



---

# A Novel Test of the Cosmological Principle

---

OLIVER OAYDA

Supervised by Prof. Geraint Lewis

October 28, 2022

## Abstract

The cosmological principle, which posits that on very large scales the universe is homogeneous and isotropic, is a fundamental assumption in modern cosmology. A key consequence of this principle is that the dipole in the cosmic microwave background, held to arise because of Earth's peculiar motion, should correspond to other cosmological observables, like the dipole in galaxy number counts across the sky. Recent studies have identified inconsistencies between these two measurements, throwing doubt on the core proposition of homogeneity and isotropy. We present a novel test of the cosmological principle: because of Earth's peculiar motion, sources with intrinsic timescales should be time dilated, imprinting a time dilation dipole on measured timescales across the sky. Agreement between this dipole and that determined from the CMB will therefore provide a new means to check the cosmological principle and address tensions in the literature. We demonstrate the feasibility of this test, giving an estimate of the number of sources required for detection of the time dilation dipole. We also explore how the timescale intrinsic to quasar light curves, assuming they follow a damped random walk, could be extracted and used as a metric for the test.

## Statement of Originality

I certify that this thesis contains work carried out by myself except where otherwise acknowledged.



**Oliver Oayda**

October 28, 2022

## Acknowledgements

I am indebted to many for their support this year. To begin with, I wish to extend my sincere and deep gratitude to my supervisor, Prof. Geraint Lewis. Geraint is an incredible explainer, and with his peerless expertise and dedication guided me through this project. Our weekly discussions were endlessly insightful and illuminating, and they helped me see cosmology in a new light. I thank him for all his time, advice and extensive feedback over the last year; his support was invaluable. Next, I am profoundly grateful for all the helpful feedback I received from other Honours students on my draft presentation and/or report—namely from Frank Scuccimarra, Max Charles and BoWen Tian. Finally, I want to extend my appreciation to my family and friends for their continual interest in my work throughout the year. They lent a friendly ear to times when I brought up my project, which was immensely encouraging and reassuring: in an aphoristic sense, we ‘propped each other up’.

## Statement of Contribution of Student

The key concept behind this investigation was conceived by Prof. Geraint Lewis, and it was raised as a possible project in email correspondence. Geraint and I had extensive discussions about how we would proceed in building a hypothetical catalogue of quasars with timescales, and then extract the signal of Earth’s motion from that catalogue by looking at time dilation. This is in order to estimate the number of quasars required for detection of that signal.

I carried out the development of the Python code to perform this task. Since this test of the cosmological principle is new, it was built from scratch but relied on Python packages/modules such as `dynesty` (Koposov *et al.*, 2022) to carry out the critical statistical computations. The code’s implementation was my decision, however Geraint offered invaluable advice and feedback to help the code be realised. As I generated a simulated sample of quasars to reach my findings and did not rely on experimental data, these Python scripts form a key part of the project.

Once we had a sketch of the results, we then discussed seeking a method to extract the timescales from quasar light curves. After some research, I decided to turn to existing code, namely `JAVELIN` (Zu *et al.*, 2013). I wrote code to construct and sample light curves based on damped random walk theory, then ran them through `JAVELIN` to investigate the recovery of their intrinsic timescales.

I created all of the figures included in this report, including scientific plots and schematics/diagrams, unless otherwise stated or acknowledged. In Appendix B, I also describe and include key extracts of the code I wrote to produce the results. Beyond those extracts, the remainder of the code is hosted on Github, with the link to the repository given in Appendix B. They are separated into different scripts depending on their function, as alluded to above.

# Contents

<b>Contents</b>	<b>4</b>
<b>1 Introduction</b>	<b>6</b>
1.1 Principles in Cosmology . . . . .	6
1.1.1 The Copernican Principle . . . . .	6
1.1.2 The Cosmological Principle . . . . .	7
1.2 Consequences of the Principles . . . . .	8
1.2.1 Hubble Flow and Doppler Shift . . . . .	8
1.2.2 The Fundamental Observer . . . . .	10
1.3 Testing the Principles . . . . .	12
1.3.1 Radio Galaxies . . . . .	13
1.3.2 Quasars . . . . .	14
1.4 A Novel Test . . . . .	15
<b>2 Method</b>	<b>16</b>
2.1 Relativistic Model . . . . .	16
2.2 Extracting Timescales . . . . .	18
2.3 Generating a Sample . . . . .	20
2.3.1 Timescales in the Population . . . . .	20
2.3.2 Sources Over the Sky . . . . .	21
2.4 Bayesian Statistics and Inference . . . . .	23
2.4.1 Bayes' Theorem . . . . .	24
2.4.2 Model Comparison . . . . .	24
2.4.3 Nested Sampling . . . . .	25
2.4.4 Application . . . . .	26
<b>3 Results</b>	<b>29</b>
3.1 Points Required . . . . .	29
3.1.1 Region B . . . . .	30
3.1.2 Region C . . . . .	31
3.1.3 Region D . . . . .	31
3.1.4 Region A . . . . .	32
3.2 Timescale Recovery . . . . .	34
3.2.1 Sampling Duration . . . . .	34
3.2.2 Cadence . . . . .	36
3.2.3 Windowing . . . . .	36
<b>4 Discussion</b>	<b>38</b>
4.1 Denouement . . . . .	38
4.2 Observational Challenges . . . . .	39
4.3 Looking to the Future . . . . .	39

---

<b>References</b>	<b>41</b>
<b>A Supplementary Results</b>	<b>46</b>
<b>B Python Code</b>	<b>47</b>
B.1 Code Summary . . . . .	47
B.2 points-required.py . . . . .	48
B.3 bayes-factors.py . . . . .	52
B.4 light-curve-MCMC.py . . . . .	56

# Chapter 1

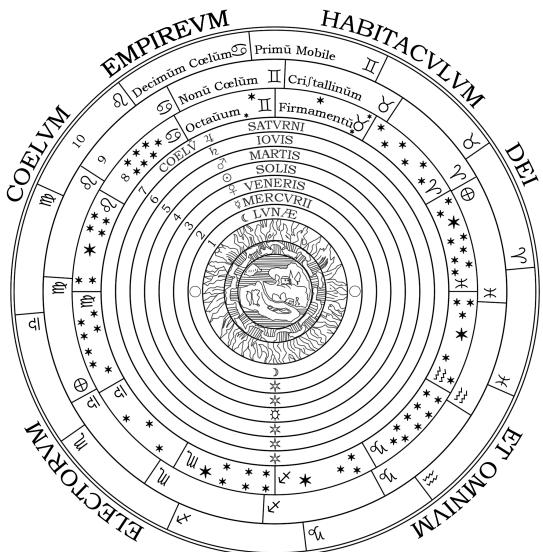
## Introduction

### 1.1 Principles in Cosmology

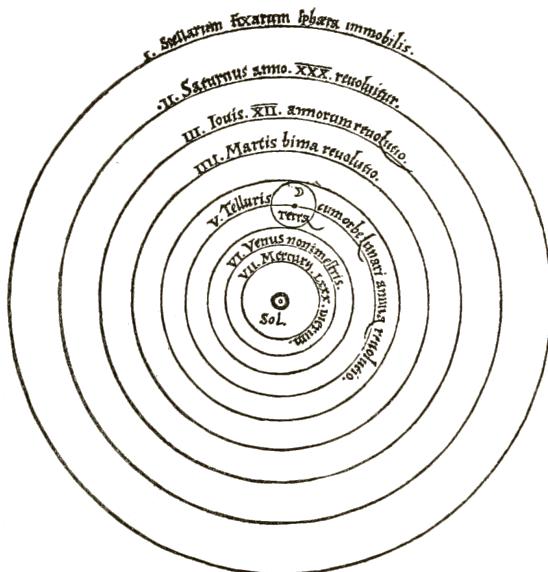
#### 1.1.1 The Copernican Principle

Philosophical assumptions about the nature of the universe have underpinned cosmology since ancient times. The earliest societies in recorded history created mythological ‘cosmogonies’—religious narratives about the universe and humanity, and the latter’s relation to the former (Kragh, 2007). In the tradition of Western philosophy, the Ionian school (the pre-Socratic thinkers) of the 6<sup>th</sup> century BCE represented the first attempt to *rationalise* the cosmos—that is, to attempt to understand it through physical principles or to *inquire into nature* (Curd, 2020). The astronomical legacy of the thinkers of the ancient world was the Aristotelian-Ptolemaic paradigm: a finite, geocentric view of the cosmos, which then only consisted of the Earth, planets, Sun and stars. In this framework, geometric arguments were used to explain the apparent motions of planets in the sky (Kragh, 2007).

This view of the universe essentially persisted through the Middle Ages until the paradigm-shift of the Copernican revolution (compare Figure 1.1a to Figure 1.1b). The Earth was displaced from its



(a) Digital reproduction of an illustration from Petrus Apianus's 1553 work *Cosmographicum liber*, as representing the Aristotelian-Ptolemaic cosmic paradigm. Credit: Cipozy (2015).



(b) Copernicus's heliocentric cosmos, as appearing in his 1543 text *De Revolutionibus*. The Sun (Sol) is shown in the centre, with Earth (Terra) and the moon on the 5th (V.) ring. Credit: Berry (2019).

Figure 1.1: *Left:* The Aristotelian-Ptolemaic paradigm. *Right:* The Copernican paradigm.

central, static position in the cosmos by the Sun. Yet this half-a-millennium-old scientific revolution suggests a more general philosophical principle which remains at the core of modern cosmology: the ‘Earth does not occupy a privileged position within the universe’ (Ellis, Maartens, *et al.*, 2012). In other words, there is nothing special about the Earth’s frame of reference. This is the *Copernican principle*.

### 1.1.2 The Cosmological Principle

The Copernican principle is closely-related to a stronger philosophical assumption: the *cosmological principle* (Ellis, Maartens, *et al.*, 2012). Cosmology experienced arguably one its most significant revolutions with Einstein’s formulation of general relativity in 1915, which ultimately linked the hitherto separate questions of the geometry of the universe and the influence of gravity on the universe’s structure (Longair, 2006). Gravity is reinterpreted geometrically as arising from the curvature of spacetime. Matter itself—or more precisely the energy density of the universe—then tells spacetime how to curve. In 1917, Einstein would use his newly-developed theory to compose a model of a static, closed universe with a spherical geometry, famously achieved with the introduction of his cosmological constant  $\Lambda$  (Einstein, 1997). However, in his 1922 paper, Alexander Friedmann (1888–1925) discovered the existence of a non-static solution to Einstein’s field equations—that is, a cosmos ‘whose spatial curvature... depend[s] on time’ (Friedmann, 2013). Here, the universe is modelled as a perfect fluid and assumed to (a) be homogeneous with uniform mass density  $\rho$  and pressure  $P$  and (b) expand isotropically (the same in all directions).

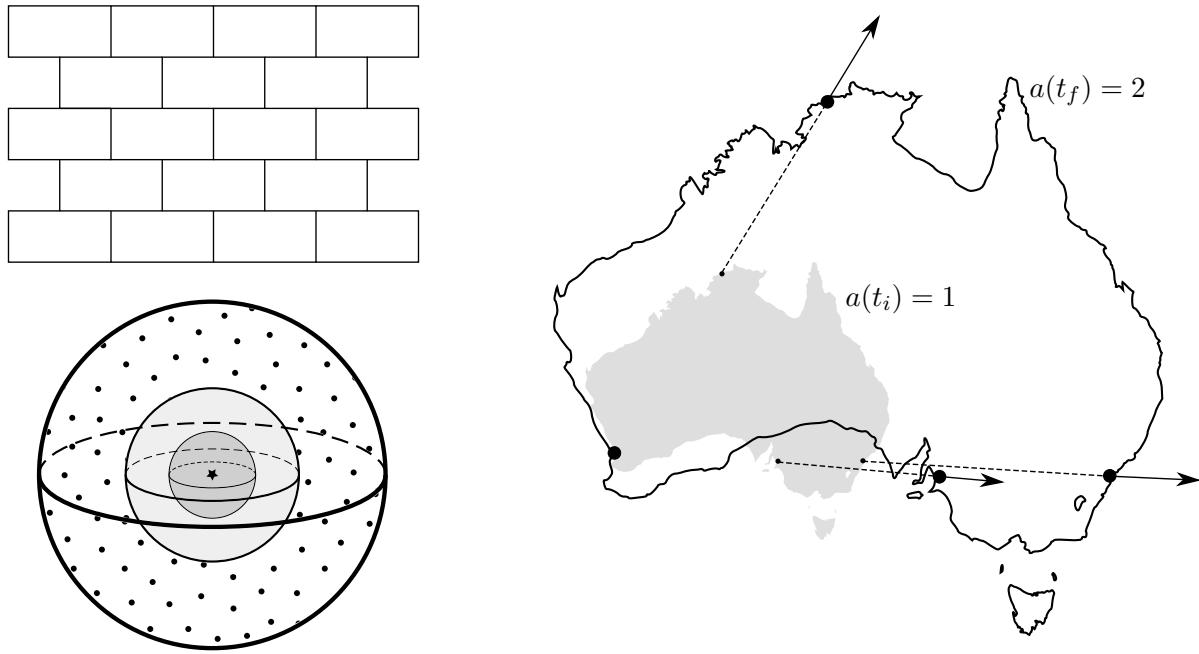
This idea that the universe can be described as *spatially* homogeneous and isotropic is the key assumption of the *cosmological principle*. The Friedmann world models are thus said to be ‘uniform isotropic models’ (Longair, 2006), and in a modern form, the ‘Friedmann equations’ describing the evolution of this universe can be written as (Mo *et al.*, 2010)

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{\kappa c^2}{a^2} + \frac{\Lambda c^2}{3} \quad (1.1)$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3} \left(\rho + \frac{3P}{c^2}\right) + \frac{\Lambda c^2}{3}. \quad (1.2)$$

Here,  $G$  is Newton’s gravitational constant,  $c$  the speed of light,  $a(t)$  the scale factor of the universe and  $\kappa$  the curvature of the universe. This ‘scale factor’ term essentially encapsulates the universe’s expansion, in that it refers to the factor by which the universe has grown in size. For instance, the scale factor changing from  $a(t_i) = 1$  to  $a(t_f) = 2$  represents the universe *doubling in scale* over the time interval  $t_i$  to  $t_f$ . This change of scale is represented in Figure 1.2b in the hypothetical scenario of the Earth expanding isotropically and homogeneously. What it means to describe something as homogeneous and isotropic, as central to the cosmological principle, is also clarified in Figure 1.2a. In this example, the wall is homogeneous in the sense that it is uniform or the same in construction. But at any point it is not isotropic, since the spacing of lines depends on the direction chosen. The concentric spheres are isotropic (but only from the centre), in the sense that is radially symmetric—the sequence of materials encountered is the same in all directions starting from the centre. However, since it consists of layers of different materials, it is not homogeneous. Thus, if we describe something as *expanding* homogeneously and isotropically, then the magnitude of expansion is the same at all locations and in every direction respectively (Carroll & Ostlie, 2017).

Most generally, a universe which is spatially homogeneous and isotropic is described by the Friedmann-Lemaître-Robertson-Walker (FLRW) metric, recognising the later independent contributions of Georges Lemaître (1894–1966), Howard Robertson (1903–1961) and Arthur Walker (1909–2001) in addition to Friedmann (Kragh, 2007; Longair, 2006). It takes the form (in the  $- + + +$



(a) *Top:* The wall is homogeneous (uniform) in its structure. *Bottom:* Only from the star at the centre are the concentric spheres isotropic, as radially outwards the sequence of materials is the same in all directions.

(b) If the Earth is expanding homogeneously and isotropically, then at some time  $t_f$  the scale factor has doubled since time  $t_i$ . An observer in Perth notices the cities of Darwin, Adelaide and Sydney receding. The further away the city, the faster it recedes. Inspired by an example in Carroll & Ostlie (2017).

Figure 1.2: *Left:* Demonstration of homogeneity and isotropy. *Right:* Analogy reflecting a universe that expands homogeneously and isotropically.

signature of the metric tensor)

$$ds^2 = -c^2 dt^2 + a^2(t) \left[ \frac{dr^2}{1 - \kappa r^2} + r^2(d\theta^2 + \sin^2 \theta d\phi^2) \right] \quad (1.3)$$

where  $r$ ,  $\theta$  and  $\phi$  are polar coordinates. This metric can be derived without Einstein's field equations, since the physics of the universe's evolution is absorbed into the time-dependent scale factor  $a(t)$  (Longair, 2006). Evidently, the terms in the square brackets on the right-hand side of (1.3) are scaled with  $a(t)$ , which can be analogised as the actual growing in size of the ruler defining the distance between two points in spacetime. Thus, the FLRW metric is essentially a tool for describing the distances between points (the spacetime interval) in an expanding universe, assuming spatial isotropy and homogeneity. In this sense, the cosmological principle is implicitly adopted by the FLRW metric. However, the assumption of homogeneity and isotropy would not be called a philosophical 'principle' until Edward Milne (1896–1950) in 1935 (Milne, 1935; Ellis, Maartens, *et al.*, 2012; Gale, 2021). At least at its conception it was taken as an *a priori* prescription for the universe, applied before substantial experimental observations could justify it.

## 1.2 Consequences of the Principles

### 1.2.1 Hubble Flow and Doppler Shift

Lemaître in 1927 acknowledged that, given a metric for a universe that is expanding homogeneously and isotropically, sources must appear as if they are moving away from an observer and accordingly be

redshifted (Lemaître, 2013). The speed at which they move away, and therefore the degree to which they are Doppler shifted, should be linearly related to their actual distance (*proper distance*) from Earth i.e.  $v = kr$ , for some constant  $k$ .

This idea of ‘Doppler shift’ is an extension of the Doppler effect for light waves. From physical experience, when an object emitting sound approaches an observer, its pitch (frequency) sounds higher to that observer; when it moves away from an observer, its pitch sounds lower. This Doppler effect occurs for light waves, with objects approaching an observer being blueshifted (higher frequency) and objects receding from an observer being redshifted (lower frequency). One key difference, however, is that generally in astronomy sources will be moving at a non-negligible velocity with respect to the speed of light, and so special relativistic effects become important.

To see how Doppler shifting is a natural consequence of a universe expanding under the cosmological principle, again refer to Figure 1.2b. To the observer in Perth, the cities furthest away seem to be receding from them the fastest. If they were to check this with an observer in Sydney, they would agree and in fact find the same relationship between distance and recessional velocity. In the context of cosmology, this linear relation was proposed independently of Lemaître by Hubble (1929) with a limited sample of galaxies and their measured Doppler shifts/recessional velocities. The result is now referred to as the Hubble-Lemaître Law, written as

$$v = H_0 d \quad (1.4)$$

for recessional velocity of a source  $v$  in  $\text{km s}^{-1}$ , proper distance to a source  $d$  in Mpc and Hubble constant  $H_0$  in  $\text{km s}^{-1} \text{Mpc}^{-1}$ , which is presently thought to be  $\approx 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$ . At the time, this was a deciding piece of evidence supporting the previously-mentioned expanding models of the universe (Kragh, 2007). The subscript in the Hubble *constant*  $H_0$  is used to avoid ambiguity, in that it refers to the value of the Hubble *parameter*  $H$  today. Since the Hubble parameter is a function of the scale factor through  $H = \dot{a}/a$ , which also appears in the left-hand side of Friedmann equation (1.1), it is in general different across the eras of the universe.

Thus, when we observe a distant astronomical source, there will be a redshift arising from the Hubble-Lemaître law (1.4), capturing the fact that the universe is expanding. But, in addition, a Doppler shift might also arise if the source is moving at some relative velocity to Earth due to gravitational motion near its neighbours. This motion is referred to as the *peculiar velocity* of a source. In other words, the velocity of the source  $\vec{v}$  could be decomposed into  $\vec{v} = \vec{v}_{\text{pec.}} + \vec{v}_{\text{Hub.}}$ , where the first term denotes the peculiar velocity and the second term denotes the velocity due to the *Hubble flow*. This is illustrated in Figure 1.3. The Hubble flow, then, refers to the apparent motion of the source due solely to the universe’s expansion, as analogous to the velocities perceived by the Perth observer in Figure 1.2b. If it were possible to factor out the peculiar motions of each source, then in principle we would only see sources moving with the Hubble flow. At first inspection, this would seem unachievable since, in general, different sources will be moving with an arbitrary peculiar motion with respect to the Hubble flow (see, for instance, the inferred flows or peculiar velocities of superclusters of galaxies near us in Figure 1.4a). However, if we reincorporate the guiding assumptions of the homogeneity and isotropy of the universe, as per the cosmological principle, then when we average out the peculiar

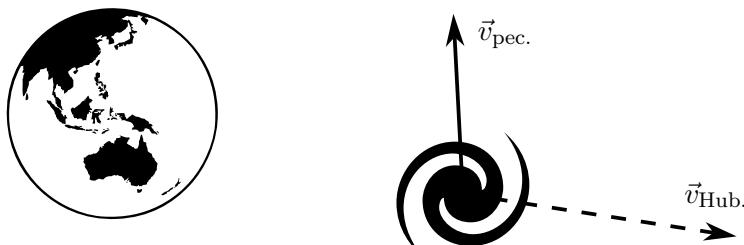
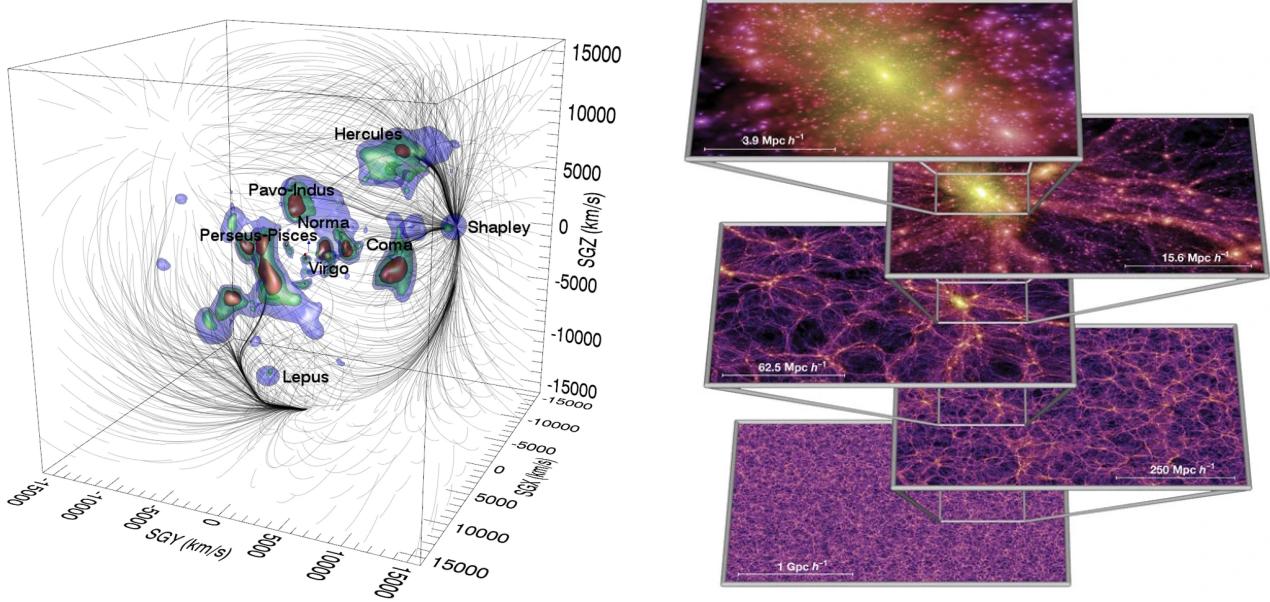


Figure 1.3: Schematic illustrating a galaxy’s peculiar (depends on local environment) and Hubble (arising from the universe’s expansion) velocities, as seen from Earth’s frame of reference.



(a) A cube of approximately 200 Mpc in side length, illustrating our neighbourhood of superclusters. We are at the origin, and the coloured regions represent the densities of different galaxy superclusters. The black threads are the inferred peculiar velocity flows at different points in space, which drain towards the Shapley supercluster. Credit: Tully *et al.* (2014).

Figure 1.4: *Left:* The Laniakea supercluster velocity flows and neighbours. *Right:* Results of the Millennium dark matter simulation.

