

## Neutrino mass priors for cosmology from random matrices

Andrew J. Long,<sup>1</sup> Marco Raveri,<sup>1,2</sup> Wayne Hu,<sup>1,3</sup> and Scott Dodelson<sup>4</sup>

<sup>1</sup>*Kavli Institute for Cosmological Physics, University of Chicago, Chicago, Illinois 60637, USA*

<sup>2</sup>*Institute Lorentz, Leiden University, PO Box 9506, Leiden 2300 RA, The Netherlands*

<sup>3</sup>*Department of Astronomy and Astrophysics, Enrico Fermi Institute,  
University of Chicago, Chicago, Illinois 60637, USA*

<sup>4</sup>*Department of Physics, Carnegie Mellon University, Pittsburgh, Pennsylvania 15312, USA*



(Received 1 December 2017; published 13 February 2018)

Cosmological measurements of structure are placing increasingly strong constraints on the sum of the neutrino masses,  $\Sigma m_\nu$ , through Bayesian inference. Because these constraints depend on the choice for the prior probability  $\pi(\Sigma m_\nu)$ , we argue that this prior should be motivated by fundamental physical principles rather than the *ad hoc* choices that are common in the literature. The first step in this direction is to specify the prior directly at the level of the neutrino mass matrix  $M_\nu$ , since this is the parameter appearing in the Lagrangian of the particle physics theory. Thus by specifying a probability distribution over  $M_\nu$ , and by including the known squared mass splittings, we predict a theoretical probability distribution over  $\Sigma m_\nu$  that we interpret as a Bayesian prior probability  $\pi(\Sigma m_\nu)$ . Assuming a basis-invariant probability distribution on  $M_\nu$ , also known as the anarchy hypothesis, we find that  $\pi(\Sigma m_\nu)$  peaks close to the smallest  $\Sigma m_\nu$  allowed by the measured mass splittings, roughly 0.06 eV (0.1 eV) for normal (inverted) ordering, due to the phenomenon of eigenvalue repulsion in random matrices. We consider three models for neutrino mass generation: Dirac, Majorana, and Majorana via the seesaw mechanism; differences in the predicted priors  $\pi(\Sigma m_\nu)$  allow for the possibility of having indications about the physical origin of neutrino masses once sufficient experimental sensitivity is achieved. We present fitting functions for  $\pi(\Sigma m_\nu)$ , which provide a simple means for applying these priors to cosmological constraints on the neutrino masses or marginalizing over their impact on other cosmological parameters.

DOI: [10.1103/PhysRevD.97.043510](https://doi.org/10.1103/PhysRevD.97.043510)

### I. INTRODUCTION

Measurements of the cosmic microwave background (CMB) [1,2] and large scale structure [3–5] are currently being used to constrain—and one day, hopefully to measure—the spectrum of neutrino masses. The two squared mass splittings are already known from terrestrial experiments [6,7], and cosmological probes are sensitive to the sum of the three neutrino masses,  $\Sigma m_\nu$ . Therefore a combination of terrestrial and cosmological measurements can be used to completely determine the neutrino mass spectrum.

When cosmological data is used to extract a limit on  $\Sigma m_\nu$ , the analysis is generally performed using Bayesian inference [8]. In this framework one is required to select a prior probability distribution  $\pi(\Sigma m_\nu)$ , which reflects one's *a priori* expectation for the sum of neutrino masses before the cosmological data is considered.

The assumed prior can have a dramatic impact on the inferred posterior probability distributions, not only for the neutrino mass itself but also for other cosmological parameters due to degeneracies. Until recently, cosmological measurements did not have enough sensitivity for the prior to matter much, and for simplicity most analyses set

$\Sigma m_\nu = 0$ . However, the enhanced precision of the Planck CMB data [1,2], meant that assuming  $\Sigma m_\nu = 0$  would lead to a shift in the value of the Hubble constant (that is a noticeable fraction of its errors) as compared to assuming  $\Sigma m_\nu \approx 0.06$  eV, which is the minimum value allowed by the terrestrial experiments. Thereafter, most analyses allowed for non-zero neutrino mass by simply shifting the delta function prior from  $\Sigma m_\nu = 0$  to  $\Sigma m_\nu \approx 0.06$  eV.

We are now moving into the era where analyses need to account for the reality that we do not know the value of  $\Sigma m_\nu$  by allowing it to vary in the fits. In choosing a prior for  $\Sigma m_\nu$  we have more options than simply the flat and logarithmic distributions; since there are three neutrino masses, there are other options that seem closer to the fundamental parameters, and these options may lead to dramatic differences in the inferred parameter constraints. For example, the authors of Ref. [9] advocate a Gaussian (hyper)prior on the logarithm of the individual masses, and they find that given the mass splittings, the prior alone strongly favors minimal masses, the normally ordered spectrum over the inverted one, and the impact on other cosmological parameters that these preferences imply (cf. Ref. [10]).

In Sec. II we review the importance of priors in establishing cosmological limits on  $\Sigma m_\nu$ . Even amongst commonly applied priors, the 95% confidence upper limit can vary by as much as 20%. More extreme logarithmic priors can lead to qualitatively different conclusions. Therefore, understanding the origin of the prior has emerged as an essential ingredient of any cosmological analysis.

Here we argue for a prior on the most fundamental parameters in the theory: the elements of the neutrino mass matrix  $M_\nu$  that appears in the Lagrangian. This approach follows in the footsteps of earlier work on the “anarchy hypothesis” or “anarchy principle” in studies of the neutrino flavor problem [11] (see also Refs. [12–21]). By specifying a probability distribution for  $M_\nu$ , we derive a probability distribution over the sum of neutrino masses  $\Sigma m_\nu$ . Since the structure of the mass terms in the Lagrangian depends on whether the neutrinos are Dirac or Majorana particles and whether the Majorana mass arises from the seesaw mechanism, the distribution over  $\Sigma m_\nu$  also depends on the physical origin for the neutrino masses.

On the one hand, this approach can be viewed as a phenomenological exercise. We are simply displacing the prior from  $\Sigma m_\nu$  to  $M_\nu$ . When our work is viewed in this way, the interesting question is whether the flat prior on  $\Sigma m_\nu$ , which is typically assumed in cosmological studies, arises from reasonable, basis-independent priors on  $M_\nu$ ; we will see that it does not. (To motivate the flat prior, note that it is the Jeffreys prior for a noise-dominated measurement).

There is an even more ambitious interpretation of this approach. It is reasonable to treat the neutrino mass matrix as random if the fundamental theory admits many vacua across which the neutrino mass spectrum varies. Then the prior on  $M_\nu$  reflects the uncertainty inherent in the underlying theory. The landscape of string theory vacua is a concrete example [22], but of course string theory is not developed to a point where the probability distribution over  $M_\nu$  can be calculated.

