gaussian. As shown in Fig. 7.9, there might be a hint for a heavier tail for positive values of $\underline{\hat{c}}$, but this does not fundamentally invalidate our Gaussian approximation. It would be easy to expand our method to consider other distributions, for example mixture models of Gaussians to describe a more complex population or a distribution with heavier tails, if non-Gaussianities in the observed distribution should make such modelling necessary. In our work, we consider the simple uni-modal Gaussians given by Eq. (7.75). If the true values discussed so far were known, then we could deterministically specify m^*_{Bi}. But we do not have the latent or true values (blue circles), we only have the corresponding observed or fitted values $\hat{z}_i, \hat{c}_i, \hat{x}_{1i}, \hat{m}^*_{Bi}$. The true redshift z_i of each SNe Ia is subject to a small amount of Gaussian noise giving us a slightly different observed redshift z_i. Likewise the true stretch, colour and maximum magnitudes are also subject to uncertainties in the fitting process, giving us the measured values $\hat{c}_i, \hat{x}_{1i}, \hat{m}^*_{Bi}$ (green circles). The problem of parameter inference we now have is how to obtain the parameters of interest $\mathscr{C}, \alpha, \beta$ given the observed or measured data $\hat{z}_i, \hat{c}_i, \hat{x}_{1i}, \hat{m}^*_{Bi}$.

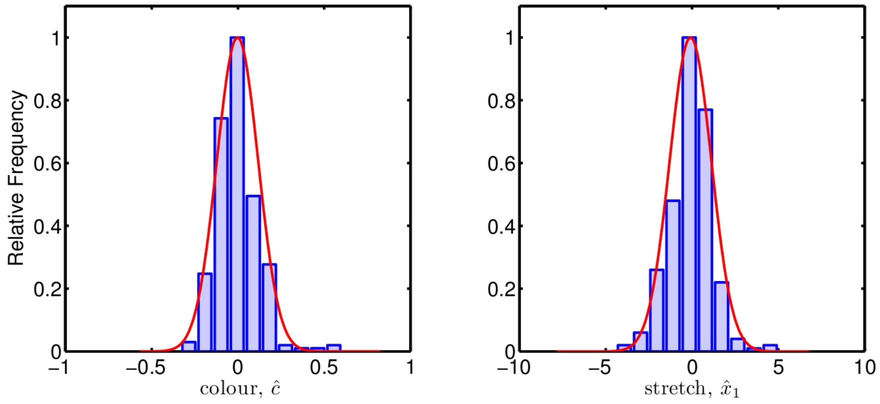


Figure 7.9: Histogram of observed stretch parameters $\hat{x}_{1i}$ and observed colour parameters $\hat{c}_i$ from the 288 SNIa from Kessler et al. (2009a), compared with a Gaussian fit (red curve).

7.6 Bayesian Hierarchical Model: Calculation

Having described the Bayesian Hierarchical Model (BHM) up in terms of the graphical representation of the Bayesian hierarchical network, I will now present the details of the calculation. The purpose of this calculation is to determine the joint posterior probability of the parameters of interest $\Theta = \{\mathscr{C}, \alpha, \beta, \sigma_\mu^{\rm int}\}$ given that we have the measured values D, that is, we wish to determine $p(\Theta|D)$.

7.6.1 A note on notation

Throughout this work we use the notation:

$$x \sim \mathcal{N}(m, \sigma^2) \tag{7.76}$$

to denote a random variable x being drawn from an underlying Gaussian distribution with mean m and variance σ^2. In vector notation, m is replaced by a vector $\underline{m}$, while σ^2 is replaced by the covariance matrix Σ.

$$\underline{x} \sim \mathcal{N}(\underline{m}, \Sigma) \tag{7.77}$$

where $\underline{x}$ has the probability density function

$$p(\underline{x}) \equiv |2\pi\Sigma|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\underline{x} - \underline{m})^T \Sigma^{-1} (\underline{x} - \underline{m})\right] \tag{7.78}$$

and we use the compressed notation to write the probability density function as:

$$\mathcal{N}_{\underline{x}}(\underline{m}, \Sigma) \equiv |2\pi\Sigma|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\underline{x} - \underline{m})^T \Sigma^{-1} (\underline{x} - \underline{m})\right] \tag{7.79}$$

7.6.2 Calculation expressed in matrix notation

We re-write Eq. (7.74) and (7.75) in matrix notation as:

$$\underline{M} \sim \mathcal{N}(\underline{M}_0, \Sigma_\Delta), \tag{7.80}$$

$$\underline{c} \sim \mathcal{N}(c_\star \cdot \mathbf{1}_n, \mathrm{diag}\big(R_c^2 \cdot \mathbf{1}_n\big)) \tag{7.81}$$

$$\underline{x}_1 \sim \mathcal{N}(x_\star \cdot \mathbf{1}_n, \mathrm{diag}\big(R_x^2 \cdot \mathbf{1}_n\big)) \tag{7.82}$$

where

$$\underline{M} = (M_1, \dots, M_n) \in \mathbb{R}^n, \tag{7.83}$$

$$\underline{M}_0 = M_0 \cdot \mathbf{1}_n \in \mathbb{R}^n, \tag{7.84}$$

$$\Sigma_\Delta = \mathrm{diag}\big((\sigma_\mu^{\rm int})^2 \cdot \mathbf{1}_n\big) \in \mathbb{R}^{n\times n}. \tag{7.85}$$

Having introduced $3n$ latent (unobserved) variables $(\underline{c}, \underline{x}_1, \underline{M})$, where n is the number of SNe in the sample, the fundamental strategy of our method is to link them to underlying population parameters via Eqs. (7.74) and (7.75), then to use the observed noisy estimates to infer constraints on the population parameters of interest (alongside the cosmological parameters), while marginalizing out the unobserved latent variables.

In doing so we are following the methodology of Gull (1989) and essentially placing a prior on the range of the latent (true) quantities $\underline{c}, \underline{x}_1, \underline{M}$.

$$p(\underline{c}|c_\star, R_c) = \mathcal{N}_{\underline{c}}(c_\star \cdot \mathbf{1}_n, \mathrm{diag}\big(R_c^2 \cdot \mathbf{1}_n\big)) \tag{7.86}$$

$$p(\underline{x}_1|x_\star, R_x) = \mathcal{N}_{\underline{x}_1}(x_\star \cdot \mathbf{1}_n, \mathrm{diag}\big(R_x^2 \cdot \mathbf{1}_n\big)) \tag{7.87}$$

It is necessary to apply this prior because two conditions are fulfilled (1) there are errors on both the measured values $\hat{\underline{c}}, \hat{\underline{x}}_1$ and (2) the errors on $\hat{\underline{c}}, \hat{\underline{x}}_1$ are large compared with the range of $\hat{\underline{c}}, \hat{\underline{x}}_1$. Failure to apply this prior on the range of the latent quantities when these two conditions are met results in a biased recovery of the SNe Ia population parameters α, β. For further discussion of this crucial step, see appendix.

We chose to model the probability of the true absolute magnitudes $\underline{M}$ also as a Gaussian.

$$p(\underline{M}|M_0, \sigma_\mu^{\mathrm{int}}) = \mathcal{N}_{\underline{M}}(M_0 \cdot \mathbf{1}_n, \mathrm{diag}\big((\sigma_\mu^{\mathrm{int}})^2 \cdot \mathbf{1}_n\big)) \tag{7.88}$$

$$\tag{7.89}$$

Notice that there are two levels of specification or choice here: (1) The choice of the underlying distributions, described by Eq. (7.80) (7.82) (7.81) and (2) The choice of the priors on those latent parameters. Throughout this work we assume that (1) are Gaussian, and when trialling the method with simulated data, we choose Gaussians for (1). We also chose to use Gaussian priors for (2). Obviously the natural choice is to match the shape of prior with the shape of the underlying distribution, this is possible when using simulated data, but with the real data the precise shape of the unknown distribution is unknown but assumed Gaussian. An interesting question is what happens when a different shaped prior is chosen from the underlying distribution e.g. what happens if a uniformly distributed population of $\underline{c}$ are used with a Gaussian prior on $\underline{c}$ - a question which I will investigate in future work.

The absolute magnitude M_i is related to the observed B-band magnitude $\hat{m}_B^*$ and the distance modulus μ by Eq. (7.73), which can be rewritten in vector notation as:

$$\underline{m_B^*} = \underline{\mu} + \underline{M} - \alpha \underline{x}_1 + \beta \underline{c} \tag{7.90}$$

The above relation is *exact*, i.e. $\underline{M}, \underline{x}_1, \underline{c}$ are here the latent variables (not the observed quantities), while $\underline{m_B^*}$ is the true value of the B-band magnitude (also unobserved). This is represented by the dotted (deterministic) arrows connecting the variables in Fig. 7.8.

We seek to determine the posterior pdf for the parameters of interest $\Theta = \{\mathscr{C}, \alpha, \beta, \sigma_\mu^{\text{int}}\}$, while marginalizing over the unknown population mean absolute magnitude, M_0. From Bayes theorem, the marginal posterior for Θ is given by:

$$p(\Theta|D) = \int \mathrm{d}M_0 p(\Theta, M_0|D) = \int \mathrm{d}M_0 \frac{p(D|\Theta, M_0)p(\Theta, M_0)}{p(D)}, \tag{7.91}$$

where $p(D)$ is the Bayesian evidence (a normalizing constant) and the prior $p(\Theta, M_0)$ can be written as

$$p(\Theta, M_0) = p(\mathscr{C}, \alpha, \beta)p(M_0, \sigma_\mu^{\text{int}}) = p(\mathscr{C}, \alpha, \beta)p(M_0|\sigma_\mu^{\text{int}})p(\sigma_\mu^{\text{int}}). \tag{7.92}$$

We take a uniform prior on the variables $\mathscr{C}, \alpha, \beta$ (on a sufficiently large range so as not to truncate likelihood, except in the case of $\Omega_m > 0$ which rules out an unphysical choice of parameter), as well as a Gaussian prior for $p(M_0|\sigma_\mu^{\text{int}})$, since M_0 is a location parameter of a Gaussian (conditional on σ_μ^{int}). Additionally, we apply a prior which excludes that part of parameter space which is the 'no Big Bang' region in the Ω_m, Ω_Λ plane.

The prior on M_0 is

$$p(M_0|\sigma_\mu^{\text{int}}) = \mathcal{N}_{M_0}(M_m, \sigma_{M_0}^2), \tag{7.93}$$

where the mean of the prior ($M_m = -19.3$ mag) is taken to be a reasonable value based on observations of nearby SNe Ia, and the variance ($\sigma_{M_0} = 2.0$ mag) is sufficiently large so that the prior is very diffuse and non-informative (the precise choice of mean and variance for this prior does not impact on our numerical results). Finally, the appropriate prior for σ_μ^{int} is a Jeffreys' prior, i.e., uniform in $\log \sigma_\mu^{\text{int}}$, as σ_μ^{int} is a scale parameter.

The likelihood $p(D|\Theta, M_0) = p(\underline{\hat{c}}, \underline{\hat{x}}_1, \underline{\hat{m}_B^*}|\Theta, M_0)$ can be expanded out an rewritten as:

$$p(\underline{\hat{c}}, \underline{\hat{x}}_1, \underline{\hat{m}}_B^* | \Theta, M_0) = \int \mathrm{d}\underline{c}\ \mathrm{d}\underline{x}_1\ \mathrm{d}\underline{M}\ p(\underline{\hat{c}}, \underline{\hat{x}}_1, \underline{\hat{m}}_B^* | \underline{c}, \underline{x}_1, \underline{M}, \Theta, M_0) p(\underline{c}, \underline{x}_1, \underline{M} | \Theta, M_0) \tag{7.94}$$

$$\begin{aligned} &= \int \mathrm{d}\underline{c}\ \mathrm{d}\underline{x}_1\ \mathrm{d}\underline{M}\ p(\underline{\hat{c}}, \underline{\hat{x}}_1, \underline{\hat{m}}_B^* | \underline{c}, \underline{x}_1, \underline{M}, \Theta) \\ &\times \int \mathrm{d}R_c\ \mathrm{d}R_x\ \mathrm{d}c_\star\ \mathrm{d}x_\star\ \ p(\underline{c} | c_\star, R_c) p(\underline{x}_1 | x_\star, R_x) p(\underline{M} | M_0, \sigma_\mu^{\mathrm{int}}) \\ &\times p(R_c) p(R_x) p(c_\star) p(x_\star) \end{aligned} \tag{7.95}$$

In the first line, we have introduced a set of $3n$ latent variables, $\{\underline{c}, \underline{x}_1, \underline{M}\}$, which describe the *true* value of the colour, stretch and absolute magnitude for each SNIa. Since these variables are unobserved, we need to marginalize over them. In the second line, we have replaced $p(\underline{c}, \underline{x}_1, \underline{M} | \Theta, M_0)$ by the priors on the latent $\{\underline{c}, \underline{x}_1, \underline{M}\}$ given by Eq. (7.88) and Eqs. (7.86–7.87), (assumed separable) and marginalized out the population parameters $\{R_c, R_x, c_\star, x_\star\}$:

$$p(\underline{c}, \underline{x}_1, \underline{M} | \Theta, M_0) = \int \mathrm{d}R_c\ \mathrm{d}R_x\ \mathrm{d}c_\star\ \mathrm{d}x_\star\ p(\underline{c} | c_\star, R_c) p(\underline{x}_1 | x_\star, R_x) \tag{7.96}$$

$$\times\, p(\underline{M} | M_0, \sigma_\mu^{\mathrm{int}}) p(R_c) p(R_x) p(c_\star) p(x_\star) \tag{7.97}$$

(we have also dropped M_0 from the likelihood, as conditioning on M_0 is irrelevant if the latent $\underline{M}$ are given). If we further marginalize over M_0 (as in Eq. (7.91), including the prior on M_0), the expression for the effective likelihood, Eq. (7.95), then becomes:

$$\begin{aligned} p(\underline{\hat{c}}, \underline{\hat{x}}_1, \underline{\hat{m}}_B^* | \Theta) &= \int \mathrm{d}\underline{c}\ \mathrm{d}\underline{x}_1\ \mathrm{d}\underline{M}\ p(\underline{\hat{c}}, \underline{\hat{x}}_1, \underline{\hat{m}}_B^* | \underline{c}, \underline{x}_1, \underline{M}, \Theta) \\ &\times \int \mathrm{d}R_c\ \mathrm{d}R_x\ \mathrm{d}c_\star\ \mathrm{d}x_\star\ \mathrm{d}M_0\ p(\underline{c} | c_\star, R_c) p(\underline{x}_1 | x_\star, R_x) \end{aligned} \tag{7.98}$$

$$\times\, p(\underline{M} | M_0, \sigma_\mu^{\mathrm{int}}) p(R_c) p(R_x) p(c_\star) p(x_\star) p(M_0 | \sigma_\mu^{\mathrm{int}}) \tag{7.99}$$

The term $p(\underline{\hat{c}}, \underline{\hat{x}}_1, \underline{\hat{m}}_B^* | \underline{c}, \underline{x}_1, \underline{M}, \Theta)$ is the conditional probability of observing values $\{\underline{\hat{c}}, \underline{\hat{x}}_1, \underline{\hat{m}}_B^*\}$ if the latent (true) value of $\underline{c}, \underline{x}_1, \underline{M}$ and of the other cosmological parameters were known. From Fig. 7.8, $\underline{m}_B^*$ is connected only deterministically to all other variables and parameters, via Eq. (7.90). Thus we can replace

$\underline{m}_B^* = \underline{\mu} + \underline{M} - \alpha \cdot \underline{x}_1 + \beta \cdot \underline{c}$ and write

$$p(\hat{\underline{c}}, \hat{\underline{x}}_1, \hat{\underline{m}}_B^* | \underline{c}, \underline{x}_1, \underline{M}, \Theta) = \prod_{i=1}^{n} \mathcal{N}_{[\hat{\underline{c}}, \hat{\underline{x}}_1, \hat{\underline{m}}_B^*]}(\mu_i + M_i - \alpha \cdot x_{1i} + \beta \cdot c_i, \hat{C}_i) \tag{7.100}$$

$$= |2\pi\Sigma_C|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}[(X - X_0)^T \Sigma_C^{-1} (X - X_0)]\right) \tag{7.101}$$

where $\mu_i \equiv \mu_i(z_i, \Theta)$ and we have defined

$$X = \{X_1, \dots, X_n\} \in \mathbb{R}^{3n}, \quad X_0 = \{X_{0,1}, \dots, X_{0,n}\} \in \mathbb{R}^{3n}, \tag{7.102}$$

$$X_i = \{c_i, x_{1,i}, (M_i - \alpha x_{1,i} + \beta c_i)\} \in \mathbb{R}^3, \quad X_{0,i} = \{\hat{c}_i, \hat{x}_{1,i}, \hat{m}_{Bi}^* - \mu_i\} \in \mathbb{R}^3, \tag{7.103}$$

as well as the $3n \times 3n$ block covariance matrix[1]

$$\Sigma_C = \begin{pmatrix} \hat{C}_1 & 0 & 0 & 0 \\ 0 & \hat{C}_2 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \hat{C}_n \end{pmatrix}. \tag{7.104}$$

Finally we explicitly include redshift uncertainties in our formalism. The observed apparent magnitude, $\hat{\underline{m}}_B^*$, on the left-hand-side of Eq. (7.100), is the value at the observed redshift, $\hat{\underline{z}}$. However, $\underline{\mu}$ in Eq. (7.100) should be evaluated at the true (unknown) redshift, $\underline{z}$. As above, the redshift uncertainty is included by introducing the latent variables $\underline{z}$ and integrating over them:

$$p(\underline{c}, \underline{x}_1, \underline{M} | \underline{c}, \underline{x}_1, \underline{M}, \Theta) = \int \mathrm{d}\underline{z} \;\; p(\underline{c}, \underline{x}_1, \underline{M} | \underline{c}, \underline{x}_1, \underline{M}, \underline{z}, \Theta) p(\underline{z}|\hat{\underline{z}}) \tag{7.105}$$

where we model the redshift errors $p(\underline{z}|\hat{\underline{z}})$ as Gaussians:

$$\hat{\underline{z}} \sim \mathcal{N}(\underline{z}, \Sigma_z) \tag{7.106}$$

$$p(\underline{z}|\hat{\underline{z}}) = \mathcal{N}_{\underline{z}}(\hat{\underline{z}}, \Sigma_z) \tag{7.107}$$

with a $n \times n$ covariance matrix:

$$\Sigma_z = \mathrm{diag}(\sigma_{z_1}^2, \dots, \sigma_{z_n}^2). \tag{7.108}$$

---

[1] Notice that we neglect correlations between different SNIa, which is reflected in the fact that Σ_C takes a block-diagonal form. It would be however very easy to add arbitrary cross-correlations to our formalism (e.g., coming from correlated systematic within survey, for example zero point calibration) by adding such non-block diagonal correlations to Eq. (7.104).

It is now necessary to integrate out all latent variables and nuisance parameters from the expression for the likelihood, Eq. (7.99). This can be done analytically, as all necessary integrals are Gaussian.

7.6.3 Integration over intrinsic redshifts

In order to perform the multi-dimensional integral over $\underline{z}$, we Taylor expand $\underline{\mu}$ around $\underline{\hat{z}}$ (as justified by the fact that redshift errors are typically small: the error from 300 km/s peculiar velocity is $\sigma_{z_i} = 0.0012$, while the error from spectroscopic redshifts from SNe themselves is $\sigma_{z_i} = 0.005$, see Kessler et al. (2009a)):

$$\begin{aligned}\mu_j &= \mu(z_j) && (7.109)\\ &= 5\log_{10}\left(\frac{D_L(z_j)}{\text{Mpc}}\right) + 25 && (7.110)\\ &\approx \mu(\hat{z}_j) + 5(\log_{10} e)\left.\frac{\partial_{z_j} D_L(z_j)}{D_L(z_j)}\right|_{\hat{z}_j}(z_j - \hat{z}_j). && (7.111)\end{aligned}$$

With this approximation we can now carry out the multi-dimensional integral of Eq. (7.105), obtaining

$$\begin{aligned}p(\underline{\hat{m}}_B^* | \underline{c}, \underline{x}_1, \underline{M}, \Theta) = {} & |2\pi\Sigma_m|^{-\frac{1}{2}}\\ & \times \exp\left[-\frac{1}{2}(\underline{\hat{m}}_B^* - (\underline{\mu} + \underline{M} - \alpha\cdot\underline{x}_1 + \beta\cdot\underline{c}))^T \Sigma_m^{-1}\right.\\ & \left.(\underline{\hat{m}}_B^* - (\underline{\mu} + \underline{M} - \alpha\cdot\underline{x}_1 + \beta\cdot\underline{c}))\right] \qquad (7.112)\end{aligned}$$

where from now on, $\underline{\mu} = \underline{\mu}(\underline{\hat{z}})$ and

$$\begin{aligned}\sigma_{m_B^* i} &\rightarrow \sigma_{m_B^* i}^{\text{raw data}} + f_i \sigma_{zi} f_i && (7.113)\\ f &= \text{diag}(f_1, \ldots, f_n) && (7.114)\\ f_i &= 5\log_{10}(e)\left.\frac{D_L'(z_i)}{D_L(z_i)}\right|_{\hat{z}_i} && (7.115)\\ &= \frac{5\log_{10}(e)}{D_L(\hat{z}_i)}\left[\frac{D_L(\hat{z}_i)}{1+z_i} + \frac{c}{H_0}(1+\hat{z}_i)\right. && (7.116)\\ &\times \text{cosn}\{\sqrt{|\Omega_\kappa|}\int_0^{\hat{z}} \mathrm{d}z' \left[(1+z')^3\Omega_m + \Omega_{\text{de}}(z) + (1+z)^2\Omega_\kappa\right]^{-1/2}\} && (7.117)\\ &\left.\times((1+z')^3\Omega_m + \Omega_{\text{de}}(z) + (1+z)^2\Omega_\kappa]^{-1/2})\right] && (7.118)\end{aligned}$$

Strictly speaking, one should integrate over redshift in the range $0 \leq z_i < \infty$, not $-\infty < z_i < \infty$, which would result in the appearance of Gamma functions in the final result. However, as long as $\frac{\sigma_{z_i}}{z_i} \ll 1$ (as is the case here), this approximation is expected to be excellent.