(b) Results of the Millennium simulation, which numerically tracks a hypothetical universe's development. The different colours show different dark matter densities (yellow is the most dense). As we zoom out to large scales (down the image), a seemingly uniform and isotropic network of filaments emerges—the *cosmic web*. Credit: Springel *et al.* (2005).

motions over very large scales, we should expect it to tend to zero—that is, if the assumptions are true. So in reality, the cosmological principle is interpreted as a statistical statement; the local universe around us, like the solar system, Milky Way, etc., may appear heterogeneous, but when we average over sufficiently large scales, that variability should give way to uniformity or sameness. This idea is reflected in Figure 1.4b in the context of a numerical universe simulation. What exactly qualifies as a ‘sufficiently large scale’ is presently a source of tension and poorly understood (Maartens, 2011), but a possible lower bound in light of galaxy counts across the sky is near 100 Mpc (approximately three hundred million lightyears; Aluri *et al.*, 2022).

### 1.2.2 The Fundamental Observer

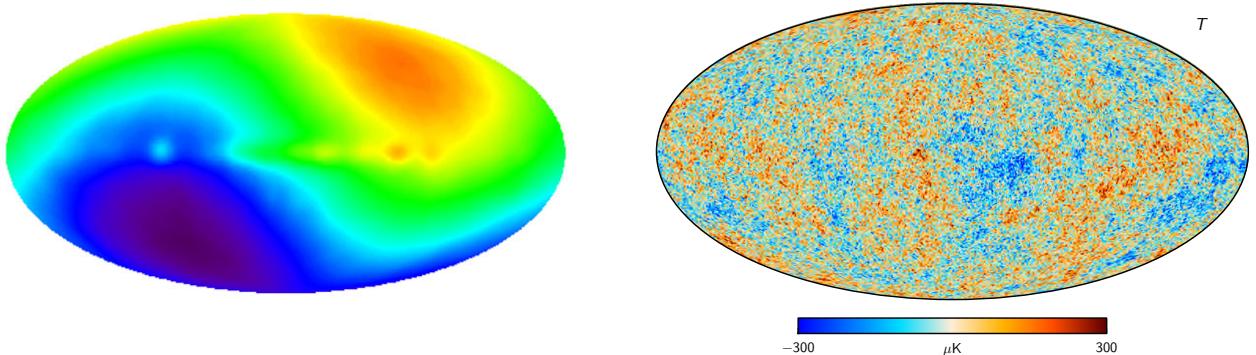
We have now recast the cosmological principle in a statistical light. But how do we understand the Copernican principle together with the cosmological principle? They appear similar, but differ by a shade of meaning. In the view of Ellis, Maartens & MacCallum (2012), the assumption that the Earth is not in a privileged position in the cosmos, as espoused by the Copernican principle, is taken to ‘justify’ the stronger and more technical requirement of (statistical) homogeneity and isotropy, as per the cosmological principle. If we take these two principles together, then they can be recast as a statement that at every location in the universe there is a set of *fundamental observers* belonging to a particular frame that sees the universe as isotropic—the Earth’s position in the cosmos being no more special than other locations. In general, *individual* observers at different spacetime points in the cosmos will be moving with a peculiar velocity with respect to this frame, or more precisely with respect to the ‘background geometry of the homogeneous isotropic’ FLRW metric (Wiltshire

(*et al.*, 2013). But the fundamental observer, by definition, would have no peculiar velocity. Moreover, based on the language introduced in Subsection 1.2.1, this frame must be where across vast scales the peculiar motions of astronomical sources has been averaged out and the universe is perceived as homogeneous and isotropic. Such a frame is termed the *cosmological rest frame*.

### Relationship to the cosmic microwave background

Yet can there be a rest frame for the matter and energy in the universe? Is it linked to a source empirically observable? In the mid 1960’s, the radio-astronomers Arno Penzias (1933–) and Robert Wilson (1936–) serendipitously discovered the cosmic microwave background (CMB) as a background ‘hiss’ coming from all directions in the sky. Despite their efforts, they could not remove it from their radio-communication telescope (Carroll & Ostlie, 2017; Kragh, 2007). This eventually confirmed the prediction of a relic blackbody radiation from the Big Bang which would gradually cool over time as the universe expands and the wavelength of the CMB photons are stretched. The very early universe was hot, consisting of free electrons which readily scatter photons, preventing them from travelling far; once the universe had cooled enough to allow the *recombination* of electrons and protons into neutral atoms ( $z \approx 1100$ ), the photons could propagate freely. In this sense, the CMB refers to the *surface of last scattering*, which emitted the earliest, non-scattered photons we can see today on Earth.

When we actually perceive the CMB (usually measured as the temperature  $T$  of photons at different patches in the sky), a dipole of amplitude  $\Delta T/T \approx 10^{-3}$  is measured, as first recorded by Smoot *et al.* (1977). One half of the sky shows a higher net temperature, and the opposite half a lower net temperature. This is shown in Figure 1.5a, where red corresponds to a higher temperature and blue a lower temperature. This temperature dipole is held to arise because of Earth’s peculiar motion. Namely, the Earth orbits the Sun; the Sun moves through an orbit in the Milky Way; the Milky Way, within the Local Group, moves in the Laniakea Supercluster at some peculiar speed with respect to the CMB (Tully *et al.*, 2014; again see Figure 1.4a for a representation of these peculiar velocity flows). This causes blueshift in the direction of that net motion and redshift in the opposite direction. The same is true of, for example, other galaxies like the Milky Way, which generally have random motions of order  $10^2\text{--}10^3 \text{ km s}^{-1}$  with respect to each other (Mo *et al.*, 2010). Observers in these galaxies would perceive dipoles in the CMB as related to each of their peculiar motions. In this sense, the CMB dipole we perceive is known as the *kinematic dipole* (Secrest, von Hausegger, *et al.*, 2021). Yet, if this dipole



(a) The *kinematic dipole* can be seen in the CMB map, as measured here by the COBE satellite. We interpret this as Earth’s net peculiar motion through the background photons, directed towards the top-right quadrant. Credit: NASA (2001).

(b) CMB map with the dipole subtracted, as measured by the Planck satellite. The temperature fluctuations are  $\approx 100$  times smaller than the anisotropy represented by the kinematic dipole. Credit: Planck Collaboration, Akrami, Y., *et al.* (2020).

Figure 1.5: *Left:* CMB with kinematic dipole. *Right:* CMB without kinematic and galaxy dipole.

is carefully subtracted from the CMB, it appears as in [Figure 1.5b](#). Here, the anisotropies are very small ( $\Delta T/T \approx 10^{-5}$ ), thus the CMB appears maximally isotropic and homogeneous.

How is this related to the cosmological rest frame? The key point is that if we proceed with the assumption of the cosmological principle and the closely-connected Copernican principle, then at each point in space the fundamental observer should perceive a CMB without a dipole. With the dipole subtracted from the CMB signal, it appears maximally isotropic and homogeneous (the cosmological principle), and as such we suppose other fundamental observers would agree on the CMB signal (the Copernican principle). These observers move *at rest with respect to the CMB*, and thus the CMB is interpreted as the cosmological rest frame. As already mentioned, in this rest frame for the matter-energy density in the universe, observers should in principle only perceive the Hubble flow i.e. motion defined by the Hubble-Lemaître law [\(1.4\)](#). It is worth mentioning that this does not threaten the principle of relativity—namely that there is no preferred frame. It is only referring to the frame where the CMB shows no dipole, and so makes no statement about whether the laws of physics are covariant.

### Concluding remark

The cosmological rest frame must be the same frame from which the Friedmann equations [\(1.1\)](#) and [\(1.2\)](#) and the FLRW metric [\(1.3\)](#) are defined. More precisely, take the FLRW metric at a singular point in time such that the scale factor  $a(t)$  is fixed. This defines a 3D slice of space represented by the variables  $r$ ,  $\theta$ , and  $\phi$ , defined within the cosmic rest frame. By definition this spatial slice has uniform or homogeneous density. Further, the way spacetime curves will depend on the spacetime curvature parameter  $\kappa$ , but suppose here  $\kappa = 0$ , representing flat spacetime. In the cosmic rest frame, the curvature of the universe will be the same in all directions (isotropic) along the three-dimensional spatial slice. Finally, the time coordinate  $t$  (*cosmic time*) is the spacetime direction orthogonal to the three spatial coordinates  $r$ ,  $\theta$  and  $\phi$ . With respect to the Friedmann equations, their derivation makes use of the energy momentum tensor  $T_{\mu\nu}$  of a perfect fluid, which introduces the variables  $\rho$  and  $P$ . Again, these are defined within the fluid's rest frame—the cosmological rest frame.

## 1.3 Testing the Principles

Hitherto the cosmological and Copernican principles have been represented as an *a priori* assumption adopted by the Friedmann equations [\(1.1\)](#) and [\(1.2\)](#) as well as the FLRW metric [\(1.3\)](#). In essence, they are simplifying assumptions which greatly alleviate the challenge of solving Einstein's field equations, reducing Friedmann equation [\(1.1\)](#) to a far more tractable ordinary differential equation (Aluri *et al.*, [2022](#)). Yet the current concordance cosmological model,  $\Lambda$ CDM, uses the FLRW metric to describe the universe (Maartens, [2011](#)). In this way, the two principles are a fundamental assumption of modern cosmology. Today, is there any empirical merit to or evidential support for these assumptions?

As already mentioned, in the cosmological rest frame, the CMB signal appears highly isotropic and homogeneous to one part in  $10^{-5}$  i.e. the temperature anisotropies  $\Delta T$  about the mean CMB temperature  $T \approx 2.7\text{ K}$  correspond to  $\Delta T/T \approx 10^{-5}$ . In this way, the CMB is described as the ‘strongest evidence for isotropy and homogeneity’, at least for the pre-Recombination era universe (Aluri *et al.*, [2022](#)). Another strong piece of evidence is from galaxy redshift surveys. These look to investigate the *large-scale structure* of the universe by probing the spatial distribution of many galaxies out to high redshifts. For instance, the results of the Two-degree-Field Galaxy Redshift Survey (2dFGRS) are shown in [Figure 1.6](#). Here, it can be seen by eye that the universe on large scales gives way to a network of filaments, sheets and voids, making the assumption of homogeneity and isotropy seem more reasonable (Knobel, [2012](#)). As already mentioned in [Subsection 1.2.1](#), it is then possible to define a physical scale at which, when averaged over, the universe is homogeneous and isotropic, though this is presently poorly defined.

The foregoing better elucidates the relationship between the cosmological and Copernican principles. Since we are restricted to our own Earth-bound frame of reference, we take observations of homogeneity and isotropy and extrapolate it to other observers in the universe. In short, our frame is not special, so we postulate others should see the same large-scale structure, the same CMB, the same Hubble-Lemaître law, etc. In this way, though the cosmological principle can be said to be empirically justifiable, the Copernican principle remains an *a priori* assumption—a limitation of the very nature of our observations. What we can still test, of course, is whether or not the cosmological principle really is a useful working assumption. In general, this is achieved by eroding the  $\Lambda$ CDM paradigm with consistency checks—if there are enough observations to suggest against its predictions, then the cosmological principle may no longer be a useful tool, or even an approximately accurate description of the universe. Next, we detail some of the key checks on the cosmological principle that have been devised and tested, as relevant to this report.

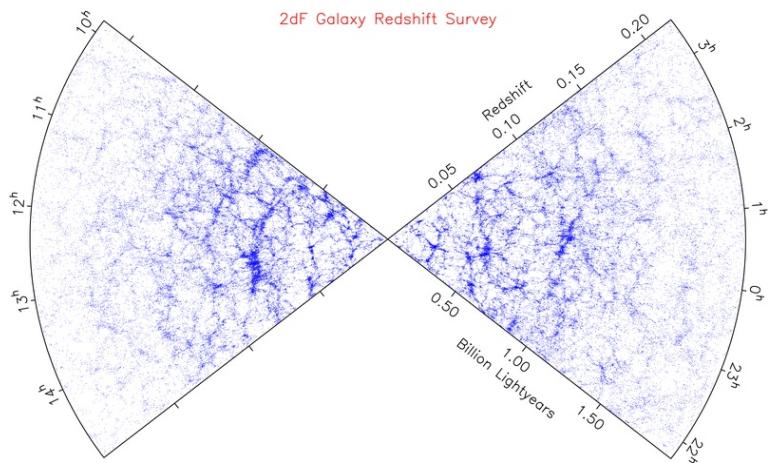


Figure 1.6: The spatial distribution of galaxies from 2dFGRS in an angular slice of sky. Each galaxy is represented by a blue dot. Earth is at the centre of the two cones. Compare this network of galaxies with the simulated dark matter network shown in Figure 1.4b. Credit: 2dFGRS Team (2015).

### 1.3.1 Radio Galaxies

One of the key assumptions of the cosmological principle is related to the CMB dipole—namely its kinematic interpretation. This was already suggested in Subsection 1.2.2. Recall that the assumption rests on homogeneity and isotropy, in the sense that the dipole-removed CMB in Figure 1.5b represents a fundamental frame where Earth’s peculiar velocity has been factored out and the CMB is maximally isotropic. Now, Ellis & Baldwin (1984) first devised a consistency test for the cosmological principle using galaxies with strong radio emission. In the cosmological rest frame, the number counts of sources and their brightness should be uniform across the sky (Aluri *et al.*, 2022). However, because of Earth’s peculiar velocity, there are essentially two effects at play, both being part of the same special relativistic picture. First, towards the direction of the sky in which the Earth is peculiarly moving, radio sources should be blueshifted; radio sources in the opposite direction, by extension, should be redshifted. In addition, Earth’s motion gives rise to *aberration*: sources in patches of the sky lying near the direction of Earth’s peculiar velocity congregate along the line of motion, and spread out at the opposite end of the direction of motion. This effect is derived in Section 2.1 of Chapter 2 and shown diagrammatically in Figure 2.3. To explain, suppose we take a patch of sky and look at the flux density  $S$  (units of  $\text{W m}^{-2} \text{Hz}^{-1}$ ) of sources. The flux density is described by a power law  $S \propto \nu^{-\alpha}$  for frequency  $\nu$  and spectral index  $\alpha$ . Looking then at the distribution of sources in this patch of sky, above some limiting flux density  $S$  it again follows a power law, where the number of sources per solid angle is  $dN/d\Omega \propto S^{-x}$  for some slope  $x$ . Sources should on net cluster along Earth’s direction of motion and spread out opposite it, leading to a dipole  $\mathcal{D}$  in the number count of sources per unit solid angle. Given  $\beta = v/c \ll 1$  (Earth’s peculiar speed  $v$  is much less than the speed of light), Ellis & Baldwin (1984) expect a kinematic dipole of amplitude

$$\mathcal{D} = [2 + x(1 + \alpha)] \beta. \quad (1.5)$$

If the kinematic interpretation of the CMB holds, as motivated by the cosmological principle, then *this dipole should correspond to the one discerned from the CMB* (see Figure 1.5a). Presently, the best estimate for the magnitude  $\beta$  and direction (in Galactic coordinates  $l, b$ ) of Earth's peculiar velocity, as derived from the CMB, is (Planck Collaboration, Aghanim, N., *et al.*, 2020)

$$\beta = \frac{v}{c} = (1.23357 \pm 0.00036) \times 10^{-3} = (369.82 \pm 0.11) \text{ km s}^{-1}$$

$$(l, b) = (264.021 \pm 0.011)^\circ, (48.253 \pm 0.005)^\circ.$$

Since  $\beta$  is of order  $10^{-3}$ , and taking  $x = 1$  and  $\alpha = 0.75$  for simplicity as in Siewert *et al.* (2021), the dipole amplitude is roughly  $\mathcal{D} \approx 4.5 \times 10^{-3}$ . In other words, the density enhancement of sources along the line of motion is small. This would prompt the need for a large catalogue of sources covering many regions of the sky so that the dipole signal can be better discerned—a theme central to this paper. Another key issue is the ‘clustering dipole’—contamination in the number count signal arises because of the clustering of galaxies in our local supercluster, as the distribution of galaxy sources locally is anisotropic. This contributes another dipole component to  $\mathcal{D}$  i.e.  $\mathcal{D} = \mathcal{D}_{\text{kin.}} + \mathcal{D}_{\text{clust.}}$ . In the words of Blake & Wall (2002), the ‘clustering dipole is the cause of our motion through the Universe [via gravity], while the velocity [kinematic] dipole is the consequence of it’. Generally this is a problem near  $z \approx 0.1$ , but not for sources at moderate redshifts ( $z \approx 1$ ) (Tiwari & Nusser, 2016).

Following the advent of large surveys of radio galaxies, for example the NRAO VLA Sky Survey (NVSS) with  $\approx 1.8$  million sources (Condon *et al.*, 1998), attempts were made to experimentally determine the dipole in radio galaxy counts. Blake & Wall (2002) used the NVSS survey and reported broad agreement between the measured radio dipole and the CMB dipole, which was also later found by Gibelyou & Huterer (2012) with additional surveys. However, Crawford (2009) recognised that the number of sources used by Blake & Wall (2002) ( $\approx 3.1 \times 10^5$  above 15 mJy) would be insufficient to yield a statistically significant result, though the author was considering a linear estimator as opposed to the spherical harmonics of the distribution used in Blake & Wall (2002). Later, Singal (2011) and Rubart & Schwarz (2013) (using additional surveys) agreed on a dipole larger in magnitude than that predicted by the CMB, but generally in agreement with its predicted direction. In many subsequent studies, though the direction of the kinematic dipole has generally been in agreement with the CMB dipole, its magnitude  $\mathcal{D}$  is typically larger (see e.g. Colin *et al.*, 2017; Bengaly *et al.*, 2018; Singal, 2019). Thus, Aluri *et al.* (2022) in their review describe at least some degree of ‘consensus’ on a radio galaxy dipole in excess of the CMB dipole.

### 1.3.2 Quasars

Radio galaxies are not the only way to determine the kinematic dipole. Active Galactic Nuclei (AGNs) refer to very bright regions of a galaxy which tend to be far more luminous than the galaxy itself. The standard paradigm is that an AGN is powered by a supermassive black hole at the centre of a host galaxy: in-falling matter on the black hole generally forms an accretion disc that heats up as gravitational potential energy is converted into kinetic energy, and subsequently radiation and thermal energy (Mo *et al.*, 2010; Kelly *et al.*, 2009). This is illustrated in Figure 1.7, which gives the key features of an AGN. Quasars—or quasi-stellar objects (QSOs)—refer to the most luminous types of AGNs, among other features. In general, quasars are measured at large cosmological distances of  $z \gtrsim 1$ , and so the clustering dipole is not a substantial issue (Aluri *et al.*, 2022).

Secrest, von Hausegger, *et al.* (2021) measured the dipolar distribution of quasars from the all-sky CatWISE2020 catalogue, as recorded by the Wide-field Infrared Survey Explorer (WISE) satellite. Following masking of the galactic plane and other regions of the sky, the final sample contained 1.36 million quasars. Employing a least-squares estimator for the dipole, the authors reported  $\mathcal{D} = 1.554 \times 10^{-2}$ , over twice as large as expected from the velocity of the Earth as determined by the CMB dipole and  $27.8^\circ$  off alignment.

In a following joint study with both radio galaxies from NVSS and quasars from WISE, Secrest, von Hausegger, *et al.* (2022) rejected the hypothesis of a dipole consistent with the kinematic CMB interpretation with a significance of  $5.2\sigma$ . The reported amplitudes for NVSS and WISE were too large by a factor of 3 and 2 respectively, and also pointed  $45^\circ$  and  $26^\circ$  away from the CMB dipole respectively. Nevertheless, Kothari *et al.* (2022) has shed doubt on at least the CATWISE2020 quasar catalogue. They found agreement with Secrest, von Hausegger, *et al.* (2021) on the number count per unit solid angle dipole, as being generally inconsistent with the CMB dipole in magnitude. Yet they also, as alternative checks on isotropy, looked at the mean spectral index and mean flux density of the sources, finding these variables to be consistent with isotropy. If, then, we cannot be confident with current surveys to robustly discern the kinematic imprint of Earth's peculiar motion, will we be able to do so in the future? And are there other means to measure this imprint in experiment, beyond the tests previously mentioned?

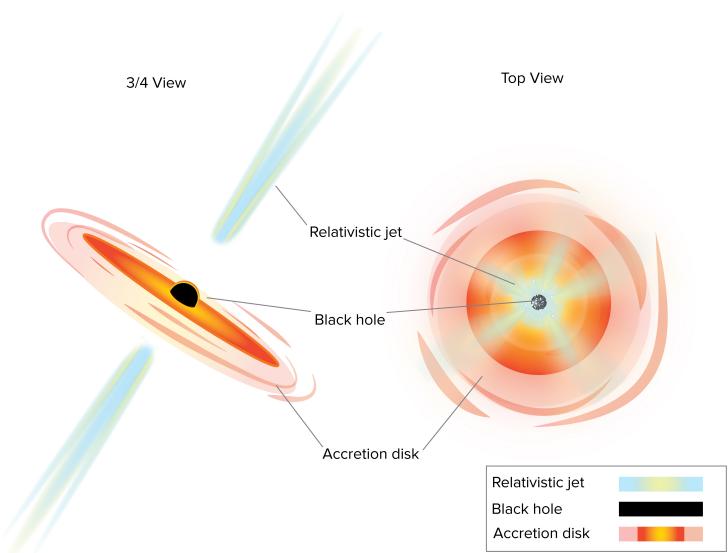


Figure 1.7: Schematic of a AGN viewed from different angles.  
Credit: Dagnello & NRAO/AUI/NSF (2019).

## 1.4 A Novel Test

On the basis of the foregoing, there is still no conclusive evidence that the kinematic interpretation of the CMB is correct i.e. the velocity in the  $\beta$  term of (1.5) is the same as determined from the CMB dipole. On the data available, the literature suggests the greatest tension is in the actual magnitude of the dipole as opposed to its direction. This is an issue since, if the tension remains, it suggests that the CMB is not a cosmological rest frame, challenging the conventional interpretation of the CMB as leftover radiation following Recombination or the use of the FLRW metric itself. In the words of Secrest, von Hausegger, *et al.* (2021):

As Ellis & Baldwin (1984) emphasized, a serious disagreement between the standards of rest defined by distant quasars and the CMB may require abandoning the standard FLRW cosmology itself. The importance of the[se] test[s]... can thus not be overstated.

In this way, the cosmological principle is being encroached upon, throwing our core assumptions into question and possibly motivating a new standard cosmological paradigm. The tensions mentioned above prompt two additional questions, both of which will be addressed by this report:

1. Is there some other way to determine the dipole arising from our peculiar motion? This would function as a consistency check with what has already been determined in the literature.
2. How many sources would in principle be needed for a meaningful, statistically significant detection of this dipole? A recurring theme in the previously-mentioned papers (Sections 1.3.1 and 1.3.2) is the need for future studies with more sources. Are there any surveys on the horizon which could provide sufficient sources?

# Chapter 2

## Method

In this chapter, we step through the method required to answer the two final questions posed by the end of [Chapter 1](#) (see [Section 1.4](#)). To do so, in [Section 2.1](#) we give the fundamental principles underlying our novel test of the cosmological principle. In [Section 2.2](#), we detail sources with timescales, like quasars, which can be used as part of the test. Finally, in [Section 2.3](#), we outline our procedure for generating a mock sample to allow evaluation of the test’s feasibility, and in [Section 2.4](#) explain how its performance is evaluated in the setting of Bayesian statistics.

### 2.1 Relativistic Model

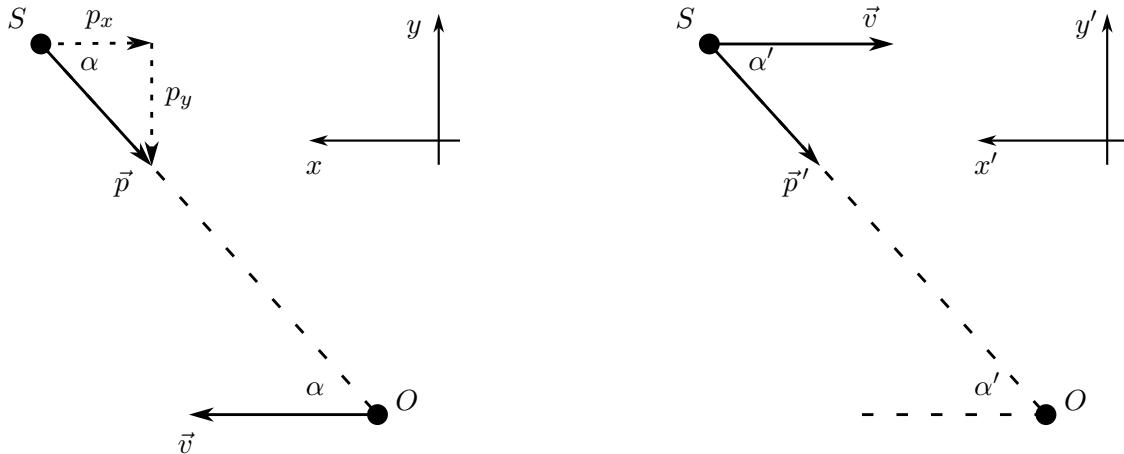
In [Section 1.3](#), it was explained how the dipole arising from Earth’s peculiar motion can be determined by noting the special relativistic blueshift in the direction of Earth’s motion and redshift in the opposite direction. Analogously, because of Earth’s peculiar velocity towards or away from astronomical sources, the *timescale of sources should be time dilated*, as predicted by special relativity. The relative velocity of the Earth and source, as determined by the angle subtended between them, will be different for each source, and thus effect the degree to which they are time dilated. This timescale could in principle refer to any measurable quantity of a source with a time duration, and so in theory, the signal of Earth’s peculiar motion can be extracted given a set of timescales at different points on the sky.