Nevertheless, if we understand that the random matrix description arises from some unknown fundamental theory, then we can restrict the form of the prior on  $M_\nu$  with some reasonable assumptions about the character of that theory. Following Ref. [12], the primary assumption in our work is that the probability distribution over  $M_\nu$  should be invariant under bi-unitary flavor transformations. In other words, if two matrices are related to one another by a change of basis, then they should be equally probable. Since this assumption is a specific implementation of the anarchy hypothesis [11], we will refer to this principle as the basis-independent anarchy hypothesis (BAH). In this work we do not consider models in which the neutrino mass matrix encodes an underlying symmetry structure, although such models are also motivated from fundamental physical principles, and we do not expect the results of our analysis to apply to these models as well. For an analysis of neutrino

mass priors in hierarchical mass models we refer the reader to Ref. [23].

We will see that the priors  $\pi(\Sigma m_\nu)$  resulting from these simple assumptions are strongly informative compared with current data. This can either be viewed as undesirable for a prior or can be interpreted as a testable target for future measurements. As a stark example, generically under the BAH, a detection of  $\Sigma m_\nu = 0.1$  eV would rule out the seesaw mechanism at approximately 99.7% CL significance.

This remainder of this article is organized as follows. We illustrate in Sec. II the importance of neutrino mass priors with a simple example. Section III introduces the reader to random matrices and explains how to calculate the probability distribution over  $\Sigma m_\nu$ . Section IV discusses how a specific probability distribution over the neutrino mass matrix  $M_\nu$  determines the probability distribution over  $\Sigma m_\nu$ . The paper concludes in Sec. V where we summarize our results and discuss the possible implications for cosmological studies and limits on  $\Sigma m_\nu$ .

## II. IMPORTANCE OF NEUTRINO MASS PRIORS

Several recent studies [9,10,24] (see also Refs. [25–31]) have investigated how much the prior on  $\Sigma m_\nu$  affects limits on cosmological parameters and odds ratios for normal versus inverted ordering. In Fig. 1 we show several recently-discussed priors. The top panel shows both a flat in  $\Sigma m_\nu$  (flat-linear) and a flat in  $\log(\Sigma m_\nu)$  (flat-log) prior in the range  $\Sigma m_\nu \in (0.06, 1)$  eV. If the likelihood is Gaussian in  $\Sigma m_\nu$  then the Jeffreys prior, which is calculated as the square root of the determinant of the Fisher information matrix, is flat-linear  $\Sigma m_\nu \in (0, \infty)$  eV [32,33] (cf. [9,24]).<sup>1</sup> The bottom panel shows priors on  $\Sigma m_\nu$  that arise when the distribution over the lightest neutrino mass  $m_L$  is assumed to be either flat-linear  $m_L \in (0, 0.33)$  eV or flat-log  $m_L \in (0.01, 0.33)$  eV. The other heavier neutrino masses are determined by fixing the squared mass splittings to their best-fit measured values, and the hierarchy is assumed to have either the normal or inverted ordering (NO or IO) where the larger splitting is between the heaviest or lightest neutrinos respectively; for additional details see Sec. III F. Note that in the flat-log case, the lower boundary is arbitrary since  $m_L$ , unlike  $\Sigma m_\nu$ , is not bounded by splitting data. We will return to this point below.

Let us now illustrate how much these prior choices affect the inferred limit on  $\Sigma m_\nu$  from a measurement  $\Sigma \hat{m}_\nu$ . We assume a Gaussian likelihood,  $\mathcal{L}(\Sigma \hat{m}_\nu | \Sigma m_\nu) = \mathcal{N} \exp[-(\Sigma \hat{m}_\nu - \Sigma m_\nu)^2 / (2\sigma^2)]$  where the normalization factor  $\mathcal{N}$

<sup>1</sup>The Jeffreys prior takes its shape from the constraining power of the data itself but is no more or less physically motivated than the other *ad hoc* priors. For example regions of parameter space where the data have no sensitivity are arbitrarily rejected by the prior. Note also, Jeffreys’ rule that a scale parameter should have a flat-log prior is not always in accord with the Jeffreys prior, especially for noise dominated measurements.

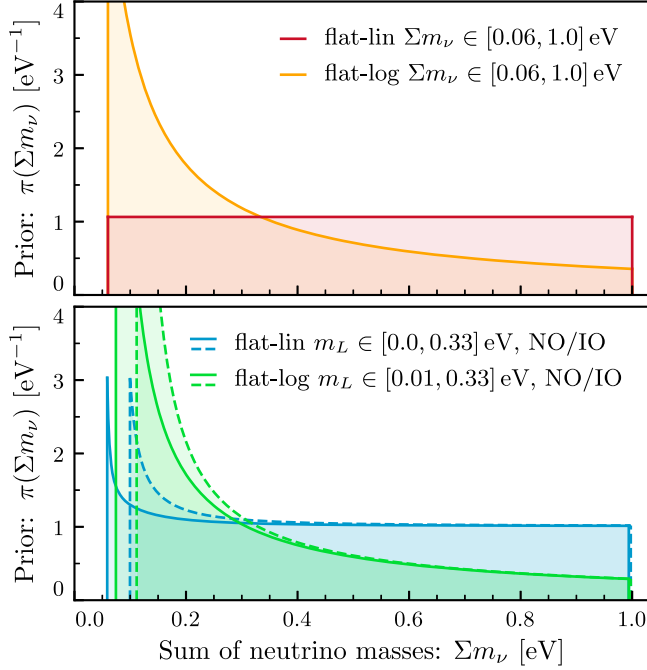


FIG. 1. *Ad hoc* choices for the prior probability on  $\Sigma m_\nu$ .

ensures that  $\int_{-\infty}^{\infty} d\Sigma \hat{m}_\nu \mathcal{L} = 1$ , and  $\sigma$  models the experimental uncertainty. We consider the six priors  $\pi(\Sigma m_\nu)$  that are shown in Fig. 1. Assuming a measurement  $\Sigma \hat{m}_\nu = 0$  and an uncertainty  $\sigma = 0.3$  eV, we calculate the posterior on  $\Sigma m_\nu$  as  $P(\Sigma m_\nu) = \mathcal{E}^{-1} \mathcal{L}(0 \text{ eV} | \Sigma m_\nu) \pi(\Sigma m_\nu)$  where the evidence  $\mathcal{E}$  ensures that  $\int_0^\infty d\Sigma m_\nu P = 1$ . The 95% CL region is estimated to be

$$\Sigma m_\nu / \text{eV} < 0.61, 0.46, 0.59, 0.60, 0.42, 0.43 \quad (1)$$

for the flat-linear prior on  $\Sigma m_\nu$ , the flat-log prior on  $\Sigma m_\nu$ , the flat-linear prior on  $m_L$  with NO, flat-linear prior on  $m_L$  with IO, flat-log prior on  $m_L$  with NO, and flat-log prior on  $m_L$  with IO. This exercise demonstrates that the 95% CL upper limit on  $\Sigma m_\nu$  varies by approximately 20% for the family of priors shown in Fig. 1.

