arXiv:1703.03425v1 [astro-ph.CO] 9 Mar 2017

Strong Evidence for the Normal Neutrino Hierarchy

Fergus Simpson,[1] **Raul Jimenez,**[1,2] **Carlos Pena-Garay,**[3,4] **Licia Verde**[1,2]

[1]ICCUB, University of Barcelona (UB-IEEC), Marti i Franques 1, Barcelona, 08028, Spain.
[2]ICREA, Pg. Lluis Companys 23, Barcelona, 08010, Spain.
[3]I2SysBio, CSIC-UVEG, P.O. 22085, Valencia, 46071, Spain.
[4]LSC, Estación de Canfranc, 22880, Spain.

E-mail: fergus2@gmail.com; raul.jimenez@icc.ub.edu; liciaverde@icc.ub.edu; penagaray@gmail.com;

Abstract. The configuration of the three neutrino masses can take two forms, known as the normal and inverted hierarchies. We compute the Bayesian evidence associated with these two hierarchies. Previous studies found a mild preference for the normal hierarchy, and this was driven by the asymmetric manner in which cosmological data has confined the available parameter space. Here we identify the presence of a second asymmetry, which is imposed by data from neutrino oscillations. By combining constraints on the squared-mass splittings [1] with the limit on the sum of neutrino masses of $\Sigma m_\nu < 0.13$ eV [2], we infer odds of 42:1 in favour of the normal hierarchy, which is classified as "strong" in the Jeffreys' scale. We explore how these odds may evolve in light of higher precision cosmological data, and discuss the implications of this finding with regards to the nature of neutrinos.

Contents

1   Introduction

Cosmological surveys provide crucial information on the absolute masses of neutrinos, information that is not accessible even to future laboratory experiments [3–5]. Recent analyses have already established a robust upper limit on the sum of neutrino masses of $\Sigma < 0.13$ eV [2] and $\Sigma < 0.12$ [6] at 95% confidence (throughout this work we shall denote Σm_ν as Σ for brevity). This limit is tantalisingly close to the lower limit for the sum of the masses in the inverted hierarchy $\Sigma = 0.0982 \pm 0.0010$ eV (68% C.L.) [1]. Because the volume covered by cosmological surveys keeps increasing, their statistical errors shrink, soon leading to either (a) a measure the total neutrino mass or (b) the exclusion of the inverted hierarchy, leaving the normal one as the only viable option. It is therefore possible to determine the neutrino hierarchy, even in the absence of a measurement of the total mass e.g., Ref. [7–9]. Future surveys may even yield a measurement of the mass differences between the lightest and heaviest eigenstates [10]. This measurement would also provide a powerful consistency check: the measured mass splitting from cosmology should be consistent with the squared mass splitting from oscillations. Given the current knowledge about mass splitting and cosmological constraints on the total mass, can anything be said about the neutrino mass hierarchy?

Since current cosmological upper limits are already close to the inverted hierarchy bound, the volume of parameter space available is heavily restricted. As such, it is useful to employ Bayesian evidence methods in order to determine which hierarchy is favoured by the data. While it seems counterintuitive that one can determine a quantity without having actually measured it, consider the following pedagogical example: if there is a thin stick of length L hidden underneath a cloth, one does not necessarily need to reveal the object in order to learn of its orientation. At the start, it is considered equally likely to be in either a horizontal or vertical position. If the cloth is gradually withdrawn in the horizontal direction, and no sign of the stick emerges, one need not wait until a residual length less than L is left unexposed to be confident of a vertical orientation. Even while

a full $2L$ of the surface remains covered, the odds are already stacked $2 : 1$ in favour of a vertical orientation.

Our main objective is to identify which hierarchy is preferred by the Bayesian evidence ratio, and understand how this is likely to evolve as new cosmological data arrives in the near future. Previous inferences on the neutrino hierarchy have been made by Refs. [11] and [12]. One crucial aspect where our work differs is the choice of the prior, which we shall explore in detail. We will also outline the consequences and implications of identifying the nature of the neutrino hierarchy.

2 Cosmology, oscillations, and statistics

In this section we provide a brief review of the three disciplines that are central to this work.

2.1 Cosmology: observational constraints and degeneracies

Massive neutrinos influence both the expansion history and growth of structure in the Universe. Neutrinos of mass $\lesssim 1$ eV become non-relativistic after the epoch of recombination probed by the Cosmic Microwave Background (CMB), and this mechanism allows massive neutrinos to alter the relationship connecting the physical matter density $\Omega_m h^2$ to the redshift of matter-radiation equality. The neutrinos' radiation-like behaviour at early times modulates the peak positions in the CMB angular power spectrum. However this effect is somewhat degenerate with other cosmological parameters. Neutrinos are also responsible for dampening the matter power spectrum below their free-streaming scale; see e.g., [13] for a review. This effect, which amounts to a scale dependent power suppression, is straightforward to compute in linear theory [14] and in the non-linear regime via numerical simulations e.g., [15–17].

Due to these mechanisms, cosmological observables are predominantly sensitive to the sum of neutrino masses Σ e.g., [13, 18–23]. While the effect of neutrino mass on the CMB is related to the physical density of neutrinos, and therefore the mass difference between eigenstates can be neglected, individual neutrino masses, however, can have an effect on the large-scale shape of the matter power spectrum. Neutrinos of different masses have different transition redshifts from relativistic to non-relativistic behaviour, and their mass splitting changes the details of the radiation-domination to matter-domination regime. A precise measurement of the matter power spectrum shape can give information on both the sum of the masses and individual masses (and thus the hierarchy), even if the second effect is much smaller than the first [10].

In this work we shall utilise cosmological constraints imposed on the sum of the three neutrino masses, Σ. The notation we shall use to denote the three neutrino mass eigenstates is m_1, m_2, and m_3. In what is known as the normal mass hierarchy, these are defined in ascending order, $m_1 \leq m_2 \leq m_3$. For the case of the inverted hierarchy, the sequence becomes $m_3 \leq m_1 \leq m_2$.

Significant progress has been made by cosmological surveys to constrain the total mass of neutrinos. For example, the latest Planck satellite results show that CMB data alone constrain $\Sigma < 0.59$ eV (95% C.L.) [24], where the upper limit is stronger than ~ 1 eV because of the extra constraining power offered by secondary effects, mostly lensing. The inclusion of late-time information about the cosmic expansion history, such as that provided by measurements of the Baryon Acoustic Oscillations (BAO), further breaks cosmological parameters degeneracies yielding a limit $\Sigma < 0.17$ eV (95% C.L.).