7.6.4 Integration over latent $\{c_\star, x_\star, \underline{M}\}$

From Eq. (7.95) and using the expression in Eq. (7.100), we wish to integrate out the latent variables

$$Y = \{Y_1, \dots, Y_n\} \in \mathbb{R}^{3n}, \tag{7.119}$$

$$Y_i = \{c_i, x_{1,i}, M_i\} \in \mathbb{R}^3, \tag{7.120}$$

$$\tag{7.121}$$

We therefore recast expression (7.100) as

$$p(\underline{\hat{c}}, \underline{\hat{x}}_1, \underline{\hat{m}}_B^* | \underline{c}, \underline{x}_1, \underline{M}, \Theta) = |\Sigma_C|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}[(AY - X_0)^T \Sigma_C^{-1} (AY - X_0)]\right) \tag{7.122}$$

where we have defined the block-diagonal matrix

$$A = \mathrm{diag}(T, T, \dots, T) \in \mathbb{R}^{3n \times 3n} \tag{7.123}$$

with

$$T = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \beta & -\alpha & 1 \end{bmatrix} \begin{bmatrix} c_i \\ x_i \\ M_i \end{bmatrix} \tag{7.124}$$

The prior terms appearing in Eq. (7.99), namely $p(\underline{c}|c_\star, R_c) p(\underline{x}_1|x_\star, R_x) p(\underline{M}|M_0, \sigma_\mu^{\mathrm{int}})$, may be written as:

$$p(\underline{c}|c_\star, R_c) p(\underline{x}_1|x_\star, R_x) p(\underline{M}|M_0, \sigma_\mu^{\mathrm{int}}) = |2\pi\Sigma_P|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}[(Y - Y_*)^T \Sigma_P^{-1} (Y - Y_*)]\right) \tag{7.125}$$

where

$$S^{-1} = \mathrm{diag}\big(R_c^{-2}, R_x^{-2}, (\sigma_\mu^{\mathrm{int}})^{-2}\big) \in \mathbb{R}^{3\times 3} \tag{7.126}$$

$$\Sigma_P^{-1} = \mathrm{diag}\big(S^{-1}, S^{-1}, \dots, S^{-1}\big) \in \mathbb{R}^{3n\times 3n} \tag{7.127}$$

$$Y_* = \underline{J} \cdot \underline{b} \in \mathbb{R}^{3n\times 1} \tag{7.128}$$

$$\underline{J} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ \vdots & \vdots & \vdots \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \in \mathbb{R}^{3n\times 3}, \tag{7.129}$$

$$\underline{b} = \begin{bmatrix} c_* \\ x_* \\ M_0 \end{bmatrix} \in \mathbb{R}^{3\times 1}. \tag{7.130}$$

Now the integral over $\mathrm{d}Y = \mathrm{d}\underline{c}\ \mathrm{d}\underline{x}_1\ \mathrm{d}\underline{M}$ in Eq. (7.99) can be performed, giving:

$$\begin{aligned} &\int \mathrm{d}Y\ p(\hat{\underline{c}},\hat{\underline{x}}_1, \hat{\underline{m}}^*_B | \underline{c}, \underline{x}_1, \underline{M}, \Theta) p(\underline{c}|c_\star, R_c) p(\underline{x}_1|x_\star, R_x) p(\underline{M}|M_0, \sigma_\mu^{\mathrm{int}}) = \\ &\qquad = |2\pi\Sigma_C|^{-\frac{1}{2}} |2\pi\Sigma_P|^{-\frac{1}{2}} |2\pi\Sigma_A|^{\frac{1}{2}} \exp\left(-\frac{1}{2}[X_0^T\Sigma_C^{-1}X_0 - Y_0^T\Sigma_A^{-1}Y_0 + Y_*^T\Sigma_P^{-1}Y_*]\right) \end{aligned} \tag{7.131}$$

where

$$\Sigma_A^{-1} = A^T\Sigma_C^{-1}A + \Sigma_P^{-1} \in \mathbb{R}^{3n\times 3n}, \tag{7.132}$$

$$\Sigma_A^{-1}Y_0 = A^T\Sigma_C^{-1}X_0 + \Sigma_P^{-1}Y_*, \tag{7.133}$$

$$Y_0 = \Sigma_A(A^T\Sigma_C^{-1}X_0 + \Sigma_P^{-1}Y_*)\Sigma_A(\Delta + \Sigma_P^{-1}Y_*), \tag{7.134}$$

$$\Delta = A^T\Sigma_C^{-1}X_0 \in \mathbb{R}^{3n\times 1}. \tag{7.135}$$

Substituting Eq. (7.131) back into Eq. (7.99) gives:

$$\begin{aligned} p(\hat{\underline{c}}, \hat{\underline{x}}_1, \hat{\underline{m}}^*_B|\Theta) = &\int \mathrm{d}R_c\ \mathrm{d}R_x\ \mathrm{d}c_\star\ \mathrm{d}x_\star\ |2\pi\Sigma_C|^{-\frac{1}{2}} |2\pi\Sigma_P|^{-\frac{1}{2}} |2\pi\Sigma_A|^{\frac{1}{2}} \\ &\times \exp\left(-\frac{1}{2}[X_0^T\Sigma_C^{-1}X_0 - Y_0^T\Sigma_A^{-1}Y_0 + Y_*^T\Sigma_P^{-1}Y_*]\right) \\ &\times p(R_c)p(R_x)p(c_\star)p(x_\star)p(M_0|\sigma_\mu^{\mathrm{int}}). \end{aligned} \tag{7.136}$$

7.6.5 Integration over population variables $\{c_\star, x_\star, M_0\}$

The priors on the population variables $\underline{b} = \{c_\star, x_\star, M_0\}$ in Eq. (7.136) can be written as:

$$\begin{aligned} p(\underline{b}) &= p(c_\star)p(x_\star)p(M_0|\sigma_\mu^{\text{int}}) \\ &= |2\pi\Sigma_0|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\underline{b}-\underline{b}_m)^T\Sigma_0^{-1}(\underline{b}-\underline{b}_m)\right) \end{aligned} \tag{7.137}$$

where

$$\Sigma_0^{-1} = \begin{bmatrix} 1/\sigma_{c_*}^2 & 0 & 0 \\ 0 & 1/\sigma_{x_*}^2 & 0 \\ 0 & 0 & 1/\sigma_{M_0}^2 \end{bmatrix} \tag{7.138}$$

and

$$\underline{b}_m = \begin{bmatrix} 0 \\ 0 \\ M_m \end{bmatrix} \in \mathbb{R}^{3\times 1} \tag{7.139}$$

Thus Eq. (7.136) can be written as:

$$\begin{aligned} p(\underline{\hat{c}}, \underline{\hat{x}}_1, \underline{\hat{m}}_B^*|\Theta) &= \int \mathrm{d}R_c\ \mathrm{d}R_x\ \mathrm{d}\underline{b}\ |2\pi\Sigma_C|^{-\frac{1}{2}}|2\pi\Sigma_P|^{-\frac{1}{2}}|2\pi\Sigma_A|^{\frac{1}{2}}|2\pi\Sigma_0|^{-\frac{1}{2}}p(R_c)p(R_x) \\ &\times \exp\left(-\frac{1}{2}[X_0^T\Sigma_C^{-1}X_0 - (\Sigma_A(\Delta + \Sigma_P^{-1}\underline{J}\cdot\underline{b}))^T\Sigma_A^{-1}(\Sigma_A(\Delta+\Sigma_P^{-1}\underline{J}\cdot\underline{b})) \right. \\ &\qquad \left. + \underline{b}^T\underline{J}^T\Sigma_P^{-1}\underline{Jb} + (\underline{b}-\underline{b}_m)^T\Sigma_0^{-1}(\underline{b}-\underline{b}_m)]\right) \\ &= \int \mathrm{d}R_c\ \mathrm{d}R_x\ |2\pi\Sigma_C|^{-\frac{1}{2}}|2\pi\Sigma_P|^{-\frac{1}{2}}|2\pi\Sigma_A|^{-\frac{1}{2}}|2\pi\Sigma_0|^{-\frac{1}{2}}p(R_c)p(R_x) \\ &\times \exp\left(-\frac{1}{2}[X_0^T\Sigma_C^{-1}X_0 - \Delta^T\Sigma_A\Delta - k_0^TK^{-1}k_0 + \underline{b}_m^T\Sigma_0^{-1}\underline{b}_m]\right) \\ &\times \int \mathrm{d}\underline{b}\ \exp\left(-\frac{1}{2}[(\underline{b}-k_0)^TK^{-1}(\underline{b}-k_0)]\right) \end{aligned} \tag{7.140}$$

where

$$
\begin{aligned}
K^{-1} &= -\underline{J}^T\Sigma_P^{-1}\Sigma_A\Sigma_P^{-1}\underline{J} + \underline{J}^T\Sigma_P^{-1}\underline{J} + \Sigma_0^{-1} \in \mathbb{R}^{3\times 3}, && (7.141)\\
K^{-1}k_0 &= \underline{J}^T\Sigma_P^{-1}\Sigma_A\Delta + \Sigma_0^{-1}\underline{b}_m \in \mathbb{R}^{3\times 1}, && (7.142)\\
k_0 &= K(\underline{J}^T\Sigma_P^{-1}\Sigma_A\Delta + \Sigma_0^{-1}\underline{b}_m). && (7.143)
\end{aligned}
$$

We can now carry out the Gaussian integral over $\underline{b}$ in Eq. (7.136), obtaining our final expression for the effective likelihood,

$$
\begin{aligned}
p(\underline{\hat{c}}, \underline{\hat{x}}_1, \underline{\hat{m}}_B^*|\Theta) = \int \mathrm{d}\log R_c \; \mathrm{d}\log R_x \; &|2\pi\Sigma_C|^{-\frac{1}{2}}|2\pi\Sigma_P|^{-\frac{1}{2}}|2\pi\Sigma_A|^{\frac{1}{2}}|2\pi\Sigma_0|^{-\frac{1}{2}}|2\pi K|^{\frac{1}{2}}\\
&\times \exp\left(-\frac{1}{2}[X_0^T\Sigma_C^{-1}X_0 - \Delta^T\Sigma_A\Delta - k_0^TK^{-1}k_0 + \underline{b}_m^T\Sigma_0^{-1}\underline{b}_m]\right),
\end{aligned}
\tag{7.144}
$$

where we have chosen an improper Jeffreys' prior for the scale variables R_c, R_x:

$$
p(R_c) \propto R_c^{-1} \Rightarrow p(R_c)\mathrm{d}R_c \propto \mathrm{d}\log R_c, \tag{7.145}
$$

and analogously for R_x. These two remaining nuisance parameters cannot be integrated out analytically, so they need to be marginalized numerically. Hence, R_c, R_x are added to our parameters of interest and are sampled over numerically, and then marginalized out from the joint posterior.

The expression for the effective likelihood given by Eq. (7.144) is the major result presented in this chapter. Having shown how this effective likelihood is motivated and arrived at, I will now present some numerical trials in which the effective likelihood is tested using simulated data.

7.7 Numerical trials with simulated data

7.7.1 Description of the real SNe Ia data sets

The simulated data sets used in the numerical trials are modeled on the (then) recent compilation of 288 SNIa from Kessler et al. (2009a), which presents analysis of new data from SDSS-II along with publicly available data from four existing surveys. The Kessler et al. (2009a) compilation comprises of:

- SDSS: 103 SNe (Kessler et al., 2009a)
- ESSENCE: 56 SNe (Miknaitis & Pignata, 2007; Wood-Vasey et al., 2007)

Parameter	Symbol	True Value
Matter energy density parameter	Ω_m	0.3
Dark energy density parameter	Ω_Λ	0.7
Dark energy equation of state	w	-1
Spatial curvature	Ω_κ	0.0
Hubble expansion rate	H_0 [km/s/Mpc]	72.0
Mean absolute magnitude of SNe	M_0 [mag]	-19.3
Intrinsic dispersion of SNe magnitude	$\sigma_\mu^{\rm int}$ [mag]	0.1
Stretch correction	α	0.13
Colour correction	β	2.56
Mean of distribution of $\underline{x}_1$	$x_\star$	0.0
Mean of distribution of $\underline{c}$	$c_\star$	0.0
s.d. of distribution of $\underline{x}_1$	R_x	1.0
s.d. of distribution of $\underline{c}$	R_c	0.1
Observational noise on m_B^*	$\sigma_{m^*_{Bi}}$	Depending on survey
Observational noise on $\underline{x}_1$	$\sigma_{x_1 i}$	Depending on survey
Observational noise on $\underline{c}$	σ_{ci}	Depending on survey
Correlation between $\underline{x}_1$ and $\underline{c}$	$\sigma_{x_1 i,ci}$	0.0

Table 7.3: Input parameter values used for the fiducial model in the generation of the simulated SNe SALT2 data sets.

- SNLS: 62 SNe (Astier & Guy, 2006)
- Nearby Sample: 33 SNe (Jha et al., 2007)
- HST: 34 SNe (Garnavich et al., 1998; Knop et al., 2003; Riess & Strolger, 2004, 2007)

The compiled set of 288 SNe Ia were analysed by (Kessler et al., 2009a) using both the SALT2 method and the MLCS method. In the following, we are exclusively employing the results of their SALT2 fits and use those as the observed data set for the purposes of our current work, as described in the previous section. More refined procedures could be adopted, for example by simulating lightcurves from scratch, using e.g. the publicly available package SNANA (Kessler et al., 2009b). In this work we chose a simpler approach, which is to simulate the SALT2 fit results in such a way to broadly match the distributions and characteristics of the real data set used in Kessler et al. (2009a).

7.7.2 Description of the simulated SNe Ia data sets

The numerical values of the parameters used for the simulated data sets are shown in Table 7.3. We adopt a flat ΛCDM cosmological model as fiducial cosmology. The α and β global fit parameters are chosen to match the best-fit values reported in Kessler et al. (2009a), while the distributional properties of the colour and stretch correction match the observed distribution of their total SN sample. For each survey, we generate a number of SNe matching the observed sample, and we model their redshift distribution as a Gaussian, with mean and variance estimated from the observed distribution within each survey. The observational error of $\underline{m}_B^*, \underline{c}, \underline{x}_1$ is again drawn from a Gaussian distribution whose mean and variance have been matched to the observed ones for each survey. Finally, the simulated data (i.e., the simulated SALT2 fits results $\underline{\hat{m}}_B^*, \underline{\hat{c}}, \underline{\hat{x}}_1$) are generated by drawing from the appropriate distributions centered around the latent variables. For simplicity, we have set to 0 the off-diagonal elements in the correlation matrix (7.104) in our simulated data, and neglected redshift errors. None of these assumptions have a significant impact on our results. In summary, our procedure for simulating data for each survey is as follows:

(i) Draw a value for the latent redshift z_i from a normal distribution with mean and variance matching the observed ones. As we neglect redshift errors in the simulated data for simplicity (since the uncertainty in z is subdominant in the overall error budget), we set $\hat{z}_i = z_i$.

(ii) Compute μ_i using the fiducial values for the cosmological parameters $\mathscr{C}$ and the above z_i from Eq. (7.5).

(iii) Draw the latent parameters x_{1i}, c_i, M_i from their respective distributions (in particular, including an intrinsic scatter $\sigma_\mu^{\rm int} = 0.1$ mag in the generation of M_i).

(iv) Compute m_{Bi}^* using x_{1i}, c_i, M_i and the SALT2 relation Eq. (7.73).

(v) Draw the value of the standard deviations $\sigma_{x_1 i}, \sigma_{c_i}, \sigma_{m_i}$, from the appropriate normal distributions for each survey type. A small, z_i-dependent stochastic linear addition is also made to $\sigma_{x_1 i}, \sigma_{c_i}, \sigma_{m_i}$, to mimic the observed correlation between redshift and error.

(vi) Draw the SALT2 fit results from $\hat{x}_{1i} \sim \mathcal{N}(x_{1i}, \sigma_{x_1 i})$, $\hat{c}_i \sim \mathcal{N}(c_i, \sigma_{c_i})$ and $\hat{m}_{Bi}^* \sim \mathcal{N}(m_{Bi}^*, \sigma_{m_i})$.

As shown in Fig. 7.10, the simulated data from our procedure have broadly similar distributions to the real ones. The two notable exceptions are the overall vertical shift observed in the distance modulus plot, and the fact that our simulated data cannot reproduce the few outliers with large values of the variances (bottom panels). The former is a consequence of the different absolute magnitude used in our simulated data (as the true one is unknown). However, the absolute magnitude is

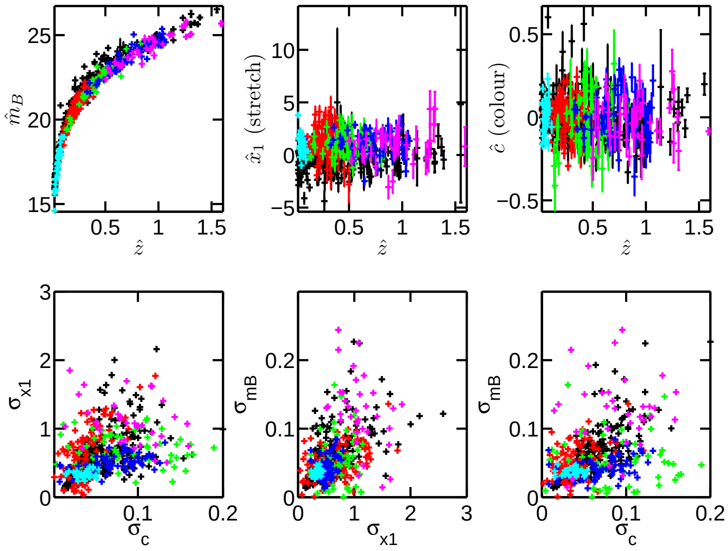


Figure 7.10: An example realization of our simulated data sets (coloured according to survey), superimposed on real data (black). Colour code for simulated data survey: nearby sample (cyan), ESSENCE (green), SDSS (red), SNLS (blue) and HST (magenta).

marginalized over at the end, so this difference has no impact on our inferences. The absence of outliers is a consequence of the fact that our simulation is a pure phenomenological description of the data, hence it cannot encapsulate such fine details. While in principle we could perform outlier detection with dedicated Bayesian procedures, we do not pursue this issue further in this paper. We stress once more that the purpose of our simulations is not to obtain realistic SNIa data. Instead, they should only provide us with useful mock data sets coming from a known model so that we can test our procedure. More sophisticated tests based on more realistically generated data (e.g., from SNANA) are left for future work.

7.7.3 Numerical sampling

After analytical marginalization of the latent variables, we are left with the following eight parameters entering the effective likelihood of Eq. (7.144):

$$\{\Omega_m, \Omega_\kappa \text{ or } w, H_0, \sigma_\mu^{\text{int}}, \alpha, \beta, R_c, R_x\}\,. \tag{7.146}$$

As mentioned above, in keeping with the literature we only consider either flat Universes with a possible $w \neq -1$ (the ΛCDM model), or curved Universes with a

Parameter	ΛCDM	wCDM
Ω_m	Uniform: $\mathcal{U}(0.0, 1.0)$	Uniform: $\mathcal{U}(0.0, 1.0)$
Ω_κ	Uniform: $\mathcal{U}(-1.0, 1.0)$	Fixed: 0
w	Fixed: -1	Uniform: $\mathcal{U}(-4, 0)$
H_0 [km/s/Mpc]	$\mathcal{N}(72, 8^2)$	$\mathcal{N}(72, 8^2)$
	Common priors	
σ_μ^{int} [mag]	Uniform on $\log \sigma_\mu^{\text{int}}$: $\mathcal{U}(-3.0, 0.0)$	
M_0 [mag]	Uniform: $\mathcal{U}(-20.3, -18.3)$	
α	Uniform: $\mathcal{U}(0.0, 1.0)$	
β	Uniform: $\mathcal{U}(0.0, 4.0)$	
R_c	Uniform on $\log R_c$: $\mathcal{U}(-5.0, 2.0)$	
R_x	Uniform on $\log R_c$: $\mathcal{U}(-5.0, 2.0)$	

Table 7.4: Priors on our model's parameters used when evaluating the posterior distribution. Ranges for the uniform priors have been chosen so as to generously bracket plausible values of the corresponding quantities.

cosmological constant ($w = -1$, the wCDM model). Of course it is possible to relax those assumptions and consider more complicated cosmologies with a larger number of free parameters if one so wishes (notably including evolution in the dark energy equation of state).

Of the parameters listed in Eq. (7.146), the quantities R_c, R_x are of no interest and will be marginalized over. As for the remaining parameters, we are interested in obtaining their marginal 1 and 2-dimensional posterior distributions. This is done by inserting the likelihood (7.144) into the posterior of Eq. (7.91), with priors on the parameters chosen in accordance with Table 7.4. We use a Gaussian prior on the Hubble parameter $H_0 = 72 \pm 8$ km/s/Mpc from local determinations of the Hubble constant (Freedman et al., 2001). However, as H_0 is degenerate with the intrinsic population absolute magnitude M_0 (which is marginalized over at the end), replacing this Gaussian prior with a less informative prior $H_0[\text{km/s/Mpc}] \sim \mathcal{U}(20, 100)$ has no influence on our results.

Numerical sampling of the posterior is carried out via a nested sampling algorithm (Skilling, 2004, 2006; Feroz & Hobson, 2008b; Feroz et al., 2009b). Although the original motivation for nested sampling was to compute the Bayesian evidence, the recent development of the MultiNest algorithm (Feroz & Hobson, 2008b; Feroz et al., 2009b) has delivered an extremely powerful and versatile algorithm that has been demonstrated to be able to deal with extremely complex likelihood surfaces in hundreds of dimensions exhibiting multiple peaks. As samples from the posterior are generated as a by-product of the evidence computation, nested sampling can also be used to obtain parameter constraints in the same run as computing the Bayesian evidence. In this paper we adopt the publicly available MultiNest algorithm (Feroz & Hobson, 2008b) to obtain samples from the posterior distribution of Eq. (7.91). We use 4000 live points and a tolerance parameter 0.1, resulting in about 8×10^5 likelihood evaluations.

We also wish to compare the performance of our BHM with the usually adopted χ^2 minimization procedure. To this end, we fit the simulated data using the χ^2 expression of Eq. (7.6). In order to mimic what is done in the literature as closely as possible, we chose a value of $\sigma_\mu^{\rm int}$ then minimize the χ^2 w.r.t. the fit parameters $\vartheta = \{\Omega_m, \Omega_\kappa \text{ or } w, H_0, M_0, \alpha, \beta\}$. We update the value of $\sigma_\mu^{\rm int}$ then repeat the minimization process as described below, until a value of $\chi^2/\text{dof} = 1$ is obtained. The steps in the process can be enumerated as follows:

(i) Select a trial value for $\sigma_\mu^{\rm int}$.

(ii) Minimise the χ^2 given in Eq. (7.6) by simultaneously fitting for the cosmology and SNe Ia parameters $\vartheta = \{\Omega_m, \Omega_\kappa \text{ or } w, H_0, M_0, \alpha, \beta\}$.

(iii) Evaluate χ^2/dof at minimum (i.e. the best fit point). If $\chi^2/\text{dof} > 1$ select a higher trial value for $\sigma_\mu^{\rm int}$, if $\chi^2/\text{dof} < 1$ select a lower trial value for $\sigma_\mu^{\rm int}$, repeat from the minimization step (ii) onwards.

(iv) Stop the process of minimization and iterative updating of $\sigma_\mu^{\rm int}$ when a value of $\chi^2/\text{dof} = 1$ is obtained.