In fact, even larger variations in the upper limit on  $\Sigma m_\nu$  can be obtained in principle. For the log prior on  $m_L$  the lower boundary of 0.01 eV is arbitrary, and it represents another *ad hoc* assumption—namely, that the lightest neutrino’s mass is not much smaller than the smaller mass splitting. However, suppose that we further decrease the lower boundary on  $m_L$ . Since most of the allowed prior volume is at small  $m_L$ , the prior probability increasingly favors the smallest possible value of  $\Sigma m_\nu$  that is consistent with the splitting data. For example if we take the limit to  $10^{-4}$  eV, the 95% CL upper limit on  $\Sigma m_\nu$  is strengthened by almost a factor of 2 from 0.42 eV to 0.26 eV for NO and from 0.43 eV to 0.30 eV for IO. This is similar to the result found in Ref. [9]. There if one assumes that the individual masses are Gaussian distributed in the log, the normal ordering is also strongly preferred. The difference between

our arbitrarily bounded flat case and the Gaussian case is that they then marginalize over the mean and width of the Gaussian so as to effectively place a lower bound that is comparable to the splittings.

In these more extreme examples, the prior assumptions qualitatively change the interpretation of the data. Without any physical principle to favor one prior over another, these changes simply represent *ad hoc* assumptions with no clear way to judge their merit or implications for neutrino physics. As such, the spread of values can be interpreted as a theoretical bias on upper limits for  $\Sigma m_\nu$  that is bounded from below only by the minimal values allowed by the splitting data.

This situation will change as the cosmological data improve beyond the  $10^{-1}$  eV level and begin to detect  $\Sigma m_\nu$ . Compared to the prior choices in Fig. 1, the data will become more informative than the prior, thereby reducing the 20% scatter. On the other hand the more extreme case of a flat prior in log  $m_L$  with an arbitrarily small lower cutoff would continue to strongly favor a minimal  $\Sigma m_\nu$ . If the data measure otherwise, we would then reject the theory that produced this prior. However to reiterate, without a physical motivation for such a prior, it is difficult to assess the implications of that rejection for neutrino physics.

Therefore in both the current and future contexts, it is important to make a stronger connection between the assumed priors and the neutrino physics that underly them. In particular we can ask the question whether any of the six *ad hoc* priors that were discussed above or the more extreme versions of the log  $m_L$  priors can be motivated by models of random neutrino mass matrices, which are more closely connected to models of neutrino mass at the Lagrangian level.

### III. RANDOM NEUTRINO MASS MATRICES

In this section we develop the formalism for lepton masses and random matrices. The reader will find Ref. [34] to be a useful resource in the mathematical theory of random matrices. For applications to lepton matrices we mostly follow the original work on the anarchy hypothesis, and the equations in this section will be familiar from Ref. [12].

#### A. Charged lepton and neutrino mass matrices

Let the mass matrix for the three flavors of charged leptons (electron, muon, and tau) be denoted by  $M_e$ . In a general basis,  $M_e$  need not be diagonal, but rather it is a 3-by-3 matrix with complex entries. We can diagonalize  $M_e$  with a singular value decomposition

$$M_e = U_e \hat{M}_e V_e^\dagger, \quad (2)$$

where  $U_e$  and  $V_e$  are unitary matrices and  $\hat{M}_e = \text{diag}(m_{e1}, m_{e2}, m_{e3})$  is the diagonal matrix of real and

non-negative singular values, which correspond to the physical particle masses.

The matrix  $M_e$  is specified by 18 real numbers, corresponding to the real and imaginary parts of each matrix element. Therefore each  $M_e$  represents a point in the 18-dimensional space  $\mathbb{R}^{18}$ , and the measure in this space is defined by

$$d^{18}M_e = \prod_{ij} d\text{Re}[(M_e)_{ij}] d\text{Im}[(M_e)_{ij}]. \quad (3)$$

The factorization in Eq. (2) specifies a change of coordinates from  $\text{Re}[M_e]$  and  $\text{Im}[M_e]$  to  $\hat{M}_e$ ,  $U_e$ , and  $V_e$ . In terms of these new variables, the measure is written as

$$d^{18}M_e = \mathcal{J}_e(\hat{M}_e) d^3\hat{M}_e d^9U_e d^9V_e / d^3\varphi_e \quad (4)$$

where

$$\begin{aligned} \mathcal{J}_e(\hat{M}_e) &= m_{e1}m_{e2}m_{e3}(m_{e1}^2 - m_{e2}^2)^2 \\ &\times (m_{e2}^2 - m_{e3}^2)^2(m_{e3}^2 - m_{e1}^2)^2 \end{aligned} \quad (5)$$

is the determinant of the Jacobian of the transformation (2). Note that  $\mathcal{J}_e(\hat{M}_e)$  only depends upon the singular values of  $M_e$ . The measure over the space of singular values is written as  $d^3\hat{M}_e = dm_{e1}dm_{e2}dm_{e3}$ . The measure over the space of unitary matrices, also known as the Haar measure, is denoted by  $d^9U_e$  for  $U_e$  and  $d^9V_e$  for  $V_e$ . The unitary matrices can be parametrized by 3 angles and 6 phases, and an explicit expression for the Haar measure in these coordinates appears in Ref. [12]. The SVD (2) is not uniquely defined, but rather it remains invariant under a  $U(1)^3$  transformation, which takes  $U_e \rightarrow U_e\Phi$  and  $V_e \rightarrow V_e\Phi$  where  $\Phi = \text{diag}(e^{i\varphi_{e1}}, e^{i\varphi_{e2}}, e^{i\varphi_{e3}})$ . Consequently, the measure (4) is modded out by  $d^3\varphi_e = d\varphi_{e1}d\varphi_{e2}d\varphi_{e3}$ .

Let us next consider the mass matrix for the three flavors of neutral leptons (electron, muon, and tau neutrinos), which is denoted as  $M_\nu$ . Currently the origin of neutrino mass remains an open question. In general, any mass generation mechanism falls into one of two categories, which have implications for the structure of  $M_\nu$ . Either the neutrinos and antineutrinos are distinct Dirac particles, implying that  $M_\nu$  is a 3-by-3 complex matrix, or the neutrinos and antineutrinos are identical Majorana particles, implying that  $M_\nu$  is a 3-by-3 complex and symmetric

matrix. The seesaw mechanism, which we discuss in Sec. III B, is one specific scenario that gives rise to massive Majorana neutrinos. We treat the seesaw model separately, because (as we discuss in Sec. III C) it is more natural to implement the probability distribution on the high-scale matrices rather than the low-energy neutrino matrix.

The symmetric mass matrix for Majorana neutrinos admits a Takagi decomposition, whereas the general mass matrix for Dirac neutrinos requires a singular value decomposition. Using these techniques, the neutrino mass matrix is factorized as

$$M_\nu = \begin{cases} U_\nu \hat{M}_\nu V_\nu^\dagger, & \text{Dirac} \\ U_\nu \hat{M}_\nu U_\nu^T, & \text{Majorana (e.g. seesaw)} \end{cases} \quad (6)$$

where  $U_\nu$  and  $V_\nu$  are unitary matrices and  $\hat{M}_\nu = \text{diag}(m_{\nu1}, m_{\nu2}, m_{\nu3})$  is the diagonal matrix of real and non-negative singular values, which correspond to the physical neutrino masses.