Including information about the shape of the matter power spectrum further tightens these constraints, even when BAO information is not used. The most recent constraint on the total mass of neutrinos has been presented by Ref. [2] where the authors have carefully examined the impact of using different galaxy populations and tracers to estimate the power spectrum of the underlying dark matter and concluded that $\Sigma < 0.13$ eV at 95% confidence. This constraint tightens to $\Sigma < 0.11$ eV if also the latest measurement of the Hubble constant [25] are included. However since this measurement is in tension with CMB data within the context of a ΛCDM model [26], we will not consider it here. Ref. [2] result is consistent with the previous constraint from a different estimate of the power spectrum of dark matter using the Lyman-α forest [6], $\Sigma < 0.12$ eV. Since all these measurements are only upper limits for which the maxima of the probability distributions are indistinguishable from zero, unless

otherwise stated, we will assume that the probability distributions for Σ obtained from cosmological data are peaked at $\Sigma = 0$.

All the above constraints assume a standard cosmological model, where the spatial geometry is flat, dark matter (excluding the neutrino component) is cold, the dark energy is a cosmological constant, general relativity is the correct description of gravity on large, cosmological scales, the initial conditions of the Universe are given by a process like inflation and thus the primordial power spectrum is a nearly scale invariant power law. In addition, it is implicitly assumed that modelling of real world effects such as tracer's bias, non-linearities, etc. is sufficiently accurate as not to induce systematic errors that are larger than the reported ones. Relaxing any of these assumptions would weaken the constraints significantly e.g., [27, 28].

Detecting the effect of neutrino masses on cosmological structure and resolving the neutrino mass scale is well within the reach of up-coming cosmological surveys (e.g., [7, 8, 29] and references therein).

2.2 Neutrino Experiments

Solar, atmospheric, reactor and accelerator neutrino experiments have observed neutrino flavor conversion driven by neutrino masses. The experiments that mostly contribute to a precise measurement of the neutrino masses are sensitive to the relative phases acquired by relativistic neutrinos on their way from the production zone to the detector. Therefore the neutrino oscillation observables are sensitive to mass splittings, specifically to differences in the square of their masses. Matter effects allow the sign of the mass splittings to be distinguished, a goal which has been achieved in the small mass squared splitting but not the large one. The degeneracy in this second case distinguishes the two hierarchy mass orderings, normal and inverted. The identification the neutrino mass hierarchy is a key objective of future neutrino oscillation experiments, in addition to a determination of the complex phase involved in flavor conversion.

The absolute neutrino mass is not accessible via flavor conversion experiments, but it can be probed using beta decay. Experiments require sufficient mass and very good energy resolution in order to measure the beta spectrum close to the endpoint, where the neutrino energy is the neutrino mass. The experiments measure the effective mass of the electron, which is the combination of neutrino masses weighted by their content in the electron flavor. The present bound derived from the electron mass is 2.3 eV ($\Sigma < 6.9\,\mathrm{eV}$) at 90% confidence. The KATRIN experiment is expected to significantly improve upon this bound, reaching approximately 0.2 eV ($\Sigma < 0.6\,\mathrm{eV}$) at 90% [3]. Meanwhile new proposals [4] like Project 8 [5] aim for yet greater sensitivity.

2.3 Bayesian Model Selection

In this work we shall employ Bayesian model selection in order to quantify the manner in which the combination of existing data, from both cosmology and neutrino experiments, should influence our belief in the normal or inverted hierarchy. Each hypothesis possesses the same number of free parameters, facilitating a fair comparison. Furthermore their low dimensionality (three masses) permits an exact evaluation of the Bayesian evidence, a process which is often highly impractical for more complex models.

In general, mechanisms which lead us to prefer one model over another can be crudely expressed as a product of two contributing factors: the maximum likelihood value, and the so-called Occam factor. If one hypothesis is capable of fitting the data better than another, then it will yield a higher maximum likelihood, and this modifies our belief accordingly. Of greater significance in this work is the second contribution to the *marginal likelihood* –here and hereafter in the sense of Ref. [30], also sometime referred to as *evidence*–, which relates to the predictivity of the model. This concept was succinctly summarised by MacKay [30] as follows:

"The Occam factor is equal to [...] the factor by which $\mathcal{H}$'s hypothesis space collapses when the data arrive."

Therefore, in the context of the present work, if a much greater fraction of one hierarchy's three-dimensional parameter volume is excluded by the data, we should favour the other hierarchy.
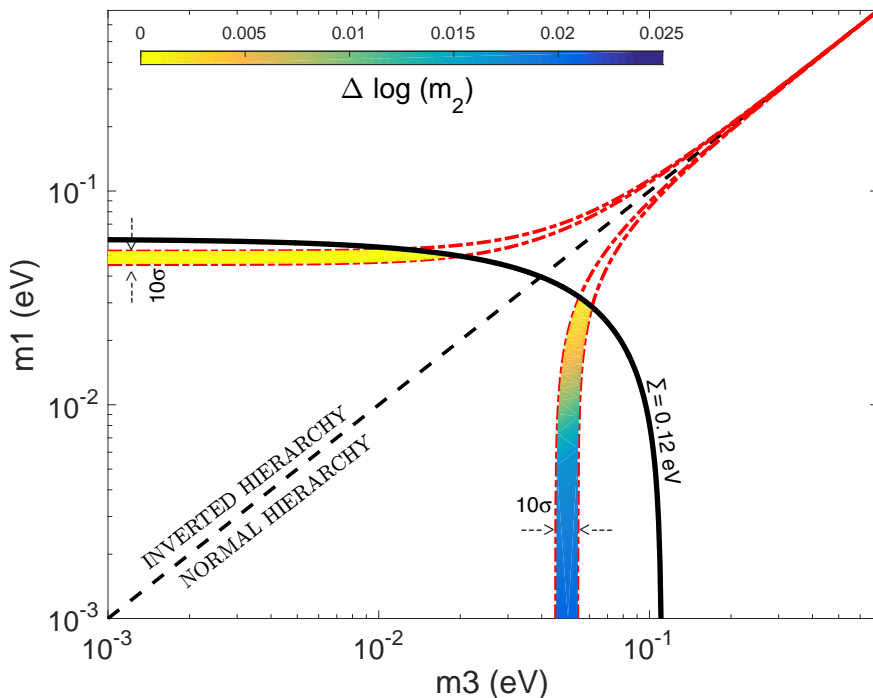
Figure 1. A visualisation of the heavily reduced parameter space available in the case of the inverted neutrino mass hierarchy, relative to the normal case. The red dash-dot contours illustrate constraints on the mass splittings, as imposed by neutrino oscillation experiments (broadened to show 10σ uncertainties for visualisation purposes). The solid black line corresponds to the combination of a cosmological upper bound on the sum of the neutrino masses $\Sigma < 0.12\,\text{eV}$ with the measurement of Δm_{12}^2. The diagonal dashed line demarcates the two hierarchies. The colouring of the shaded areas represents the amount of parameter space available in the third dimension, $\Delta(\log m_2)$.