Let us now build up the size of this special relativistic imprint of Earth’s peculiar velocity on astronomical sources. From here on, we employ units where  $c = 1$ . Consider [Figure 2.1a](#). The un-primed Cartesian  $x$ - $y$  plane (the coordinate frame) lies in the rest frame of the source  $S$ . This source subtends an angle  $\alpha$  with the velocity three-vector  $\vec{v}$  of an observer  $O$ , as measured in this frame. It also emits a photon of angular frequency  $\omega$ , which therefore has energy  $E = \hbar\omega$ . Further, its four-momentum has components  $p^\mu = (E, \vec{p})$  for three-momentum  $\vec{p}$  with magnitude  $p = \hbar\omega$ . Thus, the components of the four-momentum, given the choice of coordinates in the coordinate frame, are  $p^\mu = (\hbar\omega, -p \cos \alpha, -p \sin \alpha, 0)$ .

However, the energy of the photon  $E'$ , as measured by  $O$  (see [Figure 2.1b](#)), will correspond to the projection of the photon’s four momentum onto the time-like basis vector  $\mathbf{e}_{t'}$  in the orthonormal basis of the observer. This is simply the four-velocity  $\mathbf{u}_{\text{obs}}$ —to explain, the observer’s four-velocity only points along the time-like four-vector  $\mathbf{e}_{t'}$ , since a clock at rest in the observer’s frame moves only through time and not space. Now, in the source’s rest frame,  $u_{\text{obs}}^\mu = (\gamma, \gamma\vec{v}) = (\gamma, v\gamma, 0, 0)$  for Lorentz factor  $\gamma = (1 - v^2)^{-1/2}$ . Thus, we can project the photon’s four-momentum  $\mathbf{p}$  onto the observer’s basis through (using the  $- + + +$  signature of the metric tensor  $\eta$ ):

$$E' = -\mathbf{p} \cdot \mathbf{u}_{\text{obs}} = -\eta_{\mu\nu} p^\mu u_{\text{obs}}^\nu = +\hbar\omega\gamma(v \cos \alpha + 1) \implies \frac{\omega'}{\omega} = \gamma(\cos \alpha + 1). \quad (2.1)$$

Next, we want to relate  $\alpha \rightarrow \alpha'$ , the latter of which is the angle between the source and the observer, as measured in the observer’s rest frame. The observer chooses four unit axes  $\mathbf{e}_\mu$  with which they can measure observables, like the energy of a photon. We already have that  $\mathbf{e}_{t'} = \mathbf{u}_{\text{obs}}$ , so they are free to



(a) Observer  $O$  with three-velocity  $\vec{v}$ , as perceived in the source  $S$ 's rest frame. The emitted photon has energy  $E = \hbar\omega$ .

(b) Source  $S$  as perceived by a moving observer  $O$ . The observer measures an angular frequency  $\omega'$  and the photon is perceived to have energy  $E' = \hbar\omega'$ .

Figure 2.1: A source  $S$  emits a photon of angular frequency  $\omega$ , which is received by an observer  $O$ . *Left:* The source's rest frame (coordinate frame). *Right:* The observer's rest frame.

choose three arbitrary orthonormal spatial axes. Looking at [Figure 2.1b](#), we pick  $(\mathbf{e}_{y'})^\mu = (0, 0, 1, 0)$  and  $(\mathbf{e}_{z'})^\mu = (0, 0, 0, 1)$ . This leaves the basis vector along  $x'$  with generic form  $(\mathbf{e}_{x'})^\mu = (a, b, 0, 0)$ . Imposing the conditions that this four-vector is orthogonal to the time-like axis ( $\mathbf{e}_{t'} \cdot \mathbf{e}_{x'} = 0$ ) and that it has unit length ( $\mathbf{e}_{x'} \cdot \mathbf{e}_{x'} = 1$ ) gives the specific components  $(\mathbf{e}_{x'})^\mu = (v\gamma, \gamma, 0, 0)$ . Now that we have laid down the orthonormal basis with which the moving observer makes their observations, we can project the four-momentum  $\mathbf{p}$  onto their basis vectors to determine the component of the photon's momentum they measure along the  $x'$  direction. This is

$$p^{x'} = \mathbf{p} \cdot \mathbf{e}_{x'} = \eta_{\mu\nu} p^\mu u_{\text{obs}}^\nu = -\hbar\omega\gamma(v + \cos\alpha). \quad (2.2)$$

Accordingly, we can ascertain the relationship between  $\alpha$  and  $\alpha'$  given (2.2):

$$\cos\alpha' = -\frac{p^{x'}}{p'} = \frac{\hbar\omega\gamma(v + \cos\alpha)}{\hbar\omega'} \implies \frac{\omega'}{\omega} = \frac{\gamma(v + \cos\alpha)}{\cos\alpha'}, \quad (2.3)$$

and, equating the final result of (2.1) to the last result of (2.3),

$$\cos\alpha' = \frac{v + \cos\alpha}{v\cos\alpha + 1} \implies \cos\alpha = \frac{\cos\alpha' - v}{1 - v\cos\alpha'}. \quad (2.4)$$

(2.4) allows us to transform the angle subtended by the direction of motion and the source from frame-to-frame, and so essentially quantifies relativistic aberration, in which sources appear to congregate along the line of motion for the moving observer. Next, substituting (2.4) into (2.3) allows us to eliminate dependence on  $\alpha$ , yielding

$$\omega' = \omega [\gamma(1 - v\cos\alpha')]^{-1}. \quad (2.5)$$

This is the relationship between the frequency  $\omega$  emitted in the rest frame of the source and the frequency  $\omega'$  detected by the observer, given as a function of the angle  $\alpha'$  between the source and the observer as measured in the observer's rest frame. In this sense, it models how Earth-bound observers perceive the energy of light emitted from astronomical sources as modified by the Doppler effect.

This special relativistic picture not only induces Doppler shift, but also time dilation—they are the same effect. From (2.5), suppose  $\alpha' = 0$  (the source is head on) and  $v = 0.8c$ . Then  $\omega' = 3\omega$  i.e. the perceived frequency of light from a head-on source is three times the frequency perceived in the rest frame of the source. We can interpret this frequency change as the observer's perception of clocks in the frame of the source ticking faster; the period between peaks of the emitted EM wave defines a timescale, which appears faster to the observer. Accordingly, an event with some characteristic timescale  $\tau$  will, in this example, seem to have a shorter duration. Since  $\omega \propto 1/T$  for period  $T$ , we can quantify the perceived timescale  $\tau'$  by the observer for a source with rest timescale  $\tau$  through

$$\tau' = \tau\gamma(1 - v \cos \alpha'). \quad (2.6)$$

In summary, if Earth moves with some peculiar velocity  $v$  with respect to the CMB, then (2.6) can be used to model the directionally dependent special relativistic imprint this peculiar motion leaves on astronomical sources with intrinsic timescales—the *time dilation dipole*. So, in light of the first question of Section 1.4, determining Earth's peculiar motion through timescales could function as a novel test—a consistency check with other tests in the literature. However, which sources are good candidates for this test? We turn to that question in the following section.

## 2.2 Extracting Timescales

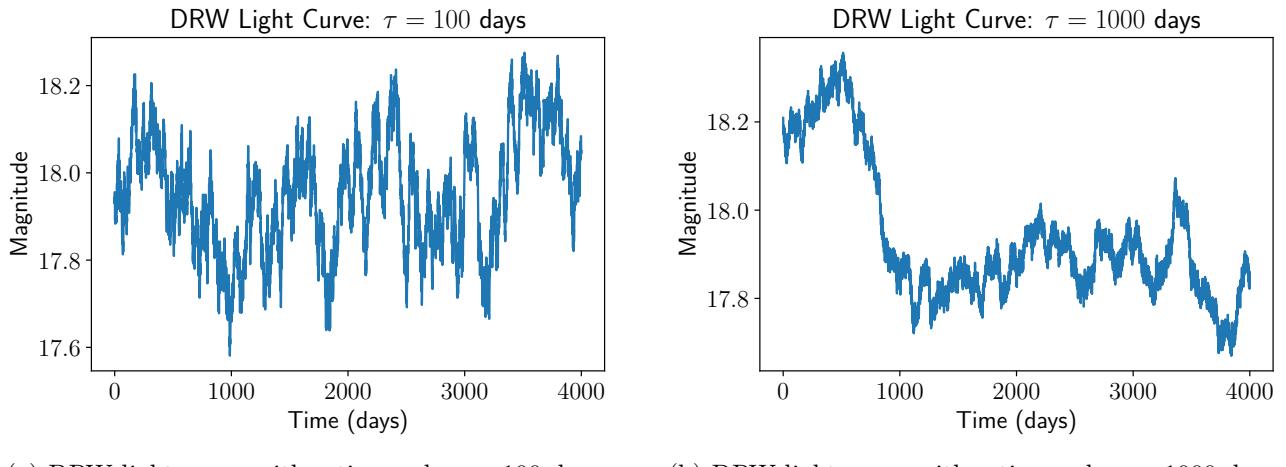
In Subsection 1.3.2, attempts to determine the kinematic dipole with quasar distributions on the sky were considered. This is purely a question of whether or not there is a source density enhancement along one direction in the sky. However, in light of the relativistic test represented by (2.6), the ‘light curve’ of quasars is another property that could be used. A light curve refers to how the intensity of light emitted by a source changes in a given measurement window. Quasars typically exhibit stochastic, aperiodic variability in intensity over time. The exact physical cause of this process is not as of yet understood, however possible mechanisms include thermal fluctuations in the accretion disc of the black hole driving the quasar (MacLeod *et al.*, 2010). Still, Kelly *et al.* (2009) proposed that quasar optical variability could be described by a *damped random walk*. This describes a light curve that takes a random step or change in brightness, but is pushed back over time towards a mean brightness. These fluctuations have a *characteristic timescale*  $\tau$ , typically given in days, which is used to describe their stochastic fluctuations. In particular, if random variables  $X$  and  $Y$  are drawn from Gaussian distributions  $\mathcal{G}(\mu, \sigma^2)$  described by  $X \sim \mathcal{G}(0, \sigma^2)$  and  $Y \sim \mathcal{G}[0, \sigma^2(1 - e^{-2\Delta t/\tau})]$  where  $\sigma^2$  is the *signal variance*, then the  $i$ -th value of the variable *quasar signal* is given by: (Kozłowski, 2017)

$$s_i = \begin{cases} X & i = 0 \\ s_{i-1}e^{-\Delta t/\tau} + Y & i \geq 1 \end{cases} \quad (2.7)$$

where  $\Delta t = t_{i+1} - t_i$  is the *observational cadence*, referring to the time interval between discrete measurements of the light curve. The final light curve is the sum of a mean magnitude  $\langle y \rangle$ , a Gaussian deviate  $\mathcal{G}(0, n_i^2)$  representing the photometric noise  $n_i$  and the points in (2.7) such that

$$y_i = s_i + \langle y \rangle + \mathcal{G}(0, n_i^2) \implies \mathbf{y}(t) = \mathbf{s}(t) + \mathbf{L}\mathbf{q} + \mathbf{n}. \quad (2.8)$$

In the last line, we have represented the light curve data  $\mathbf{y}$  as a vector of values, taken as a sum of the quasar signal vector  $\mathbf{s}$ , a matrix  $\mathbf{L}$  and vector of linear coefficients  $\mathbf{q}$  (used to add the mean magnitude) and photometric noise vector  $\mathbf{n}$ . The results of simulating two quasar light curves with different timescales using (2.7) and (2.8) are shown in Figure 2.2. The key point is that, given an observation of a light curve like in Figure 2.2, the timescale corresponding to its stochastic fluctuations can be extracted. In this way, a timescale could be determined for each quasar on the sky, though



(a) DRW light curve with a timescale  $\tau = 100$  days. (b) DRW light curve with a timescale  $\tau = 1000$  days.

Figure 2.2: Two DRW light curves with timescales  $\tau = 100$  days (*left*) and  $\tau = 1000$  days (*right*) sampled for 4000 days with cadence  $\Delta t = 0.5$  days. For both, the asymptotic amplitude  $SF_\infty = \sqrt{2}\sigma = 0.2$  and mean magnitude  $M = 18$ . The magnitude is a dimensionless variable imitating the observed change in brightness.

of course each will be different due to their varying physical properties—e.g. black hole mass. With timescales for each quasar, (2.6) could then be deployed to infer the direction and magnitude of Earth’s peculiar velocity.

Still, others sources can be used. Type Ia Supernovae have a well-defined and predictable light curve since the physical mechanism behind each supernova is thought to be the same (Perlmutter, 2003). The ‘width’ of the light curve in a plot of luminosity against time could then be defined, acting as a physical timescale. The effects of time dilation on this timescale have already been studied, but in the context of cosmological time dilation: supernovae at a higher redshift recede away from Earth faster (recall (1.4)) and so will be observed to have a greater time dilation in Earth’s frame of reference—see e.g. Blondin *et al.* (2008). The question then becomes the degree to which Earth’s *peculiar velocity* can be extracted from the overall time dilation signal. The same will necessarily be true for any source used at cosmological distances.

Putting this to the side for the moment, even if these sources are good candidates with readily measurable timescales, how many would be needed to reach a meaningful determination of Earth’s peculiar motion? As given in Subsection 1.3.1, we have  $\beta \approx 1.2 \times 10^{-3}$ , and if  $\alpha' = 0$ , then (2.6) becomes

$$\tau' = \tau\gamma(1 - v\cos\alpha') = \tau \times (1 - (1.2 \times 10^{-3})^2)^{-\frac{1}{2}} \times (1 - 1.2 \times 10^{-3}) = 0.9988 \times \tau. \quad (2.9)$$

In short, the timescale perceived in Earth’s frame  $\tau'$  will only be marginally time-dilated from the source’s rest frame timescale  $\tau$ : the signal is at best  $\approx \pm 0.1\%$  of  $\tau$ . To put it simply, the effect of Earth’s peculiar motion is subtle. This necessitates the question of whether Earth’s motion could be extracted with a sufficient number of sources which overcome the small scale of the effect, as per the second question in Section 1.4. If we additionally consider measurement uncertainty on  $\tau$ , could the signal still be discerned?

To answer this, in Section 2.3 we simulate a mock sample representative of a set of sources that could be obtained by future surveys. Following this, in Section 2.4 we apply Bayesian hypothesis fitting, verifying that the inferred parameters describing Earth’s motion better describe the simulated data than (a) the null hypothesis of no net motion and (b) the hypothesis that the CMB kinematic dipole aligns with the direction of motion determined by the simulated sample. Ultimately, this will

yield an estimate of the number of sources required to meaningfully infer Earth’s motion from the time dilation dipole. If detection is feasible, then future studies can turn to the novel test and hopefully resolve the tension between the CMB dipole and the kinematic dipole. If, even then, the tension cannot be resolved, then it may suggest that the cosmological principle can no longer be employed to describe our universe. In that case, a new cosmological paradigm is needed.

## 2.3 Generating a Sample

### 2.3.1 Timescales in the Population

#### Sampling light curves

If we desire to construct a hypothetical quasar population representative of the capabilities of future surveys, we first need to hone in on the qualities of those measurements. More precisely, how reasonably can the DRW timescales actually be extracted given quasar light curves? What degree of uncertainty is reasonable for these types of measurements?

`JAVELIN` (Zu *et al.*, 2013), standing for Just Another Vehicle for Estimating Time Lags in Nuclei, is a Python implementation which is used to model the variability of quasar light curves and extract parameters of interest. The key point for this investigation is that `JAVELIN` can be used to recover the joint posterior distribution for the timescale  $\tau$  intrinsic to a light curve, as well as its signal variance. It achieves this by modelling a light curve as if it were a DRW, in which case it can be shown that the likelihood function of the light curve data is (Kozłowski *et al.*, 2010)

$$\mathcal{L}(\mathbf{y}|\mathbf{s}, \mathbf{q}, \tau, \hat{\sigma}) = |\mathbf{C}|^{-\frac{1}{2}} \left| \mathbf{L}^T \mathbf{C}^{-1} \mathbf{L} \right|^{-\frac{1}{2}} \exp \left( -\frac{\mathbf{y}^T \mathbf{C}_\perp^{-1} \mathbf{y}}{2} \right). \quad (2.10)$$

We reconsider the role of likelihood functions in Section 2.4, but for now, it is essentially a means to determine how well the data fits a given model and its associated parameters. In this case, by optimising the parameters pertaining to the DRW model, being the timescale  $\tau$  and the variability amplitude  $\hat{\sigma} = \sigma\sqrt{2/\tau}$ , the likelihood function can be maximised. `JAVELIN` accomplishes this by exploring probability space with an MCMC (Markov Chain Monte Carlo) algorithm, as implemented in the Python package `emcee` (Foreman-Mackey *et al.*, 2013). We discuss these types of algorithms again in Section 2.4, but for now, the important point is that we can then marginalise out  $\hat{\sigma}$  in the resulting joint posterior, yielding a 1D marginal posterior distribution for  $\tau$ . This is our parameter of interest. The other terms in (2.10) are related to the data: recalling (2.8),  $\mathbf{C} = \mathbf{S} + \mathbf{N}$ , where  $\mathbf{S}$  and  $\mathbf{N}$  are the covariance matrices of the quasar signal  $\mathbf{s}$  and the photometric noise  $\mathbf{n}$  respectively.

The critical idea is that we only assume the light curve can be described by a DRW. Then, after inputting a light curve—a set of points sampled in discrete intervals of time—`JAVELIN` can optimise (2.10) and recover the probability distribution for  $\tau$ . The shape of this distribution quantifies how well  $\tau$  can be discerned, as well as the characteristic uncertainties that may be associated with its recovery.

#### Practical considerations

The way in which a quasar light curve is sampled is already known to materially impact the accuracy of the recovered value for  $\tau$ . For instance, Kozłowski (2017) found that in general light curves need to be sampled for at least  $\approx 10$  times the true rest-frame timescale. If  $\tau = 1$  year, then quasar light curves would need to be sampled over a 10 year rest-frame period. This is the case since the DRW is intrinsically a stationary process in which its moments (e.g. the mean and variance) are time invariant (Kozłowski, 2021). An insufficiently sampled light curve does not adequately capture the stationary aspect of the random walk, leading to inaccurate recovery of both  $\tau$  and  $\hat{\sigma}$ . This

was confirmed by Sheng *et al.* (2022), in which a novel machine learning algorithm was deployed to recover  $\tau$  amongst other DRW parameters. In their work, they modelled light curves which might be sampled according to observational strategies deployed by the upcoming Legacy Survey of Space and Time. There, timescales were consistently *underestimated* where the sampling duration is less than  $\times 10$  the true timescale  $\tau$ . Other practical concerns for real quasar surveys are the cadence of the observations and how they are windowed. Cadence refers to the time interval between samples (a one-day cadence means one light curve measurement per day), and windowing refers to periods over which no observation is made—e.g. a seasonal gap or period of telescope downtime.

Accordingly, to determine reasonable estimates for the uncertainty on  $\tau$ , we simulate a number of different quasar light curves using (2.8). In this case, they are assumed to have the same intrinsic properties, and so the same timescale  $\tau = 100$  days. We then sample each of these light curves differently across a number of experiments, changing the sampling duration  $t$ , cadence  $\Delta t$  and the presence of windowing (here, we have nominated to use a period of 90 days of measurements followed by 90 days of downtime with no measurements). These differently-sampled light curves are then interpreted with JAVELIN to ascertain how robustly  $\tau$  can be recovered and inform the kind of uncertainties expected for a quasar catalogue with a nominated measurement strategy. However, it will also inform the uncertainties that we should build in to our hypothetical quasar sample. This procedure is implemented in our Python script `light-curve-MCMC.py`. To investigate, plot and comment on the distribution of recovered timescales, we additionally wrote the script `light-curve-distribution.py`. Each of these are illustrated in [Appendix B](#).

### 2.3.2 Sources Over the Sky

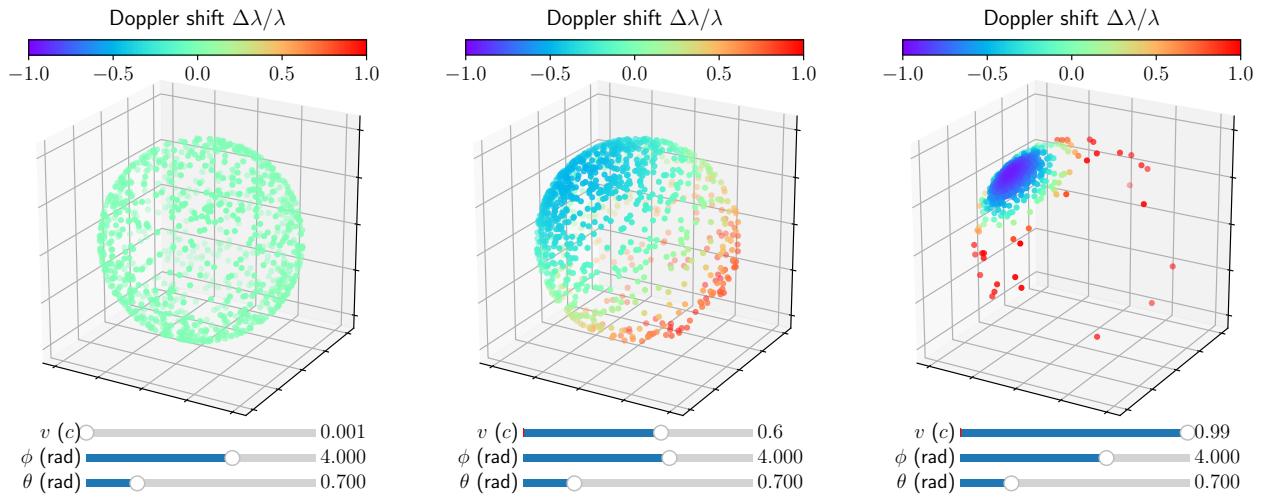
#### Point generation

We now turn to the method used to populate a hypothetical quasar catalogue. Recall that the key question is how many sources would need to be contained in this sample to detect time dilation arising from Earth’s kinematics. To proceed, we imagine that, in the source rest frame, the distribution of quasars on the sky is uniform as per the guiding assumption of the cosmological principle. This is the same frame in which Earth’s peculiar motion is factored out, and so corresponds to a cosmological rest frame. To create a uniform distribution of quasars,  $n$  points—corresponding to  $n$  quasars—are drawn from three standard Gaussian distributions  $\mathcal{G}_i(\mu = 0, \sigma^2 = 1)$  where  $i = x, y$  or  $z$ . This yields three values representing the Cartesian coordinates of each point, and each point is then normalised to have unit length. The resulting distribution of points on the surface of the unit sphere is uniform (Marsaglia, 1972). Each point is converted to spherical coordinates; since  $r = 1$ , the position of the  $i$ -th point is now only described by a polar angle  $\theta_i \in [0, \pi]$  and azimuthal angle  $\phi_i \in [0, 2\pi]$ . In essence, this gives the coordinates that a fundamental observer would use to describe the quasars they perceive across the totality of the sky. Note that, in this test, we assume that all the sources are at rest with respect to each other, such that in the source rest frame kinematic time dilation is not a factor. The  $i$ -th generated point is then assigned a timescale  $\tau_i$  and an angle  $\alpha_i$ . For an initial estimate, we assume all timescales are equal (each quasar has the same intrinsic properties), but are measured with added Gaussian noise. To be more precise, the  $i$ -th timescale is given by

$$\tau_i = \tau_0 + \mathcal{G}_i(\mu = 0, \sigma = \Delta\tau_0) \quad (2.11)$$

where  $\tau_0$  is the common timescale across all quasars and  $\Delta\tau_0$  is the common width of the normally-distributed error applied to all quasars. To find  $\alpha_i$ , we first assign a direction  $(\theta, \phi)$  to Earth’s peculiar motion and then compute the angle subtended by this vector and the coordinates of the point in the rest frame.