Having diagonalized both the charged and neutral lepton mass matrices (2) and (6), the lepton mixing matrix (PMNS matrix) is constructed as

$$U_{\text{PMNS}} = U_e^\dagger U_\nu, \quad (7)$$

which quantifies the mismatch between the unitary transformations that diagonalize  $M_e$  and  $M_\nu$ .

The space of matrices  $M_\nu$  is 18-dimensional if the neutrinos are Dirac particles, but it is only 12-dimensional if the neutrinos are Majorana particles, because the Majorana condition requires  $M_\nu$  to be symmetric. Thus the measure over  $M_\nu$  is defined by

$$\begin{cases} d^{18}M_\nu = \prod_{ij} d\text{Re}[(M_\nu)_{ij}] d\text{Im}[(M_\nu)_{ij}], & \text{D} \\ d^{12}M_\nu = \prod_{i \leq j} d\text{Re}[(M_\nu)_{ij}] d\text{Im}[(M_\nu)_{ij}], & \text{M} \end{cases} \quad (8)$$

The decompositions in Eq. (6) provide a change of coordinates, and the measure becomes

$$\begin{cases} d^{18}M_\nu = \mathcal{J}_\nu(\hat{M}_\nu) d^3\hat{M}_\nu d^9U_\nu d^9V_\nu / d^3\varphi_\nu, & \text{D} \\ d^{12}M_\nu = \mathcal{J}_\nu(\hat{M}_\nu) d^3\hat{M}_\nu d^9U_\nu, & \text{M} \end{cases} \quad (9)$$

where the Jacobian determinant is

$$\mathcal{J}_\nu(\hat{M}_\nu) = m_{\nu1}m_{\nu2}m_{\nu3} \times \begin{cases} (m_{\nu1}^2 - m_{\nu2}^2)^2(m_{\nu2}^2 - m_{\nu3}^2)^2(m_{\nu3}^2 - m_{\nu1}^2)^2, & \text{D} \\ |m_{\nu1}^2 - m_{\nu2}^2||m_{\nu2}^2 - m_{\nu3}^2||m_{\nu3}^2 - m_{\nu1}^2|, & \text{M} \end{cases} \quad (10)$$

the measure over singular values is  $d^3\hat{M}_\nu = dm_{\nu1}dm_{\nu2}dm_{\nu3}$ , and the measure over unitary matrices is defined in the same way as for  $M_e$  [see the text below Eq. (5)]. The factors of  $m_{\nu i}^2 - m_{\nu j}^2$  correspond to the phenomenon of eigenvalue repulsion, which probabilistically disfavors degeneracy; this effect is more pronounced in the Dirac model than the Majorana model.

### B. Majorana masses from the seesaw mechanism

One compelling scenario for the generation of a Majorana mass is the seesaw mechanism [35–40]. The Type-I seesaw model contains a 3-by-3 complex matrix  $M_D$  and a 3-by-3 complex, symmetric matrix  $M_M$ . In the seesaw regime where the singular values of  $M_M$  are much larger than the singular values of  $M_D$ , the neutrino mass matrix is well-approximated by

$$M_\nu = -M_D M_M^{-1} M_D^T. \quad (11)$$

Note that  $M_\nu$  is a 3-by-3 complex, symmetric matrix. The measures  $d^{18}M_D$  and  $d^{12}M_M$  are defined in analogy with the Dirac and Majorana cases of  $M_\nu$  (8). Eigenvalue repulsion in  $M_\nu$  is also strong in the seesaw case as demonstrated empirically in Ref. [12].

### C. Probability over matrices

Now we promote  $M_e$ ,  $M_\nu$ ,  $M_D$ , and  $M_M$  to be random matrices. The differential probability measure is written as

$$d\mathbb{P} = \begin{cases} f(M_e, M_\nu) d^{18}M_e d^{18}M_\nu, & \text{D} \\ f(M_e, M_\nu) d^{18}M_e d^{12}M_\nu, & \text{M} \\ F(M_e, M_D, M_M) d^{18}M_e d^{18}M_D d^{12}M_M, & \text{S} \end{cases} \quad (12)$$

which define the Dirac (D), Majorana (M), and seesaw (S) models, respectively. The probability densities,  $f$  and  $F$ , are real-valued, positive, integrable, and normalized such that  $\int d\mathbb{P} = 1$ .

Since the seesaw model is also a model of Majorana neutrinos, it may seem redundant to distinguish these cases. There are various models in which high-scale lepton-number violation gives rise to a Majorana neutrino mass matrix, but since the seesaw model is a particularly compelling and well-studied scenario, we consider this model separately. If the Majorana mass does arise from the seesaw mechanism, then it is possible to derive  $f(M_e, M_\nu)$  from  $F(M_e, M_D, M_M)$ . To make this connection concrete, first invert the seesaw relation (11) to determine the Majorana mass matrix as  $M_M = -M_D^T M_\nu^{-1} M_D$ , which defines a transformation from  $M_M$  to  $M_\nu$ . Let the Jacobian of this transformation be denoted as  $\mathcal{J}_{M_\nu}(M_D, M_\nu)$ , such that  $d^{18}M_D d^{12}M_M = \mathcal{J}_{M_\nu}(M_D, M_\nu) d^{18}M_D d^{12}M_\nu$ . Marginalizing over the unobservable matrix  $M_D$  yields

$$f(M_e, M_\nu) = \int d^{18}M_D \mathcal{J}_{M_\nu}(M_D, M_\nu) \times F(M_e, M_D, -M_D^T M_\nu^{-1} M_D), \quad (13)$$

which is the same probability density that appears in the Majorana case of Eq. (12). In practice, we find that it is easier to study the seesaw model by sampling from the  $M_D$

and  $M_M$  matrices directly rather than evaluating the Jacobian and integrals that appear in Eq. (13).

### D. Basis-independent anarchy hypothesis

The basis-independent anarchy hypothesis (BAH) provides guidance in selecting the probability densities  $f$  and  $F$  that appear in Eq. (12). In this section we clarify the meaning and implications of the BAH. Specifically, we show how the Lagrangian transforms under a basis-changing flavor transformation, and we require the probability distribution over matrices to be invariant under this transformation. A reader who is more interested in the phenomenological implications may choose to skip this section.

In writing the Lagrangian, it is convenient to work with the 2-component spinor notation [41]. The six charged leptons (electron, muon, tau, and antiparticles) are represented by the six fields  $(e_L)_i$  and  $(e_R)_i$  for  $i = 1, 2, 3$ , and the neutrinos are represented by the fields  $(\nu_L)_i$  and  $(\nu_R)_i$ . The subscripts  $L$  and  $R$  indicate left- and right-chiral Weyl spinors, respectively. The masses and weak interactions of these fields are encoded in the Lagrangian<sup>2</sup>

$$\mathcal{L} \supset \frac{g}{\sqrt{2}} \nu_L^\dagger \bar{\sigma}^\mu e_L W_\mu^+ - e_L^\dagger M_e e_R + \mathcal{O}_\nu + \text{H.c.} \quad (14)$$

where the vector field  $W_\mu^+$  represents the two charged  $W$ -bosons, and where

$$\mathcal{O}_\nu = \begin{cases} -\nu_L^\dagger M_\nu \nu_R, & \text{D} \\ -\frac{1}{2} \nu_L^T M_\nu \nu_L, & \text{M} \\ -\nu_L^\dagger M_D \nu_R - \frac{1}{2} \nu_R^T M_M \nu_R, & \text{S} \end{cases} \quad (15)$$

are the neutrino mass terms. In the seesaw model, integrating out the heavy fields  $\nu_R$  gives rise to a Majorana mass matrix  $M_\nu$  for the light neutrinos, which is of the seesaw form (11),  $M_\nu = -M_D M_M^{-1} M_D^T$ .