Once we have obtained the global best fit point, we derive 1- and 2-dimensional confidence intervals on the parameters by profiling (i.e., maximising over the other parameters) over the likelihood

$$L(\vartheta) = \exp\left(-\frac{1}{2}\chi(\vartheta)^2\right), \qquad (7.147)$$

with χ^2 given by Eq. (7.6). According to Wilks' theorem, approximate confidence intervals are obtained from the profile likelihood as the regions where the χ^2 increases by $\Delta\chi^2$ from its minimum value, where $\Delta\chi^2$ can be computed from the chi-square distribution with the number of degree of freedoms corresponding to the number of parameters of interest and is given in standard look-up tables. The appropriate $\Delta\chi^2$ values for the 1D likelihoods are shown in table 7.5. Obtaining reliable estimates of the profile likelihood using Bayesian algorithms (such as MultiNest) is a considerably harder numerical task than mapping out the Bayesian posterior. However, it has been shown that MultiNest can be successfully used for this task even in highly challenging situations (Feroz et al., 2011), provided the number of live points and

Likelihood content	68.3% (1σ)	95.4% (2σ)
$\Delta\chi^2$	1.00	4.00

Table 7.5: Change in $\Delta\chi^2$ required in 1D profile likelihood for 1σ and 2σ confidence intervals.

tolerance value used are adjusted appropriately. For our χ^2 scan, we adopt 10^4 live points and a tolerance of 0.1. We have found that those values give accurate estimates of the profile likelihood more than 2σ into the tails of the distribution for an 8 dimensional Gaussian toy model (whose dimensionality matches the case of interest here). With these MultiNest settings, we gather 1.5×10^5 samples, from which the profile likelihood is derived.

Our implementation of the χ^2 method is designed to match the main features of the fitting procedure usually adopted in the literature (namely, maximisation of the likelihood rather than marginalization of the posterior, and iterative determination of the intrinsic dispersion), although we do not expect that it exactly reproduces the results obtained by any specific implementation. Its main purpose is to offer a useful benchmark against which to compare the performance of our new Bayesian methodology.

7.7.4 Parameter reconstruction

We compare the cosmological parameters reconstructed from the standard χ^2 method and our Bayesian approach in Fig. 7.11 for a typical data realization. The left-hand-side panel shows constraints in the $\Omega_m - \Omega_\Lambda$ plane for the ΛCDM model, both from our Bayesian method (filled regions, marginalized 68.3% and 95.4% posterior) and from the standard χ^2 method (red contours, 68.3% and 95.4% confidence regions from the profile likelihood). In the right-hand-side panel, constraints are shown in the $w - \Omega_m$ plane for a flat wCDM model Universe. In a typical reconstruction, our Bayesian method produced considerably tighter constraints on the cosmological parameters of interest than the usual χ^2 approach. Our constraints are also less biased w.r.t. the true value of the parameters, an important advantage that we further characterize below.

Our BHM further produces marginalized posterior distributions for all the other parameters of the fit, including the global SNe Ia parameters α and β and the intrinsic dispersion of the SNe. The 1D marginal posteriors for those quantities are shown in Fig. 7.12. The recovered posterior means lie within 1σ of the true values. Notice that we do not expect the posterior mean to match exactly the true

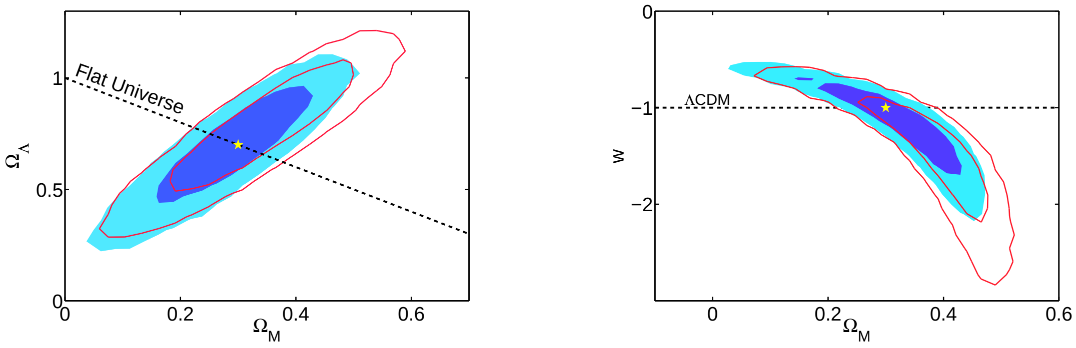


Figure 7.11: Reconstruction of cosmological parameters from a simulated data set encompassing 288 SNIa, with characteristics matching presently available surveys (including realization noise). Blue regions contain 95.4% and 68.3% of the posterior probability (other parameters marginalized over) from our BHM method, the red contours delimit 95.4% and 68.3% confidence intervals from the standard χ^2 method (other parameters maximised). The yellow star indicates the true value of the parameters. The left panel assumes $w = -1$ while the right panel assumes $\Omega_\kappa = 0$. Notice how out method produced considerably less biased constraints on the parameters.

value, because of realization noise in the simulated data. However, as shown below, our method delivers less biased estimates of the parameters, and a reduced mean squared error compared with the standard χ^2 approach. The stretch correction α is determined with 8% accuracy, while the colour correction parameter β is constrained with an accuracy better than 3%. A new feature of our method is that it produces a posterior distribution for the SN population intrinsic dispersion, $\sigma_\mu^{\rm int}$ (right-hand-side panel of Fig 7.12). This allows one to determine the intrinsic dispersion of the SNIa population to typically about 10% accuracy.

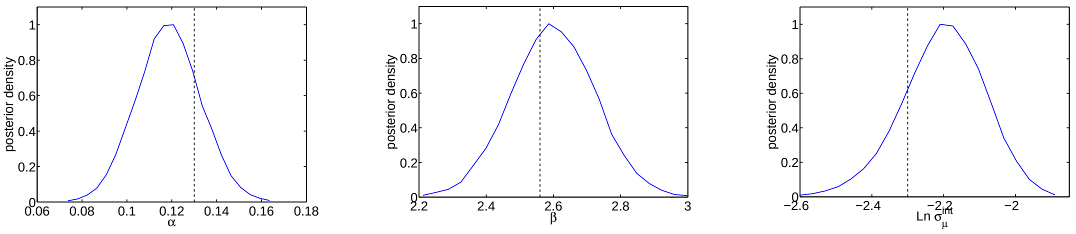


Figure 7.12: Marginalised posterior for the stretch correction α, colour correction β parameter and logarithm of the intrinsic dispersion of SNe, $\log \sigma_\mu^{\rm int}$, from a simulated data set from our Bayesian method. The vertical, dashed line gives the true value for each quantity.

7.7.5 Comparison of performance of the two methods over 100 trials

As we are dealing with a system subject to Gaussian statistical noise, the results of one single trial are not sufficient to validate the claim that the BHM method outperforms the χ^2 method. In order to demonstrate further the relative performance of the two methods we conducted a series of trials using 100 different realizations of the data, i.e. we generated 100 simulated data sets each with 288 SNe Ia, as detailed in section 7.7.2. During the numerical sampling phase of the parameter reconstruction, several trials failed due to computational problems, in these cases the results for the relevant trials for both the BHM and χ^2 methods were omitted from the final analysis. We are interested in comparing the average ability of both methods to recover parameter values that are precise, accurate and as much as possible unbiased with respect to their true values, as well as to establish the coverage properties of the credible and confidence intervals.

Coverage is defined as the probability that an interval contains (covers) the true value of a parameter, in a long series of repeated measurements. The defining property of a e.g. 95.4% frequentist confidence interval is that it should cover the true value 95.4% of the time; thus, it is reasonable to check if the intervals have the properties they claim. Coverage is a frequentist concept: intervals based on Bayesian techniques are meant to contain a given amount of posterior probability for a *single measurement* (with no reference to repeated measurements) and are referred to as credible intervals to emphasize the difference in concept. While Bayesian techniques are not designed with coverage as a goal, it is still meaningful to investigate their coverage properties. To our knowledge, the coverage properties of even the standard χ^2 method (which, being a frequentist method would ideally be expected to exhibit exact coverage) have never been investigated in the SN literature.

We generate 100 realizations of the simulated data from the fiducial model of Table 7.3 as described in section 7.7.2, and we analyse them using our BHM method and the standard χ^2 approach, using the same priors as above, given in Table 7.4. We quantify the performance of the two methods in two ways: in terms of the precision (i.e. error bar size) and in terms of accuracy (i.e. distance of reconstructed parameter value from true parameter value). For each parameter of interest θ, we compare the precision by evaluating the relative size of the posterior 68.3% range from our BHM method, $\sigma_\theta^{\mathrm{BHM}}$, compared with the 68.3% confidence interval from the χ^2 method, $\sigma_\theta^{\chi^2}$, which is summarized by the quantity $\mathcal{S}_\theta$ which shows the percentage change in error bar size with respect to the error bar derived using the χ^2 method

$$\mathcal{S}_\theta \equiv \left(\frac{\sigma_\theta^{\text{BHM}}}{\sigma_\theta^{\chi^2}} - 1 \right) \times 100. \tag{7.148}$$

A value $\mathcal{S}_\theta < 1$ means that our BHM method delivers tighter error bars on the parameter θ, so is more precise. A histogram of this quantity for the variables of interest is shown in Fig. 7.13, from which we conclude that our BHM method gives smaller error bars on Ω_m, Ω_Λ and w in almost all cases. However the uncertainty on α, β is larger from our method than from the χ^2 approach in most data realizations, as expected from the numerical trials with the toy model described in section 7.4.

Precision and tight error bars are good, but not if they come at the expense of a less accurate reconstruction. To evaluate the accuracy of each method, we build the following test statistic for each reconstruction:

$$\mathcal{T}_\theta \equiv |\overline{\theta}_{\text{BHM}}/\theta_{\text{true}} - 1| - |\theta_{\chi^2}^{\text{bf}}/\theta_{\text{true}} - 1|, \tag{7.149}$$

where $\overline{\theta}_{\text{BHM}}$ is the posterior mean recovered using our BHM method, $\theta_{\chi^2}^{\text{bf}}$ is the best-fit value for the parameter recovered using the standard χ^2 approach and θ_{true} is the true value for that parameter. $\mathcal{T}_\theta$ can be interpreted as follows: for a given data realization, if the reconstructed posterior mean from our BHM is closer to the true parameter value than the best-fit χ^2, then $\mathcal{T}_\theta < 0$, which means that our method is more accurate than χ^2. A histogram of the distribution of $\mathcal{T}_\theta$ across the 100 realizations, shown in Fig. 7.14, can be used to compare the two methods: a negative average in the histogram means that the BHM outperforms the usual χ^2. For all of the parameters considered, our BHM method is more accurate than the χ^2 method, outperforming χ^2 about 2/3 of the time. Furthermore, the reconstruction of the intrinsic dispersion σ_μ^{int} is better with our BHM method almost 3 times out of 4. We emphasize once more that our methodology also provides an estimate of the uncertainty in the intrinsic dispersion, not just a best-fit value as the χ^2 approach.

We can further quantify the improvement in the statistical reconstruction by looking at the bias, defined in EQ. (7.4.1) and mean squared error (MSE) for each parameter, defined as

$$\text{MSE} = \text{bias}^2 + \text{Var}, \tag{7.150}$$

respectively, where the expectation is taken by averaging over the observed values in our 100 simulated trials, $\hat{\theta} = \overline{\theta}_{\text{BHM}}$ ($\hat{\theta} = \theta_{\chi^2}^{\text{bf}}$) for the BHM (for the χ^2 approach) and Var is the observed parameter variance. The bias is the expectation value of the difference between estimator and true value, and tells us by how much the estimator systematically over or under estimates the parameter of interest. The MSE measures the average of the squares of the errors, i.e., the amount by which the estimator differs from the true value for each parameter. A smaller bias and a smaller MSE imply a better performance of the method. The results for the two methods are summarized in Table 7.6, which shows how our method reduces the bias by a factor $\sim 2-3$ for most parameters, while reducing the MSE by a factor of ~ 2. The only notable exception is the bias of the EOS parameter w, which is larger in our method than in the χ^2 approach.

Parameter		Bias			Mean squared error		
		Bayesian	χ^2	Improvement	Bayesian	χ^2	Improvement
ΛCDM	Ω_m	-0.0188	-0.0183	1.0	0.0082	0.0147	1.8
	Ω_Λ	-0.0328	-0.0223	0.7	0.0307	0.0458	1.5
	α	0.0012	0.0032	2.6	0.0001	0.0002	1.4
	β	0.0202	0.0482	2.4	0.0118	0.0163	1.4
	σ_μ^{int}	-0.0515	-0.1636	3.1	0.0261	0.0678	2.6
wCDM	Ω_m	-0.0177	-0.0494	2.8	0.0072	0.0207	2.9
	Ω_Λ	0.0177	0.0494	2.8	0.0072	0.0207	2.9
	w	-0.0852	-0.0111	0.1	0.0884	0.1420	1.6
	α	0.0013	0.0032	2.5	0.0001	0.0002	1.5
	β	0.0198	0.0464	2.3	0.0118	0.0161	1.4
	σ_μ^{int}	-0.0514	-0.1632	3.2	0.0262	0.0676	2.6

Table 7.6: Comparison of the bias and mean squared error for our Bayesian method and the usual χ^2 approach. The columns labelled "Improvement" give the factor by which our Bayesian method reduces the bias and the MSE w.r.t. the χ^2 approach.

Finally, in Fig. 7.15 we plot the coverage of each method for 68.3% and 95.4% intervals. Error bars give an estimate of the uncertainty of the coverage result, by giving the binomial sampling error from the finite number of realizations considered, evaluated from the binomial variance as $Np(1-p)$, where $N = 100$ is the number of trials and p is the observed fractional coverage. Both methods slightly undercover, i.e. the credible region and confidence intervals are too short, although the lack of coverage is not dramatic: e.g., the typical coverage of the 1σ (2σ) intervals from our method is $\sim$ 60% (90%). Our method shows slightly better coverage properties than the χ^2 method, while producing considerably tighter and less biased constraints (as demonstrated above). This further proves that the tighter intervals recovered by our method do not suffer from bias w.r.t the true values.

To summarise, the results from our numerical trials with simulated data show that:

(i) In general, our BHM method gives more precise constraints on the cosmological parameters, but gives less precise constraints on the SNe Ia global fit parameters α, β

(ii) In $60 - 70\%$ of trials, our BHM method recovers a more accurate value of the parameter of interest.

(iii) Our BHM method is less biased than the χ^2 method, except in the reconstruction of the w parameter

(iv) Both methods undercover, with our BHM method giving slightly better coverage.

7.8 Cosmological constraints from current SNIa data using BHM

We now apply our BHM to fitting real SN data. We use the SALT2 fits result for 288 SNIa from Kessler et al. (2009a), which have been derived from five different surveys described briefly in section 7.7.1. Our method only includes statistical errors according to the procedure described in section 7.6, coming from redshift uncertainties (arising from spectroscopic errors and peculiar velocities), intrinsic dispersion (which is determined from the data) and full error propagation of the SALT2 fit results. Systematic uncertainties play an important role in SNIa cosmology fitting, and (although not included in this study) can also be treated in our formalism in a fully consistent way. We comment on this aspect further below, though we leave a complete exploration of systematics with our BHM to a future, dedicated work.

We show in Fig. 7.16 the constraints on the cosmological parameters $\Omega_m - \Omega_\Lambda$ (left panel, assuming $w = -1$) and $w - \Omega_m$ (right panel, assuming $\Omega_\kappa = 0$) obtained with our method. All other parameters have been marginalized over. In order to be consistent with the literature, we have taken a non-informative prior on H_0, uniform in the range $[20, 100]$ km/s/Mpc. The figure also compares our results with the statistical contours from Kessler et al. (2009a), obtained using the χ^2 method. (We compare with the contours including only statistical uncertainties for consistency.) In Fig. 7.17 we combine our SNIa constraints with Cosmic Microwave Background (CMB) data from WMAP 5-yrs measurements (Komatsu et al., 2009) and Baryonic Acoustic Oscillations (BAO) constraints from the Sloan Digital Sky Survey LRG sample (Eisenstein et al., 2005), using the same method as Kessler et al. (2009a). The combined SNIa, CMB and BAO statistical constraints result in $\Omega_m = 0.28 \pm 0.02, \Omega_\Lambda = 0.73 \pm 0.01$ (for the ΛCDM model) and $\Omega_m = 0.28 \pm 0.01, w = -0.90 \pm 0.05$ (68.3% credible intervals) for the wCDM model. Although the statistical uncertainties are comparable to the results by Kessler et al. (2009a) from the same sample, our posterior mean values present shifts of up to $\sim 2\sigma$ compared to the results obtained using the standard χ^2 approach. This is a fairly significant shift, which can be attributed to our improved statistical method, which exhibits a reduced bias w.r.t. the χ^2 approach.

Fig. 7.18 shows the 1D marginalized posterior distributions for the SNe Ia global fit parameters α, β and for the intrinsic dispersion $\sigma_\mu^{\rm int}$. All parameters are well constrained by the posterior, and we find $\alpha = 0.12 \pm 0.02$, $\beta = 2.7 \pm 0.1$ and a value of the intrinsic dispersion (for the whole sample) $\sigma_\mu^{\rm int} = 0.13 \pm 0.01$ mag. Kessler et al. (2009a) find values for the intrinsic dispersion ranging from 0.08 (for SDSS-II) to 0.23 (for the HST sample), but their χ^2 method does not allow them to derive an error on those determinations. With our method, it would be easy to derive constraints on the intrinsic dispersion of each survey – all one needs to do is to replace Eq. (7.74) with a corresponding expression for each survey. This introduces one pair of population parameters $(M_0, \sigma_\mu^{\rm int})$ for each survey. In the same way, one could study whether the intrinsic dispersion evolves with redshift. A detailed study of these issues is left for future work.

The value of α found in Kessler et al. (2009a) is in the range $0.10 - 0.12$, depending of the details of the assumptions made, with a typical statistical uncertainty of order ~ 0.015. These results are comparable with our own. As for the colour correction parameter β, constraints from Kessler et al. (2009a) vary in the range $2.46 - 2.66$, with a statistical uncertainty of order $0.1 - 0.2$. This stronger dependence on the details of the analysis seems to point to a larger impact of systematic uncertainties for β, which is confirmed by evidence of evolution with redshift of the value of β (Kessler et al. (2009a), Fig. 39). Our method can be employed to carry out a rigorous assessment of the evolution with redshift of colour corrections. A possible strategy would be to replace β with a vector of parameters $\beta_1, \beta_2, \dots,$ with each element describing the colour correction in a different redshift bin. The analysis proceeds then as above, and it produces posterior distributions for the components of β, which allows to check the hypothesis of evolution. Finally, in such an analysis the marginalized constraints on all other parameters (including the cosmological parameters of interest) would automatically include the full uncertainty propagation from the colour correction evolution, without the need for further *ad hoc* inflation of the error bars. These kind of tests will be pursued in a forthcoming publication.

7.9 Bayesian Hierarchical Model Future Planned Work

Capitalising on the initial work presented in this chapter, I intend to extend and develop this new methodology in a number of different ways, both in the field of SNe Ia cosmology and also to other fields which have problems of a similar form to the SNe Ia problem. In this section I briefly outline some of the future directions of this work.

7.9.1 Applications using SNe Ia BHM

One of the primary motivations for the SNe Ia work presented in this chapter was to be able to use the SNe Ia data to be able to discriminate between different cosmological models. One scenario I plan to investigate using the new BHM and Bayesian model selection is the spatial symmetry of the Universe. Many theoretical models make statements about the inhomogeneity or otherwise of the Universe, either in terms of its expansion history or composition or both, e.g. Wiseman & Withers (2010). By using the directional information we have about the supernovae I will investigate claims for directional asymmetry. Some work has been done in this area e.g. Schwarz & Weinhorst (2007) Antoniou & Perivolaropoulos (2010) Cooke & Lynden-Bell (2010) - but the advantage of using my new method for the supernovae analysis is that we can now use Bayesian model selection to discriminate between symmetric and asymmetric models for the Universe in a quantitative manner. In particular I am planning to use the my new method to explore claims that there is a spatial asymmetry in the fine structure constant. Webb et al. (2010).

Independently of model selection problems, there are many other areas of investigation within the realm of parameter inference. One of the most interesting questions concerns the colour evolution of the SNe Ia. The colour correction is key to standardizing the SNe Ia. In this work we have assumed a global fit parameter β which is universal for all SNe Ia at all redshifts in all directions. In reality however the colour of SNe Ia may evolve with redshift as high redshift SNe Ia come from progenitor stars with a different chemical composition (lower metallicity) from those at low redshift. Work is already being carried out in investigating colour evolution with redshift. We can contribute to this effort by modifying our method to allow for evolution of β with redshift either by characterising the evolution with a simple empirical expression such as:

$$\beta(z) = \beta_0 + z\beta_1 \tag{7.151}$$

or by allowing β to take a range of different values β_i for different redshift bins z_i. Fig.7.19 shows a preliminary example of what happens when a simulated data set which incorporates β evolving with redshift as in Eq. (7.151) is fitted incorrectly using a static model $\beta(z) = \beta$ (upper panel) and then correctly fitted using the evolving model of Eq. (7.151) (lower two panels). Fig.7.20 shows what happens when a simulated data set with a static β is fitted with the correct static model (upper panel) and the incorrect evolving model. In this case the simpler, correct model is nested within the more complex model such that when $\beta_1 = 0$ the static model is regained. The lower panel of fig.7.20 shows that indeed a value of $\beta_1 = 0$ is favoured for this data set. As well as evolution with redshift, the other colour relationship to investigate is the correlation between host galaxy type and colour. Spiral and elliptical galaxies have different dust properties and hence their SNe Ia have different degrees of reddening due to host galaxy dust, important work has already begun in this area by Sullivan et al. (2006); Mandel et al. (2010); Sullivan et al. (2011) and it is an area we can easily adapt our BHM method to investigate.

In the model described in this chapter, the distribution of the colour correction terms $\hat{c}_i$ is assumed Gaussian such that there is a match between the shape of the underlying distribution described in Eq. (7.75) and the prior placed on that distribution Eq. (7.86). Another interesting question is what happens if the 'wrong' prior is used to fit the underlying distribution? In a purely abstract case one might ask what happens if an underlying uniform distribution is fitted using a Gaussian prior, numerical trials could be conducted accordingly to see how large an effect is observed on the ability on the method to recover the parameters of interest. In practice, for the real SNe Ia case there are claims within the SNe Ia community that the underlying colour distribution is not Gaussian - our question here is two fold (1) To what extent does that affect the accuracy and precision of our new BHM method and (2) Can our method be adapted to identify the true underlying distribution of the colour parameters. These investigations into the colour evolution and colour correction distribution will form the basis of our next planned paper.