3 Inference of the Neutrino Hierarchy

The Bayesian evidence for the two neutrino hierarchies has recently been considered by Ref. [11] and Ref. [12]. Here we shall argue that when dealing with parameters whose uncertainty spans several orders of magnitude, adopting a flat prior artificially favours the highest order of magnitude and may skew the interpretation of the results. We shall explore the impact of adopting a less informative prior, thus relaxing its effect on the determination of the hierarchy.

Where our work differs most significantly from others, however, is that we shall explicitly consider the full three-dimensional parameter space (m_1, m_2, m_3). Combining likelihoods in a lower dimensional space is a suboptimal procedure, and often leads to the likelihood of a model being greatly overestimated. For example, in a projected two dimensional space, two datasets may appear to possess large overlapping areas of high likelihood. Yet when viewed in their higher dimensional space, they may represent wholly disjoint volumes.

3.1 Data from Neutrino Oscillations

We shall make use of the following squared-mass splittings, as derived from observations of neutrino oscillations using NuFIT v3.0 (2016) [1]

$$m_2^2 - m_1^2 = 7.50\,(\pm 0.18) \times 10^{-5}\text{eV}^2\,, \tag{3.1}$$

$$m_3^2 - m_\ell^2 = \begin{cases} 2.524\,(\pm 0.040) \times 10^{-3}\text{eV}^2\,, & \text{(NH)} \\ -2.514\,(\pm 0.040) \times 10^{-3}\text{eV}^2\,, & \text{(IH)} \end{cases} \tag{3.2}$$

where m_ℓ denotes m_1 and m_2 for the normal (NH) and inverted (IH) hierarchies respectively, and we shall approximate the uncertainty distribution as a Gaussian.

Consider the three-dimensional volume of parameter space for our three neutrino masses: m_1, m_2, and m_3. The two pieces of data above resemble two thin slices cutting through our volume. The regions outside of these slices are excluded. The intersection of these two slices then forms two filaments, one falling on each side of the $m_1 = m_3$ plane. Each filament therefore represents the permissible regimes of what are known as the normal and inverted hierarchies. We shall seek to evaluate the posterior probability that a given filament contains the true set of masses. Before we can achieve this, we must first establish a consistent set of priors.

3.2 Three-dimensional Priors

Our prior on the three neutrino masses, $\pi(m_1, m_2, m_3)$, ought to reflect our state of belief before the data arrived. The conventional approach would be to assign an uninformative prior, and this corresponds to a uniform prior in each $\log m$. Care must be taken in monitoring the influence of the breadth of this prior, since in practice some hard bounds must be imposed. There are cases where the adoption of a linear prior may be justifiable, but this involves introducing information based on an understanding of the underlying distribution. For example, for the case of the cosmological constant (see e.g. [31]), a linear prior can be introduced if one is willing to assume that the bulk of the distribution $p(\Lambda)$ lies far above the value under consideration. Whether an equivalent assumption can be justified for the case of neutrino masses remains an open question.

An illustration of the scenario we face is given in Figure 1. The red dashed contours show the region allowed by oscillation data. They represent the two filaments discussed above, but now in projection. Their widths vary considerably since we are in the logarithmic domain (for visual clarity we use a broad 10σ contour). The complementarity of the cosmological data is highlighted by the solid black curve, which provides an upper bound on the sum of the three masses $\Sigma < 0.12$ eV, thereby truncating the extent of the contours at larger masses. Note that due to freedom in the value of m_2, the sum Σ is not uniquely defined in the $m_1 - m_3$ plane. What is of interest here is a contour of the surface $m_1 + m_2 + m_3 = 0.12$ eV which intersects the filaments. To create this we use the measurement of Δm_{12}^2 from equation (3.1) to specify m_2. From this perspective it becomes clear that even with only the oscillation data at hand, before looking at *any* cosmological data, we should not be surprised to discover that the sum of the masses is of the order 0.1 eV (i.e., towards the lower limit allowed by oscillation data). This is because towards very large masses, the uncertainty in the measurement corresponds to a vanishingly small interval in $\log m$. Therefore our posterior on the sum of the masses, Σ, is heavily penalised towards large masses. In stark contrast, if we had adopted an informative uniform prior (i.e., a uniform prior in m rather than $\log m$), we would have been repeatedly surprised as a sequence of modern cosmological surveys failed to unearth the influence of the neutrino masses.

The crucial factor here is the influence of the third mass parameter m_2, which has been marginalised out in Figure 1. The filaments vary in width in $\log m_2$, in much the same manner as can be seen in the other two dimensions. This evolution in depth, $\Delta \log m_2$, is visually represented by the colour of the shaded areas. In the regime of the normal hierarchy, and for the case of a vanishingly small m_1, the range of permissible m_2 values at any given (m_1, m_3) pair is governed by our uncertainty in the quantity Δm_{12}^2, of approximately 1.8×10^{-6} eV^2. Once we reach very low masses, $m_1 < 10^{-3}$ eV, then we settle at a minimum value of $m_2 \sim 0.0087$ eV. Thus the space available within the $\log m_2$ domain saturates at approximately 0.022. Let's compare this with the available parameter space in the inverted hierarchy. In this case, as we take the limit towards one of the neutrinos being massless, then we are left with $m_2 \sim 0.05$ eV. This corresponds to a much narrower range, approximately $\Delta \log m_2 \sim 6 \times 10^{-4}$. Thus the sheer volume of parameter space that our data has been able to exclude is highly asymmetric with respect to the $m_1 = m_3$ plane.

Our posterior belief in the two hypotheses, the inverted hierarchy and the normal hierarchy, relates to the integral of the posterior distribution over their respective areas. Provided our prior is sufficiently broad, each integral is dominated by the low mass regime, where the contours are at maximal thickness. Due to the saturation of $\Delta \log m_2$, our result becomes completely insensitive to the

breadth of our prior. The evidence ratio is well approximated by simply $\Delta_{\mathrm{NH}} \log m_2 / \Delta_{\mathrm{IH}} \log m_2 \simeq 32$. Thus we ought to favour the normal hierarchy over the inverted hierarchy with odds of $32:1$. This will strengthen further if cosmological bounds on Σ are imposed, as they penalise the inverted scenario more aggressively than the normal hierarchy, as Fig. 1 illustrates.