By the above, each quasar carries two quantities of interest, namely  $\tau_i$  and  $\alpha_i$ . We then transform these quantities into the frame of an observer on Earth moving with some peculiar speed  $v$ . Recall,



(a) Points perceived at speed  $v = 0.001c$ . At this speed, the Doppler shift is small with respect to  $\lambda$  and hard to discern with the scale chosen. The points also appear more or less uniform across the sky.

(b) Points perceived at  $v = 0.6c$ . Points are blueshifted and redshifted, forming a dipole, and congregate along the direction of motion (top left) but spread out opposite (bottom right).

(c) Extreme case where the observer is near the speed of light  $v = 0.99c$ . Doppler shift and aberration is most noticeable, with the vast majority of the points congregating in the top left along the line of motion.

Figure 2.3: Simulation representing how the distribution of quasar points ( $n = 10^3$ ) changes as an observer speeds up. Earth is at the centre of the sphere, and the distribution of points shows how an Earth-bound observer would view the sky. Earth moves to the top left but with different speeds. Doppler shift (so analogously time dilation) and aberration are discernable for the relativistic speeds.

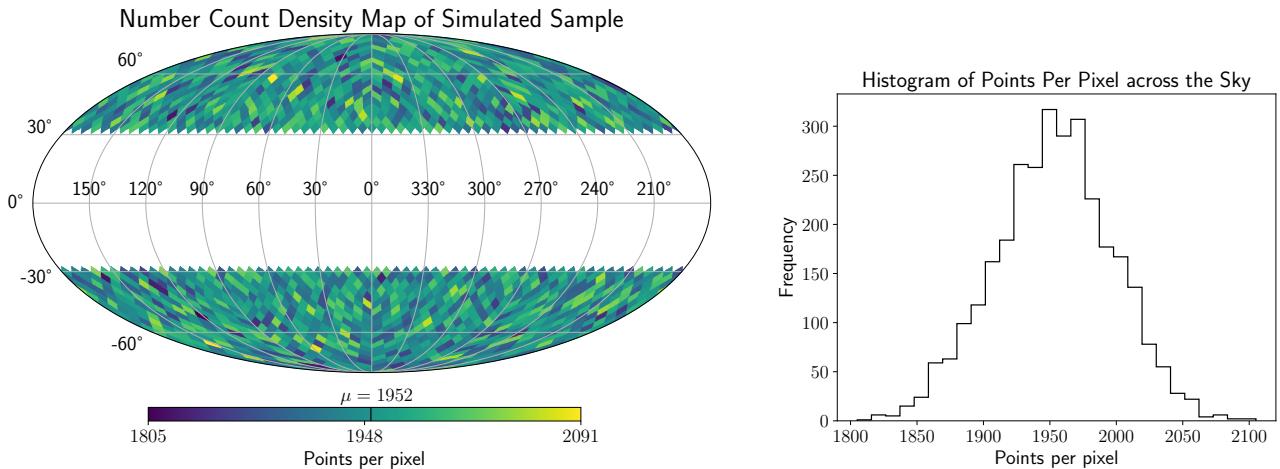
as from Section 2.1, there are two key differences between the source rest frame and the Earth-bound frame: first, there will be the effect of kinematic time dilation which takes  $\tau_i \rightarrow \tau'_i$  as given in (2.6); second, each point congregates along Earth's line of motion (aberration) as given in (2.4), which takes  $\alpha_i \rightarrow \alpha'_i$ . So, the  $i$ -th quasar  $q$  in the source rest frame is transformed to Earth's frame via  $q_i = (\tau_i, \alpha_i) \rightarrow q'_i = (\tau'_i, \alpha'_i)$ . The effects of this transformation are shown in Figure 2.3, in which as  $v$  increases, Doppler shift (and so analogously time dilation) becomes more pronounced, while sources appear to congregate along the line of motion.

## Pixelisation

With a set of  $n$  timescales on the sky as perceived by an Earth-bound observer, we now want to look for how readily a time dilation dipole can be discerned. To make the problem more computationally tractable, we divide the sky into pixels of equal area. This is achieved with the Python package `healpy`'s HEALPix<sup>1</sup> (Hierarchical Equal Area isoLatitude Pixelisation) scheme (Zonca *et al.*, 2019; Górski *et al.*, 2005). With `nside = 16`,  $\approx 3000$  pixels are generated, as shown in the Mollweide projection of Figure 2.4a. The distribution of pixel densities is then shown in the histogram of Figure 2.4b.

For each pixel  $\mathcal{P}_j$ , we compute the average  $\bar{\tau}'_j$  and standard deviation  $\sigma_j$  of the timescales for all quasar points contained within. Thus, each pixel carries information  $\mathcal{P}_j = (\bar{\tau}'_j, \sigma_j)$ . Finally, pixels  $30^\circ$  above and below the galactic plane are masked (removed from the sample) to represent a typical survey avoiding contamination from the galactic plane, as for instance in Secret, von Hausegger, *et al.* (2021). Since  $n$  will typically be of the order of ten million and the points are uniformly distributed

<sup>1</sup><https://healpix.sourceforge.io/>



(a) Mollweide projection illustrating the density of quasar points per pixel (mean density  $\mu$  labelled). Note that our coordinates  $(\theta, \phi)$  are converted into galactic coordinates  $(l, b)$ .

(b) Histogram of points per pixel for the same density map shown on the left.

Figure 2.4: Distribution of points across the sky after pixelisation. *Left:* Density map (Mollweide). *Right:* Histogram of pixel densities. In each case,  $n = 6 \times 10^6$  points before masking were used.

on the sky, this has the effect of halving the number of sources in the final sample used for analysis. The procedure detailed here is implemented in the Python script `points-required.py`, as detailed in [Appendix B](#), which calls upon key functions we have written as housed in `funcs.py`.

Now we are poised to scrutinise our mock sample of quasars: if the only information an observer on Earth has are the pixels  $\mathcal{P}_j$ , with their mean timescale and spread, could they infer the direction of Earth's peculiar motion as embedded in the time dilation dipole? To answer this question, we turn to the language of Bayesian statistics and defer to the advantages it offers in model inference and comparison.

## 2.4 Bayesian Statistics and Inference

In the context of scientific investigation, we usually seek to compare the veracity of competing hypothesis given observed data. That is to say, we wish to know which model is a better fit for the data, and by what degree it outperforms other models. The ‘orthodox’ approach to answering this question involves the use of *frequentist statistics* (Loredo, 1990). The frequentist interprets the probability of an event as its ‘long-run relative frequency of occurrence’ upon repeated, random experiments. In contrast, *Bayesian statistics* reinterprets probability as a real-valued ‘measure of the degree of belief about a proposition’ (Trotta, 2008). Instead of considering the frequency distribution of an experiment upon hypothetical repetitions, Bayesian statistics look only at the data actually observed and what inferences can be drawn from it. In this way, the Bayesian language is a ‘quantitative epistemology’: ‘a numerical encoding of one’s state of knowledge’ (Loredo, 1990). It additionally affords advantages in model comparison, like the penalisation of over-fitting and the encoding of prior states of knowledge. In our case, we will make use of the tools of Bayesian analysis in order to determine whether or not the signal of Earth’s motion, as embedded in the time dilation dipole, can be confidently extracted over the hypothesis that the Earth is not moving (null hypothesis) or that its motion aligns with the CMB dipole. First, we overview the key concepts integral to Bayesian analysis, as used in this report.

### 2.4.1 Bayes' Theorem

The titular mathematician (and Reverend) of Bayes' Theorem, Thomas Bayes (1702–1761), at least wrote down the product rule of probability but not the modern formulation of the theorem itself (Bayes & Price, 1763). This, along with the interpretation of probability as a statement of belief, is credited to the independent work of Pierre-Simon Laplace (1749–1827)—see Laplace (1902). In its modern appearance, Bayes' theorem reads:

$$P(H|D, I) = \frac{P(D|H, I)P(H|I)}{P(D|I)}. \quad (2.12)$$

In words, it states that the probability of a hypothesis  $H$  given some data  $D$  (otherwise known as the *posterior probability*) is equal to the product of the probability of the data given a hypothesis (the *likelihood*) and the probability of the hypothesis before the data (the *prior probability*), divided by the probability of the data (the *evidence* or *marginal likelihood*). The  $I$  term represents information known before the data—intuitively, to any Bayesian assessment, there is some information we already know (like possible parameter values) which guides our predictions. Because of the prior probability term  $P(H|I)$ , Bayes' theorem is a quantitative representation of the intuitive process of updating one's beliefs in light of new, incoming evidence—we can *update our priors* to accommodate this.

There is a slight recasting of Bayes' theorem which does not change the underlying logic, but does elucidate why Bayes' theorem is a powerful companion in comparing competing hypotheses. Suppose that we wish to determine the probability of a set of parameters  $\Theta$  which apply for a model  $M$ , and we are trying to fit that model to a set of observed data points  $\mathbf{D}$ . Bayes' Theorem can then be rewritten in this language: (MacKay, 2003; Speagle, 2020)

$$P(\Theta|\mathbf{D}, M) = \frac{P(\mathbf{D}|\Theta, M)P(\Theta|M)}{P(\mathbf{D}|M)} \equiv \frac{\mathcal{L}(\Theta)\pi(\Theta)}{\mathcal{Z}}. \quad (2.13)$$

Here, the likelihood function  $\mathcal{L}(\Theta) = P(\mathbf{D}|\Theta, M)$ , the prior function  $\pi(\Theta) = P(\Theta|M)$  and the evidence  $\mathcal{Z} = P(\mathbf{D}|M)$  have been introduced. The evidence  $\mathcal{Z}$  takes the role of a normalising constant for the posterior in (2.13), since the posterior probability density function needs to be normalised to 1. It is thus given by an integral over all parameters  $\Theta$  in parameter space  $\Omega_\Theta$ :

$$\mathcal{Z} = P(\mathbf{D}|M) = \int_{\Omega_\Theta} \mathcal{L}(\Theta)\pi(\Theta) d\Theta. \quad (2.14)$$

### 2.4.2 Model Comparison

In fitting a model to the data, there are two levels of *Bayesian inference*: first, assuming a particular model to be true, we optimise the model's parameters to best fit the data; second, we compare the evidence for each optimised model, favouring the model with the strongest evidence (MacKay, 2003). At the first level, optimising the parameters amounts to determining the posterior probability in (2.13) i.e. the probability of a set of parameters  $\Theta$  given data  $\mathbf{D}$  and model  $M$ . Then, at the second level, we compare the evidence  $\mathcal{Z}_{M_i}$  of model  $M_i$  to the evidence  $\mathcal{Z}_{M_j}$  for model  $M_j$  through the *Bayes factor*  $B_{ij}$ : (Trotta, 2008)

$$B_{ij} = \frac{P(\mathbf{D}|M_i)}{P(\mathbf{D}|M_j)} = \frac{\mathcal{Z}_{M_i}}{\mathcal{Z}_{M_j}}. \quad (2.15)$$

A higher evidence for model  $M_i$  means that  $B_{ij} > 1$ , which is an indication of support for this model over model  $M_j$ . In making this comparison, models are penalised for overfitting the data. To illustrate this, we make the simplification that the contribution to the evidence  $\mathcal{Z}$  integral will peak at the most probable parameters  $\Theta_{MP}$  which best optimise the likelihood function  $\mathcal{L}(\Theta_{MP})$ . As such, the evidence

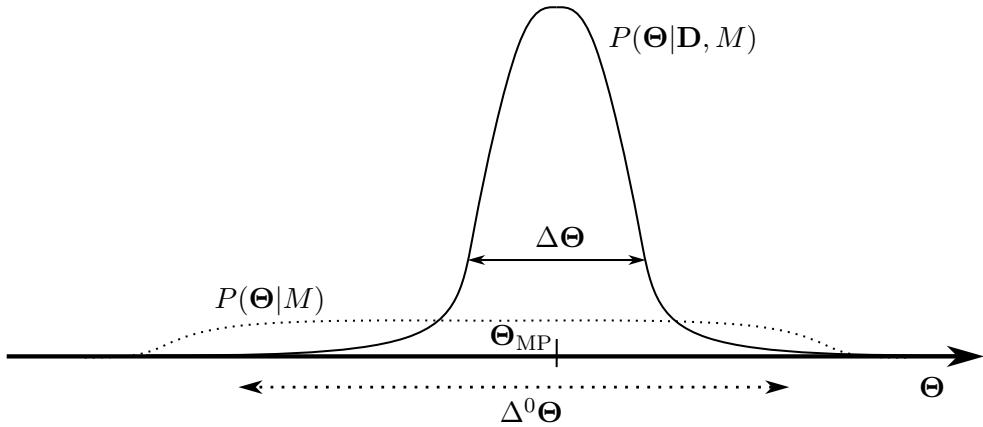


Figure 2.5: *The Occam factor.* The posterior distribution  $P(\Theta|\mathbf{D}, M)$  (solid black line) peaks at the parameters  $\Theta_{MP}$  which best optimise the likelihood function. The prior distribution  $P(\Theta|M) = \pi(\Theta)$  (dashed line) is uniformly distributed with width  $\Delta^0\Theta$ . The Occam factor is the ratio of the width of the peak and the width of the uniform prior  $\Delta\Theta/\Delta^0\Theta$ . Adapted from MacKay (2003).

integral (2.14) can be approximated as the height of the peak  $\mathcal{L}(\Theta_{MP})\pi(\Theta_{MP})$  multiplied by its width  $\Delta\Theta$  in parameter space (Laplace’s method): (MacKay, 2003)

$$\mathcal{Z} \approx \mathcal{L}(\Theta_{MP})\pi(\Theta_{MP}) \times \Delta\Theta. \quad (2.16)$$

See Figure 2.5 for an illustration of this. If we imagine that the prior  $\pi(\Theta)$  is uniformly drawn (i.e. prior parameters are drawn with equal probability) from some interval of width  $\Delta^0\Theta$ , then  $\pi(\Theta_{MP}) = 1/\Delta^0\Theta$ , and thus (2.16) becomes

$$\mathcal{Z} \approx \mathcal{L}(\Theta_{MP}) \frac{\Delta\Theta}{\Delta^0\Theta}. \quad (2.17)$$

$\Delta\Theta/\Delta^0\Theta$  is known as the *Occam factor*, named after the principle of *Occam’s razor*—‘entities should not be multiplied beyond necessity’, or, the simplest interpretation of available data is preferable (Trotta, 2008). It bears the name because of the impact it has on the evidence  $\mathcal{Z}$  depending on the complexities of a particular model. Namely, if a model is finely-tuned to the data by using excessive parameters (i.e. it has high in-set optimisation), then  $\Delta\Theta$  will be small, since there is a narrow range in parameter space over which the model is a good predictor. This decreases the Occam factor, in turn decreasing the evidence. Analogously, models which also have many parameters free to vary over a large  $\Delta^0\Theta$  are penalised as the evidence integral becomes one over more dimensions. For example, suppose we have a two parameter problem with data we expect to obey a linear relation  $y = mx + b$ : the parameters here are  $m$  and  $b$ . If we start fitting an arbitrary number of parameters  $y = c_0 + c_1x + \dots + c_nx^n$  to the data, the fit may improve as  $\mathcal{L}$  is optimised, but prior volume is wasted—the model is many-dimensional. Occam’s factor will be engaged, penalising the evidence  $\mathcal{Z}$  for the many-dimensional model but favouring the simpler linear model.

In short, the tension between how well a model explains the data and its complexity is encapsulated in the evidence  $\mathcal{Z}$ , and accordingly encapsulated in model comparison with the Bayes factor  $B_{ij}$  (2.15). The preferred model is generally the one which makes the best compromise between these competing interests. Now we turn to ways in which  $\mathcal{Z}$  is evaluated in practice.

### 2.4.3 Nested Sampling

One key issue in employing Bayesian statistics to fit models is that, in general, the posterior (2.13) and the evidence integral (2.14) cannot be solved analytically and are also challenging to find numerically.

Scientific problems tend to be many-dimensional with large domains of parameter space, rendering these quantities tough to compute. Thus, various numerical prescriptions for calculating these have been devised. For instance, the famous iterative MCMC algorithm directly samples the posterior probability distribution with walkers which wander around the parameter space  $\Omega_{\Theta}$ , spending more time in regions of greater probability (higher likelihood) and less time in regions of lower probability, thus obeying *detailed balance*. These algorithms are incorporated in Python packages such as `emcee`, and, as mentioned in Section 2.3, is used by the light curve analysis of JAVELIN.

In this paper, we make use of `dynesty` (Koposov *et al.*, 2022), a Python package which uses *Nested Sampling* (NS; Skilling, 2004; Skilling, 2006) to evaluate the Bayesian evidence integral  $\mathcal{Z}$  instead of the posterior. This confers the ability to make an immediate judgement on the strength of competing hypotheses or models through the Bayes factor (2.15). The key point behind NS is that it attempts to simplify the challenging problem of evaluating the evidence by changing the variables of the evidence integral to one over the prior volume  $X$  instead of parameter space  $\Theta$ : (Speagle, 2020)

$$\mathcal{Z} = \int_0^1 \mathcal{L}(X) dX \quad \text{where} \quad X(\lambda) = \int_{\Omega_{\Theta}: \mathcal{L}(\Theta) \geq \lambda} \pi(\Theta) d\Theta. \quad (2.18)$$

Here,  $X = X(\lambda)$  represents a cumulative prior volume containing likelihoods  $\mathcal{L}(\Theta)$  above some threshold value  $\lambda$ , normalised such that  $X(\lambda = 0) = 1$  and  $X(\lambda = \infty) = 0$ . Initially, NS starts with an ‘active’ sample of points drawn from the prior which are ranked according to their likelihoods  $\mathcal{L}$ . The lowest likelihood point is then removed from the chain, and a new point with higher likelihood than the discarded point introduced to the active sample. This iterates, with the active samples closing in on or ‘nesting’ towards peaks in the posterior distribution as lower likelihood points are discarded. These discarded points are stored to form the monotonically increasing function  $\mathcal{L}(X)$ —each  $X$  value corresponds to an iso-likelihood contour in this remapped likelihood function, which is then integrated over (via the trapezoidal rule) to obtain the evidence  $\mathcal{Z}$  and an associated uncertainty. The NS algorithm can also recover the posterior distribution as a byproduct by using the stored chain of points and their corresponding importance weights.

#### 2.4.4 Application

In essence, the NS regime of `dynesty` allows us to navigate the two levels of Bayesian inference to answer the second key question of Section 1.4. This is represented in the flowchart given in Figure 2.6.

#### Level I

At the first level, given the sample of pixels  $\mathcal{P}_j = (\bar{\tau}'_j, \sigma_j)$  generated in Section 2.3 (the data **D**), we optimise the model  $M$  described by the timescale dilation equation (2.6). This is a three parameter problem: recall that (2.6) takes in  $v$  and  $\alpha'$ , the latter of which is necessarily a function of the direction of Earth’s peculiar motion in spherical coordinates  $(\theta, \phi)$ . Thus,  $\Theta = \{v, \theta, \phi\}$ . Next, as in Bayes’ theorem (2.13), we require a likelihood function  $\mathcal{L}$  to describe how well the data fits the model and a prior function  $\pi$  to represent our understanding of what values the parameters are likely to take. For  $\mathcal{L}$ , we take the coordinates of each pixel  $\mathcal{P}_j$  on the sky and compute what the expected timescale  $\tau'_j$  should be, after factoring in time dilation through (2.6). This assumes knowledge of what the true rest-frame timescale  $\tau_j$  should be for each  $\mathcal{P}_j$ . Then, the probability of the measured timescale for each pixel is the value of the logarithm of a Gaussian  $\mathcal{G}$  at point  $\bar{\tau}'_j$ , where the Gaussian is centred at  $\tau'_j$  and has width  $\sigma_j$ . This is the log of the likelihood for a singular pixel. Expressing this mathematically,

$$\ln \mathcal{L}_j = P(\bar{\tau}'_j | v, \theta, \phi) = \mathcal{G}(\bar{\tau}'_j | \mu = \tau'_j, \sigma = \sigma_j). \quad (2.19)$$

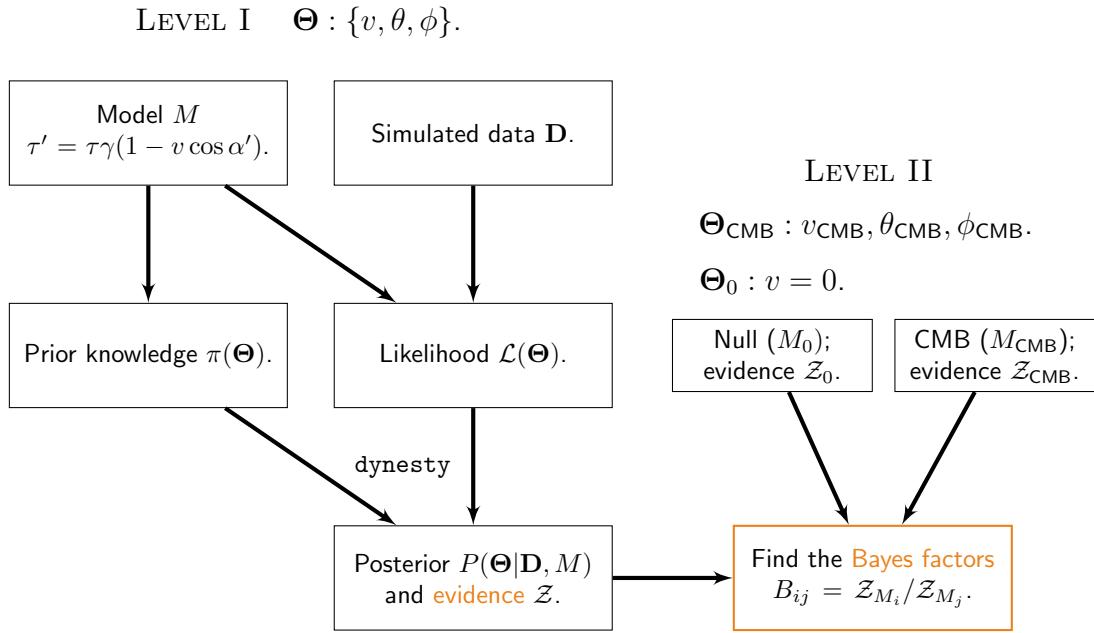


Figure 2.6: Flowchart describing how this investigation approaches the two levels of Bayesian inference. On the left hand side, at LEVEL I, the parameters which best fit Earth’s motion from the simulated data are determined and the evidence computed. At LEVEL II, the evidence for each hypothesis is compared through finding Bayes factors.

Then, the likelihood function is the sum of these values over all pixels  $\ln \mathcal{L} = \sum_j \ln \mathcal{L}_j$ . Thus, we are sampling from the natural logarithm of the likelihood function, since in general the actual value of  $\mathcal{L}$  will be very small (potentially a few negative hundred orders of magnitude). For  $\pi$ , we initially do not want to make any statement about the possible direction of Earth’s peculiar motion—that is to be determined. So we apply a uniform ignorance prior across all angles  $\theta \in [0, \pi]$  and  $\phi \in [0, 2\pi]$ : each angle is equally as likely. Similarly, a uniform prior between 0 and  $0.01c$  ( $\approx \times 10$  the speed determined from the CMB dipole) is applied for  $v$ . This information is passed through to *dynesty*, which, using the aforementioned principles, computes the evidence integral (2.14) as well as the posterior distribution for  $v$ ,  $\theta$  and  $\phi$ . This returned evidence  $Z$  will henceforth be called the evidence for the ‘fitted hypothesis’.

## Level II

This completes the first level of Bayesian inference (see again Figure 2.6). At the second level, we compute the evidence values for the null hypothesis  $Z_0$  (no peculiar motion) and the evidence values for the CMB hypothesis  $Z_{\text{CMB}}$  (peculiar motion aligns with the CMB dipole). These are determined in the same way that  $\mathcal{L}$  was computed above, except with each hypothesis’ corresponding predicted values for the true timescale  $\tau_j$  at each pixel  $\mathcal{P}_j$ . For the null hypothesis, all the pixels should not be impacted by time dilation, and for the CMB hypothesis, the effect of time dilation is predicted to align with the CMB dipole. These two hypotheses are essentially a zero-parameter problem (there is no parameter to be optimised) in that  $v = 0$  or  $(v, \theta, \phi) = (v_{\text{CMB}}, \theta_{\text{CMB}}, \phi_{\text{CMB}})$ , as labelled in Figure 2.6. In each case, then, the evidence is simply the likelihood of the data assuming no motion or assuming CMB-oriented motion. With evidences in hand, we can compute the Bayes factors (2.15). Strictly, since we sample from the logarithm of the likelihood function, the evidence is logarithmic, so we are

solving for the natural logarithm of the Bayes factors  $\ln(\mathcal{Z}_{M_i}/\mathcal{Z}_{M_j})$ . Our approach is to then compute

$$B_{12} = \ln \left( \frac{\mathcal{Z}}{\mathcal{Z}_{\text{CMB}}} \right) \quad \text{and} \quad B_{10} = \ln \left( \frac{\mathcal{Z}}{\mathcal{Z}_0} \right) \quad (2.20)$$

where 0, 1 and 2 denote the null, fitted and CMB hypotheses respectively. If  $B_{12} > 1$ , we interpret this as support for the fitted hypothesis over the CMB hypothesis. If not, the time dilation dipole cannot be resolved sufficiently from the CMB dipole. This same argument applies for  $B_{10}$ , except there the competing explanation is the absence of a dipole. This process, again as schematised in [Figure 2.6](#), is also incorporated into the script `points-required.py`.