7.9.2 Bayesian Hierarchical Network Analysis of Gamma ray-burst data

The problem of linear fitting of data in the presence of large errors is a problem found in other areas of cosmology. Our solution for the SNeIa case can be adapted to provide solutions to other problems which have a similar structure, for example, cosmology with Gamma ray bursts (GRBs). We plan to develop the work of Diaferio et al. (2011) in using Gamma-ray bursts as cosmological probes. We will:

(i) Apply the Bayesian Hieraierarchical Network methodology (hereafter, BHN) described in this chapter and March et al. (2011c) to the analysis of GRB data as cosmological probes, replacing Eq.A3 in Diaferio et al. (2011).

(ii) Re-do the joint SNe-GRB analysis described in Diaferio et al. (2011) using the BHN analysisis of the SNe data, (replacing Eq.A7 in Diaferio et al. (2011)

The motivation and aims of the proposed GRB project are to provide a fully Bayesian analysis of the GRB data in order to:

(i) Be able to use the GRB data for Bayesian model selection.

(ii) Be able to use the BHN method to better understand the systematics of the GRB data.

Adapting from Diaferio et al. (2011) Eq.44, the true bolometric peak flux P^i_{bol} for each GRB is given by:

$$\log_{10} P^i_{\text{bol}} = \sum_j \left(a_j + b_j \log_{10} Q^i_j\right) - \log_{10}[4\pi d_L^2(z^i, \mathscr{C})] + \Delta^i \tag{7.152}$$

where index $i = 1, n$ labels each of the n GRBs and index j runs from $j = 1, 4$ where $\underline{Q}^i$ is the vector of latent parameters: $\underline{Q}^i = \{\tau^i_{\text{lag}}, \tau^i_{\text{RT}}, V^i, E^i_{\text{peak}}\}$ such that:

$$\begin{aligned}\log_{10} P^i_{\text{bol}} =& a_1 + b_1 \log_{10} \tau^i_{\text{lag}} + a_2 + b_2 \log_{10} \tau^i_{\text{RT}} + a_3 + b_3 \log_{10} V^i + a_4 + b_4 \log_{10} E^i_{\text{peak}} \\ & - \log_{10}[4\pi d_L^2(z^i, \mathscr{C})] + \Delta^i \end{aligned}\tag{7.153}$$

Note that we have included the additional term Δ^i which models the intrinsic dispersion, such that:

$$\Delta_i \sim \mathcal{N}(\Delta_0, \sigma^2_{\text{int}}) \tag{7.154}$$

In the current set up, $\Delta_0 \equiv 0$, which is equivalent to saying that $\log_{10} P^i_{\text{bol}}$ is drawn from a Gaussian distribution of mean $\left(\sum_j \left(a_j + b_j \log_{10} Q^i_j\right) - \log_{10}[4\pi d_L^2(z^i, \mathscr{C})]\right)$ and standard deviation σ_{int}.

The observed bolometric peak flux $\hat{P}^i_{\text{bol}}$ is drawn from a Gaussian distribution with mean P^i_{bol} and variance σ^2_{int} such that:

$$\hat{P}^i_{\text{bol}} \sim \mathcal{N}(P^i_{\text{bol}}, \sigma^2_{\text{p}}) \tag{7.155}$$

The other observables are similarly subject to Gaussian noise:

$$\hat{\tau}^i_{\rm lag} \sim \mathcal{N}(\tau^i_{\rm lag}, \sigma^2_{\tau_{\rm lag}}) \tag{7.156}$$

$$\hat{\tau}^i_{\rm RT} \sim \mathcal{N}(\tau^i_{\rm RT}, \sigma^2_{\tau_{\rm RT}}) \tag{7.157}$$

$$\hat{V}^i \sim \mathcal{N}(V^i, \sigma^2_V) \tag{7.158}$$

$$\hat{E}^i_{\rm peak} \sim \mathcal{N}(E^i_{\rm peak}, \sigma^2_{E_{\rm peak}}) \tag{7.159}$$

$$\hat{z}^i \sim \mathcal{N}(z^i, \sigma^2_z) \tag{7.160}$$

This proposed application to the GRB data is still in the planning stage, and I intend to implement it along with my collaborators soon after the completion of my PhD.

7.10 Conclusions

The primary aim of the work presented in this chapter was to address certain deficiencies in the existing methods for extracting the cosmological parameters from SNe Ia data in conjunction with the SALT2 lightcurve fitter. The two main motivations were the lack of an appropriate framework for assessing the unknown intrinsic dispersion $\sigma_\mu^{\rm int}$ and its uncertainty, and the incompatibility of existing parameter reconstruction methods with methods of Bayesian model selection. In order to address these dual problems, we have derived a new and fully Bayesian method for parameter inference based on a Bayesian Hierachical Model, BHM.

The main novelty of our method is that it produces an effective likelihood that propagates uncertainties in a fully consistent way. We have introduced an explicit statistical modeling of the absolute magnitude distribution of the SNIa population, which for the first time allows one to derive a full posterior distribution of the SNIa intrinsic dispersion.

We have tested our method using simulated data sets and found that it compares favourably with the standard χ^2 approach, both on individual data realizations and in the long term performance. Statistical constraints on cosmological parameters are significantly improved, while in a series of 100 simulated data sets our method outperforms the χ^2 approach at least 2 times out of 3 for the parameters of interest. We have also demonstrated that our method is less biased and has better coverage properties than the usual approach.

We applied our methodology to a sample of 288 SNIa from multiple surveys. We find that the flat ΛCDM model is still in good agreement with the data, even under our improved analysis. However, the posterior mean for the cosmological parameters exhibit up to 2σ shifts w.r.t. results obtained with the conventional χ^2 approach. This is a consequence of our improved statistical analysis, which benefits from a reduced bias in estimating the parameters.

While in this chapter I have only discussed statistical constraints, our method offers a new, fully consistent way of including systematic uncertainties in the fit. As our method is fully Bayesian, it can be used in conjunction with fast and efficient Bayesian sampling algorithms, such as MCMC and nested sampling. This will allow to enlarge the number of parameters controlling systematic effects that can be included in the analysis, thus taking SNIa cosmological parameter fitting to a new level of statistical sophistication. The power of our method as applied to systematic errors analysis will be presented in a forthcoming, dedicated work.

At a time when SNIa constraints are entering a new level of precision, and with a tenfold increase in the sample size expected over the next few years, we believe it is timely to upgrade the cosmological data analysis pipeline in order to extract the most information from present and upcoming SNIa data. This work represents a first step in this direction.

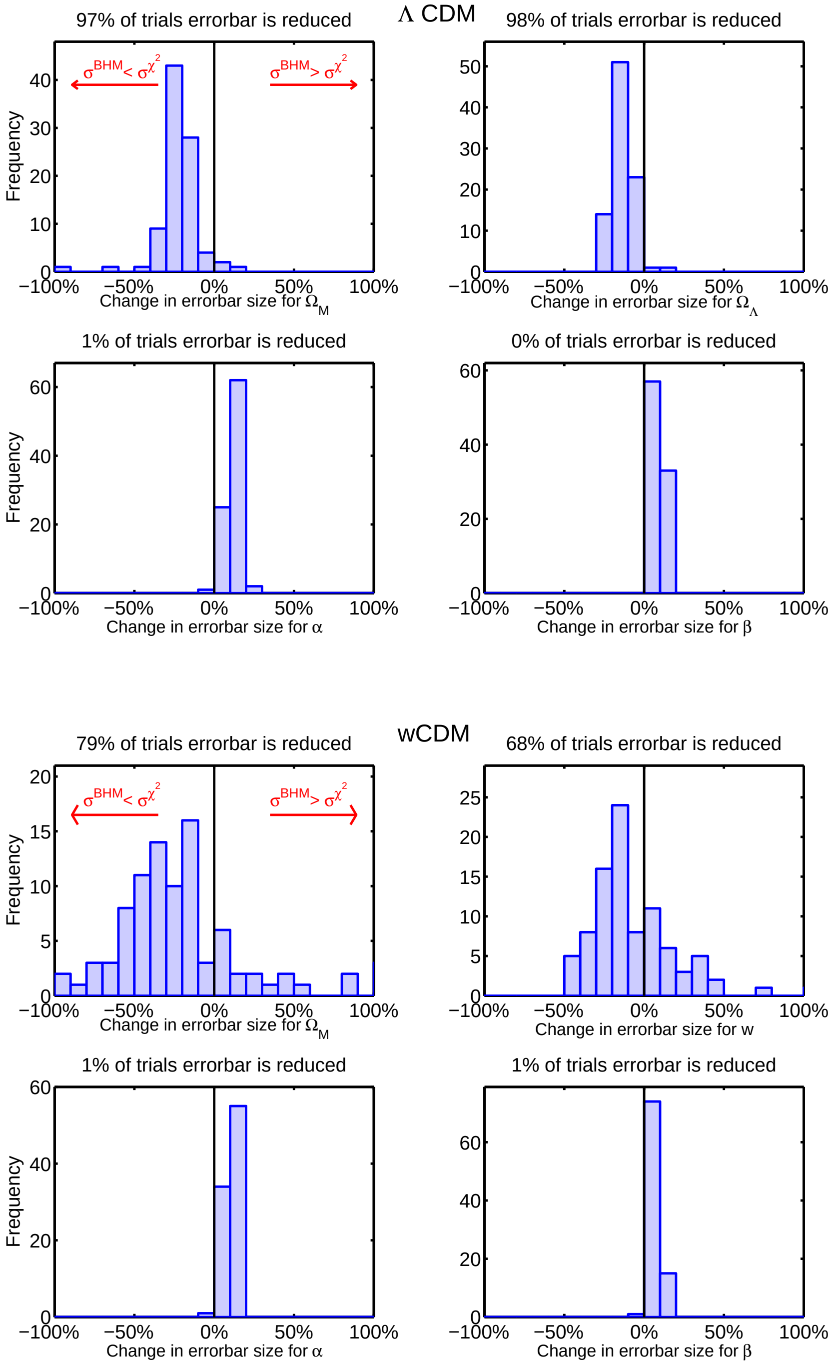


Figure 7.13: Histograms of the quantity defined in Eq. (7.148), comparing of the error bars on each parameter from our method and from the standard χ^2 approach for 100 realization, for the ΛCDM model (upper panel) and the wCDM model (lower panel). A change in error bar size of −10% indicates BHM error bars are 10% smaller than χ^2 error bars. A change in error bar size of +10% indicates BHM error bars are 10% larger than χ^2 error bars. Our BHM method generally delivers smaller errors on the cosmological parameters (top row) so is more precise, but larger errors on the SNe Ia global fit parameters α, β (bottom row) so is less precise.

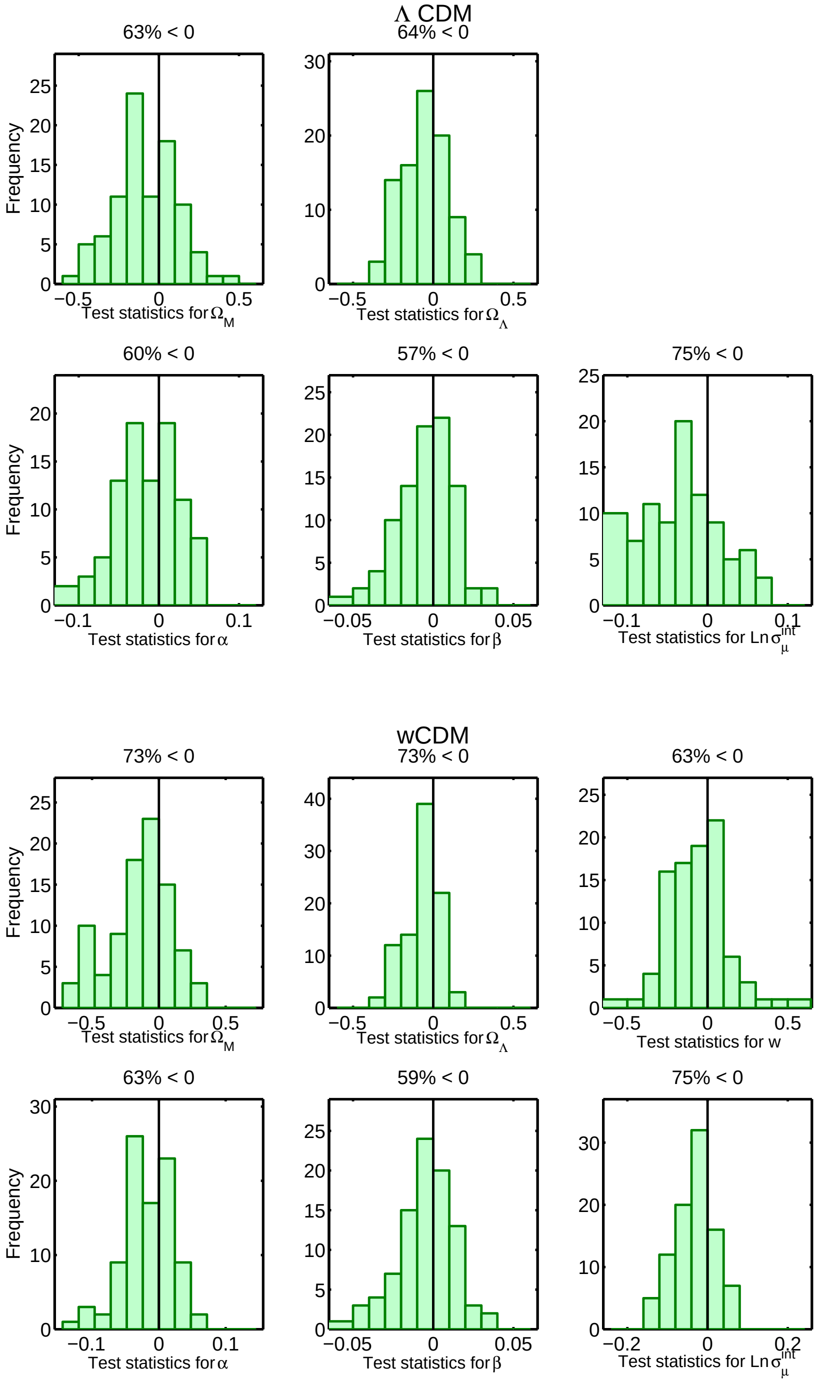


Figure 7.14: Histograms of the test statistics defined in Eq. (7.149), comparing the long-term performance of the two methods for the parameters of interest in the ΛCDM model (lower panel) and the wCDM model (upper panel). A predominantly negative value of the test statistics means that our method gives a parameter reconstruction that is closer to the true value than the usual χ^2, i.e., less biased. For the cosmological parameters (top row), our method outperforms χ^2 about 2 times out of 3.

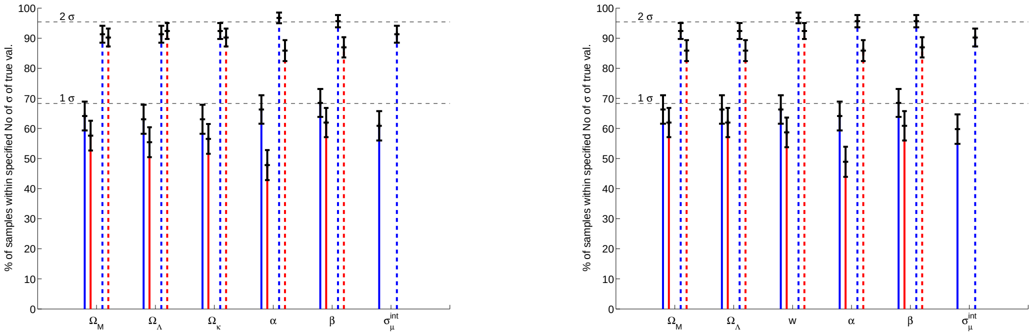


Figure 7.15: Coverage of our method (blue) and standard χ^2 (red) for 68.3% (solid) and 95.4% (dashed) intervals, from 100 realizations of simulated data for the ΛCDM model (left) and the wCDM model (right). While both methods show significant undercoverage for all parameters, our method has a comparable coverage to the standard χ^2, except for w. Coverage values for the intrinsic dispersion $\sigma_\mu^{\mathrm{int}}$ are not available from the χ^2 method, as it does not produce an error estimate for this quantity.

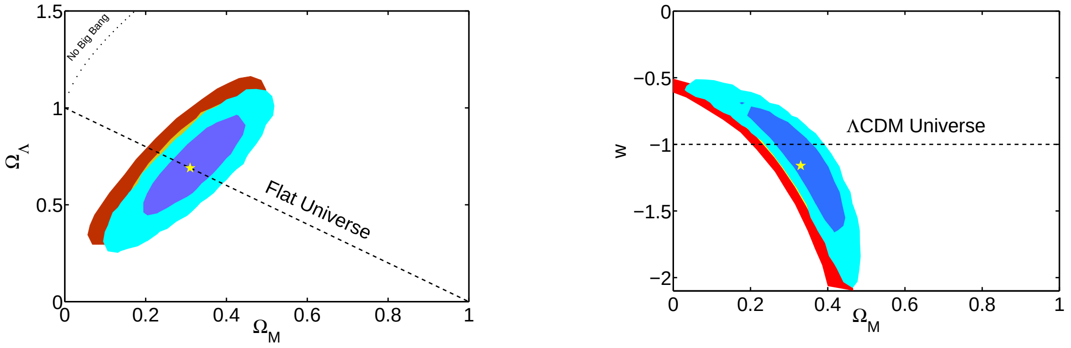


Figure 7.16: Constraints on the cosmological parameters Ω_m, Ω_Λ (left panel, assuming $w = -1$) and w, Ω_m (right panel, assuming $\Omega_\kappa = 0$) from our Bayesian method (light/dark blue regions, 68.3% and 95.4% marginalized posterior), compared with the statistical errors from the usual χ^2 approach (yellow/red regions, same significance level; from Kessler at al. (2009a)). The yellow star gives the posterior mean from our analysis.

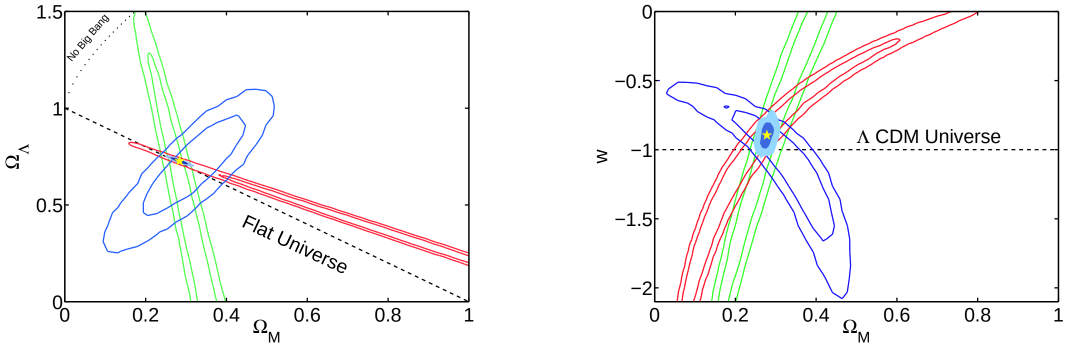


Figure 7.17: Combined constraints on the cosmological parameters Ω_m, Ω_Λ (left panel, assuming $w = -1$) and w, Ω_m (right panel, assuming $\Omega_\kappa = 0$) from SNIa, CMB and BAO data. Red contours give 68.3% and 95.4% regions from CMB alone, green contours from BAO alone, blue contours from SNIa alone from our Bayesian method. The filled regions delimit 68.3% and 95.4% combined constraints, with the yellow star indicating the posterior mean.

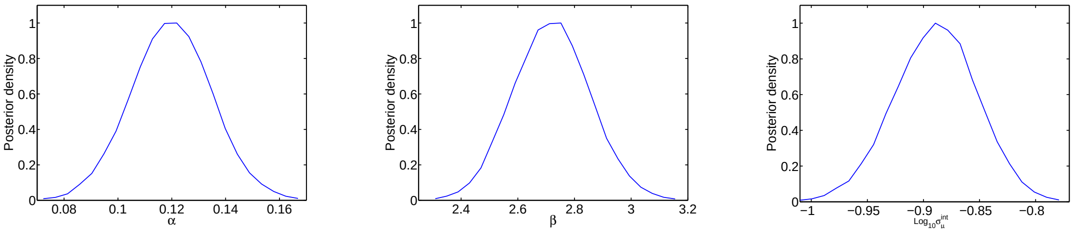


Figure 7.18: Marginalised posterior for the stretch correction α, colour correction β parameter and logarithm of the intrinsic dispersion of SNe, $\log \sigma_\mu^{\rm int}$ from current SNIa data.

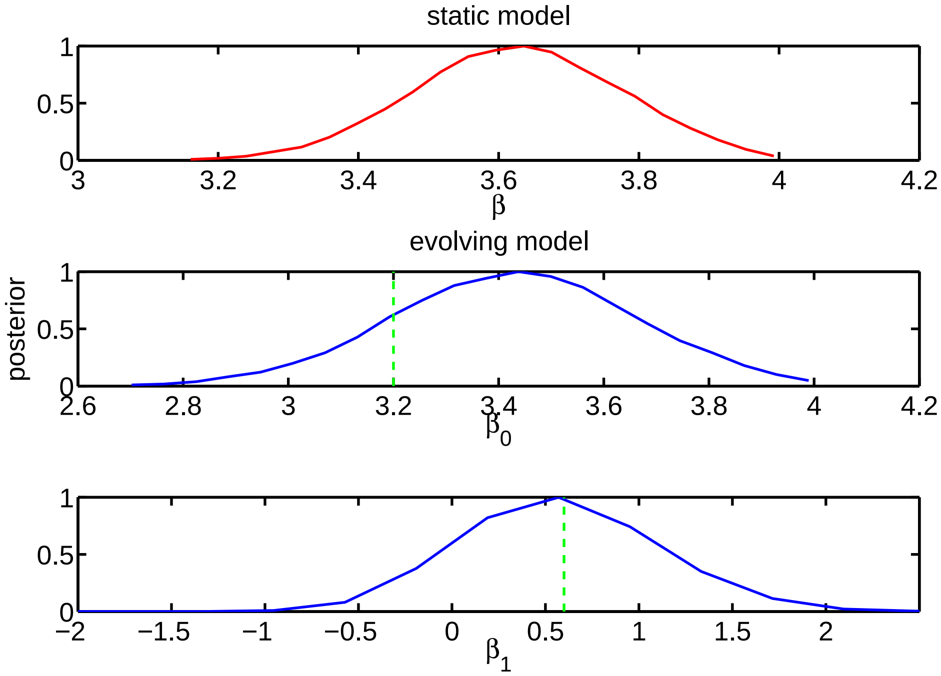


Figure 7.19: A simulated data set in which β evolves with z such that $\beta(z) = \beta_0 + z\beta_1$. Upper panel: fitted using an the incorrect static model (red). Lower panels: fitted using the correct evolving model (blue). Green dashed lines show location of true parameter.