Now consider a basis-changing flavor transformation. The fields in the new basis (primed) can be expressed as

$$e'_L = U_{e_L} e_L, \quad \nu'_L = U_{\nu_L} \nu_L, \quad e'_R = U_{e_R} e_R, \\ \nu'_R = U_{\nu_R} \nu_R, \quad \text{and} \quad W_\mu^{+'} = W_\mu^+ \quad (16)$$

where  $U_{e_L}$ ,  $U_{\nu_L}$ ,  $U_{e_R}$ , and  $U_{\nu_R}$  are unitary matrices ( $UU^\dagger = 1$ ). The mass terms in Eq. (14) take the same form in the new basis provided that we identify the mass matrices in the new basis (primed) as

<sup>2</sup>The mapping onto 4-component spinor notation is explained in Ref. [41]. For instance,  $\nu_L^\dagger \bar{\sigma}^\mu e_L W_\mu^+ + \text{H.c.} = \bar{\Psi}_\nu \gamma^\mu P_L \Psi_e W_\mu^+ + \text{H.c.}$  and  $e_L^\dagger M_e e_R + \text{H.c.} = \bar{\Psi}_e M_e \Psi_e$  and  $\nu_L^T M_\nu \nu_L + \text{H.c.} = \bar{\Psi}_\nu^c M_\nu \Psi_\nu$ .

$$\begin{aligned}
M'_e &= U_{e_L} M_e U_{e_R}^\dagger, & M'_D &= U_{\nu_L} M_D U_{\nu_R}^\dagger, \\
M'_M &= U_{\nu_R}^* M_M U_{\nu_R}^\dagger, \\
\text{and } M'_\nu &= \begin{cases} U_{\nu_L} M_\nu U_{\nu_R}^\dagger, & \text{D} \\ U_{\nu_L} M_\nu U_{\nu_L}^T, & \text{M} \\ U_{\nu_L} M_\nu U_{\nu_L}^T, & \text{S} \end{cases} \quad (17)
\end{aligned}$$

The weak interaction [first term in Eq. (14)] is not invariant in general, but instead it transforms to  $(g/\sqrt{2})(\nu'_L)^\dagger \bar{\sigma}^\mu (U_{\nu_L} U_{e_L}^\dagger) e'_L W_\mu^+$ .

The physical motivation for the anarchy hypothesis is that some unspecified high energy physics provides an ensemble of theories, i.e. vacua, across which the lepton masses vary. The principle of basis-independence further asserts that mass matrices related by a basis-changing flavor transformation should be equally probable. In other words, basis-independence restricts the allowed form of the probability densities  $f(M_e, M_\nu)$  and  $F(M_e, M_D, M_M)$  that we introduced in Eq. (12). In general there are various ways to implement basis independence:

**BAH:** Require the probability densities to satisfy  $f(M_e, M_\nu) = f(M_e, M'_\nu)$  and  $F(M_e, M_D, M_M) = F(M_e, M'_D, M'_M)$ , which enforces basis-independence only on the neutrino mass matrices.

**B2:** Require the probability densities to satisfy  $f(M_e, M_\nu) = f(M'_e, M'_\nu)$  and  $F(M_e, M_D, M_M) = F(M'_e, M'_D, M'_M)$  where the primed matrices are given by Eq. (17) with  $U_{e_L} = U_{\nu_L}$ .

**B3:** Require the probability densities to satisfy  $f(M_e, M_\nu) = f(M'_e, M'_\nu)$  and  $F(M_e, M_D, M_M) = F(M'_e, M'_D, M'_M)$  for arbitrary  $U_{e_L}$  and  $U_{\nu_L}$ .

The basis-independent anarchy hypothesis of Ref. [12] is denoted by BAH, whereas B2 and B3 correspond to alternative strategies for enforcing basis independence.

As argued in Ref. [11] the large lepton mixing angles and the comparable neutrino masses does not point to any underlying (flavor symmetry) structure in the neutrino sector. This motivated the authors of Ref. [11] to propose the anarchy hypothesis. In the charged lepton sector, on the other hand, the masses are very hierarchical ( $m_e : m_\mu : m_\tau = 1 : 200 : 3500$ ), which suggests that a different physical mechanism is at play, e.g. the Froggatt-Nielsen mechanism. Hence, basis independence is imposed only on the neutrino mass matrix through the BAH, rather than on both lepton mass matrices as with either B2 or B3.

The BAH requires that the probability densities are constructed from basis-invariant functions of  $M \in \{M_D, M_M, M_\nu\}$  that include the determinant  $d = |\det M|$  and the traces  $t_n = \text{Tr}[(M^\dagger M)^n]$ . Note that these functions only depend upon the singular values of the corresponding matrix. If we use these functions to build the probability densities, it follows that the distributions over angular variables, such as the phases and mixing angles in  $U_{\text{PMNS}}$ , are simply given by the Haar measure. In this way, the

assumption of basis-independence leads to a powerful prediction for the Pontecorvo-Maki-Nakagawa-Sakata (PMNS) matrix parameters [12].

Let us close this section by remarking upon the distinction between BAH, B2, and B3. If the stochastic nature of the mass matrices arises from high energy physics above the weak scale, then the interactions of the leptons should respect  $SU(2)_L$  gauge invariance. This is implemented by combining  $e_L$  and  $\nu_L$  into the lepton doublet  $L = (\nu_L, e_L)$  and requiring them to transform in the same way under flavor transformations. Hence in this context it is only reasonable to expect  $f$  and  $F$  to be invariant under flavor transformations that obey  $U_{e_L} = U_{\nu_L}$ , i.e. functions satisfying B2. The set of functions that satisfy B3 also satisfy BAH and B2. In fact BAH and B2 imply B3 but not all functions that satisfy B2 also satisfy BAH, for instance  $\text{Tr}[M_e M_e^\dagger M_\nu M_\nu^\dagger]$ . Upon factorizing  $M_e$  and  $M_\nu$  with Eqs. (2) and (6), we can write this function as  $\text{Tr}[\hat{M}_e^2 U_{\text{PMNS}} \hat{M}_\nu^2 U_{\text{PMNS}}^\dagger]$ , which depends on both the singular values ( $\hat{M}_e$  and  $\hat{M}_\nu$ ) and the basis-invariant PMNS matrix (7). Thus if we were to enforce basis independence with B2 rather than the BAH, we would find that the probability density may depend directly on the PMNS matrix. However, we chose instead to enforce basis independence with the BAH, because the hierarchical nature of the charged lepton masses points to an underlying symmetry principle for their mass generation, as we have already discussed above.