We are drawn to this conclusion because the oscillation data have eliminated a highly asymmetric region of parameter space in the logarithmic domain. This may not seem particularly intuitive, so it is instructive to consider similar scenarios where our posterior belief is driven by the volume of the available parameter space. One example mentioned earlier is the identifying the orientation of a concealed stick - this is straightforward to verify numerically. Another analogy which may be helpful to consider is the abundance of measurements whose numerical values start with a "1" or a "9". An intuitive response might be that they arise with equal frequency, but the reality is that there is almost a seven-fold difference in their respective abundances (in any field of science)[1]. The most straightforward interpretation of this phenomenon is, again, due to the availability of parameter space in the logarithmic domain.

3.3 Mass Degeneracy

It is interesting to consider a hypothetical special case: What would happen if Δm_{21}^2 had been perfectly measured to be identically zero? This approximation is often used in cosmology since the solar mass splitting has virtually no observational consequences for cosmological data. Then there is no third dimension to worry about in Figure 1. The value of m_2 is defined at each point in the plane, so the shaded area would all be the same colour. And we therefore no longer have any asymmetry between the two hierarchies if we neglect any cosmological information. Our conclusion within this hypothetical scenario would be that both hierarchies are equally probable. Adding a cosmological bound breaks the symmetry and, in particular shows a preference for NH with odds 1.8:1 when $\Sigma < 0.12$.

4   Hierarchical model

In the previous section we presented a pedagogical calculation illustrating how the Occam factor strongly favours the normal hierarchy. However the framework was a little too simplistic to allow robust conclusions to be drawn. This is most readily seen by considering a conditional slice through the prior of the form $\pi(m_1|m_2, m_3)$. One would presume that values of m_1 which are in close proximity to m_2 and m_3 are favoured, yet a simple broad prior does not account for this. A broad three-dimensional prior would be appropriate if the masses could be treated as independent, uncorrelated variables, but this is unlikely to be the case.

Rather than restrict ourselves to a single pre-determined prior, of fixed position and breadth, it is advantageous to work with a *hyperprior*. In practice we introduce an entire family of priors and effectively marginalises over them. Hyperpriors are more flexible and make the posterior less sensitive to the prior itself. We shall operate in a five-dimensional parameter space, consisting of the three neutrino masses (for brevity we shall denote the mass vector $\mathbf{m}$), and two *hyperparameters* μ and σ (see e.g.,[33–35]), as follows

$$p(D, \mathbf{m}, \mu, \sigma | \mathcal{H}) = p(D|\mathbf{m}, \mathcal{H}) p(\mathbf{m}|\mu, \sigma, \mathcal{H}) \pi(\mu, \sigma | \mathcal{H}) . \qquad (4.1)$$

Here D represents the data vector and $\mathcal{H}$ the hypothesis i.e., normal or inverted hierarchy, and $\pi(\mu, \sigma | \mathcal{H})$ is the hyperprior.

To reiterate, the prior ought to reflect our state of belief before the data arrived. The three masses are indistinguishable before the data arrives, so our prior *must* reflect this symmetry. This feature is also known as exchangeability. So we shall consider that three masses, m_i, m_j, m_k are drawn from this prior. Based on their relative sizes, these three masses are then assigned to m_1, m_2 and m_3, as defined earlier. In order to specify the position and breadth of the prior, we shall introduce the variables μ and σ which control the mean and standard deviation respectively. Note that this is

---

[1]This is known as Benford's law, which applies to many real-life sets of numerical data [32].

the prior on each mass *before any ordering takes place*. Thus the prior propagated through to m_1 will tend to prefer lighter masses values than the effective prior on m_3. Further we assume that

$$p(\log \mathbf{m}|\mu,\sigma) \sim \mathcal{N}(\mu,\sigma)\,, \tag{4.2}$$

where $\mathcal{N}$ denotes the normal, Gaussian, distribution. For the purposes of performing Bayesian model selection, the quantity which we seek to evaluate is the evidence, $p(D|\mathcal{H})$. The inference procedure may then be decomposed into the following tiers, as is characteristic of Bayesian hierarchical models.

$$p(D|\mathcal{H}) = \iint p(D|\mu,\sigma,\mathcal{H})\pi(\mu,\sigma)d\sigma d\mu\,, \tag{4.3}$$

$$p(D|\mu,\sigma,\mathcal{H}) = \int p(D|\mathbf{m},\mathcal{H})p(\mathbf{m}|\mu,\sigma,\mathcal{H})d\mathbf{m}\,. \tag{4.4}$$

Note that our analysis differs significantly from that of Ref. [11]. Although not stated explicitly, they appear to make the following identification for the case of the normal hierarchy (where m_1 takes on the role of the lightest neutrino),

$$p(D|m_1)\pi(m_1) = \iint p(D|\mathbf{m})\pi(\mathbf{m})\delta(m_2^2 - m_1^2 - \Delta m_{12}^2)\delta(m_3^2 - m_1^2 - \Delta m_{13}^2)dm_2 dm_3\,; \tag{4.5}$$

δ denotes the Dirac delta function. The relation above implicitly assumes that *informative* uniform priors are imposed upon both m_2 and m_3. This disproportionately favours masses with a value close to an imposed upper bound. To illustrate the potential issues which can arise with informative priors, imagine applying the same technique to other sets of particles within the Standard Model. How should we proceed if we did not know the absolute values of any quark or lepton masses, but were given approximate values of their mass splittings and the sums of their masses. In each case, our estimation of the up, down, and electron masses would be heavily biased, since they lie several orders of magnitude below their heavier counterparts. Likewise, in the current circumstance, the lightest neutrino could be several orders of magnitude below 0.1 eV. Theoretical considerations set the minimum viable neutrino mass at approximately 10^{-13} eV [36].

4.1 Methodology

Here we briefly outline the procedure we use to numerically evaluate the likelihood of the hyperparameters $p(D|\mu,\sigma,\mathcal{H})$.

1. For a given value of μ and σ, this defines a probability distribution from which we draw $N = 10^7$ sets of three samples. For each set of three, the masses are ranked from smallest to largest, here we shall designate these ranks as S, M, and L.

2. For each set we compute the following four quantities: $L^2 - M^2$; $L^2 - S^2$; $M^2 - S^2$; and $\Sigma \equiv S + M + L$.