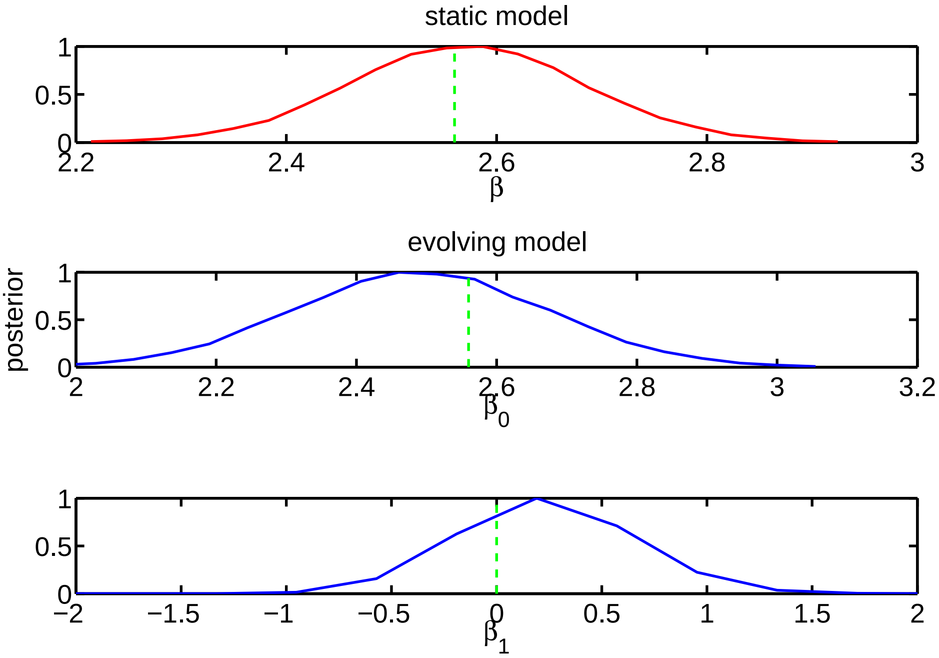


Figure 7.20: A simulated data set in which β is static such that $\beta(z) = \beta$. Upper panel: fitted using an the correct static model (red). Lower panels: fitted using the incorrect evolving model (blue) $\beta(z) = \beta_0 + z\beta_1$. Parameter inference for the incorrect case correctly shows that the additional degree of freedom is not necessary and the evolving contribution β_1 may be set to zero. Here the true model with $\beta(z) = \beta$ is nested with in the more complex model. Green dashed lines show location of true parameter.

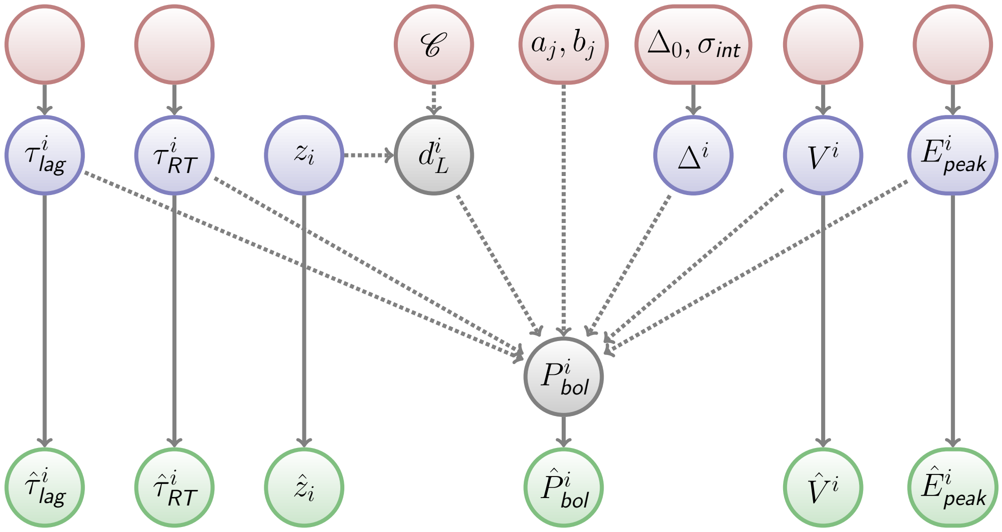


Figure 7.21: Bayesian hierarchical network showing the deterministic (dashed) and probabilistic (solid) connections between variables in our Bayesian hierarchical model (BHM). Variables of interest are in red, latent (unobserved) variables are in blue and observed data (denoted by hats) are in green. Note the additional parameter Δ^i which models the intrinsic dispersion. In this case $\Delta_0 \equiv 0$

Chapter 8

Robustness to Systematic Error for Future Dark Energy Probes

The Fisher matrix formalism outlined in section 5.3 gives a useful methodology for predicting the error ellipses around the fiducial model on the parameters of interest for future proposed astrophysical probes of cosmology. One of the most common Figures of Merit (FoM) used as a metric for comparing proposed probes is the inverse area of the error ellipse derived from the Fisher matrix formalism (Albrecht et al., 2006, 2009) which gives a measure of the expected statistical power or ability of a probe to be able to constrain the parameters of interest. Alternative FoMs in higher dimesnions are given by (Huterer & Turner, 2001; Albrecht & Bernstein, 2007; Wang, 2008; Crittenden et al., 2009; Mortonson et al., 2010), and a more general Bayeisan approach to FoMs is given in (Trotta et al., 2010). The purpose of the FoM is to evaluate in advance the expected statistical power of future probes. Survey parameters can be adjusted in order to maximise the statistical power of a particular probe, and proposed probes can be ranked by their FoM. This ranking can then assist in the decision making process of how to allocate limited resources to get the best science return.

One of the limitations of the inverse area FoM is that it does not take into consideration what would happen if either the current or future proposed probe were biased - these inverse area FoMs are evaluated at the fiducial model in parameter space. But what happens if the future proposed probe is biased, that is shifted somewhat in parameter space with respect to the fiducial model due to unforeseen systematic errors? How can we evaluate the possible impact of bias inducing systematic errors which shift the proposed probe with respect to the fiducial model? How can we quantify which probes are more or less robust to systematic errors? In order to address these questions and provide a method for quantifying the robustness to systematic errors of future proposed probes, we have extended the FoM formalism to give a new 'Robustness' statistic.

In this chapter I first give an overview of the standard statistical FoM and then introduce our new Robustness FoM, discussing its properties with reference to a Gaussian linear model. I then present a case study in which I apply the Robustness FoM to two future dark energy probes, supernovae type Ia (SNe Ia) and Baryon Acoustic Oscillations (BAO). This chapter follows closely the paper March et al. (2011b).

8.1 Figures of Merit for Future Dark Energy Probes

8.1.1 Gaussian linear model

Suppose there are two different dark energy probes, whose likelihood function is assumed to be Gaussian and is characterized by a Fisher matrix (i.e. inverse covariance matrix) L_i ($i = 1, 2$), i.e.

$$\mathcal{L}_i(\Theta) \equiv p(d_i|\Theta) = \mathcal{L}_0^i \exp\left(-\frac{1}{2}(\mu_i - \Theta)^t L_i(\mu_i - \Theta)\right) \tag{8.1}$$

where Θ are the parameters of interest and μ_i is the location of the maximum likelihood value in parameter space. In the absence of systematic errors, the maximum likelihood point, μ_i, is located at the true value of the parameters (i.e. fiducial values), which here we take to be the origin. Here we neglect realization noise, i.e. such that μ_i is interpreted as the expectation value of the maximum likelihood estimator averaged over many data realizations. However, the presence of unforeseen systematic errors could introduce a non-zero shift in μ_i. In sections 8.4.2 and 8.4.1 we shall show how plausible but unforeseen systematics in the observables of two different dark energy probes translate into a systematic shift in μ_i in the w_0, w_a parameter space. The posterior distribution for the parameters, $p(\Theta|D)$, is obtained by Bayes theorem as

$$p(\Theta|D) = \frac{p(\Theta)p(D|\Theta)}{p(D)}, \tag{8.2}$$

where $p(\Theta)$ is the prior, $p(D)$ the Bayesian evidence and D are the data being used. If we assume a Gaussian prior centered on the origin with Fisher matrix Σ, the posterior from each probe is also a Gaussian, with Fisher matrix

$$F_i = L_i + \Sigma \quad (i = 1, 2) \tag{8.3}$$

and posterior mean

$$\overline{\mu}_i = F_i^{-1}(L_i \mu_i) \tag{8.4}$$

If we combine the two probes, we obtain a Gaussian posterior with Fisher matrix

$$F = L_1 + L_2 + \Sigma \tag{8.5}$$

and mean

$$\mu = F^{-1} \sum_{i=1}^{2} L_i \mu_i. \tag{8.6}$$

The precision of the posterior, as characterised by itsFisher matrix, does not depend on the degree of overlap of the likelihoods of the individual probes. In terms of their associated error ellipses, the area of the joint posterior ellipse depends only on the size and relative orientation of the error ellipses associated with the individual probes, it does not depend on their degree of overlap or relative position in parameter space. A FoM which depends only on the the joint posterior Fisher matrix assigns equal merit to an unbiased proposed future probe, and a biassed future probe subject to systematic error, as both would have the same FoM. This prompts us to define an additional FoM, which we call 'Robustness' which quantifies the effect of bias due to systematic error.

8.1.2 The Statistical Figure of Merit

As described in section 5.3, the reciprocal of the area of the error ellipse is often used to describe the statistical power of a future dark energy probe. This FoM or measure of statistical performance, widely known as the DETF FoM (Albrecht et al., 2006; Huterer & Turner, 2001), is usually defined (up to multiplicative constants) as

$$|L_i|^{1/2} \tag{8.7}$$

As an alternative to the standard statistical FoM described by Eq. (8.7) we suggest to adopt a more statistically motivated measure of the information gain, namely the Kullback-Leibler divergence (KL) between the posterior and the prior, representing the information gain obtained when upgrading the prior to the posterior via Bayes theorem:

$$D_{KL} \equiv \int p(\Theta|D) \ln \frac{p(\Theta|D)}{p(\Theta)} d\Theta \tag{8.8}$$

The KL divergence measures the relative entropy between the two distributions: it is a dimensionless quantity which expresses the information gain obtained via the likelihood. For an example case study with Gaussian prior:

$$p(\Theta) = \mathcal{N}_\Theta(\theta_\pi, \Sigma) \tag{8.9}$$

and a Gaussian likelihood

$$\mathcal{L}(\Theta) = \mathcal{L}_0 \exp\left(-\frac{1}{2}(\theta_0 - \Theta)^t L(\theta_0 - \Theta)\right) \tag{8.10}$$

the evidence for data d is given by

$$p(d) \equiv \int \mathrm{d}\Theta p(d|\Theta)p(\Theta) \tag{8.11}$$

$$= \mathcal{L}_0 \frac{|\Sigma|^{1/2}}{|F|^{1/2}} \exp\left[-\frac{1}{2}\left(\theta_0^t L\theta_0 + \theta_\pi^t \Sigma\theta_\pi - \overline{\theta}^t F\overline{\theta}\right)\right] \tag{8.12}$$

where $F = L + \Sigma$ and $\overline{\theta} = F^{-1}(L\theta_0 + \Sigma\theta_\pi)$

the information gain (w.r.t. the prior Σ) from the combination of both probes is given by

$$D_{KL} = \frac{1}{2}\left(\ln|F| - \ln|\Sigma| - \mathrm{tr}[1 - \Sigma \mathrm{F}^{-1}]\right) \tag{8.13}$$

In section 8.1.3 we shall be interested in assessing the statistical performance of future dark energy probes, in a context where probe 1 is taken to represent present-day constraints on dark energy parameters, while probe 2 is a future proposed dark energy mission. We normalize the KL divergence for the combination of probe 1 and probe 2, given by Eq. (8.13), w.r.t. the case where probe 2 is assumed to be a hypothetical experiment that would yield identical dark energy constraints as the existing ones (probe 1). This is not meant to represent a realistic dark energy probe, but merely to give a benchmark scenario for the normalization of the information gain. This choice of normalization has the added advantage of cancelling out most of the prior dependence in Eq. (8.13). After exponentiating the normalized KL divergence, we therefore suggest to adopt as a statistical FoM the dimensionless quantity

$$\begin{aligned} S \equiv & \frac{|L_1 + L_2 + \Sigma|^{1/2}}{|2L_1 + \Sigma|^{1/2}} \\ & \times \exp\left(\frac{1}{2}\mathrm{tr}[\Sigma((\mathrm{L}_1 + \mathrm{L}_2 + \Sigma)^{-1} - (2\mathrm{L}_1 + \Sigma)^{-1})]\right). \end{aligned} \tag{8.14}$$

Thus far we have not dealt with the issue of systematic errors leading to bias, but have merely suggested an alternative to the existing statistical FoM, in the next section we shall develop the formalism which leads to the additional Robustness FoM which deals with systematic errors.

8.1.3 Robustness of Dark Energy Probes

In order to quantify the robustness to potential systematics of a combination of probes, we wish to derive a measure of the degree of consistency between them. The fundamental idea underlying our new robustness FoM is that our confidence in the robustness of a new dark energy probe is increased if it returns constraints which overlap significantly with previously existing probes constraints. If on the contrary the new probe has a small degree of consistency with previous experiments, this might point to either a failure of the underlying theoretical model or to the presence of unforseen and unmodelled systematics in the new probe (or both). In the following, we focus on the latter hypothesis.

The idea is to perform a Bayesian model comparison between two hypotheses, namely $\mathcal{H}_0$, stating that the data D are all compatible with each other and the model, versus $\mathcal{H}_1$, purporting that the observables are incompatible and hence tend to pull the constraints in different regions of parameter space. The Bayes factor between the two hypotheses, giving the relative probabilities (odds) between $\mathcal{H}_0$ and $\mathcal{H}_1$ is given by

$$R = \frac{p(D|\mathcal{H}_0)}{\prod_{i=1}^{2} p(d_i|\mathcal{H}_1)}, \tag{8.15}$$

where the Bayesian evidence for a given hypothesis $\mathcal{H}$ is

$$p(d|\mathcal{H}) = \int \mathrm{d}\Theta p(d|\Theta, \mathcal{H}) p(\Theta|\mathcal{H}). \tag{8.16}$$

If $R \gg 1$, this is evidence in favour of the hypothesis $\mathcal{H}_0$ that the data are compatible. If instead $R \ll 1$ the alternative hypothesis $\mathcal{H}_1$ is preferred (namely, that the data are incompatible). Examples of the application of the statistics R introduced above can be found in Hobson et al. (2002); Feroz et al. (2008) – see the Appendix of Feroz et al. (2009) for a toy model illustration. For a review of Bayesian methods in cosmology, and in particular of model selection techniques, see Trotta (2008).

We can restrict our considerations to just two probes, hence $D = \{d_1, d_2\}$. Then the criterium of Eq. (8.15) can be written as (omitting for brevity the explicit conditioning on hypotheses)

$$R = \frac{p(d_1, d_2)}{p(d_1)p(d_2)} = \frac{p(d_2|d_1)p(d_1)}{p(d_1)p(d_2)} = \frac{p(d_2|d_1)}{p(d_2)}. \tag{8.17}$$

The conditional evidence for d_2 given dataset d_1 can be calculated as

$$p(d_2|d_1) = \int p(d_2|\Theta)p(\Theta|d_1)d\Theta, \tag{8.18}$$

where the first term is the likelihood for the second probe and the second term is the posterior from the first probe. By using the likelihood (8.1), and making use of Eq. (8.12) we obtain:

$$\begin{aligned} p(d_2|d_1) = \mathcal{L}_0^{(2)} & \frac{|F_1|^{1/2}}{|F|^{1/2}} \\ & \exp\left[-\frac{1}{2}\left(\mu_2^t L_2 \mu_2 + \overline{\mu}_1^t F_1 \overline{\mu}_1 - \mu^t F \mu\right)\right], \end{aligned} \tag{8.19}$$

where μ is given by Eq. (8.6), F by Eq. (8.5) and $\overline{\mu}_1$ by Eq. (8.4). Using again Eq. (8.12) we obtain for the denominator in Eq. (8.17)

$$p(d_2) = \mathcal{L}_0^{(2)} \frac{|\Sigma|^{1/2}}{|F_2|^{1/2}} \exp\left[-\frac{1}{2}\left(\mu_2^t L_2 \mu_2 - \overline{\mu}_2^t F_2 \overline{\mu}_2\right)\right], \tag{8.20}$$

so that we obtain

$$\begin{aligned} R = & \frac{|F_1|^{1/2}|F_2|^{1/2}}{|F|^{1/2}|\Sigma|^{1/2}} \\ & \exp\left[-\frac{1}{2}\left(\overline{\mu}_1^t F_1 \overline{\mu}_1 + \overline{\mu}_2^t F_2 \overline{\mu}_2 - \mu^t F \mu\right)\right]. \end{aligned} \tag{8.21}$$

Therefore we can recast Eq. (8.15) into

$$\begin{aligned} \ln R = & \frac{1}{2}\mu^t F \mu - \frac{1}{2}\sum_{i=1}^{2} \overline{\mu}_i^t F_i \overline{\mu}_i \\ & - \frac{1}{2}\ln\frac{|F|}{|\Sigma|} + \frac{1}{2}\sum_{i=1}^{2}\ln\frac{|F_i|}{|\Sigma|}. \end{aligned} \tag{8.22}$$

We shall use below the robustness R to define a new FoM. For now, let us notice that it is the product of two terms: the terms involving determinants of the Fisher matrices add up to an Occam's razor factor, which is always > 0. The second part (summing over quadrating forms involving the various Fisher matrices) expresses the degree of overlap of the posteriors from the two probes. This term will reduce R if the posteriors from the two probes are significantly displaced from the posterior obtained using the combined data set (a smoking gun for systematic bias). The generalization of Eq. (8.22) to an arbitrary number of probes is derived in section 8.2.

8.1.4 The Robustness Figure of Merit

We now specialize to the situation where probe 1 describes our current knowledge about dark energy parameter, while probe 2 represents a proposed future dark energy mission. Notice that probe 1 does not need to be a single experiment (i.e., just SN Ia or just BAO), but it can be interpreted as being the effective joint constraint from a combination of all available present-day dark energy probes. Without loss of generality, we assume that the current constraints are unbiased, i.e. we set $\mu_1 = 0$ in the following, and we wish to evaluate the robustness of a future probe, as defined in Eq. (8.22), which might be subject to systematic bias.

Let us assume for the moment being that we can estimate the bias b in parameter space which probe 2 might be subject to. A procedure to achieve this will be presented below for the specific cases of SN Ia and BAO observations. For now, we remain completely general, and assume that the maximum likelihood estimate for the dark energy parameters from probe 2 is displaced from their true value by a bias vector b, i.e. $\mu_2 = b$. This, together with the assumption that probe 1 is unbiased (i.e., $\mu_1 = 0$) gives $\bar{\mu}_2 = F_2^{-1} L_2 b$ and the joint posterior mean from both probes is

$$\mu = F^{-1} L_2 b. \tag{8.23}$$

Then we can write for the robustness R, Eq. (8.22)

$$\begin{aligned} \ln R = &\frac{1}{2}(F^{-1}L_2 b)^t F (F^{-1} L_2 b) - \frac{1}{2}(bL_2)^t F_2^{-1}(L_2 b) \\ &- \frac{1}{2}\ln\frac{|F|}{|\Sigma|} + \frac{1}{2}\sum_{i=1}^{2}\ln\frac{|F_i|}{|\Sigma|}, \end{aligned} \tag{8.24}$$

which can be rewritten as

$$\begin{aligned}\ln R &= -\frac{1}{2}(bL_2)^t(F_2^{-1} - F^{-1})(L_2 b) + \mathcal{R}_0 \\ &= -\frac{1}{2}b^t F^* b + \mathcal{R}_0\end{aligned} \tag{8.25}$$

where we have defined

$$F^* \equiv L_2(F_2^{-1} - F^{-1})L_2 \tag{8.26}$$

$$\mathcal{R}_0 \equiv -\frac{1}{2}\ln\frac{|F|}{|F_1|}\frac{|\Sigma|}{|F_2|}. \tag{8.27}$$

If the prior Σ is negligible with respect to L_2 we have $F_2 = L_2$ and $F^* = F_2 - F_2 F^{-1} F_2$.

In order to normalize the value of R, we adopt the 'repeated experiment' procedure we used for the normalization of the statistical FoM. This is defined as the hypothetical case where the new experiment (probe 2) yields exactly the same Fisher matrix as the existing probe 1, and is unbiased, i.e. $F_1 = F_2$ and $b = (0, 0)$. For this identically repeated case the robustness of the two probes is given by

$$R_* = \frac{|F_1|}{(|2L_1 + \Sigma||\Sigma|)^{1/2}}. \tag{8.28}$$

Normalizing R from Eq. (8.25) to the above value means that $R/R_* = 1$ is the robustness that one would achieve by carrying out a new dark energy measurement that would yield exactly the same constraints as we currently have, and no bias. We therefore define the quantity

$$R_N \equiv \frac{R}{R_*} \tag{8.29}$$

as our robustness FoM, which expresses the robustness of probe 2 under the assumption that it will be affected by a bias b.

The robustness FoM above represents a "worst case scenario" (for a given b) for probe 2, because we are assuming that it will be for sure systematically biased. A more balanced approach is to average the robustness along the direction defined by the systematic bias vector b. This gives an "average robustness", which accounts for different possible sizes in the strength of the bias[1]. In order to perform the average,

---

[1] Notice that as we average R along b we do not re-evaluate the Fisher matrix of the probe as a function of b, but we simply translate the Fisher matrix found at the fiducial point (i.e., the true parameters values). The Fisher matrix typically depends only weakly on the fiducial model chosen, as long as we consider the models within the parameter confidence region. If the bias vector is not much larger than the statistical errors we can therefore approximate the Fisher matrix at the

we rotate the coordinate axes so that the new x-axis is aligned with the vector b (assuming here a 2-dimensional parameter space for simplicity):

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \Lambda b \tag{8.30}$$

where Λ is a suitable rotation matrix, and $\Lambda^t = \Lambda^{-1}$. Then the average robustness along the direction defined by b is given by

$$\langle R \rangle \equiv \int W(x) \frac{R}{R_*} dx = \frac{e^{\mathcal{R}_0}}{R_*} \int W(x) e^{-\frac{1}{2} D_{11} x^2} dx \tag{8.31}$$

where

$$D \equiv \Lambda F^* \Lambda^t \tag{8.32}$$

and $W(x)$ is a suitable weighting function. A natural choice for W is a Gaussian with characteristic scale for the bias given by the length of the bias vector, $|b|$,

$$W(x) = \frac{1}{\sqrt{2\pi}|b|} e^{-\frac{1}{2}\frac{x^2}{|b|^2}}, \tag{8.33}$$

so that Eq. (8.31) becomes

$$\begin{aligned} \langle R \rangle &= \frac{e^{\mathcal{R}_0}}{R_* \sqrt{2\pi}|b|} \int e^{-\frac{1}{2}x^2 \left[D_{11} + |b|^{-2} \right]} dx \\ &= \frac{(|F_2||2L_1 + \Sigma|)^{1/2}}{|FF_1|^{1/2}} ||b|^2 D_{11} + 1|^{-1/2}, \end{aligned} \tag{8.34}$$

The Gaussian weight is centered at the unbiased parameter values, but it also has a tail that stretches above the characteristic scale of the bias, $|b|$, in order to account for a potentially much larger bias. We have checked that the use of other weight functions (e.g., a top-hat weight out to a maximum bias value given by the size of the bias vector) give a qualitatively similar result. We define the quantity given by Eq. (8.34) as the "average robustness" FoM.