## E. Probability densities after BAH

Following Refs. [11,12] let us now adopt the BAH assumption to write a simplified expression for the probability density  $f(M_e, M_\nu)$  that appears in Eq. (12). In this section we focus on only the Dirac and Majorana models, since the corresponding expressions for the seesaw model are more cumbersome and less illuminating. The BAH implies that the distribution over the neutrino mass matrix is independent from the distribution over the charged lepton matrix, and it follows that  $f(M_e, M_\nu)$  can be written as a sum of factorizable components. We focus on the simplest case for which

$$f(M_e, M_\nu) = p_e(M_e) p_\nu(\hat{M}_\nu) \quad (18)$$

where  $p_\nu$  only depends on  $M_\nu$  through the singular values  $\hat{M}_\nu$ . Then upon using the matrix measure decomposition in Eq. (9) the probability measure (12) becomes  $d^{\mathbb{P}} = d^{\mathbb{P}_e} d^{\mathbb{P}_\nu}$  where  $d^{\mathbb{P}_e} = p_e(M_e) d^{18} M_e$  and

$$d^{\mathbb{P}_\nu} = \begin{cases} p_\nu(\hat{M}_\nu) \mathcal{J}_\nu(\hat{M}_\nu) d^3 \hat{M}_\nu d^9 U_\nu d^9 V_\nu / d^3 \varphi_\nu, & \text{D} \\ p_\nu(\hat{M}_\nu) \mathcal{J}_\nu(\hat{M}_\nu) d^3 \hat{M}_\nu d^9 U_\nu, & \text{M} \end{cases} \quad (19)$$

The  $d^{\mathbb{P}}$ 's are separately normalized such that  $\int d^{\mathbb{P}} = 1$ . Noting that  $d^{18} M_e \supset d^9 U_e$  and  $d^9 U_e d^9 U_\nu = d^9 U_e d^9 U_{\text{PMNS}}$ , we see that the distribution over the PMNS angles and

phases is simply given by the Haar measure  $d^9 U_{\text{PMNS}}$  [12]. Since we are primarily interested in the distribution over masses, we integrate over the angular variables to obtain

$$d\mathbb{P}_\nu = \begin{cases} p_\nu(\hat{M}_\nu) \mathcal{J}_\nu(\hat{M}_\nu) d^3 \hat{M}_\nu \mathcal{V}_2, & \text{D} \\ p_\nu(\hat{M}_\nu) \mathcal{J}_\nu(\hat{M}_\nu) d^3 \hat{M}_\nu \mathcal{V}_1, & \text{M} \end{cases} \quad (20)$$

where  $\mathcal{V}_1 = 2\pi^6/3$  and  $\mathcal{V}_2 = \pi^9/3$  are the volumes of the compact angular spaces.

### F. Summed mass and squared splittings

The neutrino mass spectrum is not known, but rather the sum of the masses,  $\Sigma m_\nu = m_{\nu_1} + m_{\nu_2} + m_{\nu_3}$ , has been constrained, and the squared mass splittings have been measured. Conventionally the mass splittings are defined as  $\delta m^2 = m_{\nu_2}^2 - m_{\nu_1}^2$  and  $\Delta m^2 = m_{\nu_3}^2 - (m_{\nu_1}^2 + m_{\nu_2}^2)/2$  where the three mass eigenvalues are identified by the flavor content of the corresponding neutrino eigenstate (e.g., conventionally  $m_{\nu_3}$  is the mass of the neutrino with the smallest  $\nu_e$  content). However, as discussed in Sec. III C, we are interested in basis-independent probability measures over the space of neutrino mass matrices that lead to probability measures over the mass eigenvalues that are independent of the mixing angles (flavor composition). In other words,  $m_{\nu_1}$ ,  $m_{\nu_2}$ , and  $m_{\nu_3}$  have the same statistics. Consequently, the data cannot be implemented as  $\delta m^2 = m_{\nu_2}^2 - m_{\nu_1}^2 \sim 10^{-5} \text{ eV}^2$ , for instance, since this expression singles out  $\nu_2$  and  $\nu_1$ . In fact, the data tells us that the splitting between *any two* masses must be  $10^{-5} \text{ eV}^2$ ; it could be  $m_{\nu_2}^2 - m_{\nu_1}^2 \sim 10^{-5} \text{ eV}^2$ , but it could also be  $m_{\nu_1}^2 - m_{\nu_2}^2 \sim 10^{-5} \text{ eV}^2$  or  $m_{\nu_3}^2 - m_{\nu_1}^2 \sim 10^{-5} \text{ eV}^2$  just as well.

To identify a basis-independent parametrization of the spectrum, let us begin by defining  $m_H$ ,  $m_M$ , and  $m_L$  to be the masses of the heaviest, medium, and lightest neutrinos, respectively. These masses are basis-independent functions of the spectrum. Next let us define

$$\Sigma m_\nu = m_L + m_M + m_H \quad (21a)$$

$$\Delta m_{HM}^2 = m_H^2 - m_M^2 \quad (21b)$$

$$\Delta m_{ML}^2 = m_M^2 - m_L^2 \quad (21c)$$

where  $\Delta m_{HM}^2$  ( $\Delta m_{ML}^2$ ) is the squared mass splitting between the two heavier (lighter) mass eigenstates. It is useful to define  $h = \text{sign}[\Delta m_{HM}^2 - \Delta m_{ML}^2]$  that lets us distinguish the two scenarios

$$\text{NO: } h = +1, \quad \Delta m_{HM}^2 > \Delta m_{ML}^2 \quad (22a)$$

$$\text{IO: } h = -1, \quad \Delta m_{HM}^2 < \Delta m_{ML}^2, \quad (22b)$$

which define normal ordering (smaller splitting on bottom) and inverted ordering (smaller splitting on top).

To connect with measurements of the neutrino mass spectrum, it is useful to introduce the squared mass splittings  $\delta m^2$  and  $|\Delta m^2|$ . Let us define,

$$\delta m^2 = \text{Min}[\Delta m_{HM}^2, \Delta m_{ML}^2] \quad (23a)$$

$$|\Delta m^2| = \text{Max}[\Delta m_{HM}^2, \Delta m_{ML}^2] + \delta m^2/2 \quad (23b)$$

such that  $\delta m^2$  is the smaller of the two squared mass splittings, and  $|\Delta m^2|$  is the larger of the two squared mass splittings plus half the smaller splitting. For example, consider a spectrum for which  $m_{\nu_1} < m_{\nu_2} < m_{\nu_3}$  and the smaller splitting is on the bottom,  $m_{\nu_2}^2 - m_{\nu_1}^2 < m_{\nu_3}^2 - m_{\nu_2}^2$ . In this case, we can write  $\delta m^2 = m_{\nu_2}^2 - m_{\nu_1}^2$  and  $|\Delta m^2| = m_{\nu_3}^2 - (m_{\nu_1}^2 + m_{\nu_2}^2)/2$ . Alternatively, consider a spectrum  $m_{\nu_3} < m_{\nu_1} < m_{\nu_2}$  with the smaller splitting on the top,  $m_{\nu_2}^2 - m_{\nu_1}^2 < m_{\nu_2}^2 - m_{\nu_3}^2$ . Then,  $\delta m^2 = m_{\nu_2}^2 - m_{\nu_1}^2$  and  $|\Delta m^2| = |m_{\nu_3}^2 - (m_{\nu_1}^2 + m_{\nu_2}^2)/2|$ . However, as we discussed at the start of this section, we do not let these relations with  $(m_{\nu_1}, m_{\nu_2}, m_{\nu_3})$  define the squared mass splittings, because they are not basis-independent.