3. In order to determine the likelihood of the two hyperparameters, we compute the inner integral of equation (4.3). For the case of the normal hierarchy this becomes

$$\begin{aligned} p(D|\mu,\sigma,\mathcal{H}_{NH}) &= \int p(D|\mathbf{m},\mathcal{H}_{NH})p(\mathbf{m}|\mu,\sigma,\mathcal{H}_{NH})d\mathbf{m} \\ &\simeq \frac{1}{N}\sum_{i=1}^{N} p(D_{L^2-M^2}|\mathbf{m}_i)p(D_{M^2-S^2}|\mathbf{m}_i)p(D_\Sigma|\mathbf{m}_i)\,. \end{aligned} \tag{4.6}$$

Here $\mathbf{m}_i$ denotes the mass vector of the ith random sample, while the D_x terms represent the different components of the data vector. For example, D_Σ represents the measurement of Σ as derived from cosmological density perturbations.
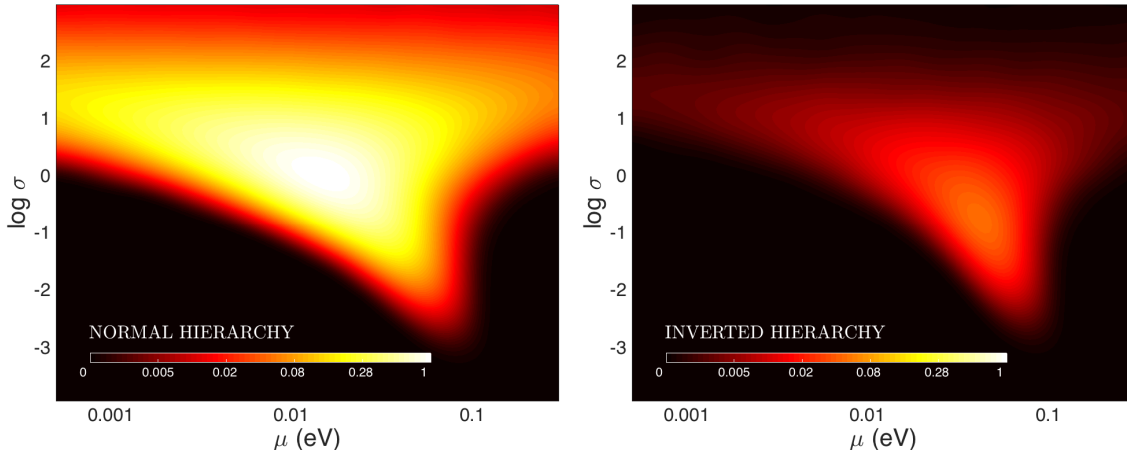
Figure 2. Marginal likelihoods of the hyperparameters μ and σ, as defined in equation (4.4). In each panel the same data is being used, namely the two measurements of the mass splittings as given in equations (3.2) and (3.1), and an upper bound is imposed on the sum of the neutrino masses $\Sigma < 0.15$ (95% C.L.). The left hand panel corresponds to the case of a normal neutrino mass hierarchy, while the right hand panel reflects the inverted ordering scheme. Note that a logarithmic colour scheme is employed in order to enhance the visibility of the right hand panel. The ratio of the two integrated regions yields a Bayes factor, K, of 31. Results for other values of Σ are presented in Table 1.

4. Repeat Step 3 for the case of the inverted hierarchy. This involves a reinterpretation of the squared mass splittings as follows

$$p(D|\mu, \sigma, \mathcal{H}_{IH}) \simeq \frac{1}{N} \sum_i p(D_{L^2-S^2}|\mathbf{m}_i)p(D_{L^2-M^2}|\mathbf{m}_i)p(D_\Sigma|\mathbf{m}_i) \,.$$

5. Repeat Steps 1-4 for 100 x 100 different values of μ and σ, across a logarithmically spaced grid. This yields the joint probability distribution shown in Figure 2. The prior is uniform across this region, therefore all of the visible structure has been imposed by the dataset.

In order to test our numerical precision, we repeated this procedure with a smaller number of samples, $N = 10^6$. This was found to modify the resultant Bayes factors by less than 1%.

5   Results

Figure 2 illustrates the two-dimensional marginal likelihood of the hyperparameters, $p(D|\mu, \sigma)$. The left hand panel corresponds to the case of the normal hierarchy. Low values of σ are disfavoured by the requirement of a significant mass splitting. Meanwhile very large values of σ are penalised as they are less predictive. Cosmological information provides the third boundary, eliminating the largest values of μ. Nonetheless a substantial area of parameter space retains a high likelihood. By contrast, in the right hand panel we see the region of parameter space associated with the inverted hierarchy. For a linear colour scale the features are barely visible, so here we introduce a logarithmic scale. The likelihood suffers both in terms of its peak value (the maximum is a factor of fifteen lower than in the normal hierarchy) and in terms of its confinement to a more restricted region of parameter space.

Another striking feature of Figure 2 is that ratio between the two panels, at any given point in the $\mu - \sigma$ plane, favours the NH. So we should anticipate that our results are not particularly sensitive to our choice of hyperprior, $\pi(\mu, \sigma)$. For comparison, in Fig. 3 we show the corresponding distribution without imposing the cosmological bound. The comparison of Figs. 2 and 3, clearly illustrates how cosmology has wiped out the space available for the inverted hierarchy.
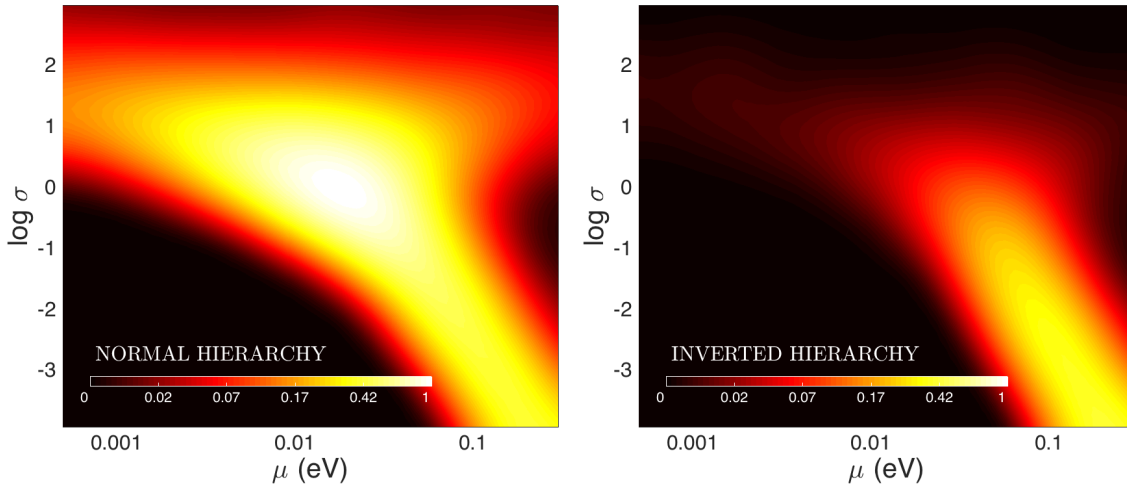
Figure 3. As Fig. 2 but without the cosmological upper bound imposed on the sum of the neutrino masses. Comparison with Fig. 2 shows how cosmology has wiped out the space available for the inverted hierarchy.