### Metric for success

*The two Bayes factors in (2.20) are the key metric for success.* As  $n$ , the number of quasars in the original sample, is varied, the intuitive corollary is that the Bayes factors will increase; more data smooths out the contribution of uncertainty in the simulated sample. The question then is—at what value of  $n$  will the Bayes factors consistently show support for the fitted hypothesis?

We can turn to a picture of this by projecting the 2D marginal posterior for  $\theta$  and  $\phi$  onto the sky, recalling that `dynesty` also evaluates the posterior distribution as a byproduct after finding the evidence. This is shown in [Figure 2.7](#). There, the grey region represents the 2D posterior, which shows the uncertainty in the discerned location of the time dilation dipole from the simulated data. Namely, 85% of the probability volume lies within the region demarcated by the outermost grey contour. The true direction of peculiar motion (red cross) is imprinted onto the quasar sample upon generation, as described in [Section 2.3](#). In this example, the true direction (red cross) lies within (at the 65% level) the region of uncertainty, and the best estimate for the true direction is shown with the yellow cross. However, the direction of the CMB dipole lies well outside the grey probability volume. What does this mean for the sensitivity of our novel test? Analysis of the quasar sample shows stronger support for a direction of motion not in alignment with the CMB dipole. So, in this mock example with  $n = 6 \times 10^6$  before masking (a net total of  $n = 3 \times 10^6$  sources) and  $\Delta\tau_0 = 10\%$ , the Bayes factor  $B_{12}$  will be greater than one, meaning there is reasonable evidence for inconsistency with the CMB dipole.

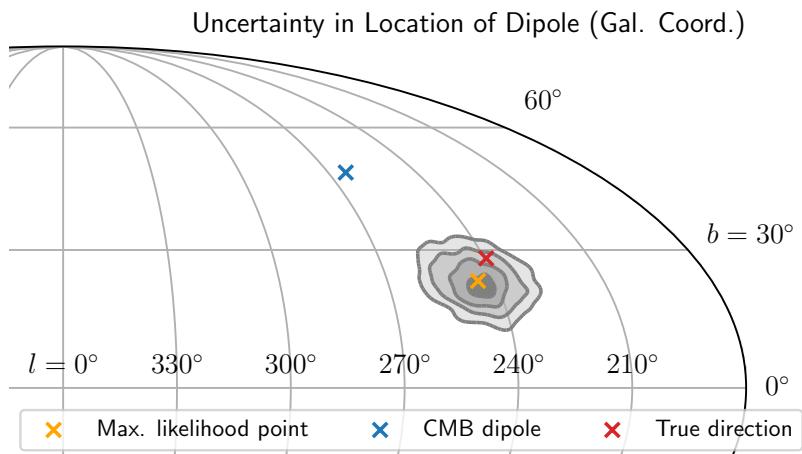


Figure 2.7: 2D marginal posterior for  $\theta$  and  $\phi$  projected onto the sky (Mollweide). Here,  $n = 6 \times 10^6$  before masking and  $\Delta\tau_0 = 0.1$  (10% uncertainty on  $\tau$ ). The contours represent intervals containing 10%, 40%, 65% and 85% of the posterior density function.

### Key approach

Our approach is to repeat the foregoing steps for different numbers of  $n$  and see how  $B_{12}$  and  $B_{10}$  change. If, on repeated tests with regenerated data, the Bayes factors consistently indicate a certain level of support (with a set  $n$ ) for the fitted hypothesis over the CMB and null hypotheses, then the time dilation dipole can be used to (in)validate the cosmological principle. In the following chapter, we step through whether or not that is the case in light of the results obtained.

# Chapter 3

## Results

In this chapter, our key findings are presented. In [Section 3.1](#), we interpret the Bayes factors and ascertain the number of quasars  $n$  required to resolve the true time dilation dipole over the hypothesis of no motion or CMB-oriented motion. In [Section 3.2](#), we then turn to how robustly JAVELIN has recovered timescales for given quasar light curves, and conclude with an estimate on the kinds of uncertainties expected in a typical quasar catalogue.

### 3.1 Points Required

In [Figure 3.1](#), we illustrate different regions on the sky that have been tested, converted from our coordinates  $(\theta, \phi)$  into Galactic coordinates  $(l, b)$ . Each yellow cross refers to a location that is imprinted into the sample; if that direction were the true direction of Earth's peculiar motion, then the effect of time dilation is attributed to each quasar point accordingly using [\(2.6\)](#). We have deliberately investigated certain regions, shown by the grey boxes in [Figure 3.1](#), to understand the sensitivity of the data analysis process. More precisely, if the true direction of Earth's motion lies within the masked region (Region B) or on the edge (Region D), does this hinder our ability to recover that direction? What if, as is more physically likely, the true direction is within  $\approx 30^\circ$  of the CMB dipole (Region A)? Or on the opposite side of the sky (Region C)? To enable a fair comparison of the regions, we have set Earth's peculiar speed  $\beta = 10^{-3}$ , near the value derived from the CMB ( $\beta = 1.2 \times 10^{-3}$ ). We now

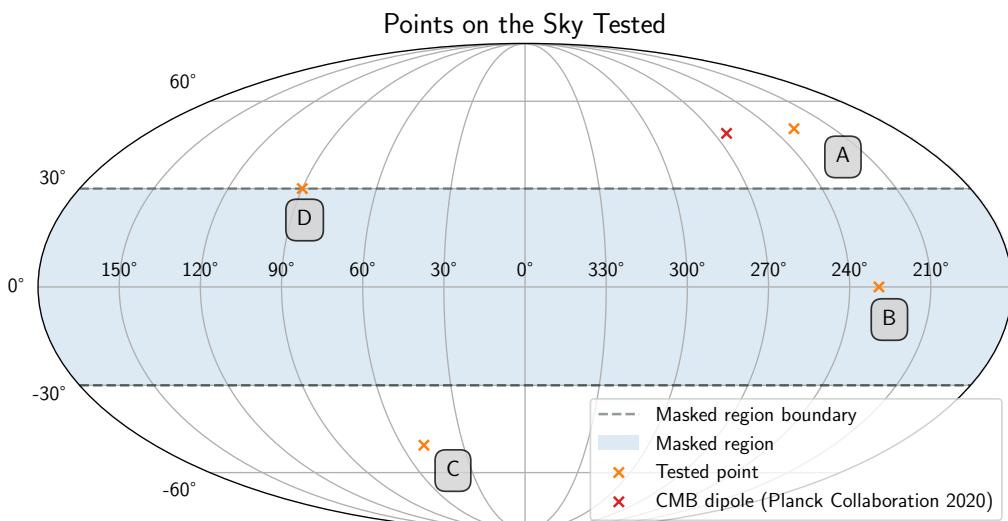


Figure 3.1: True directions of motion (yellow crosses) tested in our analysis with respect to the CMB dipole (red cross). The masked (removed) region of the sample is represented by the light blue shading. The labels applied to each of the directions of motion give clarity in the following analysis.

investigate each of these regions in turn.

### 3.1.1 Region B

Turning first to Region B of Figure 3.1, the key results are given in Figure 3.2. Recall that at each region, we generate distinct samples with varying  $n$  and compute the Bayes factors across repeated trials ( $\approx 20$  trials per  $n$ ). The  $n$  in this plot and subsequent ones refers to the number of sources before masking, so the net number of sources post-masking is half this, as explained in Subsection 2.3.2. Also, each point represents the mean of the natural logarithm of the Bayes factor after repeated trials, and the error bars are  $1\sigma$  of the log Bayes factors across those trials. Finally, the blue points give  $\ln B_{10}$ , in which the fitted hypothesis is compared to the null hypothesis, while the yellow points give  $\ln B_{12}$ , where the fitted hypothesis is compared to the CMB hypothesis.

What conclusions can be made? In Table 3.1, we give the conventional qualitative interpretation of the Bayes factors. It is important to note that these interpretations are not definitive; the values quoted reflect attempts to extrapolate an abstract value, which in reality exists on a continuous spectrum, to discrete normative categories. Yet, they are useful in understanding what values the Bayes factors generally take and what types of evaluative judgments can be drawn in light of them. Where  $\ln B_{ij}$  is between 0 and 1 (so  $B_{ij}$  is between 1 and  $e$ ), then there is only a slight indication of support for model  $M_i$  ‘not worth more than a bare mention’. Greater than these values, support appears with increasing degrees of strength. However, where the Bayes factors are negative,  $M_j$  is a more favourable hypothesis or model over  $M_i$ . Looking at Figure 3.2, by the third blue point from the left the log Bayes factors  $\ln B_{10}$ , on net, tend to be positive. At that point,  $n = 2.4 \times 10^6$  and  $\ln B_{10} = 7 \pm 5$ . For the case of  $\ln B_{12}$  (the yellow points), this occurs at the second point from the left where  $n = 1.45 \times 10^6$  and  $\ln B_{12} = 14 \pm 8$ .

A few comments are in order. First, the number of points required before clear support emerges is less when comparing the fitted to the CMB hypothesis, in contrast with the fitted to the null hypothesis. We see this in the sense that the yellow points rise more sharply than the blue points, reflecting that, at Region B, motion aligned with the CMB has far less explanatory power given the

$\ln B_{ij}$	Level of support
$< -5$	Very strong evidence for $M_j$
$-5$ to $-3$	Strong evidence for $M_j$
$-3$ to $-1$	Positive support for $M_j$
$-1$ to $0$	Bare mention of support for $M_j$
$0$ to $1$	Bare mention of support for $M_i$
$1$ to $3$	Positive support for $M_i$
$3$ to $5$	Strong support for $M_i$
$> 5$	Very strong support for $M_i$

Table 3.1: Qualitative interpretation of Bayes factors, as adapted from Kass & Raftery (1995). Note that their original paper quotes  $2 \ln B_{ij}$ , and here, their values are extrapolated into the negatives. Now, suppose  $B_{10} > 5$ . The interpretation then would be that there is very strong support for the fitted hypothesis over the null hypothesis.

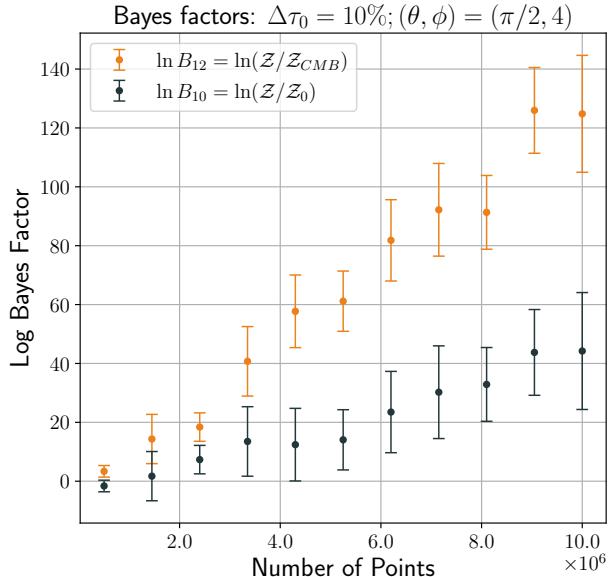


Figure 3.2: Log Bayes factors  $\ln B_{10}$  and  $\ln B_{12}$  against number of points  $n$  in sample for Region B. Uncertainty  $\Delta\tau_0 = 10\%$ .

data. Intuitively, this is because Region B is spatially distant from the CMB direction, as can be gleaned from [Figure 3.1](#). Second, since both Bayes factors illustrate support for the fitted hypothesis after a threshold  $n$ , it suggests that the time dilation dipole can be robustly extrapolated from the simulated sample, insofar that it is a stronger hypothesis than no motion or CMB-aligned motion assuming sufficient data. This is despite the penalisation from the Occam factor: as explained in [Subsection 2.4.2](#), a model with more complexity (more parameters) tends to be penalised over simpler explanations in the evidence integral [\(2.14\)](#). Here, the fitted hypothesis is a three-parameter problem, but the CMB/null hypotheses are zero-parameter problems. Even then, the latter are poorer explanations of the data than the more complex fitted model. This is still the case despite most of the signal being in the masked region: inasmuch that Earth’s peculiar motion is simulated towards the direction of Region B in [Figure 3.1](#), both poles of the time dilation dipole will be inside the mask. We therefore conclude that the prevailing impact of a mask is to limit the number of points available for analysis, but not obscure the time dilation dipole so as to totally inhibit its detection.

### 3.1.2 Region C

Turning now to Region C, the key results are shown in [Figure 3.3](#). Hitherto we explored trends with a fixed uncertainty (recall how this uncertainty is incorporated into the sample at [\(2.11\)](#)). Now, at Region C, we varied the uncertainties on the timescale  $\Delta\tau_0$  across multiple experiments, as shown in Figures [3.3a–3.3d](#). The most immediate observation is that increasing  $\Delta\tau_0$  corresponds to increasing the errors at fixed  $n$ . More precisely, the spread of the Bayes factors at a set  $n$  after repeated trials is larger. Now, this is less of an issue for  $\ln B_{12}$  (the yellow points)—irrespective of the large errors, the Bayes factors are consistently positive and always unambiguously show support for the fitted hypothesis over the CMB hypothesis. Yet, for the null hypothesis, it is more difficult to ascertain at what value of  $n$  strong support emerges with larger uncertainties. True it may be that for [Figure 3.3a](#) with  $\Delta\tau_0 = 1\%$ , support emerges immediately at  $n = 10^6$  where  $\ln B_{10} = 5 \pm 1$  and  $\ln B_{12} = 46 \pm 1$ . But take, for example, [Figure 3.3d](#) with  $\Delta\tau_0 = 50\%$ —even though the mean log Bayes factor  $\ln B_{10}$  indicates extreme support (well past the discrete categories of [Table 3.1](#)) for the fitted hypothesis over the null for all  $n$ , the spread is so large that, for some trials, the Bayes factors are negative. In particular,  $n = 5 \times 10^6$  yields Bayes factors of  $\ln B_{10} \approx 100 \pm 200$ , and at the left-most blue point of  $n = 10^6$  we have  $B_{10} \approx 30 \pm 50$ .

How do we account for this variation of the Bayes factors across trials at fixed  $n$ ? Since the data is regenerated each time, in which it is randomly sampled (see [Subsection 2.3.2](#)), some trials may preferentially favour the null hypothesis of no motion over the fitted hypothesis. Accordingly, this variation is characteristic of the sensitivity of the data analysis process to the actual input data. The net effect of this is that it becomes more challenging to give a concrete estimate of the required  $n$  for the fitted hypothesis to *consistently* be favoured. It is still possible, however, to argue that in the vicinity of  $n \gtrsim 1.5 \times 10^7$ , the Bayes factors  $\ln B_{10}$  for [Figure 3.3d](#) are consistently positive—at these values of  $n$  and greater the lower extremity of the errors are marginally below the line of  $\ln B_{10} = 0$ . We can also say that for  $\Delta\tau_0 = 10\%$  ([Figure 3.3b](#)), support emerges for  $n \gtrsim 3 \times 10^6$ , and for  $\Delta\tau_0 = 30\%$  ([Figure 3.3c](#)), support emerges where  $n \gtrsim 10^7$ .

### 3.1.3 Region D

A similar conclusion can be reached for Region D, where the signal lies on the edge of the mask. We defer these results to [Figure A.1](#) in [Appendix A](#) for the interested reader, since they only further evince the key messages espoused above. We direct our attention in the following section to Region A where the signal is close the CMB dipole. This is the most physically relevant experiment in light of, for example, the results of Secrest, von Hausegger, *et al.* (2021), which to reiterate suggest a spatial misalignment between the CMB and quasar number count dipole of  $\approx 30^\circ$ .

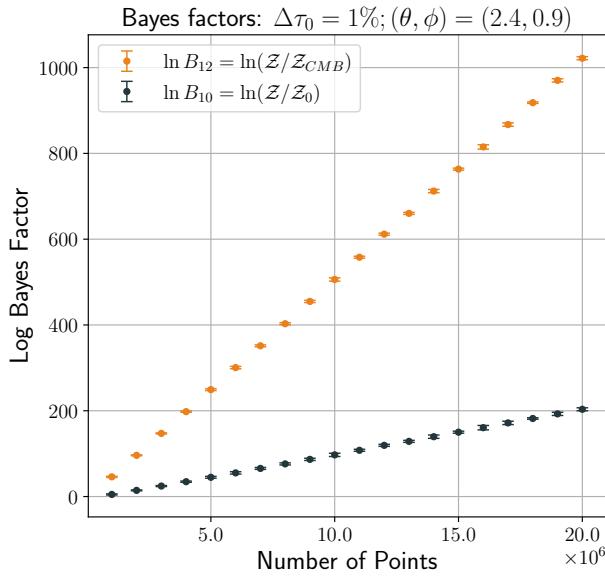
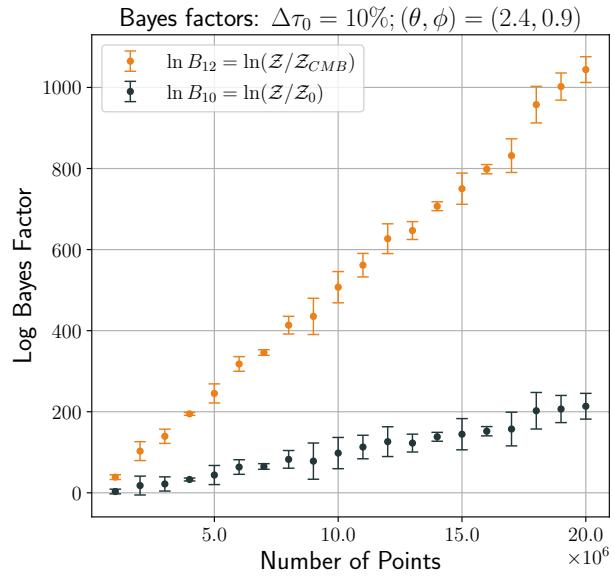
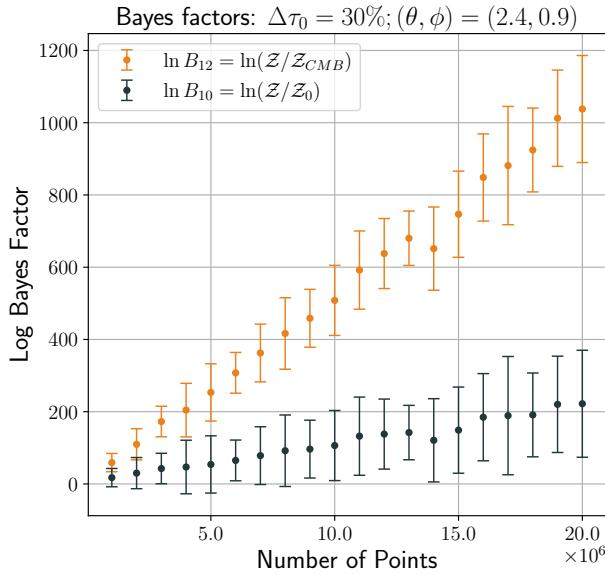
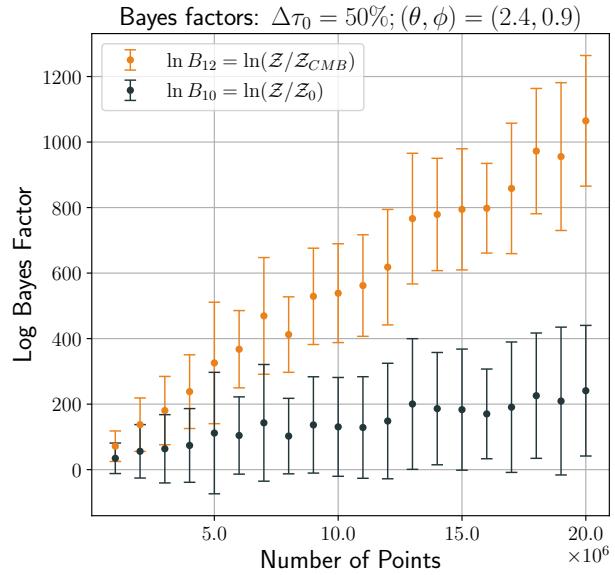
(a) Bayes factors against  $n$  points with  $\Delta\tau_0 = 1\%$ .(b) Bayes factors against  $n$  points with  $\Delta\tau_0 = 10\%$ .(c) Bayes factors against  $n$  points with  $\Delta\tau_0 = 30\%$ .(d) Bayes factors against  $n$  points with  $\Delta\tau_0 = 50\%$ .

Figure 3.3: Bayes factors against number of points  $n$  at Region C for different uncertainties on the timescale  $\Delta\tau_0$ . Note that a larger uncertainty corresponds to larger error bars—a greater spread of Bayes factors across repeated trials.

### 3.1.4 Region A

The key results for this region are given in Figure 3.4. Here, we have imprinted uncertainties of  $\Delta\tau_0 = 10\%$  and  $\Delta\tau_0 = 30\%$  onto the sample. Note that the colours have not been flipped—it turns out that, for this region, the CMB hypothesis has far more explanatory power than the null hypothesis. That accords with our intuition: insofar that the true direction of motion is now close to the CMB dipole, the CMB hypothesis becomes more likely, especially in light of the Occam factor. Conversely, there is always unambiguous support for the fitted hypothesis over the null hypothesis. The question

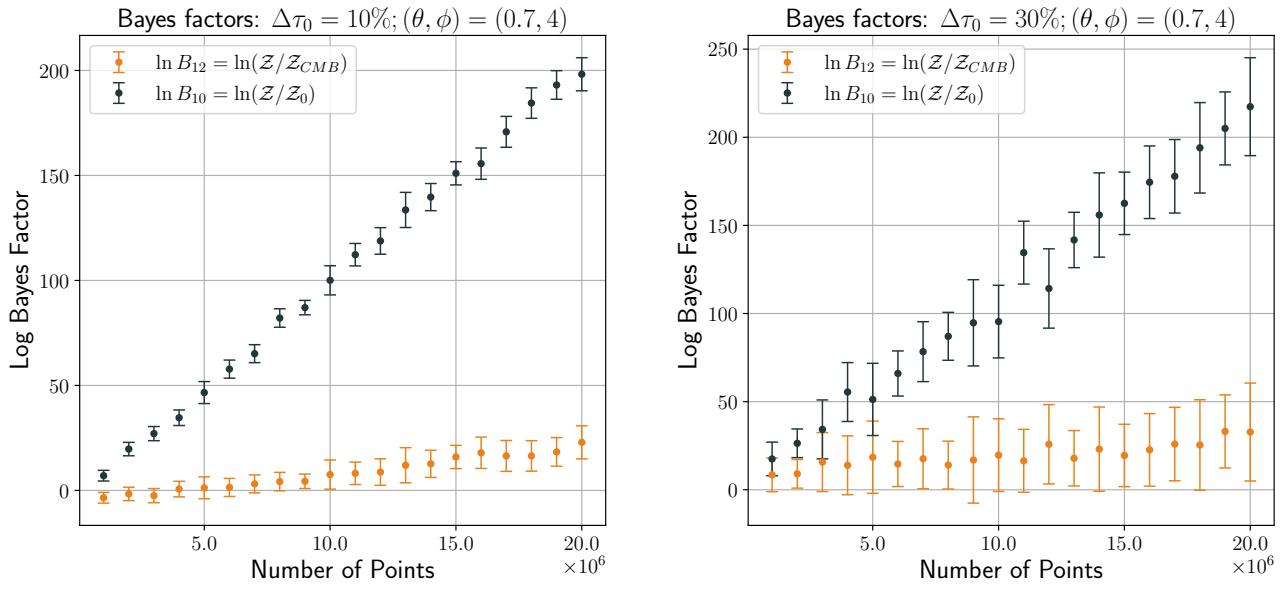


Figure 3.4: Bayes factors against number of points  $n$  at Region A at two different uncertainties on the timescale  $\Delta\tau_0$ . This is the primary region of interest to test the feasibility of detection of a time dilation dipole.

then becomes: at what point does the evidence favour the fitted hypothesis over the CMB hypothesis?