Finally, we can also combine the statistical and robustness FoMs to obtain an overall FoM expressing both the statistical power and the robustness to systematic

---

biased parameters values with the one evaluated at the fiducial point.

bias of the probe as

$$T_N \equiv R_N \times S, \tag{8.35}$$

$$\langle T\rangle \equiv \langle R\rangle \times S, \tag{8.36}$$

where S is given by Eq. (8.14), while R_N and $\langle R\rangle$ by Eqs. (8.29) and (8.31), respectively.

8.2 Generalization to arbitrary number of probes

The generalization of Eq. (8.22) to an arbitrary number of probes proceeds as follows. First, we notice that one can always summarize current constraints from several observations in one single joint posterior. Let us call the data from the combination of all available present-day probes d_0 (with Fisher matrix F_0). If one wishes to consider N future probes, we can ask whether all of the N probes are mutually compatible[2]. Eq. (8.17) gives in this case

$$\begin{aligned} R_{\text{all}} &= \frac{p(d_N d_{N-1}\dots d_1|d_0)}{\prod_{j=1}^N p(d_j|d_0)} && (8.38)\\ &= \prod_{j=1}^N \frac{p(d_j|d_{j-1}\dots d_1 d_0)}{p(d_j|d_0)} && (8.39)\\ &= \prod_{j=2}^N \frac{p(d_j|d_{j-1}\dots d_1 d_0)}{p(d_j|d_0)} && (8.40)\end{aligned}$$

where in the last line we have cancelled out the very last term in both the numerator and the denominator, so that the sum starts with $j = 2$. We now refer to Eq. (8.19)

---

[2] An alternative test would be to check whether the N-th probe is compatible with the previous $N-1$ (assuming those are already available and they are free of systematics themselves). In this case the relevant quantity is

$$R_{\text{N}} = \frac{p(d_N|d_{N-1}\dots d_1)}{p(d_N)p(d_{N-1}\dots d_1)} \tag{8.37}$$

which can be computed by appropriate substitutions in Eq. (8.22).

to obtain

$$\begin{aligned} p(d_j|d_{j-1}\dots d_1 d_0) = \mathcal{L}_0^{(j)} \frac{|F_{012\dots(j-1)}|^{1/2}}{|F_{012\dots j}|^{1/2}} \\ \times \quad \exp\left[-\frac{1}{2}\left(\mu_j^t L_j \mu_j\right.\right. \\ + \quad \mu^t_{012\dots(j-1)} F_{012\dots(j-1)} \mu_{012\dots(j-1)} \\ - \quad \left.\left.\mu^t_{012\dots j} F_{012\dots j} \mu_{012\dots j}\right)\right], \end{aligned} \tag{8.41}$$

where the definitions correspond to those before, so that $F_{012\dots j} \equiv F_0 + \sum_{i=1}^{j} L_i$, and in particular $F_{012\dots N} \equiv F$. Notice already that most terms in the numerator of Eq. (8.40) will cancel. Similarly, following Eq. (8.20)

$$\begin{aligned} p(d_j|d_0) &= \mathcal{L}_0^{(2)} \frac{|F_0|^{1/2}}{|F_{0j}|^{1/2}} \\ &\times \quad \exp\left[-\frac{1}{2}\left(\mu_j^t L_j \mu_j + \mu_0^t F_0 \mu_0 - \mu_{0j}^t F_{0j} \mu_{0j}\right)\right], \end{aligned} \tag{8.42}$$

Now one can evaluate Eq. (8.40) with the help of Eqs. (8.41) and (8.42):

$$\begin{aligned} R_{\mathrm{all}} = \frac{|F_{01}|^{1/2}}{|F|^{1/2}} \prod_{j=2}^{N} \frac{|F_0|^{-1/2}}{|F_{0j}|^{-1/2}} \times \exp\left[-\frac{1}{2}(\mu_{01}^t F_{01} \mu_{01} - \mu^t F \mu\right. \\ \left. - (N-1)\mu_0^t F_0 \mu_0 + \sum_{j=2}^{N} \mu_{0j}^t F_{0j} \mu_{0j})\right] \end{aligned} \tag{8.43}$$

We thus obtain for the robustness

$$\begin{aligned} \ln R_{\mathrm{all}} = \frac{1}{2}\left(\sum_{i=1}^{N} \ln|F_{0i}| - (N-1)\ln|F_0| - \ln F\right) \\ - \frac{1}{2}(\sum_{i=1}^{N} \overline{\mu}_{0i}^t F_{0i} \overline{\mu}_{0i} - (N-1)\mu_0^t F_0 \mu_0 - \mu^t F \mu), \end{aligned} \tag{8.44}$$

which generalizes Eq. (8.22).

8.3 Properties of the Robustness FoM

Before applying the above formalism to future dark energy probes, we wish to gain some further insight into the behaviour of our robustness FoM by considering it in the context of a Gaussian toy model. We start with the normalized expression for the average robustness, Eq. (8.34) and assume now that the confidence regions of the two probes are identical up to a roto-translation (and therefore the determinants of F_1, F_2 are equal). If moreover the prior is very weak we can approximate the posterior with the likelihood, hence

$$\langle R \rangle \approx 2\frac{|F_1|}{|F|}||b|^2 D_{11} + 1|^{-1/2}. \tag{8.45}$$

Let us further assume that probes 1 and 2 are aligned, i.e., they have a degeneracy direction lying along the same straight line. This means also that their Fisher matrices are simultaneously diagonalizable (i.e. they commute) and that F is also diagonalizable. Since the bias vector b by definition connects the maximum likelihood points of the two probes, its direction is also aligned with one of the principal axis of the probes in this particular example. Then we can write

$$\begin{aligned} D &= \Lambda(F_2 - F_2 F^{-1} F_2)\Lambda^{-1} & (8.46)\\ &= F_2^D - F_2^D (F_1^D + F_2^D)^{-1} F_2^D & (8.47) \end{aligned}$$

where the superscript D denotes the diagonalized version of a matrix. The last step follows because for any matrix A diagonalized by Λ and any power k one has

$$\Lambda A^k \Lambda^{-1} = (A^D)^k. \tag{8.48}$$

Now let us denote the length of the j-th semiaxis of the i-th probe by $\sigma_{i,j}$ where (after diagonalization) the semiaxis $j = 1(2)$ lies along the abscissa (ordinate) . Then we have

$$D_{11} = \sigma_{2,1}^{-2}\left(1 - \frac{\sigma_{1,1}^2}{\sigma_{1,1}^2 + \sigma_{2,1}^2}\right) \tag{8.49}$$

and therefore

$$\langle R \rangle \approx \frac{2(\sigma_{2,1}\sigma_{2,2})}{(\sigma_{1,2}^2 + \sigma_{2,2}^2)^{1/2}(|b|^2 + \sigma_{1,1}^2 + \sigma_{2,1}^2)^{1/2}}. \tag{8.50}$$

This expression shows that the average robustness is invariant with respect to rescaling of the axes: in fact, if the distances along the abscissa, $\sigma_{1,1}, \sigma_{2,1}, |b|$, are rescaled by an arbitrary factor, $\langle R \rangle$ does not change; the same applies in the y-direction.

Since we assumed the ellipses to be congruent, we have two qualitatively different cases: orthogonal ellipses ($\perp$), i.e. $\sigma_{2,2} = \sigma_{1,1}$ and $\sigma_{2,1} = \sigma_{1,2}$; and parallel ellipses ($\|$), i.e. $\sigma_{1,1} = \sigma_{2,1}$ and $\sigma_{1,2} = \sigma_{2,2}$. In the orthogonal case we obtain

$$\langle R \rangle^{\perp} = \frac{2r}{1+r^2}\left(1 + \frac{|b|^2 r^2}{\sigma_{2,1}^2(1+r^2)}\right)^{-1/2} \tag{8.51}$$

where $r = \sigma_{2,1}/\sigma_{2,2}$ measures the elongatedness of the ellipses. In the parallel case we obtain instead for any r

$$\langle R \rangle^{\|} = \left(1 + \frac{|b|^2}{2\sigma_{2,1}^2}\right)^{-1/2} \tag{8.52}$$

From these expressions we can derive some general consequences. Because of our choice of normalization, unbiased identical probes have unity robustness. In general, if the bias length is small with respect to the statistical errors of the second probe, then parallel probes are more robust than orthogonal ones. If the second probe is very elongated (degenerated) along the bias direction, i.e. $r \gg 1$, then again parallel probes are more robust than orthogonal ones. If instead the degeneracy of the second probe lies orthogonally to the bias direction, $r \ll 1$, there are two cases: parallel probes are more robust if the bias is smaller than the long axis ($|b|^2 < \sigma_{2,2}$), but less robust in the opposite case. Of course the general case, with arbitrary bias direction and length and arbitrary sizes and orientation of the probes cannot be reduced to such simple conclusions.

Armed with the above intuition, we now consider in Fig. 8.1, a numerical illustration of 4 different cases for the relative orientation of the two probes (orthogonal or parallel) and the direction of the bias vector (along the short or long semiaxis). The two sets of iso-likelihood contours enclose 68% and 95% confidence levels; as above, the second probe (blue contours) has the same area as the first (i.e., $L_1 = L_2$), but its degeneracy direction can be rotated, and its maximum likelihood value is displaced from the true value by a systematic bias (of fixed length in all cases), given by the green vector. The first probe (red contours) is assumed to be unbiased. The prior is $\Pi = \text{diag}(1, 1)$ (i.e. a prior of 1.0 in each parameter with no correlations imposed). For each case, we give the corresponding statistical FoM, Eq. (8.14), the robustness FoMs, (8.29) and (8.31), and the total FoM (for the averaged robustness), Eq. (8.36).

The robustness FoM (with or without averaging) depends both on the direction along which the bias is directed and on the relative orientation of the degeneracy directions of the two probes. When the bias is directed along the degeneracy direction of probe 1 and probe 2 is aligned along that direction (lower left panel), the robustness is maximal. It decreases if the two probes are orthogonal to each other, since this reduces the degree of overlap between them (upper panels). Finally, robustness is smallest when the two probes are aligned but the bias is direct orthogonally w.r.t the degeneracy direction (lower right panel), as argued above. Looking ahead to the application of the robustness formalism to the dark energy equation of state parameters in the next section, we can anticipate here that the most relevant case is the one where the two probes are similarly oriented (bottom panels of Fig. 8.1). This is because different dark energy probes are typically degenerate in the equation of state parameters along quite similar directions. Therefore, their relative robustness can be expected to depend mainly on the orientation of the bias w.r.t. the main degeneracy direction.

The statistical FoM is largest when the probes are orthogonal to each other, as expected. Notice that the statistical FoM is unaffected by the bias, and only depends on the relative alignment of the two probes. For a given orientation and size of the bias vector, the total FoM allows one to decide which configuration for probe 2 is to be preferred. For the example of Fig. 8.1, if the bias vector points along the degeneracy direction of probe 1 (left-hand side panels), one would prefer probe 2 to be aligned with probe 1 ($\langle T \rangle = 0.71$) as opposed to probe 2 being orthogonal to probe 1 ($\langle T \rangle = 0.61$). If instead the bias is orthogonal to the degeneracy of probe 1 (right-hand side panels), then the best choice for probe 2 is for it to be orthogonal to probe 1 ($\langle T \rangle = 0.62$ compared to $\langle T \rangle = 0.44$).

We can also ask what is the optimal orientation of probe 2 with respect to probe 1 if one wanted to maximise its robustness, given a bias direction. In Fig. 8.2, we plot both the statistical and the average robustness FoMs as a function of the rotation angle between the principal direction of the two probes. The average robustness is evaluated for 3 different directions of the bias (coloured vectors in the top panel). We notice once more that the statistical FoM is maximised when the probes are orthogonal. However, the robustness FoM is maximised when the degeneracy direction of probe 2 is close to being aligned with the direction of the bias vector, as this maximizes the overlap with probe 1 even when probe 2 suffers from a systematic error. Finally, increasing the length of the bias by a factor of 2 (fainter green vector in the top panel) reduces the overall average robustness.

In summary, the robustness of a future probe is a function of its statistical properties (i.e., the direction along which its main degeneracy is aligned, compared with the degeneracy direction of probe 1) as well as of the direction and size of the systematic bias. The performance of a future probe should be assessed by considering simultaneously its statistical power but also its robustness to systematics.

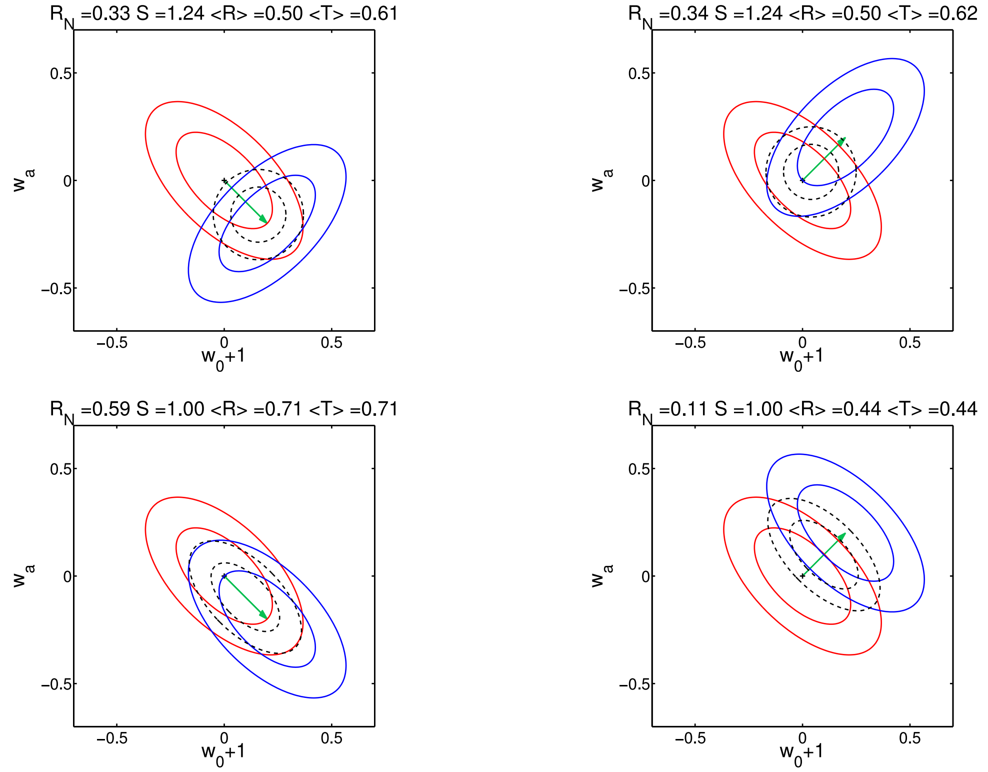


Figure 8.1: Illustration of statistical and robustness FoM for a future probe (blue ellipses, 68% and 95% C.L.) which is systematically biased w.r.t. the present-day constraints (red ellipses) in the direction given by the green bias vector. The black dotted ellipses represent the combined constraints. Notice that the statistical FoM (S) does not change in the presence of a systematic bias.

Optimizing a future dark energy experiment in terms of its statistical errors alone would generically lead to an experiment which is less robust, for a given overall level of plausible systematics. Any optimization procedure should therefore involve the complementary criteria of statistical strength and robustness to systematic bias.

We now turn to applying the above concept to the concrete scenario of two classes of future dark energy missions, namely type Ia SN and BAO measurements.

8.4 Robustness of Future Dark Energy Probes

We consider a simple and widely used phenomenological description of an evolving dark energy model, where the equation of state is $w(z) = w_0 + w_a z/(1+z)$, characterized by the two free parameters (w_0, w_a) (Chevallier & Polarski, 2001; Linder, 2003). For probe 1 (representing current constraints on w_0, w_a) we take a Gaussian approximation to the joint likelihood resulting from the combination of Union 2

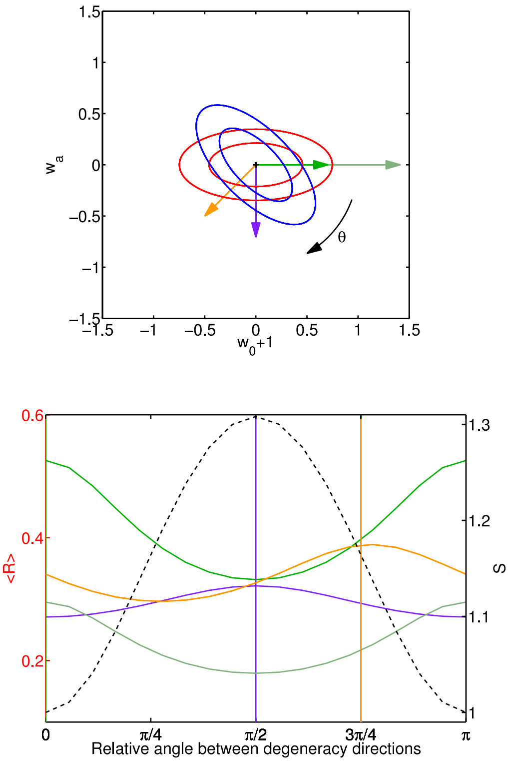


Figure 8.2: Dependency of FoMs on the angle between the degeneracy direction of the two probes. Upper panel: the red (blue) ellipses represent the 68% and 95% likelihood contours of probe 1 (probe 2, which is potentially biased). The degeneracy direction of probe 2 is offset by an angle θ w.r.t probe 1. The three vectors gives three possible directions for the bias. Lower panel: value of statistical FoM S (black dashed line, right-hand axis), and average robustness FoM, $\langle R \rangle$, (coloured solid lines, left-hand axis, colour and thickness matching the bias vectors in the upper panel), as a function of the relative angle θ. Vertical coloured lines give the angle of each bias vector.

SNe Ia data (Amanullah et al., 2010), SDSS BAO (Percival et al., 2010), WMAP7 measurements of the shift parameters (Komatsu et al., 2010), and SHOES measurements of the Hubble constant (Riess et al., 2009). We further assume a flat Universe. For the prior on the dark energy parameters, we take a Gaussian centered at $(w_0 + 1, w_a) = (0, 0)$ with Fisher matrix $\Pi = \text{diag}(1, 1/100)$. In the following, when we look at the Fisher matrix we mean the 2D marginalised Fisher matrix (i.e., marginalised down to the dark energy parameter space). Although in the rest of this paper we focus exclusively on the robustness FoM for dark energy parameters, we note that our robustness formalism is equally applicable to any other cosmological parameter one is interested in.

In order to evaluate robustness, we need to specify the bias vector b. There are several plausible ways of doing this, and the outcome will depend on what one thinks a possible systematic bias might be due to. In our case, in order to illustrate our new FoM, we determine b by assuming a possible systematic bias in the probe's observables, and then projecting the resulting systematic shift onto the dark energy parameter space of interest, as described in detail below. We stress that this is by no means the only procedure by which one can estimate b. Other assumptions about the origin of systematic errors will in general lead to a different b, and therefore to a different value for the robustness of the future probe.

8.4.1 Future SN Ia Measurements

The data, analysis and text in this section 8.4.1 are the work of Dragan Huterer and are not my own.

We consider a survey dedicated to observing type Ia supernovae from space, with a redshift distribution like the one expected from SNAP, with 2000 SNe distributed as in Kim et al. (2004), plus a low-z sample of 300 SNe distributed uniformly in the redshift range $0.03 < z < 0.08$. The projected SNAP magnitude errors include both statistical and systematic components, and are modelled as follows:

$$\sigma_b^2 = \left[\frac{0.15^2}{N_b} + A_{\text{syst}}^2 \left(\frac{1+z_b}{1+z_{\text{max}}} \right)^2 \right], \tag{8.53}$$

where N_b is the number of SNe in each bin centered at z_b and of width $dz = 0.1$. The second term on the right-hand side of Eq. (8.53) models a systematic floor that increases linearly with z up to a maximum of A_{syst} mag per $dz = 0.1$ bin at $z_{\text{max}} = 1.7$ (Linder & Huterer, 2003). In order to evaluate the robustness of SNa data for different levels of systematics, we will consider values of $A_{\text{syst}} = 0.01, 0.02, 0.05$.