5.1 Bayesian Evidence

Our posterior belief in the two hypotheses, the normal and inverted orderings, is given by

$$\frac{p(\mathcal{H}_{NH}|D)}{p(\mathcal{H}_{IH}|D)} = \frac{p(\mathcal{H}_{NH})}{p(\mathcal{H}_{IH})}\frac{p(D|\mathcal{H}_{NH})}{p(D|\mathcal{H}_{IH})} . \tag{5.1}$$

In principle we are free to introduce a bias in the ratio $p(\mathcal{H}_{NH})/p(\mathcal{H}_{IH})$, but here we consider the prior probabilities of the two models to be equal. The important quantity in influencing the strength of our posterior belief is therefore the ratio of the marginal likelihoods, $p(D|\mathcal{H}_{NH})/p(D|\mathcal{H}_{IH})$, also known as the Bayes factor K. This quantity is evaluated using equation (4.3), while again employing an uninformative prior $\pi(\mu,\sigma)$.

In Table 1 we present the Bayesian evidence for a range of different degrees of precision for the sum of the neutrino masses. Hereafter, for the classification of the evidence, we refer to the slightly modified Jeffreys' scale [37] as discussed in [38]. The current robust limit $\Sigma < 0.13$ eV from [2] already provides odds of approximately 42:1, which translate into strong evidence in favour the normal hierarchy. Note that, if we instead use the Ly-α limit, $\Sigma < 0.12$ eV, the odds increase to 90:1.

5.2 Sensitivity Analysis

In order to explore our sensitivity to the choice of prior $\pi(\mu,\sigma)$, we shall now impose an informative prior on σ. Figure 4 shows the impact of adopting a hyperprior of the form $p(\sigma) \propto \sigma^{-2}$. This favours narrower distributions, which in turn slightly weakens the extent to which the NH is favoured over IH. For the illustrated example we set $\Sigma < 0.15$ eV, leaving us with odds of $17 : 1$. Meanwhile $\Sigma < 0.12$ eV returns odds of $33 : 1$.

We conducted a further sensitivity test by replacing the normal distribution with a top hat distribution, described by the same set of parameters μ and σ. This was found to have a minimal impact on the evidence ratios.

6 Discussion and conclusions

By performing a likelihood analysis in its full three-dimensional parameter space, we demonstrate that the inverted hierarchy suffers from a much more extensive exclusion of its parameter space than the normal hierarchy. This can be largely attributed to the extra freedom in the mass of m_2 in the logarithmic domain, which differs between the two hierarchies by a factor of 30. Despite some
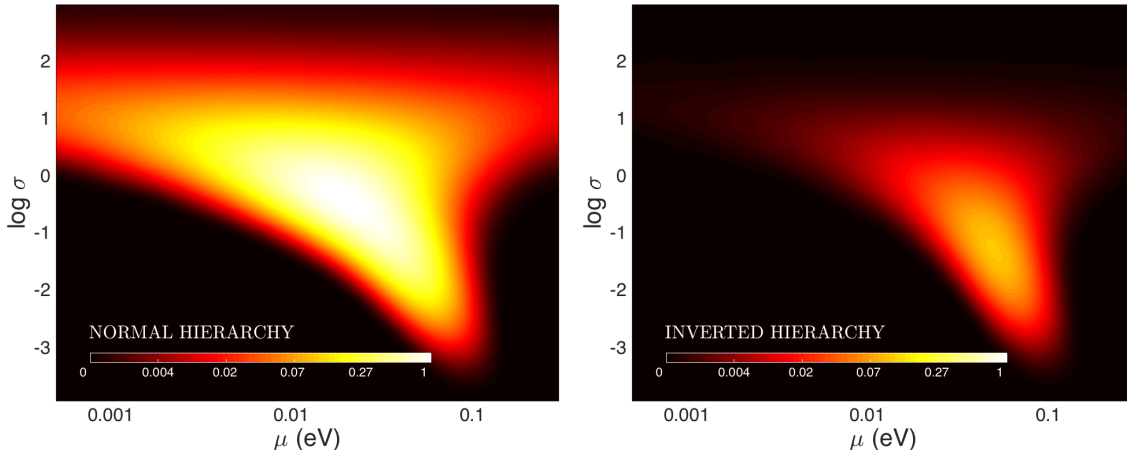
Figure 4. Exploring our sensitivity to the choice of hyperprior. The above panels are in a similar format to Fig. 2, but here we show the posterior distribution on the hyperparameters when imposing an informative prior of the form $\pi(\sigma) \propto \sigma^{-2}$. This enforces a bias in favour of closely clustered neutrino masses, which leads to a slight weakening of the evidence ratio, yielding odds of $17 : 1$ for $\Sigma < 0.15$ eV (95% C.L.), and $33 : 1$ for $\Sigma < 0.12$ eV (95% C.L.).

inevitable dependence on the adopted priors, we can draw some robust conclusions which have several broad implications.

6.1   Implications in neutrino physics

The determination of the nature of neutrinos, whether Majorana or Dirac fermions, is a major long-term unsolved problem in particle physics, with double beta decay experiments being the most sensitive probes to solve this riddle. These experiments could prove the existence of a rare double beta decay of nuclei without neutrino emission, a decay which is mediated by the internal exchange of the light majorana neutrinos (other beyond the standard model physics may also contribute to the decay). In the minimal case, the expected half-lives of neutrino-less double beta decays of nuclei depend on nuclear physics dynamics and on the neutrino mass scale and hierarchy. In fact, the neutrino mass hierarchy plays a key role in identifying the largest half-lives to be explored [39]. In our opinion, the results derived in this work are very relevant to the development of next generation double beta decay future experiments. While it is true that, independently of the neutrino mass hierarchy, the discovery could be close to the upper bounds, for example 10^{27} yr for Xe [40], the strong evidence for the normal neutrino mass hierarchy favours experimental techniques with potential to reach multi-ton active mass detectors and meet the requirements of very low background with, for example, topological signatures[41, 42].

to reach multi-ton active mass detectors and meet the requirements of very low background with, for example, topological signatures[].

Our results have also important implications for neutrino oscillation experiments and neutrino mass models. Neutrino mass hierarchy is among the physics goals of several proposed neutrino detectors [43–47]. Experiments whose major scientific goal is the identification of the neutrino mass hierarchy are particularly affected by the conclusions of this paper. In particular, experiments with observables more sensitive to normal mass hierarchy have better chances to confirm the Bayesian evidence. We also found a Bayesian preference for the lightest mass smaller than 5 meV. These results, light smallest mass and normal hierarchy, agree well with predictions derived by, among other models, minimal SO(10) grand unified models [48].