Observations of neutrino flavor oscillations (see [6] for a review) have furnished measurements of the squared mass splittings  $\delta m^2 = (7.37 \pm 0.17) \times 10^{-5} \text{ eV}^2$  and  $|\Delta m^2| = (2.50 \pm 0.04) \times 10^{-3} \text{ eV}^2$  (NO) and  $|\Delta m^2| = (2.46 \pm 0.04) \times 10^{-3} \text{ eV}^2$  (IO) [42]. The data admits two scenarios for mass ordering (22) depending on whether the smaller splitting is on the bottom (NO) or on the top (IO). Let  $P_{\text{data}}(\Delta m_{HM}^2, \Delta m_{ML}^2)$  denote a probability density over the squared mass splittings that encodes these measurements. Using Eq. (23) the mass splitting measurements imply that  $P_{\text{data}}$  is bi-modal with one peak at  $\Delta m_{HM}^2 \approx |\Delta m^2|$  and  $\Delta m_{ML}^2 \approx \delta m^2$  for  $h = +1$  (NO), and a second peak at  $\Delta m_{HM}^2 \approx \delta m^2$  and  $\Delta m_{ML}^2 \approx |\Delta m^2|$  for  $h = -1$  (IO).

Equation (21) defines a transformation from  $(m_{\nu_1}, m_{\nu_2}, m_{\nu_3})$  to  $(\Sigma m_\nu, \Delta m_{HM}^2, \Delta m_{ML}^2)$ . Let the corresponding Jacobian determinant be denoted as  $\mathcal{J}_{123}(\Sigma m_\nu, \Delta m_{HM}^2, \Delta m_{ML}^2)$  such that  $dm_{\nu_1} dm_{\nu_2} dm_{\nu_3} = \mathcal{J}_{123} d\Sigma m_\nu d\Delta m_{HM}^2 d\Delta m_{ML}^2$ . In terms of these new variables, the differential probability measure (20) becomes

$$d\mathbb{P}_\nu = P(\Sigma m_\nu, \Delta m_{HM}^2, \Delta m_{ML}^2) \times d\Sigma m_\nu d\Delta m_H^2 d\Delta m_{ML}^2 \quad (24)$$

where the probability density is given by

$$P(\Sigma m_\nu, \Delta m_{HM}^2, \Delta m_{ML}^2) = p_\nu(\hat{M}_\nu) \mathcal{J}_\nu(\hat{M}_\nu) \mathcal{J}_{123}(\Sigma m_\nu, \Delta m_{HM}^2, \Delta m_{ML}^2), \quad (25)$$

and it is normalized such that  $\int d\mathbb{P}_\nu = 1$ .

Since the transformation from  $(m_{\nu_1}, m_{\nu_2}, m_{\nu_3})$  into  $(m_H, m_M, m_L)$  involves minimization and maximization functions, it is nonanalytic when any of the two masses are

equal. This makes it difficult to calculate the Jacobian across the full parameter space. However, the problem simplifies significantly for probability densities  $p_\nu(\hat{M}_\nu)$  that respect basis-independence. For a given triplet  $(m_L, m_M, m_H)$  there are six possible ways to identify  $(m_{\nu_1}, m_{\nu_2}, m_{\nu_3})$ . Thanks to the basis-independence of the prior, each of these parameter combinations must be equally probable. Thus, the full parameter space  $(m_{\nu_1}, m_{\nu_2}, m_{\nu_3})$  can be subdivided into six ‘‘copies.’’ Without loss of generality, we can work in the region where  $m_{\nu_1} < m_{\nu_2} < m_{\nu_3}$ , which lets us identify  $m_L = m_{\nu_1}$ ,  $m_M = m_{\nu_2}$ , and  $m_H = m_{\nu_3}$ . Working in this wedge of parameter space, the probability density is multiplied by 6 to account for a sum over the copies.

With the procedure described above it is straightforward to calculate  $\mathcal{J}_{123}$ , but the expression is unwieldy, and we will not reproduce it here. However, it is important to understand how  $\mathcal{J}_{123}$  scales with  $\Sigma m_\nu$ . Note that  $\mathcal{J}_{123}$  has mass dimension  $-2$ , and one can verify that it scales as  $\mathcal{J}_{123} \sim (\Sigma m_\nu)^{-2}$  in the regime where  $\Sigma m_\nu \gg \sqrt{|\delta m^2|}$ ,  $\sqrt{|\Delta m^2|}$ . For comparison, we determine the scaling of  $\mathcal{J}_\nu(\hat{M}_\nu)$  by inspecting Eqs. (10) and (21). In the regime where the splittings are small, we have

$$\mathcal{J}_\nu \sim \begin{cases} (\Sigma m_\nu)^3 (\delta m^2)^2 |\Delta m^2|^4, & \text{D} \\ (\Sigma m_\nu)^3 (\delta m^2) |\Delta m^2|^2, & \text{M} \end{cases}. \quad (26)$$

The factors of  $\delta m^2$  and  $|\Delta m^2|$  correspond to the eigenvalue repulsion that we discussed below Eq. (10). Then, the probability density in Eq. (25) goes as  $P \sim p_\nu(\hat{M}_\nu) \times (\Sigma m_\nu)^1$ , which is linearly rising in  $\Sigma m_\nu$  if  $p_\nu$  is flat. We will make use of this fact in the following sections.

Equation (25) gives the joint probability density over the summed neutrino mass and the squared mass splittings. We are also interested in the conditional probability over  $\Sigma m_\nu$  after requiring the squared mass splittings to match their measured values and also requiring the hierarchy to be either normal or inverted. Thus we calculate

$$\begin{aligned} \pi(\Sigma m_\nu | h) &= \frac{6}{\mathcal{E}_{\text{NO}} + \mathcal{E}_{\text{IO}}} \\ &\times \int d\Delta m_{HM}^2 d\Delta m_{ML}^2 P(\Sigma m_\nu, \Delta m_{HM}^2, \Delta m_{ML}^2) \\ &\times P_{\text{data}}(\Delta m_{HM}^2, \Delta m_{ML}^2) \\ &\times \begin{cases} \Theta(\Delta m_{ML}^2 < \Delta m_{HM}^2), & h = +1 \text{ (NO)} \\ \Theta(\Delta m_{HM}^2 < \Delta m_{ML}^2), & h = +1 \text{ (IO)} \end{cases}. \end{aligned} \quad (27)$$

The bimodal probability distribution  $P_{\text{data}}$ , which is defined below Eq. (23), implements the mass splitting measurements. The step function, which satisfies  $\Theta(x) = 1$  for  $x \geq 0$  and  $\Theta(x) = 0$  for  $x < 0$ , lets us condition on the

hierarchy. The normalization factor  $\mathcal{E}_{\text{NO}} + \mathcal{E}_{\text{IO}}$  is the sum of the evidences for the NO and IO cases separately so that  $\sum_{h=\pm 1} \int_0^\infty d\Sigma m_\nu \pi(\Sigma m_\nu | h) = 1$ .