We assume a flat universe with four parameters relevant for this analysis, matter density relative to critical Ω_M, equation of state today w_0, its variation with scale factor w_a, and a nuisance offset in the Hubble diagram $\mathcal{M}$. Marginalizing over Ω_M and $\mathcal{M}$ and assuming $A_{\text{syst}} = 0.02$, we find that our fiducial survey produces statistical errors of $\sigma_{w_0} = 0.075$ and $\sigma_{w_a} = 0.30$, corresponding to the black 68% C.L. ellipse in Fig. 8.3. The bias in the dark energy parameters, b, reconstructed from SN measurements induced by an arbitrary bias in the observed magnitudes $\delta m(z)$ can be derived from the Fisher matrix for SNe (e.g. Knox et al. (1998); Huterer &

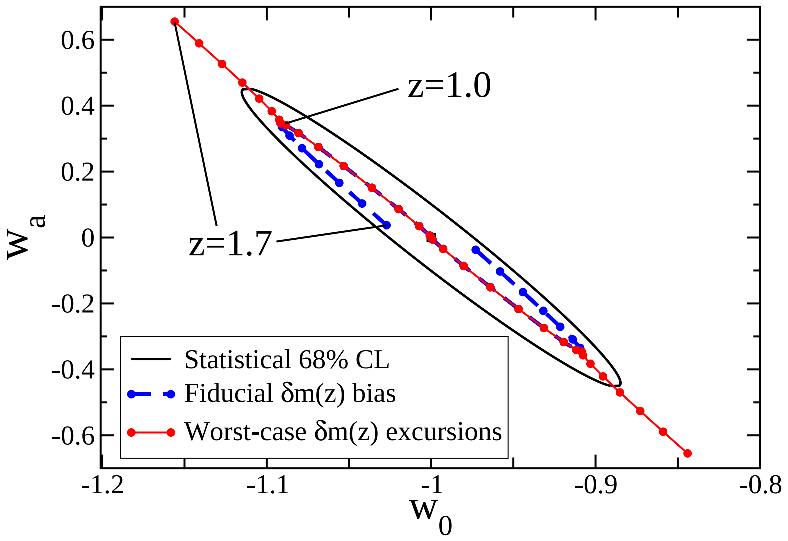


Figure 8.3: Systematic bias in the w_0-w_a plane for future SNIa data. The square denotes our fiducial value and the ellipse gives the 68% CL statistical contour from future SNIa data. The blue curve shows the systematic bias given by Eq. (8.56), with points showing cumulative contributions from each of the 16 redshift bins – that is, cumulative value of the sum in Eq. (8.55).For clarity, we explicitly label bias contributions accumulated by redshifts $z = 1$ and $z = 1.7$. The red segments denote the worst-case bias, where the sign of $\delta m(z)$ at each redshift bin conspires to shift the (w_0, w_a) value away from the true value ("Maximum excursion bias"); see Eq. (8.57). For clarity, we have also plotted the biases with the opposite sign relative to the fiducial model parameter values. *This figure is the work of Dragan Huterer.*

Turner (2001)), and is given by

$$b_i = \sum_j (F^{-1})_{ij} \sum_\alpha \frac{dm(z_\alpha)}{d\mu_j} \frac{1}{\sigma_\alpha^2} \delta m_{\rm sys}(z_\alpha) \tag{8.54}$$

$$\equiv \sum_\alpha c_\alpha^{(i)} \delta m_{\rm sys}(z_\alpha) \tag{8.55}$$

where μ_i are the cosmological parameters, $c_\alpha^{(i)} \equiv \sum_j (F^{-1})_{ij}(dm(z_\alpha)/d\mu_j)/\sigma_\alpha^2$ and where α runs over redshift bins. We adopt a systematic bias of the same form as the "floor" that was previously included in the total *statistical* error:

$$\delta m_{\rm sys}(z_\alpha) = A_{\rm syst} \left(\frac{1+z_\alpha}{1+z_{\rm max}} \right) \tag{8.56}$$

Bias of this magnitude leads to the bias on cosmological parameters which can be calculated using Eqs. (8.55) and (8.56), and is shown as the blue curve in Fig. 8.3. Each point on the curve shows cumulative contributions to the excursion in w_0 and w_a around the fiducial model (with $(w_0, w_a) = (-1, 0)$) for each of the 16 redshift bins we consider, $z \in [0.1, 0.2], \ldots, z \in [1.6, 1.7]$. In other words, points on the blue curve show cumulative contribution of the sum in Eq. (8.55).

But this form of the bias assumes that the excursions in $\delta m(z_\alpha)$ are of the same sign (taken to be positive), and equal to the maximum allowed value in Eq. (8.56). The worst-case bias to the dark energy parameters is obtained if $\delta m(z_\alpha)$ changes sign in each redshift bin just so as to maximize the excursion in w_0 and w_a. Such a worst-case bias can be straightforwardly calculated (Huterer & Takada, 2005)

$$b_i^{\text{worst}} = \sum_\alpha |c_\alpha^{(i)}|\, \delta m_{\text{sys}}(z_\alpha) \tag{8.57}$$

for a single dark energy parameter μ_i, where $\delta m_{\text{sys}}(z_\alpha) > 0$ was taken by default. In other words, the systematic error takes a plus or minus sign equal to that of the coefficient $c_\alpha^{(i)}$ in each redshift bin [3]. Such a worst-case excursion in the (w_0, w_a) plane is shown as the red curve with points in Fig. 8.3. We call this scenario the "maximum excursion bias" (MEB), and use it as an estimate for the bias vector b in the computation of our robustness FoM.

8.4.2 Future Baryonic Acoustic Oscillations Measurements

The data, analysis and text in this section 8.4.2 are the work of Luca Amendola and are not my own.

The second class of future probe we consider consists of a full-sky spectroscopic redshift survey modelling a future spacee mission with specifications close to WFIRST or Euclid (or a Stage-IV mission in the language of the DETF). The probe is fully specified by choosing a number of redshift bins and giving the expected number densities of galaxies per bin and the sky coverage, assumed here to be 20,000 square degrees. Table 8.1 gives the redshift binning and the galaxy number densities, taken from the data published by the Euclid collaboration (Laureijs et al., 2009). We assume however that only half of these galaxies can be effectively employed (efficiency $\epsilon = 0.5$), corresponding to one half of values in the Table.

In order to forecast the statistical errors on dark energy parameters, we adopt

---

[3] For multiple parameters, there is ambiguity to define the worst-case error, since a sign of $\delta m_{\text{sys}}(z_\alpha)$ that makes excursion in w_0 positive may actually make the w_a excursion negative or vice versa. We make a choice that the excursion in w_0 is positive in a given redshift bin, which determines the sign of $\delta m_{\text{sys}}(z_\alpha)$; then the excursion in w_a in that bin is simply $c_\alpha^{(w_a)}\, \delta m_{\text{sys}}(z_\alpha)$.

z	$n(z) \times 10^{-3}$
0.5-0.7	3.56
0.7-0.9	2.42
0.9-1.1	1.81
1.1-1.3	1.44
1.3-1.5	0.99
1.5-1.7	0.55
1.7-1.9	0.29
1.9-2.1	0.15

Table 8.1: Expected galaxy number densities per redshift bin in units of $(h/\mathrm{Mpc})^3$ for the Euclid survey.

the Fisher matrix method of Seo & Eisenstein (2003, 2007), also employed in Amendola et al. (2005). Here we give a short summary of the method and refer to these papers for the implementation details. In the limit where the survey volume $V_{\rm survey}$ is much larger than the scale of any features in $P_{\rm obs}(k)$, it has been shown that the redshift survey Fisher matrix in a redshift bin Δz can be approximated as (Tegmark, 1997)

$$\begin{aligned} F_{ij} &= \int_{-1}^{1}\int_{k_{\rm min}}^{k_{\rm max}} \frac{\partial \ln P_{\rm obs}(k,\mu)}{\partial \mu_i}\frac{\partial \ln P_{\rm obs}(k,\mu)}{\partial \mu_j} \\ &\times \ V_{\rm eff}(k,\mu)\frac{2\pi k^2 dk d\mu}{2(2\pi)^3} \end{aligned} \tag{8.58}$$

Here, k, μ are the wavevector moduls and direction cosine with respect to the line of sight, respectively, and the derivatives are evaluated on the parameters μ_i of the fiducial model. The upper cut-off $k_{\rm max}$ is chosen so as to avoid the non-linear regime, while the large-scale cut-off $k_{\rm min}$ is set to $0.001h/\mathrm{Mpc}$ but its precise value has a very weak impact. $V_{\rm eff}$ is the effective volume of the survey:

$$V_{\rm eff}(k,\mu) = \left[\frac{n_{\rm g}P_{\rm g}(k,\mu)}{n_{\rm g}P_{\rm g}(k,\mu)+1}\right]^2 V_{\rm survey}, \tag{8.59}$$

where $V_{\rm survey}$ is the 20,000 square degrees survey volume contained in a given redshift bin. The galaxy comoving number density $n_{\rm g}(z)$ is assumed to be spatially constant within a redshift bin, while P_g is the galaxy spectrum defined below. The total Fisher matrix is obtained by summing over all the redshift bins of Table 8.1. The matter power spectrum in any given cosmology can be written in terms of the spectrum in the fiducial (or "reference", subscript "ref") cosmology as

$$P_{\rm obs}(k_{{\rm ref}\perp}, k_{{\rm ref}\parallel}, z) = \frac{D(z)^2_{\rm ref} H(z)}{D(z)^2 H(z)_{\rm ref}} P_{\rm g}(k_{{\rm ref}\perp}, k_{{\rm ref}\parallel}, z) + P_{\rm shot}\,, \tag{8.60}$$

where

$$P_{\rm g}(k_{{\rm ref}\perp}, k_{{\rm ref}\parallel}, z) = b(z)^2 \left[1 + \beta(z) \frac{k^2_{{\rm ref}\parallel}}{k^2_{{\rm ref}\perp} + k^2_{{\rm ref}\parallel}}\right]^2 P_{\rm matter}(k, z)\,. \tag{8.61}$$

In Eq. (8.60), $H(z)$ and $D(z)$ are the Hubble parameter and the angular diameter distance, respectively, and the prefactor encapsulates the geometrical distortions due to the Alcock-Paczynski effect (Seo & Eisenstein, 2003, 2007). $k_\perp$ and $k_\parallel$ are the wave-numbers across and along the line of sight in the given cosmology, and they are related to the wave-numbers calculated assuming the reference cosmology by $k_{{\rm ref}\perp} = k_\perp D(z)/D(z)_{\rm ref}$ and $k_{{\rm ref}\parallel} = k_\parallel H(z)_{\rm ref}/H(z)$. $P_{\rm shot}$ is the unknown white shot noise that remains even after the conventional shot noise of inverse number density has been subtracted (Seo & Eisenstein, 2003, 2007). In Eq. (8.61), $b(z)$ is the *linear bias* factor between galaxy and matter density distributions, $f_g(z)$ is the linear growth rate, $\beta(z) = f_g(z)/b(z)$ is the linear redshift-space distortion parameter (Kaiser, 1987) and $P_{\rm matter}$ is the linear matter power spectrum. The fiducial values for the bias and the growth factor are $b(z) = 1$ and $f_g = \Omega_M^{0.545}$, respectively. In Di Porto et al. (2011) it has been shown that the precise fiducial value of $b(z)$ does not have a large impact on the results.

This method employs all the information contained in the power spectrum, including the redshift distortion, not just the position of the baryonic wiggles. As above, we choose a flat fiducial cosmology with $\Omega_M = 0.24$, $h = 0.7$, $\Omega_{DE} = 0.737$, $\Omega_K = 0$, $\Omega_b h^2 = 0.0223$, $n_s = 0.96$, $w_0 = -1, w_a = 0$. Unlike in the SN case, we do not impose an explicit systematic floor in the forecasted BAO errors; the finite sky coverage and density of galaxies provide an effective floor for any given BAO survey. As mentioned above, beside the cosmological parameters, for each redshift bin we also include as free parameters to be differentiated (and then marginalized) in the Fisher matrix a matter-galaxy bias factor and an additive shot noise term in the power spectrum (for details see Amendola et al. (2005)). These terms act as additional effective systematic floors.

The systematic effect we assume for the redshift survey is a fractional error in estimating the value of the Hubble function $H(z_i)$ of magnitude $A_{\rm syst} = 0.001, 0.002, 0.005$ in each bin i. Such a bias in $H(z)$ propagates to a bias in the angular diameter distance $D(z)$, as well, if the standard flat-space Friedman-Robertson-

Walker relation

$$D(z) = (1+z)^{-1} \int_0^z \frac{dz'}{H(z')} \tag{8.62}$$

holds true, which we assume here. The angular diameter distance bias is then related to the Hubble function bias by

$$\delta(\ln D) = -\delta(\ln H) \frac{H(z)}{(1+z)D(z)} \int_0^z \frac{dz'}{H^2(z')}. \tag{8.63}$$

where we have used the assumption that the bias in $\ln H$ is redshift-independent. This simple choice for modelling systematic errors in BAO is meant to approximately capture a possible systematic shift in the baryon peak position due to e.g. the presence of isocurvature modes (Zunckel et al., 2011) or non-linear effects, of the kind described e.g. in Seo et al. (2008). A more realistic choice of systematic errors is difficult to model accurately (as, for example, a bias in $H(z)$ and/or $D(z)$ also modifies in general the whole spectrum behavior and the redshift distortions), and it is left for future work. Our present choice is meant as a simple illustration of the method and a first step towards evaluating the robustness FoM.

If instead of the true matter power spectrum, $P(k)$, we measure a spectrum that contains a systematic error $\delta s_\alpha = \delta(\ln H_\alpha)$ or $\delta s_\alpha = \delta(\ln D_\alpha)$ in the value of $H(z_\alpha)$ and $D(z_\alpha)$ (where the systematic shifts are related by Eq. (8.63)), the maximum likelihood estimate for the i-th parameter will be shifted w.r.t. its true value by a bias given by (see e.g. Taylor et al. (2007))

$$\begin{aligned} \delta\mu_i &= F_{ij}^{-1} \left[\frac{1}{8\pi^2} \int d\mu k^2 dk \frac{\partial \ln P}{\partial \mu_j} \frac{\partial \ln P}{\partial s_\alpha} \right] \delta s_\alpha \\ &\equiv c_\alpha^{(i)} \delta s_\alpha \end{aligned} \tag{8.64}$$

(sum over repeated indexes). Analogously to the previous subsection we have defined

$$c_\alpha^{(i)} \equiv F_{ij}^{-1} \left[\frac{1}{8\pi^2} \int d\mu k^2 dk \frac{\partial \ln P}{\partial \mu_j} \frac{\partial \ln P}{\partial s_\alpha} \right] . \tag{8.65}$$

In this particular case, however, the i-th parameters coincide with $\delta s_i = \delta(\ln H_i), \delta(\ln D_i)$ and therefore the matrix $c_\alpha^{(i)}$ is the identity matrix. We can then directly project the systematic bias onto the dark energy parameters (w_0, w_a), obtaining a bias vector b

of the form

$$b_l = \sum_\beta \left(\frac{\partial w_l}{\partial \ln H(z_\beta)} - \right. \\ \left. \frac{H(z_\beta)}{(1+z_\beta)D(z_\beta)} \int_0^{z_\beta} \frac{dz'}{H^2} \frac{\partial w_l}{\partial \ln D(z_\beta)} \right) \delta \ln H(z_\beta) \tag{8.66}$$

where $\delta \ln H(z_\beta) = 0.001, 0.002, 0.005$, the subscript β runs over the redshift bins, and $l = 0, a$. We have chosen to consider systematic shifts in the range of 0.1% to 0.5% to reflect ballpark estimates of what BAO systematic errors due e.g. to residual non-linear corrections might be. We stress once more that this is a simplified treatment used here mainly for illustration purposes of our method.

We evaluate b_l for each redshift bin and then estimate the maximum bias by following the same method discussed in the previous subsection. Here it happens that the contributions to b_l are always positive and therefore automatically select the worst case scenario, i.e., the maximum excursion bias. The resulting maximum excursion bias for different levels of systematics is shown as the green vectors in the bottom panel of Fig. 8.4, together with the statistical errors from our BAO probe (blue ellipses) and current constraints (red ellipses), plotted for comparison.

8.5 Results

Our results for the statistical and robustness FoMs are summarized in Table 8.2, where we give the values of our robustness, statistical and total FoM. We also show the value of the DETF FoM (normalized to the value obtained from the current probes) for comparison.

First, by inspecting Fig. 8.4, we notice that the systematic bias projected onto the (w_0, w_a) plane is much better aligned with the degeneracy direction of the probes for SNIa than for BAO. From our discussion in section 8.3, this leads to expect a higher value for the robustness FoM for SNIa than for BAO. Also, the size of the bias vectors in the dark energy parameters is roughly comparable for SNIa and BAO, although in the latter case we have adopted a bias in the observables (H and D) which is a factor of 10 smaller than for the SNIa observables (the magnitudes). Table 8.2 shows that indeed both the robustness and the average robustness FoMs are slightly larger for SNIa than for BAO across the range of systematic error levels we adopted for each probe. This is a consequence of the fact that the BAO bias leads to a smaller degree of overlap of the BAO constraints with the present-day constraints, which is a more serious lack of robustness than for the SNIa. In the latter

			BAO			SNIa		
			Maximum excursion bias			Maximum excursion bias		
Fom	Symbol	Defined in	0.1%	0.2%	0.5%	1%	2%	5%
Robustness	R_N	Eq. (8.29)	1.4	0.83	0.026	1.7	1.3	0.24
Average robustness	$\langle R \rangle$	Eq. (8.34)	1.4	1.1	0.54	1.7	1.4	0.81
Statistical FoM	S	Eq. (8.14)	2.7			7.0		
Total FoM	T_N	Eq. (8.35)	3.6	2.2	0.070	11.9	9.1	1.7
Total average FoM	$\langle T \rangle$	Eq. (8.36)	3.8	2.9	1.4	11.9	9.8	5.7
Bias length	$\lvert b \rvert$	caption	0.18	0.36	0.90	0.34	0.68	1.7
DETF FoM		Eq. (8.7)	4.4			13		

Table 8.2: Robustness and statistical Figure of Merits for future BAO and SNIa surveys, for different levels of systematic errors in the observables. We also give the DETF FoM for comparison (normalized to its value from current constraints). We also give the length of the bias vector b in the (w_0, w_a) plane. The Maximum excursion bias errors refer to both $D(z_b)$ and $H(z_b)$ in the case of BAO, and $m(z_b)$ in the case of SNIa.

case, although the bias vectors are slightly larger in the dark energy parameters (typically by a factor of 2, cf Table 8.2), the bias direction is well aligned with the statistical degeneracy, and therefore the reduction in the overlap between the present constraints and future SNIa constraints is less severe, translating in a higher robustness. For the highest level of systematic error in each case (0.5% for BAO and 5% for SNIa), we find that the robustness FoM for BAO is about a factor of 10 smaller than for SNIa. The average robustness of BAO is also smaller, but only by about 1/3, which reflects the more balanced assessment given by the average robustness. Thus, for our particular choice of systematics, our findings run against the general lore that BAO observations are more robust to systematics than SNIa.

In terms of our statistical FoM, the SNIa survey is better by a factor of about 3, in good agreement with the result obtained from the usual DETF FoM. Taken together, the better values of both the statistical and robustness FoM for SNIa lead to a higher value of the total FoM for SNIa than for BAO.

It is important to stress that our robustness results above are *not* a generic feature of SNIa and BAO observations. Rather, they reflect our specific choices for the systematic bias in the observables for BAO and SNIa. Other choices of systematic bias are possible and will in general give a different results for the robustness, which we shall explore in a dedicated paper.

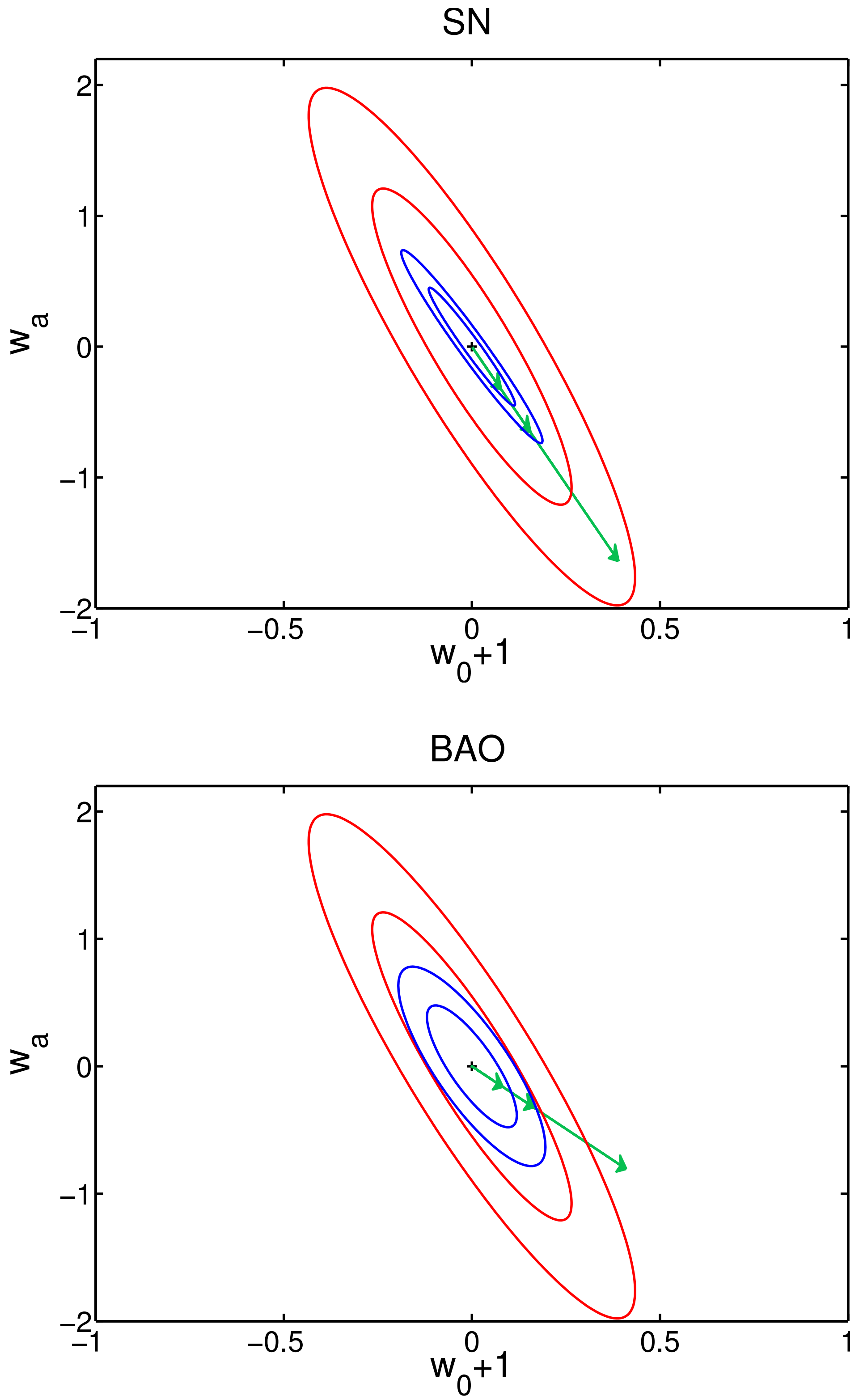


Figure 8.4: Construction of the robustness FoM for a future SNIa survey (top panel) and a future BAO Euclid-like survey (bottom panel). Red ellipses show current 68%, 95% constraints (in a Gaussian approximation) from a combination of all available probes, blue ellipses show projected constraints from the future probe at the fiducial point (assumed to be ΛCDM). The green vectors show the systematic maximum excursion bias (MEB) for systematic erros of 1%, 2% and 5% for SNIa and 0.1%, 0.2% and 0.5% for BAO.

8.6 Conclusions

We have introduced a new formalism to quantify the robustness of future dark energy probes to systematic bias, and argued that this important new quantity should be taken into account when evaluating the performance of future surveys. In constrast to usual measures of statistical performance, our robustness FoMs depend on the direction and size of the systematic bias induced in the dark energy parameters by residual systematics in the observables. We have thus described an approach to include the effect of systematic errors in the dark energy figures of merit.

We have applied this formalism to future SNIa and BAO probes by developing a simple phenomenological model of possible residual systematic errors. Our results indicate that – for the specific choice of systematics adopted here – SNIa are slightly more robust to systematics than BAO, despite having assumed a systematic shift in the observables for SNIa which is a factor of 10 larger than for BAO. Coupled with the higher statistical performance of SNIa, this would lead to prefer SNIa over BAO in terms of their overall FoM. It is clear however that this particular result cannot be generalized beyond our choice of systematics and surveys. In a future work we will investigate how this result change by adopting more refined descriptions of the systematic bias for each probe.

Chapter 9

Summary and Conclusions

This thesis sought to provide practical solutions to the question of how we translate observations of astrophysical phenomena into probabilistic statements about the parameters of the underlying physical theories such as dark energy; how we quantify our degree of belief about those fundamental cosmological models; how we discriminate between different cosmological models.