6.2   Interpreting results from upcoming surveys

Fig. 5 serves as guidance for the interpretation of forthcoming results from cosmology. In order to facilitate a comparison with conventional cosmological analyses, we present our results in terms of the

Table 1. Evidence ratios for the Normal and Inverted orderings and classification according to Jeffreys' [37, 38], for a range of hypothetical cosmological data. Fig. 2 illustrates the marginal likelihood surface for the case $\Sigma < 0.15$ eV. Since our prior odds are equal, then the Bayes factor K matches the quoted posterior odds.

95% Upper Bound on Σ (eV)	< 0.05	< 0.1	< 0.12	< 0.13	< 0.15	< 0.2
Odds (NH/IH)	6500:1	95:1	53:1	42:1	31:1	18:1
$\log K$	9	4.5	4.0	3.7	3.4	2.9
Classification	Very Strong	Strong	Strong	Strong	Strong	Positive

Table 2. The same as Table 1, but now for the scenario where $p(D_\Sigma|\Sigma)$, the marginal likelihood for Σ from cosmological data, peaks at 0.05 eV.

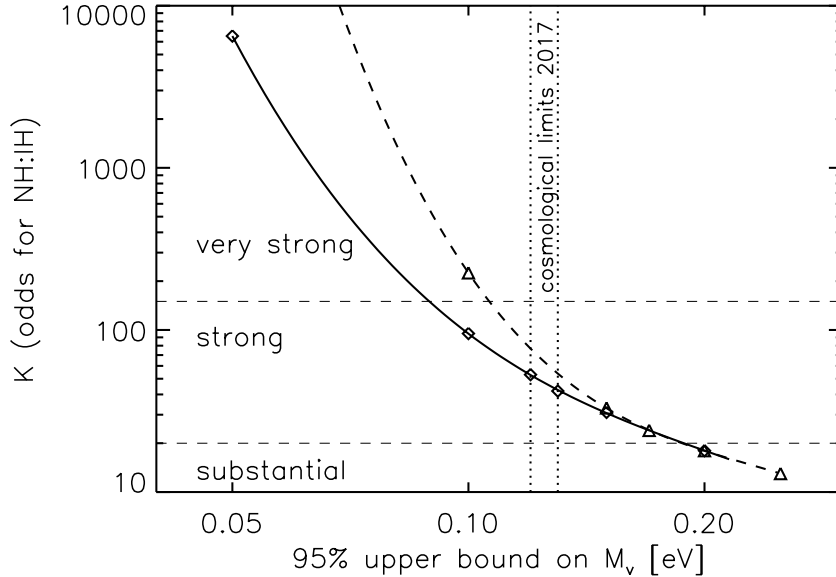| 95% Upper Bound on Σ (eV)	< 0.1	< 0.15	< 0.17	< 0.2	< 0.25
Odds (NH/IH)	225:1	33:1	24:1	18:1	13:1
$\log K$	5.4	3.5	3.2	2.9	2.5
Classification	Very Strong	Strong	Strong	Positive	Positive

Figure 5. Odds as a function of cosmological upper limits on the sum of neutrino masses Σ. The solid line corresponds to the (actual) case where the maximum of the Σ distribution is indistinguishable from zero. The dashed line correspond to a case where maximum of the Σ distribution is at 0.05 eV. The symbols correspond to the values reported in Tab. 1 and 2. Also indicated (vertical dotted lines) are the current limits from CMB and clustering of galaxies [2] or Lyman α forest [6]. Jeffreys' interpretation of the Bayes factor values are also reported.

95% confidence bounds which would be achieved when imposing a uniform prior on Σ, and without utilising external information from neutrino oscillation experiments. The solid line corresponds to the family of Gaussian likelihoods $p(D_\Sigma|\Sigma)$, of varying standard deviation but each with maxima located at 0 eV, as motivated by the findings of recent analyses. Meanwhile the dashed line demonstrates the impact of shifting the maximum likelihood value to 0.05 eV. These lines illustrate the degree of belief we would reach in the normal neutrino hierarchy, when this hypothetical cosmological data is combined with the squared mass splittings presented in Section 3.1. They were computed by interpolating the

values of Tab. 1, shown as symbols. What if these hypothetical future measurements provide instead a likelihood for Σ which is centred at a value close to the current 95% upper limit? In this case, if this value is e.g., $0.1\,\text{eV}$, the preference for normal hierarchy is much weaker but still a preference remains. For example for $1\text{-}\sigma$ errors of $\{0.005, 0.01, 0.025, 0.05\}\,\text{eV}$ corresponding to 95% upper limits of $\{0.11, 0.12, 0.15, 0.2\}\,\text{eV}$, one would obtain odds of $\{3.7, 5.3, 11, 12\}$ to 1 in favour of NH.

6.3 Conclusions

We have presented a joint analysis of data on the masses of the three neutrinos, using Bayesian hierarchical modelling. We find that the combination of laboratory and cosmological measurements already provide strong evidence in favour of a normal hierarchy.

The extent to which our belief in the inverted hierarchy diminishes is related to the collapse in its available parameter space. Within the context of the inverted hierarchy, only contrived combinations of neutrino masses remain compatible with the data. While this effect has been known to exist in a one-dimensional context, here we have demonstrated that the effect is significantly amplified when accounting for the full three-dimensional domain. Our results differ (quantitatively, not qualitatively) from those of Refs. [11, 12] who use an informative uniform prior and find that the odds for NH:IH are about 2:1 or 3:1, only, depending on the dataset combination. Such a prior tends to favour high masses. When dealing with parameters whose uncertainty spans several orders of magnitude, such as in this case, we argue that a less informative prior should be adopted.

In this case, due to the differences in parameter space, the first significant digit of a given neutrino mass is much more likely to be a '1' than a '9', in analogy with the Benford's law. This may at first seem unintuitive, yet it is incontrovertible. One does not require knowledge of the formation mechanism in order to make this statement. In a similar vein, the mass hierarchy is very much more likely to be normal rather than inverted.

Is there a means by which our conclusion can be evaded? As always, our posterior belief is to some extent influenced by our choice of prior. We have demonstrated that modest modifications to the prior impart only a minor impact on the evidence ratios. One could construct a more contrived set of priors which strongly favour inverted orderings. Or equivalently, one could amplify the prior on the hypothesis itself $p(\mathcal{H}_{IH})$. However there appears no clear justification for either of these actions.

References

[1] M. C. Gonzalez-Garcia, M. Maltoni, and T. Schwetz, JHEP **11**, 052 (2014), 1409.5439.