Turning to Figure 3.4a, it is comparatively clearer to see where support emerges: after  $n \approx 10^7$ , the Bayes factors  $\ln B_{12}$  are consistently positive—e.g, at  $n = 1.1 \times 10^7$ , we have  $\ln B_{12} = 8 \pm 5$ . We can therefore conclude that, if  $\Delta\tau_0 = 10\%$ , then we can resolve an anomaly between the directions of the time dilation dipole and the CMB dipole with about 10 million points before masking, corresponding to a net total of about 5 million points after masking. But if we turn to Figure 3.4b where  $\Delta\tau_0 = 30\%$ , it is harder to make that evaluative judgement given the errors. Again, if our metric for success were solely the mean Bayes factor, we would find support immediately at  $n = 10^6$  where  $\ln B_{12} \approx 10 \pm 10$ . But we are searching for *consistent, unambiguous support*—if in practice the time dilation dipole of a real quasar sample has large errors, then the novel test is not robust. Since we are faced with that challenge, from Figure 3.4b the evidence favours the fitted hypothesis over the CMB hypothesis consistently where  $n \gtrsim 1.5 \times 10^7$  before masking, so the net total is  $n \gtrsim 7.5 \times 10^6$  i.e. 7.5 million points after masking. This estimate is more tentative than the one given for  $\Delta\tau_0 = 10\%$ , but the key point that there appears to be a threshold  $n$  for support is nonetheless the same.

### Concluding remark

The foregoing analysis illustrates that, even if the true direction of Earth’s motion were within  $\approx 30^\circ$  of the CMB dipole, it could still be resolved from the latter with a requisite  $n$ . Looking at the results above, this is in the vicinity of 5 million to 7.5 million sources post-masking. The key parameter which alters the number of points required is the uncertainty  $\Delta\tau_0$ , as defined in (2.11). Here, we investigated uncertainties which ranged from 1%–50%. Which values within this range are more realistic than others? To answer that question, we consider how JAVELIN recovered the true timescales from quasar light curves in the following section.

## 3.2 Timescale Recovery

Recall from Section 2.3 that JAVELIN takes in a light curve as a set of points sampled in discrete intervals of time, potentially irregularly, and then recovers estimates for the timescale  $\tau$  and variability amplitude  $\hat{\sigma}$ . These parameters are intrinsic to the quasar which generates the light curve, and assume that the light curve can be described by a DRW. We now evaluate JAVELIN’s efficacy in recovering  $\tau$ , so as to better inform and interpret the results of the previous section. In all cases, we have assumed the photometric noise is drawn from  $\mathcal{G}(n_i = 0.01)$ —see (2.8)—and each light curve has a variability amplitude  $\hat{\sigma} = \text{SF}_\infty / \sqrt{\tau} = 0.2/100 = 2 \times 10^{-3}$ .

### 3.2.1 Sampling Duration

Here, we present our findings on the impact of sampling duration  $t$  on the recovery of the timescale  $\tau$ . As an initial test, we generated  $n = 300$  quasar light curves, and then sampled them over a duration  $t = 4000$  days, 40 times longer than the intrinsic DRW timescale. This was with a cadence of  $\Delta t = 10$  days. The key plots are given in Figure 3.5. Recall that, having taken in a light curve, in this case similar to the one in Figure 2.2a, JAVELIN explores the joint 2D posterior of  $\tau$  and  $\hat{\sigma}$  with an MCMC. We are only interested in  $\tau$ , so  $\hat{\sigma}$  is marginalised out. We thus show the 1D  $\tau$  posterior distribution for one of the 300 light curves in Figure 3.5a. From this distribution, we extract a median  $\tau$ , as labelled in that figure ( $\tau = 105.1^{+40.9}_{-21.0}$ ). This is because the mean is more susceptible to the long tail of the skewed distribution, which from our findings biases the timescale toward values larger than the truth. The distribution of those 300 medians is then shown in Figure 3.5b. There, the true timescale ( $\tau = 100$  days) and median of the distribution  $\tau = 98.2^{+26.1}_{-18.7}$  days is labelled, where the upper and lower limits represent 16%/84% intervals of posterior density. This same quantity will be labelled on all following plots of the timescale distribution.

There are two points to note here. First, the true value of  $\tau$  is recovered reasonably well from the distribution in Figure 3.5b, in the sense that the truth  $\tau = 100$  days and median  $\tau = 98.2$  days

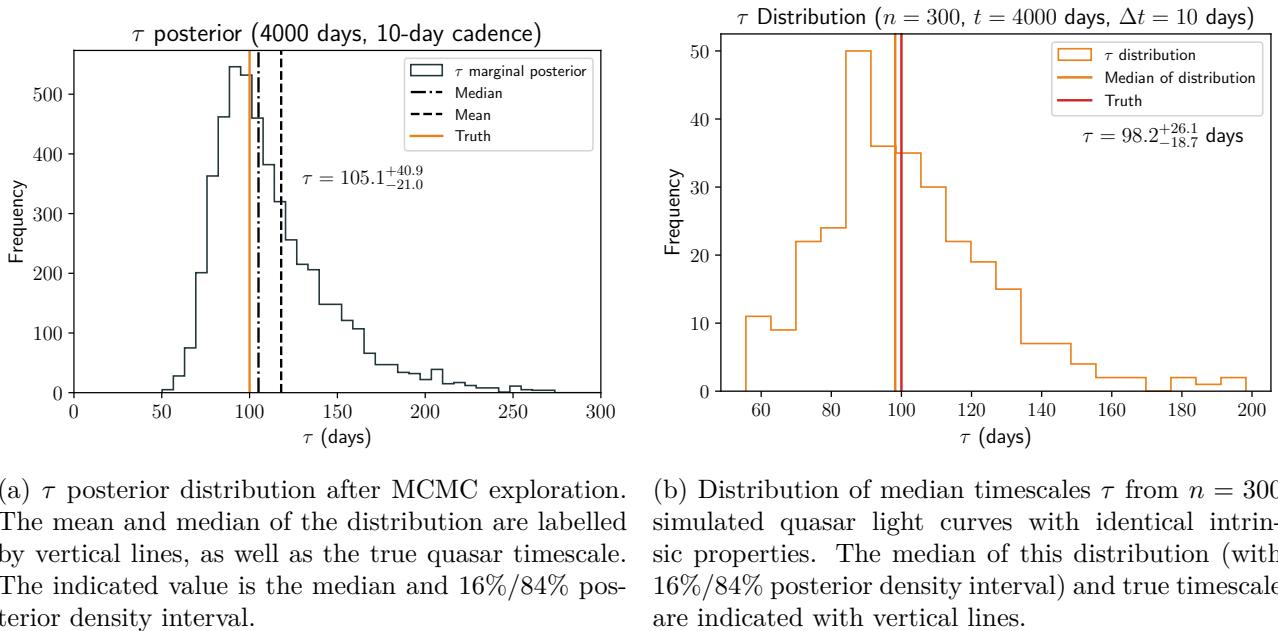
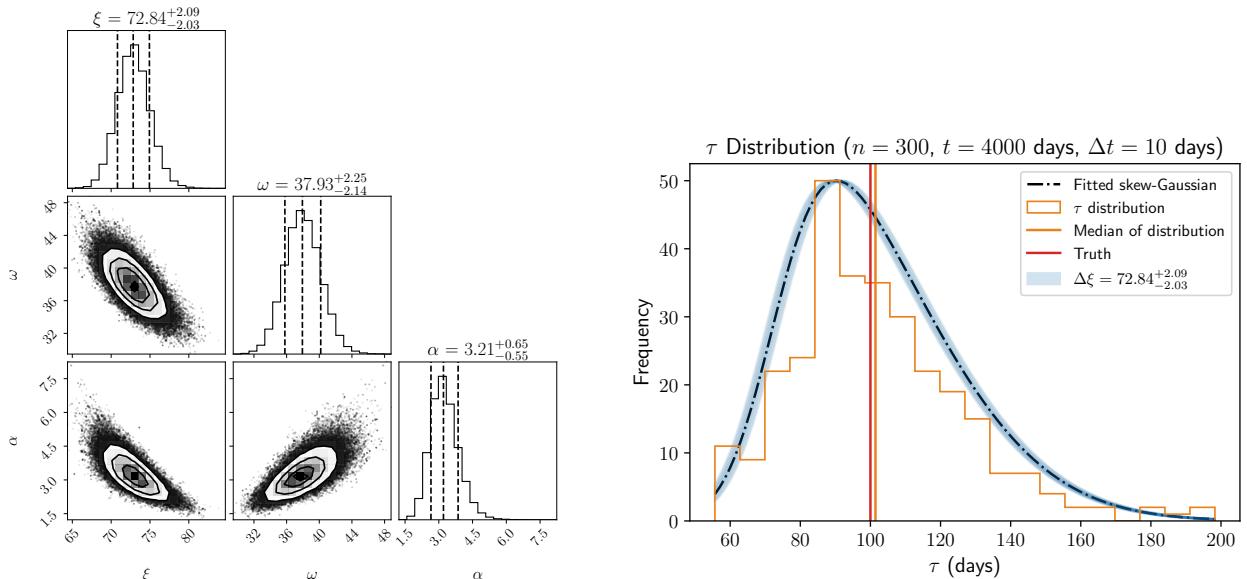


Figure 3.5: Recovery of the timescale of  $n = 300$  quasar light curves, sampled over 4000 days with a 10-day cadence. *Left:* Posterior distribution for  $\tau$  for one light curve within the 300. *Right:* Distribution of the median of all 300  $\tau$  posteriors.



(a) Corner plot of  $\xi$ ,  $\omega$  and  $\alpha$ , as pertaining to a skewed normal distribution, following MCMC exploration. This illustrates the 1D (the 3 histograms on the diagonal) and the 2D (the 3 inner scatter plots) marginalised posteriors of the parameters.

(b) The same distribution from Figure 3.5b but with a skewed normal distribution overlayed following fitting via MCMC, arbitrarily normalised so that both maxima coincide. The blue region reflects the uncertainty in the location parameter  $\Delta\xi$ .

Figure 3.6: Fitting a skewed normal to the same distribution in Figure 3.5b. *Left:* Corner plot of the parameters  $\xi$ ,  $\omega$  and  $\alpha$ . *Right:* Fitted skewed normal distribution overlayed.

are within  $\approx 2\%$  of each other. Second, in Figure 3.5b, the interval quoted in the median value of  $\tau$  characterises the distribution, but *is not a metric reflecting the uncertainty in  $\tau$*  i.e. it is not a one-to-one analogue of  $\Delta\tau_0$ . To explain, in Figure 3.6 we have fitted a skewed Gaussian to the distribution of Figure 3.5b. This is (neglecting normalisation) described by three parameters  $\xi$ ,  $\omega$  and  $\alpha$  representing the location, scale and skewness of the distribution respectively. To fit the curve, we optimised those three parameters in an MCMC run using the `emcee` Python package, as implemented in the script `light-curve-distribution.py` (see Appendix B). The results of that optimisation are shown in the corner plot of Figure 3.6a, made using the package `corner` (Foreman-Mackey, 2016). To optimise the parameters, our likelihood function was described by  $\ln \mathcal{L} = \sum_i \ln P_i(\tau_i | \xi, \omega, \alpha)$ , being simply the sum of the probabilities of all  $\tau$  if they lied on a PDF described by the skewed Gaussian. This gives us an uncertainty  $\Delta\xi$ ,  $\Delta\omega$  and  $\Delta\alpha$ , but in truth we are only interested in  $\Delta\xi$ .

When we actually observe, say, an area of sky with some number of quasars, even if they have different intrinsic properties, a large enough sample gives rise to a characteristic or typical distribution of timescales. We expound this point in Chapter 4, but for now, time dilation in (2.6) will impact all quasars in a patch of sky similarly, shifting the distribution of timescales  $\tau$ . So, some statistic like the mean or median will be offset from the value expected in the rest frame. Accordingly, the limit to how well we can measure this offset, as emblematic of the relativistic imprint of Earth's peculiar motion, is given by the uncertainty on our recovery of some arbitrary location parameter of the distribution of timescales i.e.  $\Delta\xi$ . The arbitrary normalisation we have chosen for the fitted distribution in Figure 3.6b, namely that the maxima of both distributions coincide, is therefore unimportant. As shown in that same figure, the fact that  $\xi = 72.84^{+2.09}_{-2.03}$  days means that the location parameter is well-defined with respect to the actual value of  $\tau$  i.e. the uncertainty in  $\xi$  corresponds to 2% of  $\tau$ . This is encouraging, since it suggests a degree of precision to which we can know the median or mean of some distribution of timescales at different points on the sky, as more analogous to  $\Delta\tau_0$ .

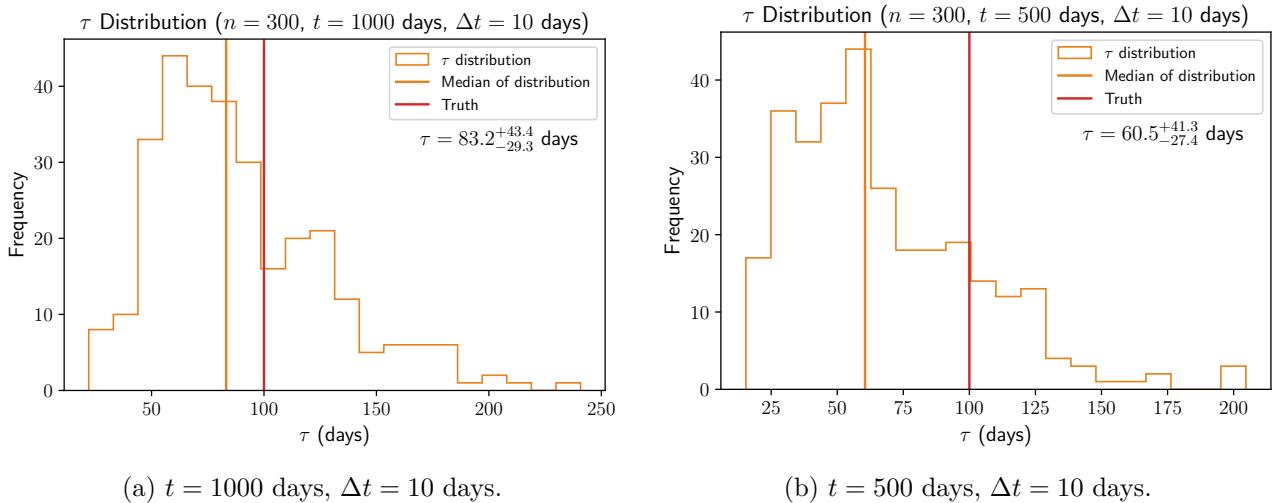


Figure 3.7: Distribution of median timescales  $\tau$  from  $n = 300$  simulated quasar light curves where the cadence  $\Delta t = 10$  days is maintained but the sampling duration  $t$  is varied. *Left:*  $t = 1000$  days. *Right:*  $t = 500$  days. Again, the median and 16%/84% posterior density interval is labelled for each.

With that established, what impact does the sampling duration have on the distribution of timescales? In Figures 3.7a–3.7b, we give the distribution of timescales for  $n$  quasars sampled for 1000 days and 500 days respectively. Again, the true timescale  $\tau = 100$  days in each plot is labelled by the red line. From those plots, we ascertain that a shorter sampling duration biases the recovered timescales to increasingly lower values. This corroborates the results of Kozłowski (2017) and Sheng *et al.* (2022) as mentioned in Section 2.3, namely the requirement that the sampling duration  $t$  should be at least  $\times 10$  the rest-frame timescale of the quasar for accurate recovery. Yet here, even with  $t = 1000$  days for  $\tau = 100$  days (Figure 3.7b), the recovery is poorer (the truth and median are inconsistent by 15%) than for  $t = 4000$  days. By  $t = 500$  days (Figure 3.7b), that error is  $\approx 40\%$ . From our experiments, we would therefore qualify that requirement, inasmuch as having a sampling duration *greater than*  $10\tau$  is ideal.

### 3.2.2 Cadence

How does cadence  $\Delta t$  impact the recovery of  $\tau$ ? Kozłowski (2017) additionally posited that high cadence is unnecessary; if the duration of sampling  $t$  is sufficient, then  $\Delta t$  will not impact the recovery of  $\tau$ , assuming the cadence is better than the intrinsic timescale  $\tau$ . We confirm those findings, noting that where  $\Delta t = 50$  days and  $\Delta t = 75$  days (Figures 3.8a and 3.8b respectively), the median and true  $\tau$  lie within  $\approx 5\%$  of each other. At  $\Delta t = \tau = 100$  days in Figure 3.8c, the median timescale is biased towards longer durations and the discrepancy is  $\approx 30\%$ . Further, in that same plot, none of the timescales in the distribution lie within  $\approx 20\%$  of the truth, suggesting that this bias is consistent across each MCMC run.

### 3.2.3 Windowing

Finally, as an estimation of how windowing impacts the recovery of  $\tau$ , we simulate an observational strategy which involves a 90-day ‘on season’, followed by 90 days of downtime where no observation of the light curve is made. This essentially has the effect of dividing the data into 90-day groups equally spaced in time. The key finding is shown in Figure 3.8d, where we have also fitted a skewed Gaussian to enable comparison with Figure 3.6b. We note that the discrepancy between the truth and median of the distribution is now  $\approx 5\%$ , whereas for the same observational conditions without

windowing ( $t = 4000$  days and  $\Delta t = 10$  days in Figure 3.5b) it was  $\approx 2\%$ . In reality, then, we can at best comment on a marginal deterioration in the accuracy of the recovered  $\tau$ , which is less substantial than other observational changes mentioned above. We also note that the uncertainty in the location parameter  $\Delta\xi$  has deteriorated marginally with  $\xi = 68.18^{+2.66}_{-2.42}$  days.

### Concluding remark

On the basis of the foregoing, we conclude that the key limiting factor for the accuracy of the recovered timescale  $\tau$  is the duration of sampling  $t$ . We find a requirement of  $t > 10\tau$  for the error between the median  $\tau$  and truth to be within 5%. Cadence and windowing, at least for the conditions we tested, have marginal impact. Namely, the accuracy of the determined value of  $\tau$  is not substantially diminished with less frequent cadences, as long as it less than  $\tau$ , and where a uniform 90-day on/90-day off window is used.

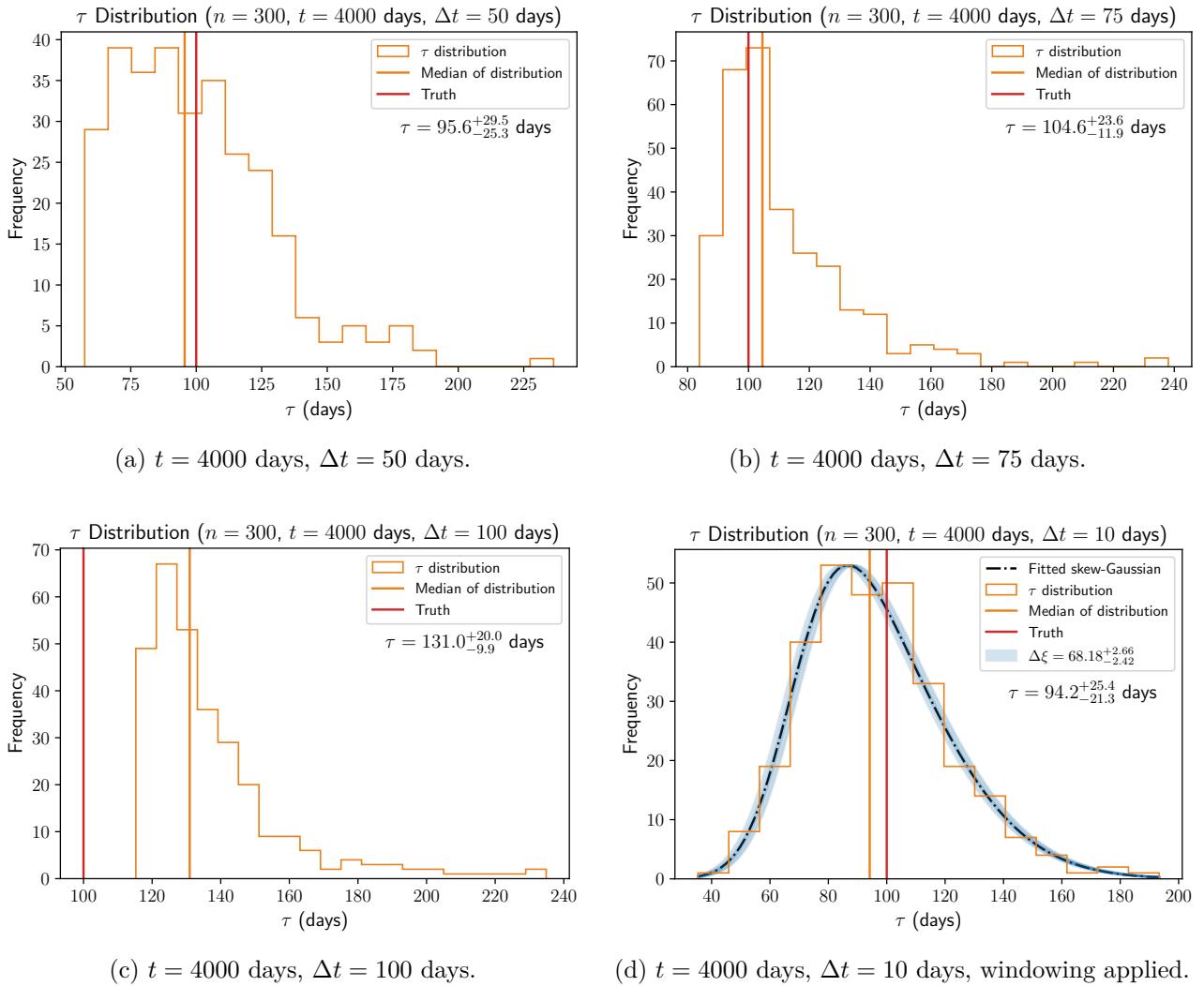


Figure 3.8: Distribution of recovered timescales  $\tau$  for  $n = 300$  quasars of identical properties where the cadence  $\Delta t$  and the use of windowing are varied, as labelled in each individual subfigure.

# Chapter 4

## Discussion

### 4.1 Denouement

The bedrock of modern models of the universe is laid down by the cosmological principle. In assuming spatial isotropy and homogeneity, we make use of symmetry to greatly simplify the otherwise complex task of solving Einstein’s field equations of general relativity. Armed with the FLRW metric and the Friedmann equations, we describe the structure of spacetime which suffuses the universe and how that structure expands with the passage of time. So it is that the  $\Lambda$ CDM model, the prevailing paradigm of cosmology, has at its core the cosmological principle.

In recent years, following the formulation of experiments to test the key proposition of the cosmological principle, it has come under threat. For instance, the findings of Secrest, von Hausegger, *et al.* (2021), corroborated in the follow-up study of Secrest, von Hausegger, *et al.* (2022), suggest a tension between experimental data and the *a priori* assumptions we have of the universe. Recall that, under the cosmological principle, the dipole in the CMB should align with a dipole in, for instance, source number count per unit solid angle. It is here that a discrepancy has emerged. In the words of Dr von Hausegger, (University of Oxford – Department of Physics, 2022)

If distant sources are not isotropic in the rest frame in which the CMB is isotropic, it implies a violation of the cosmological principle... So, we must now seek corroborating evidence to understand what causes this unexpected result.

We have proposed a novel means to search for that corroborating evidence. As predicted by special relativity, sources with intrinsic timescales should appear time dilated in Earth’s frame of reference; the ‘time dilation dipole’ that arises from Earth’s peculiar motion can then be compared to the CMB dipole to verify if there is any inconsistency. The final questions left at the end of Chapter 1: can we detect this dipole, and if so, how many sources with timescales would be required for that detection to be meaningful—that is, statistically robust?

The results in Section 3.1 of Chapter 3 illustrate that (a) the dipole is discernable in a simulated catalogue of quasars and (b) the number of sources required depends on how well we can measure timescales at different locations across the sky. There, it was shown that the number of sources pre-masking is in the vicinity of  $10^7$ , but a more precise number is dependent on the uncertainty  $\Delta\tau_0$  embedded into each source. If we assume that the true dipole is misaligned by  $\approx 30^\circ$  with the CMB dipole and uncertainties  $\Delta\tau_0$  are between 10%–30%, we would need an all-sky survey in the vicinity of  $n = 10^7$  to  $n = 1.5 \times 10^7$  sources before masking, which amounts to a net total of  $n = 5 \times 10^6$  to  $n = 7.5 \times 10^6$  sources following masking of the galactic plane. These last two values are the final key estimates for the number of points required.