In this thesis I have presented our work on extensions to Bayesian model selection, i.e. Bayesian Doubt (chapter 6), cosmological parameter inference from the supernovae type Ia (chapter 7) and a new figure of merit, Robustness (chapter 8) for planning future surveys.

In our work on Bayesian Doubt we developed a new methodology for assigning an absolute value to the quality of a model, rather than just having the relative measure of the goodness of a model. In the work presented in chapter 6 I described the concept of Bayesian Doubt and applied it to the question of whether the ΛCDM model is absolutely the best possible model given the data, or whether there was room in model space for a better as yet un-thought of model. The analysis supported the claim that the ΛCDM model should not be doubted, however, as more data are gathered then the level of doubt may be re-assessed, possibly leading to model discovery.

In the future I want to use Bayesian model selection and Bayesian Doubt in conjunction with fast Nested Sampling algorithms such as MultiNest to discriminate between the other types of model offered in explanation for late time acceleration. I also want to apply Bayesian Doubt to different categories of cosmological model on offer (e.g. modified gravity) to provide a benchmark ideal yet realistic model within each category against which the various models can be compared. This would give an absolute measure of how well any one model within the class performs, and would indicate whether there is room with the current model space for an additional as yet unknown model.

The work on cosmological parameter inference presented in chapter 7 developed a new Bayesian Hierarchical Model approach to parameter inference which successfully overcomes the shortcomings of the standard χ^2 approach, now analysing the SNe Ia data in a way which is fully consistent with Bayesian model selection techniques. This methodology allows for many other applications to the SNe Ia data, for example investigating the colour evolution of the SNe Ia. The Bayesian Hierarchical Model developed can also be applied to similar problems in cosmology which suffer from the same problem as the SNe Ia data of having large error bars compared with the range of the data points. We are currently developing this method for use in inferring the cosmological parameters from the Gamma Ray Burst data.

Strategic planning for future missions or astrophysical probes of cosmology is an area which requires careful statistical analysis to predict which missions and which configurations of those missions give the best returns for the investment. Previously, the ability of a particular mission to be able to constrain the parameters of interest was quantified by a figure of merit which was typically the reciprocal area of the Fisher matrix of those parameters. However, this standard figure of merit did not take into account the possible effect of unforeseen systematic errors. We introduced a new figure of merit, ‘Robustness’ which does quantify by how much a particular proposed astrophysical probe would be affected by unforeseen systematic errors. We evaluated the Robustness of a plausible future SNe Ia mission, and a future BAO mission and found that in this particular example the SNe Ia were more robust than the BAO.

In the future I intend to use the Robustness statistic to quantify how well different proposed survey configurations perform, for different types of dark energy probes, in order to identify which are the best configurations to use. As new types of dark energy probe are developed I will use Robustness to quantify to what degree the proposed new probe is compatible with existing probes, thereby providing strategic information, which can be used to decide which are the most robust probes to invest in.

References

Albrecht A., et al., 2006, astro-ph/0609591

—, 2009, arXiv:0901.0721

Albrecht A. J., Bernstein G., 2007, Phys. Rev., D75, 103003

Amanullah, et al., 2010, Ap.J., 716, 712

Amendola L., Quercellini C., Giallongo E., 2005, MNRAS, 357, 429

Amendola L., Tsujikawa S., 2010, Dark Energy: Theory and Observations, Amendola, L. & Tsujikawa, S., ed. Cambridge University Press

Antoniou I., Perivolaropoulos L., 2010, JCAP, 12, 12

Astier P., Guy J., 2006, A&A, 447, 31

Baade W., Zwicky F., 1934, Proceedings of the National Academy of Science, 20, 254

Bassett B. A., Fantaye Y., Hlozek R., Kotze J., 2009a, arXiv:0906.0993

—, 2009b, arXiv0906.0974

Bellhouse D., 2004, Stat. Sci., 19, 3

Blake C., Glazebrook K., 2003, ApJ, 594, 665

Bondi H., 1947, MNRAS, 107, 410

Bordag M., Mohideen U., Mostepanenko V. M., 2001, Physics Reports, 353, 1

Box G., Tiao G., 1992, Bayesian inference in statistical analysis, Wiley classics library. Wiley

Bridle S. L., Lewis A. M., Weller J., Efstathiou G., 2003, MNRAS, 342, L72

Bucher M., Moodley K., Turok N., 2000, PhysRevD, 62, 083508

Caldwell R. R., 2002, Physics Letters B, 545, 23

Casimir H. B. G., 1948, Indag. Math., 10, 261

Chevallier M., Polarski D., 2001, Int. J. Mod. Phys., D10, 213

Chevallier M., Polarski D., 2001, International Journal of Modern Physics D, 10, 213

Clarkson C., 2009, arXive:0911.2601

Clifton T., Ferreira P. G., Padilla A., Skordis C., 2011, ArXiv e-prints

Cooke R., Lynden-Bell D., 2010, MNRAS, 401, 1409

Copeland E. J., Sami M., Tsujikawa S., 2006, International Journal of Modern Physics D, 15, 1753

Crittenden R. G., Pogosian L., Zhao G.-B., 2009, JCAP, 0912, 025

D'Agostini G., 1995, arXiv:hep-ph/9512295

—, 2005, arXiv:0511182

Di Porto C., Amendola L., Branchini E., 2011, arXive:1101.2453

Diaferio A., Ostorero L., Cardone V. F., 2011, arXive:astro-ph/1103.5501

Dodelson S., 2003, Modern cosmology. Amsterdam (Netherlands): Academic Press

Dunkley J., Komatsu E., Nolta M. R., Spergel D. N., Larson D., Hinshaw G., Page L., Bennett C. L., Gold B., Jarosik N., Weiland J. L., Halpern M., Hill R. S., Kogut A., Limon M., Meyer S. S., Tucker G. S., Wollack E., Wright E. L., 2009, ApJS, 180, 306

Eisenstein D. J., Zehavi I., Hogg D. W., Scoccimarro R., Blanton M. R., Nichol R. C., Scranton R., Seo H.-J., Tegmark M., Zheng Z., Anderson S. F., Annis J., Bahcall N., Brinkmann J., Burles S., Castander F. J., Connolly A., Csabai I., Doi M., Fukugita M., Frieman J. A., Glazebrook K., Gunn J. E., Hendry J. S., Hennessy G., Ivezić Z., Kent S., Knapp G. R., Lin H., Loh Y.-S., Lupton R. H., Margon B., McKay T. A., Meiksin A., Munn J. A., Pope A., Richmond M. W., Schlegel D., Schneider D. P., Shimasaku K., Stoughton C., Strauss M. A., SubbaRao M., Szalay A. S., Szapudi I., Tucker D. L., Yanny B., York D. G., 2005, ApJ, 633, 560

Enqvist K., 2008, General Relativity and Gravitation, 40, 451

Fang W., Hu W., Lewis A., 2008b, PhysRevD, D78, 087303

Fang W., et al., 2008a, PhysRevD, D78, 103509

Feroz F., Cranmer K., Hobson M., de Austri R. R., Trotta R., 2011, arXiv:1101.3296

Feroz F., Hobson M. P., 2008a, MNRAS, 384, 449

—, 2008b, MNRAS, 384, 449

Feroz F., Hobson M. P., Bridges M., 2009a, MNRAS, 398, 1601

—, 2009b, MNRAS, 398, 1601

Feroz F., Hobson M. P., Roszkowski L., Ruiz de Austri R., Trotta R., 2009, arXive:0903.2487

Feroz F., et al., 2008, JHEP, 10, 064

Filippenko A. V., 1997, Annual Review of Astronomy and Astrophysics, 35, 309

Freedman W. L., Madore B. F., Gibson B. K., Ferrarese L., Kelson D. D., Sakai S., Mould J. R., Kennicutt Jr. R. C., Ford H. C., Graham J. A., Huchra J. P., Hughes S. M. G., Illingworth G. D., Macri L. M., Stetson P. B., 2001, ApJ, 553, 47

Friedmann A., 1924, Zeitschrift fur Physik, 21, 326

Frieman J. A., Turner M. S., Huterer D., 2008, ARA&A, 46, 385

Frodesen A., Skjeggestad O., Tøfte H., 1979, Probability and statistics in particle physics, Probability and Statistics in Particle Physics No. v. 1. Universitetsforl.

Gamezo V. N., Khokhlov A. M., Oran E. S., 2005, ApJ, 623, 337

Garnavich P. M., Kirshner R. P., Challis P. a., 1998, ApJL, 493, L53+

Gelman A., Rubin D., 1992, Statistical Science, 7, 457

Gilfanov M., Bogdán Á., 2010, Nature, 463, 924

Gordon C., Trotta R., 2007, MNRAS, 382, 1859

Green J., Schechter P., Baltay C., Bean R., Bennett D., Brown R., Conselice C., Donahue M., Gaudi S., Lauer T., Perlmutter S., Rauscher B., Rhodes J., Roellig T., Stern D., Sumi T., Tanner A., Wang Y., Wright E., Gehrels N., Sambruna R., Traub W., 2011, arXive:1108.1374s

Gregory P. C., 2005, Bayesian Logical Data Analysis for the Physical Sciences: A Comparative Approach with 'Mathematica' Support. Cambridge University Press

Gull S., 1989, Maximum Entropy and Bayesian Methods, 511

Guth A. H., 1981, PhysRevD, 23, 347

Guy J., Astier P., Baumont S., Hardin D., Pain R., Regnault N., Basa S., Carlberg R. G., Conley A., Fabbro S., Fouchez D., Hook I. M., Howell D. A., Perrett K., Pritchet C. J., Rich J., Sullivan M., Antilogus P., Aubourg E., Bazin G., Bronder J., Filiol M., Palanque-Delabrouille N., Ripoche P., Ruhlmann-Kleider V., 2007, A&A, 466, 11

Guy J., Astier P., Nobili S., Regnault N., Pain R., 2005, A&A, 443, 781

Hastings W. K., 1970, Biometrika, 57, 97

Hillebrandt W., Niemeyer J. C., 2000, ARA&A, 38, 191

Hobson M. P., Bridle S. L., Lahav O., 2002, MNRAS, 335, 377

Hobson, M. P., Jaffe, A. H., Liddle, A. R., Mukeherjee, P., & Parkinson, D., ed., 2010, Bayesian Methods in Cosmology. Cambridge University Press

Howell D. A., 2011, Nature Communications, 2

Howell D. A., Sullivan M., Brown E. F., Conley A., Le Borgne D., Hsiao E. Y., Astier P., Balam D., Balland C., Basa S., Carlberg R. G., Fouchez D., Guy J., Hardin D., Hook I. M., Pain R., Perrett K., Pritchet C. J., Regnault N., Baumont S., LeDu J., Lidman C., Perlmutter S., Suzuki N., Walker E. S., Wheeler J. C., 2009, ApJ, 691, 661

Hu W., 2008, Phys. Rev., D77, 103524

Hu W., Sawicki I., 2007, PhysRevD, D76, 104043

Hu W., Sugiyama N., 1995, ApJ, 444, 489

Hubble E., 1929, Proceedings of the National Academy of Science, 15, 168

Huterer D., Takada M., 2005, APP, 23, 369

Huterer D., Turner M. S., 1999, PhysRevD, 60, 081301

Huterer D., Turner M. S., 2001, PhysRevD, D64, 123527

Jaynes E. T., Baierlein R., 2004, Physics Today, 57, 100000

Jha S., Riess A. G., Kirshner R. P., 2007, ApJ, 659, 122

Kaiser N., 1987, MNRAS, 227, 1

Kasen D., Röpke F. K., Woosley S. E., 2009, Nature, 460, 869

Kessler R., Becker A. C., Cinabro, 2009a, ApJS, 185, 32

Kessler R., Bernstein J. P., Cinabro D., Dilday B., Frieman J. A., Jha S., Kuhlmann S., Miknaitis G., Sako M., Taylor M., Vanderplas J., 2009b, PASP, 121, 1028

Khokhlov A. M., 1991, A&A, 245, 114

Kim A. G., Linder E. V., Miquel R., Mostek N., 2004, MNRAS, 347, 909

Knop R. A., Aldering G., Amanullah R., Astier P., 2003, ApJ, 598, 102

Knox L., Scoccimarro R., Dodelson S., 1998, PhysRevLet, 81, 2004

Kolb E. W., Matarrese S., Notari A., Riotto A., 2005, PhysRevD, 71, 023524

Komatsu, et al., 2010, arXiv:1001.4538

Komatsu E., Dunkley J., Nolta M. R., Bennett C. L., Gold B., Hinshaw G., Jarosik N., Larson D., Limon M., Page L., Spergel D. N., Halpern M., Hill R. S., Kogut A., Meyer S. S., Tucker G. S., Weiland J. L., Wollack E., Wright E. L., 2009, ApJS, 180, 330

Kowalski M., Rubin D., Aldering G., Agostinho R. J., Amadon A., 2008, ApJ, 686, 749

Krause O., Tanaka M., Usuda T., Hattori T., Goto M., Birkmann S., Nomoto K., 2008, Nature, 456, 617

Kunz M., Sapone D., 2006, PhysRevD, 74, 123503

Laureijs R., et al., 2009, arXiv:0912.0914

Lewis A., Bridle S., 2002, PhysRevD, 66, 103511

Lewis A., Challinor A., Lasenby A., 2000, ApJ, 538, 473

Li M., Li X., Zhang X., 2009, arXive:0912.3988

Liddle A. R., Mukherjee P., Parkinson D., 2006a, Astron. Geophys., 47, 4.30

Liddle A. R., Mukherjee P., Parkinson D., Wang Y., 2006b, PhysRevD, D74, 123506

Linde A., 2005, arXiv:hep-th/050320

Linde A. D., 1984, Reports on Progress in Physics, 47, 925

Linder E. V., 2003, PhysRevLet, 90, 091301

Linder E. V., Huterer D., 2003, PhysRevD, D67, 081303

Loredo T. J., 1990, From laplace to supernova sn 1987a: Bayesian inference in astrophysics

Mackay D. J. C., 2003, Information Theory, Inference and Learning Algorithms, Mackay, D. J. C., ed.

Mandel K. S., Narayan G., Kirshner R. P., 2010, arXiv:1011.5910

Mandel K. S., Narayan G., Kirshner R. P., 2011, ApJ, 731, 120

Mandel K. S., Wood-Vasey W. M., Friedman A. S., Kirshner R. P., 2009, ApJ, 704, 629

Maor I., Brustein R., Steinhardt P. J., 2001, PhysRevLet, 86, 6

March M. C., Starkman G. D., Trotta R., Vaudrevange P. M., 2011a, MNRAS, 410, 2488

March M. C., Trotta R., Amendola L., Huterer D., 2011b, MNRAS, 415, 143

March M. C., Trotta R., Berkes P., Starkman G. D., Vaudrevange P. M., 2011c, arXive:1102.3237

Marra V., Kolb E. W., Matarrese S., Riotto A., 2007, PhysRevD, 76, 123004

Metropolis N., Rosenbluth A. W., Rosenbluth M. N., Teller A. H., Teller E., 1953, J. Chem .Phys., 21, 1087

Miknaitis G., Pignata G., 2007, ApJ, 666, 674

Misner C. W., Thorne K. S., Wheeler J. A., 1973, Gravitation. San Francisco: W.H. Freeman and Co.

Mortonson M. J., Huterer D., Hu W., 2010, PhysRevD, D82, 063004

Mukhanov V., 2005, Physical foundations of cosmology. Cambridge University Press

Mukherjee P., Parkinson D., Liddle A. R., 2006, ApJL, 638, L51

Nolta M. R., Dunkley J., Hill R. S., Hinshaw G., Komatsu E., Larson D., Page L., Spergel D. N., Bennett C. L., Gold B., Jarosik N., Odegard N., Weiland J. L., Wollack E., Halpern M., Kogut A., Limon M., Meyer S. S., Tucker G. S., Wright E. L., 2009, ApJS, 180, 296

Nomoto K., Sugimoto D., Neo S., 1976, Astrophysics and Space Science, 39, L37

Nomoto K., Thielemann F.-K., Yokoi K., 1984, ApJ, 286, 644

NRC, 2010, New worlds, new horizons in astronomy and astrophysics

Parkinson D., Mukherjee P., Liddle A. R., 2006, PhysRevD, 73, 123523

Peacock J. A., 1999, Cosmological Physics. Cambridge University Press

Peebles P. J., Ratra B., 2003, Reviews of Modern Physics, 75, 559

Percival W. J., et al., 2010, MNRAS, 401, 2148

Perlmutter S., Aldering G., Goldhaber G., Knop R. A., Nugent P., Castro P. G., Deustua S., Fabbro S., Goobar A., Groom D. E., Hook I. M., Kim A. G., Kim M. Y., Lee J. C., Nunes N. J., Pain R., Pennypacker C. R., Quimby R., Lidman C., Ellis R. S., Irwin M., McMahon R. G., Ruiz-Lapuente P., Walton N., Schaefer B., Boyle B. J., Filippenko A. V., Matheson T., Fruchter A. S., Panagia N., Newberg H. J. M., Couch W. J., The Supernova Cosmology Project, 1999a, ApJ, 517, 565

Perlmutter S., Turner M. S., White M., 1999b, PhysRevLet, 83, 670

Peskin M., Schroeder D., Martinec E., 1996, An Introduction to Quantum Field Theory. Addison-Wesley Advanced Book Program

Phillips M. M., 1993, ApJL, 413, L105

Protassov R., van Dyk D. A., Connors A., Kashyap V. L., Siemiginowska A., 2002, ApJ, 571, 545

Rasanen S., 2010, arXive:1012.0784

Reichardt C. L., Ade P. A. R., Bock J. J., Bond J. R., Brevik J. A., Contaldi C. R., Daub M. D., Dempsey J. T., Goldstein J. H., Holzapfel W. L., Kuo C. L., Lange A. E., Lueker M., Newcomb M., Peterson J. B., Ruhl J., Runyan M. C., Staniszewski Z., 2008, arXive:0801.1491

Riess A. G., Filippenko A. V., Challis P., Clocchiatti A., Diercks A., Garnavich P. M., Gilliland R. L., Hogan C. J., Jha S., Kirshner R. P., Leibundgut B., Phillips M. M., Reiss D., Schmidt B. P., Schommer R. A., Smith R. C., Spyromilio J., Stubbs C., Suntzeff N. B., Tonry J., 1998, AJ, 116, 1009

Riess A. G., Macri L., Casertano S., Sosey M., Lampeitl H., Ferguson H. C., Filippenko A. V., Jha S. W., Li W., Chornock R., Sarkar D., 2009, ApJ, 699, 539

Riess A. G., Strolger L., 2004, ApJ, 607, 665

—, 2007, ApJ, 659, 98

Riley Hobson B., 2004, Mathmatical Methods for Physics and Engineering. Cambridge University Press

Ruiz-Lapuente, P., ed., 2010, Dark Energy: Observational and Theoretical Approaches. Cambridge University Press

Schneider P., 2006, Extragalactic Astronomy and Cosmology, An Introduction. Springer

Schwarz D. J., Weinhorst B., 2007, A&A, 474, 717

Seljak U., Zaldarriaga M., 1996, ApJ, 469, 437

Seo H., Eisenstein D. J., 2007, ApJ, 665, 14

Seo H., Siegel E. R., Eisenstein D. J., White M., 2008, ApJ, 686, 13

Seo H.-J., Eisenstein D. J., 2003, ApJ, 598, 720

Sivia D., Skilling J., 2006, Data Analysis, A Bayesian Tutorial. Oxford University Press

Skilling J., ed., 2004, American Institute of Physics Conference Series, Vol. 735, Nested Sampling, Skilling J., ed.

Skilling J., 2006, Bayesian Analysis, 1, 833

Smith R. E., Peacock J. A., Jenkins A., White S. D. M., Frenk C. S., Pearce F. R., Thomas P. A., Efstathiou G., Couchman H. M. P., 2003, MNRAS, 341, 1311

Starkman G. D., Trotta R., Vaudrevange P. M., 2008, ArXiv e-prints

Sullivan M., et al., 2006, Astrophys. J., 648, 868

—, 2011, Astrophys. J., 737, 102

Taylor A. N., Kitching T. D., Bacon D. J., Heavens A. F., 2007, MNRAS, 374, 1377

Tegmark M., 1997, Physical Review Letters, 79, 3806

Tegmark M., Eisenstein D. J. e. a., 2006, PhysRevD, 74, 123507

Tolman R. C., 1934, Proceedings of the National Academy of Science, 20, 169

Tripp R., 1998, A&A, 331, 815

Trotta R., 2008, Contemporary Physics, 49, 71

Trotta R., 2008, Contemp. Phys., 49, 71

Trotta R., Feroz F., Hobson M. P., Roszkowski L., Ruiz de Austri R., 2008, JHEP, 12, 024

Trotta R., Kunz M., Liddle A. R., 2010, arXive:1012.319

Trotta R., Riazuelo A., Durrer R., 2001, PhysRevLet, 87, 231301

Tsujikawa S., 2010, ArXiv e-prints

Väliviita J., Enqvist K., Kurki-Suonio H., 2001, in American Institute of Physics Conference Series, Vol. 555, Cosmology and Particle Physics, Durrer R., Garcia-Bellido J., Shaposhnikov M., eds., p. 320

Vardanyan M., Trotta R., Silk J., 2009, MNRAS

Wang L., Goldhaber G., Aldering G., Perlmutter S., 2003, ApJ, 590, 944

Wang Y., 2008, PhysRevD, D77, 123525

Wang Y., Mukherjee P., 2007, PhysRevD, 76, 103533

Webb J. K., King J. A., Murphy M. T., Flambaum V. V., Carswell R. F., Bainbridge M. B., 2010, ArXiv e-prints

Weinberg S., 1989, Reviews of Modern Physics, 61, 1

—, 2008, Cosmology. Oxford University Press

Wilks S. S., 1938, The Annals of Mathematical Statistics, 9, pp. 60

Williams D., 2001, Weighing the odds: a course in probability and statistics. Cambridge University Press

Wiseman T., Withers B., 2010, ArXiv e-prints

Wood-Vasey W. M., Friedman A. S., Bloom J. S., Hicken M., Modjaz M., Kirshner R. P., Starr D. L., Blake C. H., Falco E. E., Szentgyorgyi A. H., Challis P., Blondin S., Mandel K. S., Rest A., 2008, ApJ, 689, 377

Wood-Vasey W. M., Miknaitis G., Stubbs C. W., 2007, ApJ, 666, 694

Young G., Smith R., 2005, Essentials of statistical inference: G.A. Young, R.L. Smith, Cambridge series on statistical and probabilistic mathematics. Cambridge University Press

Zunckel C., Okouma P., Kasanda S. M., Moodley K., Bassett B. A., 2011, Phys. Lett., B696, 433

Zwicky F., 1940, Reviews of Modern Physics, 12, 66