[2] A. J. Cuesta, V. Niro, and L. Verde, Physics of the Dark Universe **13**, 77 (2016), 1511.05983.

[3] A. Osipowicz *et al.* (KATRIN) (2001), hep-ex/0109033.

[4] B. Monreal and J. A. Formaggio, Phys. Rev. **D80**, 051301 (2009), 0904.2860.

[5] P. J. Doe *et al.* (Project 8), in *Proceedings, Community Summer Study 2013: Snowmass on the Mississippi (CSS2013): Minneapolis, MN, USA, July 29-August 6, 2013* (2013), 1309.7093, URL https://inspirehep.net/record/1255861/files/arXiv:1309.7093.pdf.

[6] N. Palanque-Delabrouille, C. Yèche, J. Baur, C. Magneville, G. Rossi, J. Lesgourgues, A. Borde, E. Burtin, J.-M. LeGoff, J. Rich, M. Viel, and D. Weinberg, JCAP **11**, 011 (2015), 1506.05976.

[7] C. Carbone, L. Verde, Y. Wang, and A. Cimatti, JCAP **1103**, 030 (2011), 1012.2868.

[8] B. Audren, J. Lesgourgues, S. Bird, M. G. Haehnelt, and M. Viel, JCAP **1301**, 026 (2013), 1210.2194.

[9] J. Hamann, S. Hannestad, and Y. Y. Y. Wong, JCAP **1211**, 052 (2012), 1209.1043.

[10] R. Jimenez, T. Kitching, C. Pena-Garay, and L. Verde, JCAP **1005**, 035 (2010), 1003.5918.

[11] S. Hannestad and T. Schwetz, JCAP **11**, 035 (2016), `1606.04691`.

[12] S. Vagnozzi, E. Giusarma, O. Mena, K. Freese, M. Gerbino, S. Ho, and M. Lattanzi, ArXiv e-prints (2017), `1701.08172`.

[13] J. Lesgourgues and S. Pastor, Phys. Rept. **429**, 307 (2006), `astro-ph/0603494`.

[14] J. R. Bond, G. Efstathiou, and J. Silk, Phys. Rev. Lett. **45**, 1980 (1980).

[15] J. Brandbyge and S. Hannestad, JCAP **1001**, 021 (2010), `0908.1969`.

[16] C. Wagner, L. Verde, and R. Jimenez, Astrophys. J. **752**, L31 (2012), `1203.5342`.

[17] S. Bird, M. Viel, and M. G. Haehnelt, Mon. Not. Roy. Astron. Soc. **420**, 2551 (2012), `1109.4416`.

[18] W. Hu, D. J. Eisenstein, and M. Tegmark, Physical Review Letters **80**, 5255 (1998), `astro-ph/9712057`.

[19] J. Lesgourgues, S. Pastor, and L. Perotto, PRD **70**(4), 045016 (2004), `hep-ph/0403296`.

[20] S. Wang, Z. Haiman, W. Hu, J. Khoury, and M. May, Physical Review Letters **95**(1), 011302 (2005), `astro-ph/0505390`.

[21] M. Takada, E. Komatsu, and T. Futamase, PRD **73**(8), 083520 (2006), `astro-ph/0512374`.

[22] S. Hannestad and Y. Y. Y. Wong, JCAP **7**, 004 (2007), `astro-ph/0703031`.

[23] T. D. Kitching, A. F. Heavens, L. Verde, P. Serra, and A. Melchiorri, PRD **77**(10), 103008 (2008), `0801.4565`.

[24] P. A. R. Ade *et al.* (Planck) (2015), `1502.01589`.

[25] A. G. Riess, L. M. Macri, S. L. Hoffmann, D. Scolnic, S. Casertano, A. V. Filippenko, B. E. Tucker, M. J. Reid, D. O. Jones, J. M. Silverman, R. Chornock, P. Challis, *et al.*, ApJ **826**, 56 (2016), `1604.01424`.

[26] L. Verde, P. Protopapas, and R. Jimenez, Physics of the Dark Universe **2**, 166 (2013), `1306.6766`.

[27] R. de Putter, E. V. Linder, and A. Mishra, Phys. Rev. **D89**(10), 103502 (2014), `1401.7022`.

[28] T. Namikawa, S. Saito, and A. Taruya, JCAP **1012**, 027 (2010), `1009.3204`.

[29] R. Jimenez, T. Kitching, C. Peña-Garay, and L. Verde, JCAP **5**, 035 (2010), `1003.5918`.

[30] D. J. MacKay, *Information theory, inference and learning algorithms* (Cambridge university press, 2003).

[31] S. Weinberg, Physical Review Letters **59**, 2607 (1987).

[32] B. Benford, Proc. Am. Philos. Soc **78**, 551572 (1938).

[33] G. C. Tiao and G. E. Box, The American Statistician **27**(1), 12 (1973).

[34] D. J. MacKay, in *Maximum entropy and Bayesian methods* (Springer, 1996), pp. 43–59.

[35] M. Hobson, S. Bridle, and O. Lahav, Monthly Notices of the Royal Astronomical Society **335**(2), 377 (2002).

[36] S. Davidson, G. Isidori, and A. Strumia, Physics Letters B **646**(2), 100 (2007).

[37] H. Jeffrey, *The theory of probability* (Oxford University Press, 1961).

[38] K. R. E. and R. A. E., Journal of the American Statistical Association **430**, 773 (1995).

[39] H. Murayama and C. Pena-Garay, Phys. Rev. **D69**, 031301 (2004), `hep-ph/0309114`.

[40] A. Gando *et al.* (KamLAND-Zen), Phys. Rev. Lett. **117**(8), 082503 (2016), [Addendum: Phys. Rev. Lett.117,no.10,109903(2016)], `1605.02889`.

[41] F. Granena *et al.* (NEXT) (2009), `0907.4054`.

[42] A. E. Chavarria, C. Galbiati, X. Li, and J. A. Rowlands (2016), `1609.03887`.

[43] K. Abe *et al.* (Hyper-Kamiokande Proto-Collaboration), PTEP **2015**, 053C02 (2015), `1502.05199`.

[44] R. Acciarri *et al.* (DUNE) (2015), `1512.06148`.

[45] F. An *et al.* (JUNO), J. Phys. **G43**(3), 030401 (2016), `1507.05613`.

[46] M. G. Aartsen *et al.* (IceCube) (2016), 1607.02671.

[47] S. Adrian-Martinez *et al.* (KM3Net), J. Phys. **G43**(8), 084001 (2016), 1601.07459.

[48] F. Buccella, M. Chianese, G. Mangano, G. Miele, S. Morisi, and P. Santorelli (2017), 1701.00491.