If we assume that those sources are quasars, then in Section 3.2 of Chapter 3 we find that the timescale associated with the damped random walk of a quasar light curve can be recovered to within  $\approx 5\%$  accuracy. To achieve this degree of accuracy, the observational strategy deployed needs to sample the quasar for ten times the true timescale  $\tau$  and with a cadence shorter than that timescale.

The fact that the data is windowed (at least uniformly) is unlikely to substantially impact that level of accuracy. Thus, the most efficient observational strategy, with an end to extracting quasar timescales for this novel test, would prioritise sampling duration over a high frequency of observation.

## 4.2 Observational Challenges

Our results represent an order of magnitude estimate for the requisite data required to carry out this test. In reality, an actual survey may be faced with additional observational challenges, beyond the duration, cadence and windowing requirements mentioned above. In truth, though, our proposal is not restricted to quasar light curves. As mentioned in [Chapter 2](#), other variable sources on the sky with a discernable timescale could be used. Type Ia supernovae have a characteristic width to their light curve, which effectively acts as a clock. There is also, in principle, no reason why different families of sources with measurable timescales can be used in conjunction i.e. quasars and Type Ia supernovae.

To explain, as detailed in [Section 2.3](#), we have assumed that all sources in our sample have the same intrinsic timescale, are at rest with respect to each other, that we have some means to know their true rest-frame timescale and that the sample is all-sky. Yet, in practice, a more realistic sample would divide regions of the sky into cells (potentially with incomplete sky coverage) and bin over a redshift range. This is a practical necessity since, if we were for example observing quasars, their timescales would be time dilated due to the expansion of the universe (cosmological time dilation) in addition to kinematic time dilation as in [\(2.6\)](#). Thus, in each cell bin at some position on the sky, the impact due to cosmological time dilation would be the same. So, it may be that each quasar sub-sample has a different intrinsic clock to other quasars, or is considered alongside entirely different sources with different timescales—the key point is that, with a large enough set of sources in the sample, we should expect a distribution of timescales which is similar to other cell bins at different locations *but for the effect of kinematic time dilation*. This would impact the distribution by shifting some location parameter, like the mean or median, from the expected value in the rest-frame.

Thus, as in [Section 3.2](#), our metric for the uncertainty on  $\tau$  is the uncertainty in assigning a location parameter  $\xi$  to the distribution of timescales. It is changes to this parameter across the sky that the novel test will, in reality, investigate. In that case, a future mock sample to be used in the novel test could be more representative of observational reality, where for instance each quasar is given a different timescale, redshift, luminosity, peculiar velocity, etc. Then, the procedure would be similar to that of Chapters [2](#) and [3](#). This would more tightly constrain our estimate for the uncertainty of the timescale pertaining to some location parameter of a distribution of quasars, and, in turn, better constrain an estimate for the number of points required to robustly discern the time dilation dipole.

## 4.3 Looking to the Future

Even if this novel test is feasible and can be grounded in the realities of observation, are there surveys on the horizon which possess the requisite data to detect the time dilation dipole?

The Vera Rubin Observatory, as pictured in [Figure 4.1](#) and still under construction atop Cerro Pachón, Chile, will spearhead the Legacy Survey of Space and Time (LSST). Described as the ‘most ambitious survey currently planned in the optical’, LSST will upon completion of its decade-long plan have a catalogue of quasars exceeding 10 million ([Ivezić \*et al.\*, 2019](#)), all measured at a  $\approx 2$  night cadence ([Panda & Czerny, 2020](#)). In light of the sources required and the cadence/duration requirements explored in this report, this is highly encouraging; the LSST catalogue will likely be *the same order of magnitude* as the points needed to reach a statistically significant detection of a time dilation dipole. So, looking to the future, will LSST shed light on the reality of Earth’s cosmic motion? And how will this stand in light of the cosmological principle?

We began this report by examining the historical part that philosophical assumptions have played



Figure 4.1: The Vera Rubin telescope at Cerro Pachón, Chile (April 2021). Will the LSST reveal the truth of the Earth’s motion through the universe? Credit: Rubin Obs/NSF/AURA ([2022](#)).

in the course of cosmology, from the ancient world to modernity. In [Section 1.1](#), one way we framed this was in the context of the Copernican revolution, which overthrew paradigms that had prevailed for over a millennia. To call upon the words of Harrison ([2000](#)): ‘Each society in each age constructs a different cosmic picture that is like a mask fitted on the face of the unknown Universe’.

If we cannot unmask and reveal the true *Universe*, then we can refine our models of the *universe*, shedding light on the way in which assumptions shape our understanding—our paradigms. The cosmological principle is one such assumption, and the novel test of it we have devised herein is practically realisable. In the near future, we may be able to determine if the CMB dipole coincides with the time dilation dipole by turning to vast catalogues of quasar light curves, amongst other sources. In doing so, the tensions in the literature, as mentioned in [Section 1.3](#), can be approached in a new light with a view to resolving them. In this fashion, we may critically re-examine century-old assumptions about the nature of the universe.

# References

- 2dFRGS Team. *The 2dFRGS Image Gallery* <http://www.2dfgrs.net/> (2022).
- Aluri, P. K. *et al.* Is the Observable Universe Consistent with the Cosmological Principle? *arXiv e-prints*, arXiv:2207.05765. arXiv: [2207.05765 \[astro-ph.CO\]](https://arxiv.org/abs/2207.05765) (July 2022).
- Bayes, T. & Price, n. LII. An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, F. R. S. communicated by Mr. Price, in a letter to John Canton, A. M. F. R. S. *Philosophical Transactions of the Royal Society of London* **53**, 370–418. eprint: <https://royalsocietypublishing.org/doi/pdf/10.1098/rstl.1763.0053>. <https://royalsocietypublishing.org/doi/abs/10.1098/rstl.1763.0053> (1763).
- Bengaly, C. A., Maartens, R. & Santos, M. G. Probing the Cosmological Principle in the counts of radio galaxies at different frequencies. *Journal of Cosmology and Astroparticle Physics* **2018**, 031–031. <https://doi.org/10.1088/1475-7516/2018/04/031> (Apr. 2018).
- Berry, A. *A Short History of Astronomy* eng. <https://www.gutenberg.org/files/59212/59212-h/59212-h.htm> (Project Gutenberg, 2019).
- Blake, C. & Wall, J. A velocity dipole in the distribution of radio galaxies. *Nature* **416**, 150–152. arXiv: [astro-ph/0203385 \[astro-ph\]](https://arxiv.org/abs/astro-ph/0203385) (Mar. 2002).
- Blondin, S. *et al.* Time Dilation in Type Ia Supernova Spectra at High Redshift. *The Astrophysical Journal* **682**, 724–736. <https://doi.org/10.1086/589568> (Aug. 2008).
- Carroll, B. W. & Ostlie, D. A. *An Introduction to Modern Astrophysics* Second Edition. ISBN: 9781108422161 (Cambridge University Press, 2017).
- Cipozy. *Ptolemaic System - Digital Reproduction* This file is licensed under the Creative Commons Attribution-Share Alike 4.0 International license. 2015. [https://commons.wikimedia.org/wiki/File:Ptolemaic\\_System\\_-\\_Digital\\_Reproduction.png#file](https://commons.wikimedia.org/wiki/File:Ptolemaic_System_-_Digital_Reproduction.png#file).
- Colin, J., Mohayaee, R., Rameez, M. & Sarkar, S. High-redshift radio galaxies and divergence from the CMB dipole. *Monthly Notices of the Royal Astronomical Society* **471**, 1045–1055. arXiv: [1703.09376 \[astro-ph.CO\]](https://arxiv.org/abs/1703.09376) (Oct. 2017).
- Condon, J. J. *et al.* The NRAO VLA Sky Survey. *The Astronomical Journal* **115**, 1693–1716 (May 1998).
- Crawford, F. Detecting the Cosmic Dipole Anisotropy in Large-Scale Radio Surveys. *The Astrophysical Journal* **692**, 887–893. arXiv: [0810.4520 \[astro-ph\]](https://arxiv.org/abs/0810.4520) (Feb. 2009).
- Curd, P. *Presocratic Philosophy*. *The Stanford Encyclopedia of Philosophy* (ed Zalta, E. N.) <https://plato.stanford.edu/archives/fall2020/entries/presocratics/>. Metaphysics Research Lab, Stanford University, 2020.
- Dagnello, S. & NRAO/AUI/NSF. *AGNs Seen at Different Angles* 2019. <https://public.nrao.edu/gallery/agns-seen-at-different-angles/>.
- Einstein, A. *The Collected Papers of Albert Einstein. The Berlin Years: Writings, 1914-1917. (English translation supplement)* trans. by Engel, A., 421–32 (Princeton University Press, 1997).

- Ellis, G. F. R. & Baldwin, J. E. On the expected anisotropy of radio source counts. *Monthly Notices of the Royal Astronomical Society* **206**, 377–381 (Jan. 1984).
- Ellis, G. F. R., Maartens, R. & MacCallum, M. A. H. *Relativistic Cosmology* ISBN: 0521381150 (Cambridge University Press, Cambridge, 2012).
- Foreman-Mackey, D. corner.py: Scatterplot matrices in Python. *The Journal of Open Source Software* **1**, 24. <https://doi.org/10.21105/joss.00024> (June 2016).
- Foreman-Mackey, D., Hogg, D. W., Lang, D. & Goodman, J. emcee: The MCMC Hammer. *Publications of the Astronomical Society of the Pacific* **125**, 306–312. <https://doi.org/10.1086%2F670067> (Mar. 2013).
- Friedmann, A. in *A Source Book in Astronomy and Astrophysics, 1900–1975* (eds Lang, K. R. & Gingerich, O.) 838–843 (Harvard University Press, 2013). <https://doi.org/10.4159/harvard.9780674366688.c134>.
- Gale, G. *Cosmology: Methodological Debates in the 1930s and 1940s. The Stanford Encyclopedia of Philosophy* (ed Zalta, E. N.) <https://plato.stanford.edu/archives/fall2021/entries/cosmology-30s/>. Metaphysics Research Lab, Stanford University, 2021.
- Gibelyou, C. & Huterer, D. Dipoles in the sky. *Monthly Notices of the Royal Astronomical Society* **427**, 1994–2021. arXiv: [1205.6476 \[astro-ph.CO\]](https://arxiv.org/abs/1205.6476) (Dec. 2012).
- Górski, K. M. et al. HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere. *The Astrophysical Journal* **622**, 759–771. eprint: [arXiv:astro-ph/0409513](https://arxiv.org/abs/astro-ph/0409513) (Apr. 2005).
- Harris, C. R. et al. Array programming with NumPy. *Nature* **585**, 357–362. <https://doi.org/10.1038/s41586-020-2649-2> (Sept. 2020).
- Harrison, E. R. *Cosmology : the science of the universe* 2nd ed. eng. ISBN: 052166148X (Cambridge University Press, 2000).
- Hubble, E. A relation between distance and radial velocity among extra-galactic nebulae. *Proceedings of the National Academy of Sciences* **15**, 168–173. eprint: <https://www.pnas.org/doi/pdf/10.1073/pnas.15.3.168>. <https://www.pnas.org/doi/abs/10.1073/pnas.15.3.168> (1929).
- Hunter, J. D. Matplotlib: A 2D graphics environment. *Computing in Science & Engineering* **9**, 90–95 (2007).
- Ivezić, Ž. et al. LSST: From Science Drivers to Reference Design and Anticipated Data Products. *The Astrophysical Journal* **873**, 111. <https://dx.doi.org/10.3847/1538-4357/ab042c> (Mar. 2019).
- Kass, R. E. & Raftery, A. E. Bayes Factors. *Journal of the American Statistical Association* **90**, 773–795. ISSN: 01621459. <http://www.jstor.org/stable/2291091> (1995).
- Kelly, B. C., Bechtold, J. & Siemiginowska, A. ARE THE VARIATIONS IN QUASAR OPTICAL FLUX DRIVEN BY THERMAL FLUCTUATIONS? *The Astrophysical Journal* **698**, 895–910. <https://doi.org/10.1088/0004-637x/698/1/895> (May 2009).
- Knobel, C. *An Introduction into the Theory of Cosmological Structure Formation* 2012. <https://arxiv.org/abs/1208.5931>.
- Koposov, S. et al. *joshspeagle/dynesty: v1.2.3* version v1.2.3. June 2022. <https://doi.org/10.5281/zenodo.6609296>.
- Kothari, R., Panwar, M., Singh, G., Tiwari, P. & Jain, P. *A study of Dipolar Signal in distant Quasars with various observables* 2022. <https://arxiv.org/abs/2208.14397>.

- Kozłowski, S. A Survey Length for AGN Variability Studies. *Acta Astronomica* **71**, 103–112. arXiv: [2109.03896 \[astro-ph.GA\]](https://arxiv.org/abs/2109.03896) (June 2021).
- Kozłowski, S. Limitations on the recovery of the true AGN variability parameters using damped random walk modeling. *A&A* **597**, A128. <https://doi.org/10.1051/0004-6361/201629890> (2017).
- Kozłowski, S. *et al.* Quantifying Quasar Variability as Part of a General Approach to Classifying Continuously Varying Sources. *The Astrophysical Journal* **708**, 927–945. arXiv: [0909.1326 \[astro-ph.CO\]](https://arxiv.org/abs/0909.1326) (Jan. 2010).
- Kragh, H. *Conceptions of cosmos from myths to the accelerating universe : a history of cosmology* eng. ISBN: 9780191706219 (Oxford University Press, Oxford, 2007).
- Laplace, P.-S. *A Philosophical Essay on Probabilities* First Edition. Trans. by Truscott, F. W. & Emery, F. L. (John Wiley & Sons, 1902).
- Lemaître, G. Republication of: A homogeneous universe of constant mass and increasing radius accounting for the radial velocity of extra-galactic nebulae. *General Relativity and Gravitation* **45**, 1635–46. <https://doi.org/10.1007/s10714-013-1548-3> (2013).
- Longair, M. S. *The cosmic century : a history of astrophysics and cosmology* ISBN: 0521474361 (Cambridge University Press, Cambridge, 2006).
- Loredo, T. J. in *Maximum Entropy and Bayesian Methods* (ed Fougère, P. F.) 81–142 (Springer Netherlands, Dordrecht, 1990). ISBN: 978-94-009-0683-9. [https://doi.org/10.1007/978-94-009-0683-9\\_6](https://doi.org/10.1007/978-94-009-0683-9_6).
- Maartens, R. Is the Universe homogeneous? *Philosophical Transactions: Mathematical, Physical and Engineering Sciences* **369**, 5115–5137. ISSN: 1364503X. <http://www.jstor.org/stable/23057256> (2011).
- MacKay, D. J. C. *Information theory, inference, and learning algorithms* eng. ISBN: 0521642981 (Cambridge University Press, Cambridge, UK ; 2003).
- MacLeod, C. L. *et al.* MODELING THE TIME VARIABILITY OF SDSS STRIPE 82 QUASARS AS A DAMPED RANDOM WALK. *The Astrophysical Journal* **721**, 1014–1033. <https://doi.org/10.1088/0004-637x/721/2/1014> (Sept. 2010).
- Marsaglia, G. Choosing a Point from the Surface of a Sphere. *The Annals of Mathematical Statistics* **43**, 645–646. <https://doi.org/10.1214/aoms/1177692644> (1972).
- McKinney, W. *Data Structures for Statistical Computing in Python* in *Proceedings of the 9th Python in Science Conference* (eds van der Walt, S. & Millman, J.) (2010), 56–61.
- Milne, E. A. *Relativity, Gravitation and World-structure* (Oxford University Press, 1935).
- Mo, H., van den Bosch, F. & White, S. *Galaxy Formation and Evolution* ISBN: 9780521857932 (Cambridge University Press, 2010).
- NASA. *Astronomy Picture of the Day* <https://apod.nasa.gov/apod/ap010128.html> (2022).
- Panda, S. & Czerny, B. Generating AGN spectra for LSST: First steps in XXXIX Polish Astronomical Society Meeting (eds Małek, K. *et al.*) **10** (Oct. 2020), 285–287. arXiv: [1911.04803 \[astro-ph.HE\]](https://arxiv.org/abs/1911.04803).
- Perlmutter, S. Supernovae, Dark Energy, and the Accelerating Universe. *Physics Today* **56**, 53–60. eprint: <https://doi.org/10.1063/1.1580050>. <https://doi.org/10.1063/1.1580050> (2003).
- Planck Collaboration, Aghanim, N., *et al.* Planck 2018 results - I. Overview and the cosmological legacy of Planck. *A&A* **641**, A1. <https://doi.org/10.1051/0004-6361/201833880> (2020).

- Planck Collaboration, Akrami, Y., *et al.* Planck 2018 results - IV. Diffuse component separation. *A&A* **641**, A4. <https://doi.org/10.1051/0004-6361/201833881> (2020).
- Rubart, M. & Schwarz, D. J. Cosmic radio dipole from NVSS and WENSS. *Astronomy & Astrophysics* **555**, A117. arXiv: [1301.5559 \[astro-ph.CO\]](https://arxiv.org/abs/1301.5559) (July 2013).
- Rubin Obs/NSF/AURA. *Rubin Observatory Gallery* <https://gallery.lsst.org/bp/#/>. 2022.
- Secrest, N. J., von Hausegger, S., *et al.* A Test of the Cosmological Principle with Quasars. *The Astrophysical Journal Letters* **908**, L51. arXiv: [2009.14826 \[astro-ph.CO\]](https://arxiv.org/abs/2009.14826) (Feb. 2021).
- Secrest, N. J., von Hausegger, S., Rameez, M., Mohayaee, R. & Sarkar, S. *A Challenge to the Standard Cosmological Model* 2022. <https://arxiv.org/abs/2206.05624>.
- Sheng, X., Ross, N. & Nicholl, M. Legacy Survey of Space and Time cadence strategy evaluations for active galactic nucleus time-series data in Wide-Fast-Deep field. *Monthly Notices of the Royal Astronomical Society* **512**, 5580–5600. ISSN: 0035-8711. eprint: <https://academic.oup.com/mnras/article-pdf/512/4/5580/43378775/stac803.pdf>. <https://doi.org/10.1093/mnras/stac803> (Mar. 2022).
- Siewert, T. M., Schmidt-Rubart, M. & Schwarz, D. J. Cosmic radio dipole: Estimators and frequency dependence. *A&A* **653**, A9. <https://doi.org/10.1051/0004-6361/202039840> (2021).
- Singal, A. K. Large disparity in cosmic reference frames determined from the sky distributions of radio sources and the microwave background radiation. *Physical Review D* **100**, 063501. arXiv: [1904.11362 \[physics.gen-ph\]](https://arxiv.org/abs/1904.11362) (Sept. 2019).
- Singal, A. K. Large Peculiar Motion of the Solar System from the Dipole Anisotropy in Sky Brightness due to Distant Radio Sources. *The Astrophysical Journal Letters* **742**, L23. arXiv: [1110.6260 \[astro-ph.CO\]](https://arxiv.org/abs/1110.6260) (Dec. 2011).
- Skilling, J. *Nested Sampling in Bayesian Inference and Maximum Entropy Methods in Science and Engineering: 24th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering* (eds Fischer, R., Preuss, R. & Toussaint, U. V.) **735** (Nov. 2004), 395–405.
- Skilling, J. Nested sampling for general Bayesian computation. *Bayesian Analysis* **1**, 833–859. <https://doi.org/10.1214/06-BA127> (2006).
- Smoot, G. F., Gorenstein, M. V. & Muller, R. A. Detection of Anisotropy in the Cosmic Blackbody Radiation. *Physical Review Letters* **39**, 898–901 (Oct. 1977).
- Speagle, J. S. DYNESTY: a dynamic nested sampling package for estimating Bayesian posteriors and evidences. *Monthly Notices of the Royal Astronomical Society* **493**, 3132–3158. arXiv: [1904.02180 \[astro-ph.IM\]](https://arxiv.org/abs/1904.02180) (Apr. 2020).
- Springel, V. *et al.* Simulations of the formation, evolution and clustering of galaxies and quasars. *Nature* **435**, 629–636. <https://doi.org/10.1038%2Fnature03597> (June 2005).
- The Pandas Development Team. *pandas-dev/pandas: Pandas version latest*. Feb. 2020. <https://doi.org/10.5281/zenodo.3509134>.
- Tiwari, P. & Nusser, A. Revisiting the NVSS number count dipole. *Journal of Cosmology and Astroparticle Physics* **2016**, 062. arXiv: [1509.02532 \[astro-ph.CO\]](https://arxiv.org/abs/1509.02532) (Mar. 2016).
- Trotta, R. Bayes in the sky: Bayesian inference and model selection in cosmology. *Contemporary Physics* **49**, 71–104. arXiv: [0803.4089 \[astro-ph\]](https://arxiv.org/abs/0803.4089) (Mar. 2008).
- Tully, R. B., Courtois, H., Hoffman, Y. & Pomarède, D. The Laniakea supercluster of galaxies. *Nature* **513**, 71–73. arXiv: [1409.0880 \[astro-ph.CO\]](https://arxiv.org/abs/1409.0880) (Sept. 2014).

University of Oxford – Department of Physics. ‘*Lopsided*’ Universe could mean revision of standard cosmological model <https://www.physics.ox.ac.uk/news/lopsided-universe-could-mean-revision-standard-cosmological-model>. 2022.

Virtanen, P. *et al.* SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods* **17**, 261–272 (2020).

Wiltshire, D. L., Smale, P. R., Mattsson, T. & Watkins, R. Hubble flow variance and the cosmic rest frame. *Physical Review D* **88**, 083529. arXiv: [1201.5371 \[astro-ph.CO\]](https://arxiv.org/abs/1201.5371) (Oct. 2013).

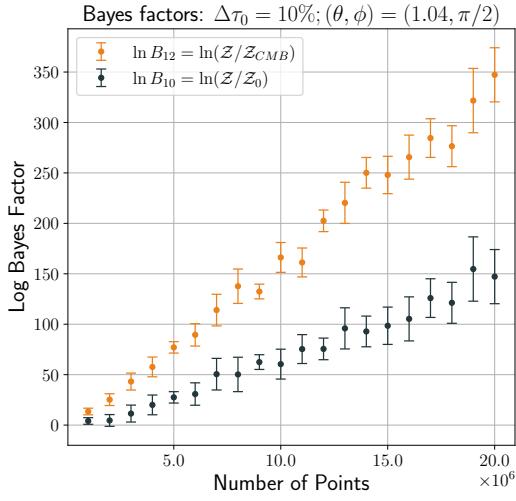
Zonca, A. *et al.* healpy: equal area pixelization and spherical harmonics transforms for data on the sphere in Python. *Journal of Open Source Software* **4**, 1298. <https://doi.org/10.21105/joss.01298> (Mar. 2019).

Zu, Y., Kochanek, C. S., Kozłowski, S. & Udalski, A. Is Quasar Optical Variability a Damped Random Walk? *The Astrophysical Journal* **765**, 106. arXiv: [1202.3783 \[astro-ph.CO\]](https://arxiv.org/abs/1202.3783) (Mar. 2013).

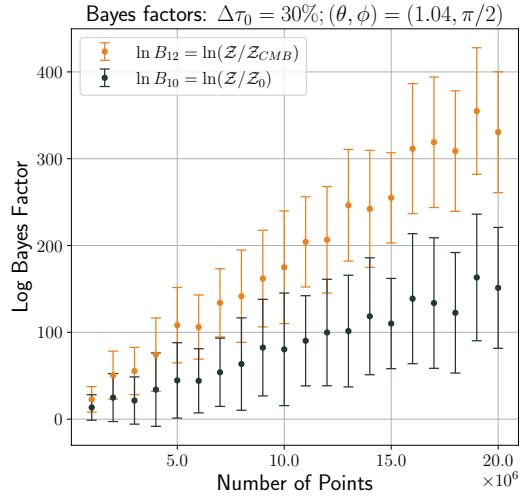
# Appendix A

## Supplementary Results

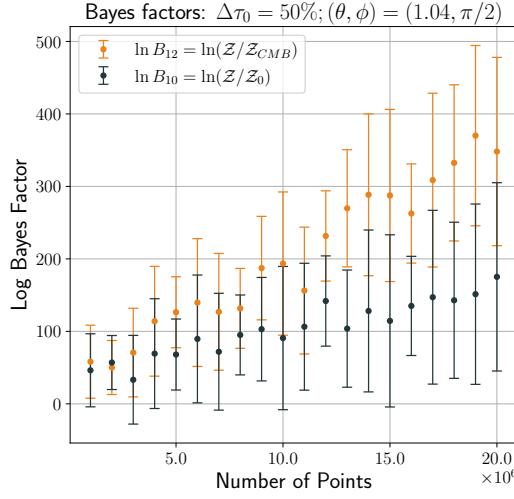
As explained in [Subsection 3.1.3](#), [Figure A.1](#) further illustrates the key trends when  $n$  is varied. Note that these plots apply for Region D of [Figure 3.1](#).



(a) Bayes factors against  $n$  points with  $\Delta\tau_0 = 1\%$ .



(b) Bayes factors against  $n$  points with  $\Delta\tau_0 = 10\%$ .



(c) Bayes factors against  $n$  points with  $\Delta\tau_0 = 30\%$ .

Figure A.1: Bayes factors against  $n$  at Region D for different uncertainties on the timescale  $\Delta\tau_0$ .

# Appendix B

## Python Code

### B.1 Code Summary

In this appendix chapter, we extract some of the key code used in this investigation. The entirety of the code is hosted on a Github repository at <https://github.com/o-oayda/Project-Code>, in which the top-level directory includes the key scripts, and the sub-folders include various results, like the evidences evaluated after each trial as well as the report figures. For a quick breakdown of the core functions of each script as related to this report, see [Table B.1](#). Note that all the code used in this investigation is written in Python 3, and has been run with Python 3.9.10.

In addition to `dynesty`, `corner`, `emcee`, `JAVELIN` and `healpy` as already mentioned earlier, this code made significant use of the Python packages `numpy` (Harris *et al.*, 2020), `matplotlib` (Hunter, 2007), `pandas` (The Pandas Development Team, 2020; McKinney, 2010) and `scipy` (Virtanen *et al.*, 2020).

Script	Function
<code>points-required.py</code>	Computes the evidences via <code>dynesty</code> for each hypothesis after generating a sample of $n$ quasars with intrinsic timescale $\tau$ and uncertainty $\Delta\tau_0$ .
<code>bayes-factors.py</code>	‘Computes the Bayes factors for each experiment given evidences. Also responsible for producing the plots of $\ln B$ against $n$ in <a href="#">Section 3.1</a> .’
<code>funcs.py</code>	Houses core functions e.g. implementation of equations, sampling, etc.
<code>plotting.py</code>	Houses core plotting functions, responsible for producing <a href="#">Figure 2.4</a> and <a href="#">Figure 2.7</a> .
<code>light-curve-MCMC.py</code>	‘Responsible for generating light curves via the DRW process and extracting $\tau$ and $\hat{\sigma}$ through <code>JAVELIN</code> .’
<code>light-curve-distribution.py</code>	Responsible for inspecting the distribution of timescales and producing the $\tau$ distribution plots in <a href="#">Section 3.2</a> .
<code>points-visualisation.py</code>	Responsible for producing <a href="#">Figure 2.3</a> .
<code>locations-tested.py</code>	Responsible for producing <a href="#">Figure 3.1</a> .

Table B.1: Name of Python scripts and their associated function.

## B.2 points-required.py

```

1 import healpy as hp
2 import numpy as np
3 import time
4 import math
5 import scipy.constants as sc
6 import pandas as pd
7 import scipy.stats as sts
8 import dynesty
9 from dynesty import plotting as dyplot
10 import multiprocessing
11 import matplotlib
12 from matplotlib import rc
13 from funcs import *
14 from scipy.constants import speed_of_light
15 from plotting import *
16 import traceback
17
18 ##### DYNESTY MODEL FITING FUNCTIONS #####
19
20 def model(v, az, pol):
21     observerVector = sph2cart((pol, az))
22     observerVector2 = np.asarray([observerVector])
23     alphaDash = angle(pixelVectors, observerVector2)
24     model = tau_true_pixels * \
25         (1-v*np.cos(alphaDash))/(np.sqrt(1-v**2)) # assumes knowledge of actual time-scale
26     return model
27
28 def prior_transform(uTheta):
29     uV, uAz, uPol = uTheta
30     v = 0.01*uV # uniform between 0 and 0.01
31     az = 2*np.pi*uAz # uniform between 0 and 2*pi # np.pi + np.pi*uAz
32     pol = uPol*np.pi # uniform between 0 and pi
33     return v, az, pol
34
35 def lnlike(Theta):
36     v, az, pol = Theta
37     tau_val = model(v, az, pol)
38     return sum(sts.norm.logpdf(m_new, loc=tau_val, scale=a_new))
39
40 ##### DETERMINING POINTS REQUIRED #####
41
42 t0 = time.time()
43 newDir = input('Enter directory name: ')
44
45 ##### Key variables
46 obsSpeed = 0.001
47 obsPolar = (0.7,4)
48 sigma_range = [0.3] # intrinsic uncertainty for each time-scale
49 nside = 16 # defines pixel density upon pixelisation
50 points_range = np.linspace(10**6,2*10**7,num=20).astype(int) # range of points
51 trials = range(1,10) # number of trials to run for
52
53 ##### Defining other variables
54 restLambda = ang_freq_to_lambda(1) # define ang_freq to be 1.
55 observerVector = sph2cart(obsPolar)
56 observerVector2 = np.asarray([observerVector])
57 pol_true, az_true = obsPolar
58 v_true = obsSpeed

```

```

59
60 ##### Now determine evidences over a range of points
61 for sigma in sigma_range:
62     for k in trials:
63         iteration = 0
64         ### Create vectors to store evidence values
65         fitted_evidence = np.zeros(len(points_range))
66         fitted_evidence_unc = np.zeros(len(points_range))
67         null_evidence = np.zeros(len(points_range))
68         cmb_evidence = np.zeros(len(points_range))
69         for n in points_range:
70             while True:
71                 try:
72                     # sample points from spherical distribution
73                     print('Sampling points, n = ' + str(n) + '...')
74                     xi, yi, zi = sample_spherical(n)
75                     print('Done.')
76                     initialPointsVectors = np.asarray([xi, yi, zi]).T
77
78                     # Apply aberration to points
79                     print('Rotating points...')
80                     rotatedPoints = transformedPoints(
81                         obsSpeed,
82                         observerVector2,
83                         restLambda,
84                         initialPointsVectors,
85                         DopplerShift=False)
86                     print('Done.')
87                     # Place timescales on the sky and dilate depending on point
88                     # All rest-frame timescales set to unity
89                     tau_true = np.ones(len(rotatedPoints))
90                     rotatedPointVectors = np.asarray(
91                         [rotatedPoints[0:len(rotatedPoints), 0],
92                          rotatedPoints[0:len(rotatedPoints), 1],
93                          rotatedPoints[0:len(rotatedPoints), 2]]).T # as above
94
95                     # use timescale dilation equation
96                     alphaDash = angle(rotatedPointVectors, observerVector2)
97                     tau_dilated = timeDilation(tau_true, alphaDash, obsSpeed)
98
99                     # Add measurement uncertainty to time-scales
100                    tau_error = np.random.normal(loc=0, scale=sigma, size=len(tau_true))
101
102                    # actual measured time-scales in observer frame
103                    tau_new = tau_dilated + tau_error
104                    tau_uncertainties = 0.1*np.ones(len(tau_new)) # uncertainty assumed to be 10%
105
106                    print('Averaging points...')
107                    # Create pixels and average time-scales with pixelAverage function
108                    npix = hp.nside2npix(nside) # number of pixels
109                    pixelVectors, m, a, hpx_map = pixelAverage(
110                         nside,
111                         rotatedPoints,
112                         tau_new,
113                         tau_uncertainties)
114                    print('Done.')
115
116                    # Apply mask (set points to zero inside mask)
117                    theta = np.zeros(len(pixelVectors))
118                    for i in range(0, len(pixelVectors)):

```

```

119     theta[i] = np.arccos(pixelVectors[i, 2]/math.sqrt(pixelVectors[i, 0]
120                           ** 2 + pixelVectors[i, 1]**2 + pixelVectors[i, 2]**2))
121     m[(theta >= np.deg2rad(60)) & (theta <= np.deg2rad(120))] = 0
122     a[(theta >= np.deg2rad(60)) & (theta <= np.deg2rad(120))] = 0
123
124     # Remove masked pixels which have tau = 0 for model fitting
125     # the new pixels with mask applied are stored in m_new and a_new
126     a_new = a[m!=0]
127     m_new = m[m!=0]
128     pixelVectors = pixelVectors[np.where(m!= 0)]
129     tau_true_pixels = np.ones(npix)
130     tau_true_pixels = tau_true_pixels[np.where(m!= 0)]
131
132     # Fit parameters using dynesty with multiprocessing
133     with multiprocessing.Pool(8) as pool:
134         dsampler = dynesty.NestedSampler(
135             lnlike, prior_transform, ndim=3, pool=pool, queue_size=8)
136         dsampler.run_nested()
137         dresults = dsampler.results
138
139     # Plot and save cornerplot generated by dynesty
140     fig, axes = dyplot.cornerplot(
141         dresults,
142         truths=[v_true,az_true,pol_true],
143         show_titles=True,
144         title_kwargs={'y': 1.04, 'fontsize' : 17},
145         label_kwargs={'fontsize' : 16},
146         labels=['$v$', '$\phi$', '$\theta$'],
147         title_fmt='.3f')
148
149     plt.savefig('Results/' + newDir + '/corner-n' + str(n) + '-nside' + str(nside)
150     → + '.pdf')
151
152     # Store evidence for fitted hypothesis
153     lnZ = dresults.logz[-1]
154     lnZ_err = dresults.logzerr[-1]
155     fitted_evidence[iteration] = lnZ
156     fitted_evidence_unc[iteration] = lnZ_err
157     print("log(Z) = {0:1.4f} ± {1:1.4f}".format(lnZ, lnZ_err))
158
159     # Compute and store evidence for null hypothesis
160     tau_val_null = 1
161     log_Z_0 = sum(sts.norm.logpdf(m_new, loc=tau_val_null, scale=a_new))
162     null_evidence[iteration] = log_Z_0
163     print("log(Z_0) = {0:1.1f}".format(log_Z_0))
164
165     # Compute and store evidence for CMB hypothesis
166     pol, az, v = np.pi/2 - (48.253*np.pi/180), 264.021 * \
167     np.pi/180, (369.82*1000)/speed_of_light
168     CMBDipoleVector = sph2cart((pol, az))
169     CMBDipoleVector2 = np.asarray([CMBDipoleVector])
170     alphaDash_CMB = angle(pixelVectors, CMBDipoleVector2)
171     tau_val_CMB = tau_true_pixels * \
172         (1-v*np.cos(alphaDash_CMB))/(np.sqrt(1-v**2))
173     log_Z_CMB = sum(sts.norm.logpdf(m_new, loc=tau_val_CMB, scale=a_new))
174     cmb_evidence[iteration] = log_Z_CMB
175     print("log(Z_CMB) = {0:1.1f}".format(log_Z_CMB))
176
177     ### Call plotting script to execute plots
uncertaintyOnSkyPlot()

```

```

178             nside,
179             dresults,
180             az_true,
181             pol_true,
182             n,
183             newDir,
184             topRightQuad=True)
185
186             numberCountHist(hpx_map,n,newDir)
187             numberCountDensity(hpx_map,m,n,newDir)
188
189             # Show number of points after mask
190             print('Points (after mask): {}'.format(str(sum(hpx_map))))
191
192             iteration += 1
193             break
194         except ValueError:
195             print(traceback.format_exc())
196             pass
197
198         # Unify evidences into dataframe after loop completion and save as csv
199         evidence_matrix =
200             np.c_[points_range,null_evidence,cmb_evidence,fitted_evidence,fitted_evidence_unc]
201         df =
202             pd.DataFrame(evidence_matrix,columns=['Points','ln(Z_0)','ln(Z_CMB)','ln(Z)','ln(Z_err)'])
203         df.to_csv('Results/' + str(newDir) + '/evidences-trials' + str(k) + 'sigma' + str(sigma) +
204             '.csv')
205         print(df)
206
207         # Print time taken to run
208         t1 = time.time()
209         total_time = t1-t0
210         print('Time taken (secs): {}'.format(str(total_time)))

```

### B.3 bayes-factors.py

An extract of `bayes-factors.py` is given below, without the header which involved reading in relevant data files:

```

170 ##### COMPUTE BAYES FACTORS #####
171 # Read in csv files containing evidences and combine into a dataframe
172 frames = []
173 for filename in test_files:
174     df = pd.read_csv(filename)
175     frames.append(df)
176
177 points_range = df['Points']
178 total_frame = pd.concat(frames)
179
180 # find unique points across trials
181 unique_points = sorted(list(set(total_frame['Points'])))
182
183 # extract all rows corresponding to a certain number of points
184 bayes_cmb_fitted = []
185 bayes_cmb_fitted_error = []
186 bayes_fitted_null = []
187 bayes_fitted_null_error = []
188 for k in unique_points:
189     active_frame = total_frame.loc[total_frame['Points'] == k]
190     factors = active_frame['ln(Z)'] - active_frame['ln(Z_CMB)']
191     bayes_cmb_fitted.append(np.mean(factors))
192     bayes_cmb_fitted_error.append(np.std(factors))
193     factors2 = active_frame['ln(Z)'] - active_frame['ln(Z_0)']
194     bayes_fitted_null.append(np.mean(factors2))
195     bayes_fitted_null_error.append(np.std(factors))
196
197 bayes_cmb_fitted = np.asarray(bayes_cmb_fitted,dtype=float)
198 bayes_cmb_fitted_error = np.asarray(bayes_cmb_fitted_error,dtype=float)
199 bayes_fitted_null = np.asarray(bayes_fitted_null,dtype=float)
200 bayes_fitted_null_error = np.asarray(bayes_fitted_null_error,dtype=float)
201
202 # convert to string with the \pm symbol for LaTeX table
203 factor_cmb_fitted = []
204 factor_fitted_null = []
205 for i in range(0,len(bayes_cmb_fitted)):
206     factor_cmb_fitted.append(conv2PlusMinus(
207         np.asarray(bayes_cmb_fitted[i]),
208         np.asarray(bayes_cmb_fitted_error[i])))
209     factor_fitted_null.append(conv2PlusMinus(
210         np.asarray(bayes_fitted_null[i]),
211         np.asarray(bayes_fitted_null_error[i])))
212
213 # determine mean and 1 sigma across trials for fitted, null and CMB hypotheses
214 means_fitted = []
215 devs_fitted = []
216 fitted_evidence = []
217 for j in unique_points:
218     active_frame = total_frame.loc[total_frame['Points'] == j]
219
220     means_fitted.append(np.mean(active_frame['ln(Z)']))
221     fitted_evidence.append(active_frame['ln(Z)'])
222     devs_fitted.append(np.std(active_frame['ln(Z)']))
223
224 means_cmb = []

```

```

225 devs_cmb = []
226 cmb_evidence = []
227 for j in unique_points:
228     active_frame = total_frame.loc[total_frame['Points'] == j]
229
230     means_cmb.append(np.mean(active_frame['ln(Z_CMB)']))
231     cmb_evidence.append(active_frame['ln(Z_CMB)'])
232     devs_cmb.append(np.std(active_frame['ln(Z_CMB)']))
233
234 means_null = []
235 devs_null = []
236 for j in unique_points:
237     active_frame = total_frame.loc[total_frame['Points'] == j]
238
239     means_null.append(np.mean(active_frame['ln(Z_0)']))
240     devs_null.append(np.std(active_frame['ln(Z_0)']))
241
242 fitted = np.asarray(means_fitted,dtype=float)
243 devs_fitted = np.asarray(devs_fitted,dtype=float)
244 cmb = np.asarray(means_cmb)
245 devs_cmb = np.asarray(devs_cmb)
246 null = np.asarray(means_null)
247 devs_null = np.asarray(devs_null)
248
249 # use conv2PlusMinus to create string with \pm symbol as errors
250 errors_fitted = []
251 errors_null = []
252 errors_cmb = []
253 for i in range(0,len(fitted)):
254     errors_fitted.append(
255         conv2PlusMinus(
256             np.asarray(fitted[i]),
257             np.asarray(devs_fitted[i])))
258
259     errors_null.append(
260         conv2PlusMinus(
261             np.asarray(null[i]),
262             np.asarray(devs_null[i])))
263
264     errors_cmb.append(
265         conv2PlusMinus(
266             np.asarray(cmb[i]),
267             np.asarray(devs_cmb[i])))
268
269 # dataframe to be exported to csv --- replaced with unique points
270 export =
271     → np.c_[unique_points,errors_fitted,errors_cmb,errors_null,factor_cmb_fitted,factor_fitted_null]
271 results1 = (pd.DataFrame(export))
272
273 # write all evidence values and Bayes factors to .tex table
274 table_name = 'table' + str(test_index)
275 filename = str(table_name) + '-build'
276 with open('Results/' + str(table_name) + '.tex','w') as tex_file:
277     tex_file.write(results1.to_latex(
278         na_rep='',
279         index=False,
280         header = ['$\\$',
281                 "{$\\ln \\mathcal{Z} \\$}",
282                 "{$\\ln \\mathcal{Z}_\\text{CMB} \\$}",
283                 "{$\\ln \\mathcal{Z}_0 \\$}'],

```

```

284     '\${\$\\ln \\left( \\mathcal{Z} / \\mathcal{Z}_{\\text{CMB}} \\right)}$',
285     '\${\$\\ln \\left( \\mathcal{Z} / \\mathcal{Z}_0 \\right)}$',
286     column_format='''S[table-format=8,round-precision=0]
287     S[table-format=6(3),round-mode=uncertainty,round-precision=1]
288     S[table-format=6(3),round-mode=uncertainty,round-precision=1]
289     S[table-format=6(3),round-mode=uncertainty,round-precision=1]
290     S[table-format=1(1),round-mode=uncertainty,round-precision=1]
291     S[table-format=1(1),round-mode=uncertainty,round-precision=1]'''',
292     escape=False))
293
294 # put table in standalone tex document
295 with open("Results/" + str(filename) + ".tex", "w") as tex_file:
296     tex_file.write(r"""
297 \documentclass[varwidth=\maxdimen]{standalone}
298 % \documentclass{memoir}
299 \usepackage{booktabs}
300 \usepackage[table, svgnames]{xcolor}
301 \usepackage[round-mode=places,separate-uncertainty=true,table-align-uncertainty=true]{siunitx}
302 \setlength{\extrarowheight}{1mm}
303
304 \rowcolors{1}{black!10}{}
305 \setlength{\aboverulesep}{0pt}
306 \setlength{\belowrulesep}{0pt}
307
308 \begin{document}
309
310 \begin{table}
311     \input{" + str(table_name) + "}"
312 \end{table}
313
314 \end{document}""")
315
316 plt.figure(figsize=(6.4,6))
317
318 # plot fitted to CMB Bayes factors
319 plt.errorbar(
320     unique_points,
321     bayes_cmb_fitted,
322     yerr=bayes_cmb_fitted_error,
323     label=r'$\\ln B_{12} = \\ln ( \\mathcal{Z} / \\mathcal{Z}_{\\text{CMB}} )$',
324     capsize=4,
325     markersize=4,
326     lw=1,
327     fmt='o',
328     c="#eb811b") # met orange
329
330 # plot fitted to null Bayes factors
331 plt.errorbar(
332     unique_points,
333     bayes_fitted_null,
334     yerr=bayes_fitted_null_error,
335     label=r'$\\ln B_{10} = \\ln ( \\mathcal{Z} / \\mathcal{Z}_0 )$',
336     capsize=4,
337     markersize=4,
338     lw=1,
339     fmt='o',
340     c="#23373b") # met blue
341
342 plt.xlabel('Number of Points', fontsize=18)
343 plt.ylabel('Log Bayes Factor', fontsize=18)

```

```
344 plt.grid(True)
345 plt.legend(fontsize=14)
346 plt.title(title,fontsize=18)
347 plt.ticklabel_format(useMathText=True)
348
349 ax = plt.gca()
350 fig = plt.gcf()
351
352 ax.xaxis.set_major_formatter(00MFormatter(6, "%1.1f"))
353
354 ax.tick_params(labelsize=14)
355 ax.xaxis.offsetText.set_fontsize(14)
356 plt.savefig('Results/' + 'table-' + test_index + '-plot.pdf',bbox_inches='tight')
357
358 plt.show()
359
360 # change to results directory and build standalone tex document
361 os.chdir('/Users/oliveroayda/Documents/GitHub/Project-Code/Results')
362 os.system('pdflatex ' + str(filename) + '.tex')
```

## B.4 light-curve-MCMC.py

```

1 from cmath import inf
2 import numpy as np
3 import pandas as pd
4 from javelin.lcmodel import Cont_Model
5 from javelin.zylc import get_data
6 from funcs import window
7
8 n_iters = 300
9 duration = 4000 # in days
10 c = 10 # observational cadence in days
11 windowing = True
12 season_length = 90 # in days
13 season_gap = 90 # in days
14
15 # generate true light curve
16 for i in range(1,n_iters):
17     tau = 100 # signal decorrelation timescale in days
18     sf = 0.2 # structure function
19     sigma_hat = sf/np.sqrt(tau) # modified variability amplitude
20     sigma = sf/np.sqrt(2)
21     delta_t = 0.5 # intrinsic cadence delta_t = t_{i+1} - t_i in days
22     mean_mag = 18 # please keep
23     n_epochs = int(duration/delta_t)
24     s = np.zeros(n_epochs)
25     ts = np.zeros(n_epochs)
26
27     for t in range(0, n_epochs):
28         if t == 0:
29             # for the 0th point, draw from Gaussian deviate of width sigma
30             s[t] = np.random.normal(loc=0, scale=sigma)
31             ts[t] = t*delta_t
32         else:
33             s[t] = s[t-1]*np.exp(-delta_t/tau) + np.random.normal(
34                 loc=0,scale=sigma*np.sqrt(1-np.exp(-2*delta_t/tau)))
35             ts[t] = t*delta_t
36
37     photometric_noise = np.random.normal(loc=0, scale=0.01, size=len(s))
38
39     # light curve points are s + Gaussian noise + mean magnitude
40     y = s + mean_mag + photometric_noise
41
42     # down sample light curve based on cadence
43     ts_cadenced = ts[0::int(c/delta_t)]
44     y_cadenced = y[0::int(c/delta_t)]
45     photometric_noise_cadenced = photometric_noise[0::int(c/delta_t)]
46
47     # apply windowing
48     if windowing:
49         ts_final = window(ts_cadenced,
50                           season_length=season_length,
51                           season_gap=season_gap,
52                           cadence=c)
53         y_final = window(y_cadenced,
54                           season_length=season_length,
55                           season_gap=season_gap,
56                           cadence=c)
57         photometric_noise_final = window(photometric_noise_cadenced,
58                                         season_length=season_length,

```

```

59         season_gap=season_gap,
60         cadence=c)
61     else:
62         # unwindowed data
63         ts_final = ts_cadenced
64         y_final = y_cadenced
65         photometric_noise_final = photometric_noise_cadenced
66
67     # store into data matrix
68     the_data = np.zeros((len(ts_final), 3))
69     the_data[:, 0] = ts_final
70     the_data[:, 1] = y_final
71     the_data[:, 2] = photometric_noise_final
72
73     # MCMC saved light curves
74     d = duration
75     df = pd.DataFrame(the_data)
76     df.to_csv(
77         'DRW Results/Multi Curve/curve-d' +\
78             str(d) + '-c' + str(c) + '-iter' + str(i+1) + '.dat'\
79             if not windowing else 'DRW Results/Multi Curve/cad-curve-d' +\
80             str(d) + '-c' + str(c) + '-iter' + str(i+1) + '.dat',
81             header=False,
82             index=False,
83             sep=' ')
84
85     javdata = get_data(
86         ["DRW Results/Multi Curve/curve-d" +\
87             str(d) + "-c" + str(c) + '-iter' + str(i+1) + ".dat"]\
88             if not windowing else ["DRW Results/Multi Curve/cad-curve-d" +\
89             str(d) + "-c" + str(c) + '-iter' + str(i+1) + ".dat"],
90             names=["$\tau = 100$ days light curve", ])
91
92     cont = Cont_Model(javdata)
93     cont.do_mcmc(
94         fchain="DRW Results/Multi Curve/" + "chain-d" +
95             str(d) + "-c" + str(c) + '-iter' + str(i+1) + ".dat"\
96             if not windowing else "DRW Results/Multi Curve/" + "cad-chain-d" +\
97             str(d) + "-c" + str(c) + '-iter' + str(i+1) + ".dat",
98         flogp="DRW Results/Multi Curve/logp" + "-d" +
99             str(d) + "-c" + str(c) + '-iter' + str(i+1) + ".dat"\
100             if not windowing else "DRW Results/Multi Curve/cad-logp" + "-d" +\
101             str(d) + "-c" + str(c) + '-iter' + str(i+1) + ".dat",
102             threads=8)

```