#### IV. BASIS-INDEPENDENT ANARCHY PRIORS ON $\Sigma m_\nu$

In the context of random mass matrices, the neutrino mass prior  $\pi(\Sigma m_\nu)$  is precisely the marginalized probability from (27). How does the BAH guide us in choosing a prior? We first show in Sec. IV A, that the common *ad hoc* priors on  $\Sigma m_\nu$  discussed in Sec. II appear highly contrived in the BAH context. Arguably the three most natural choices for probability distributions over  $M_\nu$  are flat, Gaussian, and logarithmic distributions over each matrix element. Interestingly the logarithmic distribution is forbidden by the BAH. The Gaussian distribution has a similar behavior to the flat distribution (see Sec. A), and so we focus on the latter in Sec. IV C.

##### A. Relationship to *ad hoc* priors

In Sec. II we looked at a few *ad hoc* priors on  $\Sigma m_\nu$ , such as flat-linear and flat-log, that are commonly used for cosmological data analyses. In this section we discuss whether these priors can arise in the framework of random neutrino matrices subject to the BAH condition.

In constructing the distribution  $\pi(\Sigma m_\nu | h)$ , the only freedom is to specify the probability density  $p_\nu(\hat{M}_\nu)$  that appears in Eq. (25). Recall that the requirement of basis-independence (BAH) means that we must construct  $p_\nu(\hat{M}_\nu)$  from functions of the neutrino mass matrix  $M_\nu$  that are left invariant under the basis-changing flavor transformation; see Sec. III D. This severely constraints the functional forms that we can use; examples of viable functions include the determinant,  $d = |\det M_\nu| = m_{\nu_1} m_{\nu_2} m_{\nu_3}$ , and the traces,  $t_n = \text{Tr}[(M_\nu^\dagger M_\nu)^n] = (m_{\nu_1}^2)^n + (m_{\nu_2}^2)^n + (m_{\nu_3}^2)^n$  for positive, integer  $n$ .

Cosmological studies typically set a flat-linear prior on  $\Sigma m_\nu$ , and we can ask whether this choice is reasonable from the context of random neutrino matrices. Note that taking  $p_\nu(\hat{M}_\nu) \propto [\mathcal{J}_\nu(\hat{M}_\nu)]^{-1}$  in Eq. (20) results in  $d\mathbb{P} \propto dm_1 dm_2 dm_3$ , which is flat in each of the three neutrino masses. To obtain a distribution that is flat in  $\Sigma m_\nu$  we should consider Eqs. (24) and (25). Recall that  $\mathcal{J}_{123} \sim (\Sigma m_\nu)^{-2}$  for  $\Sigma m_\nu \gg \Delta m_{HM}^2, \Delta m_{ML}^2$ , as we discussed above Eq. (26). Thus, we should take  $p_\nu(\hat{M}_\nu) \propto [\mathcal{J}_\nu(\hat{M}_\nu)]^{-1} (m_1 m_2 m_3)^{2/3}$ , because  $m_1 \sim m_2 \sim m_3 \sim \Sigma m_\nu / 3$  in the degenerate regime. This results in a prior that is flat in  $\Sigma m_\nu$  for  $\Sigma m_\nu \gg \Delta m_{HM}^2, \Delta m_{ML}^2$ .

To understand whether  $p_\nu(\hat{M}_\nu) \propto [\mathcal{J}_\nu(\hat{M}_\nu)]^{-1} (\det \hat{M}_\nu)^{2/3}$  is a reasonable prior, let us express it in terms of the matrix invariants. The Jacobian factor  $\mathcal{J}_\nu(\hat{M}_\nu)$  appears in Eq. (10) where it is expressed as a function of the singular values  $m_{\nu_i}$ . Writing  $\mathcal{J}_\nu$  instead in terms of the matrix invariants,  $d$  and  $t_n$



defined above, it takes the form  $\mathcal{J}_\nu = \mathcal{C}d$  for Dirac neutrinos and  $\mathcal{C}^{1/2}d$  for Majorana neutrinos where  $\mathcal{C} = (-27d^4 + 5d^2t_1^3 - \frac{1}{4}t_1^6 - 9d^2t_1t_2 + t_1^4t_2 - \frac{5}{4}t_1^2t_2^2 + \frac{1}{2}t_2^3)$ . Thus taking  $p_\nu \propto [\mathcal{J}_\nu]^{-1}d^{2/3}$  leads to a distribution that is asymptotically flat in  $\Sigma m_\nu$ , but it gives  $p_\nu$  a highly non-trivial dependence on  $M_\nu$ . Similar arguments hold for all of the six *ad hoc* priors that were discussed in Sec. II.

### B. Log prior on matrix elements

It is customary to use a logarithmic prior for dimensional ‘‘scale’’ parameters. Therefore it is natural to consider imposing such a prior for the neutrino mass matrix. In order to place a logarithmic prior on the real and imaginary parts of each matrix element, we should choose the probability density  $f(M_e, M_\nu)$  from Eq. (12) to be proportional to  $[\Pi_{ij}(\text{Re}M_\nu)_{ij}(\text{Im}M_\nu)_{ij}]^{-1}$  with some cutoffs at small and large mass to ensure integrability. However, this product of matrix elements cannot be expressed in terms of the basis-independent matrix invariants,  $d$  and  $t_n$  from Sec. IV A. Consequently a logarithmic prior on the matrix elements is not consistent with the BAH. This is an example of how theory-motivated assumptions like the BAH provide guidance in choosing a physically-motivated prior.

### C. Flat prior on matrix elements

In this section we introduce a probability density that will be the focus of our extensive numerical study in the following sections. Consider the probability density that is flat in the real and imaginary parts of each element of the neutrino mass matrix  $M_\nu$ . In terms of the probability density  $p_\nu(\hat{M}_\nu)$  from Eqs. (12) and (18), this distribution is expressed as

$$p_\nu(\hat{M}_\nu|\mu_\nu) = \mathcal{N}\Theta(\mu_\nu^2 - \text{Tr}[\hat{M}_\nu^\dagger \hat{M}_\nu]) \quad (28)$$

for the Majorana and Dirac models. Here  $\Theta(x)$  is the unit step that evaluates to  $\Theta = 1$  for  $x \geq 0$  and  $\Theta = 0$  otherwise. The parameter  $\mu_\nu$  controls the neutrino energy scale. In the seesaw model the probability density  $F(M_e, M_D, M_M)$  from Eq. (12) is chosen to be

$$F(M_e, M_D, M_M|\mu_D, \mu_M) = \mathcal{N}\Theta(\mu_D^2 - \text{Tr}[M_D^\dagger M_D])\Theta(\mu_M^2 - \text{Tr}[M_M^\dagger M_M]). \quad (29)$$

Since we are only interested in the spectrum of light neutrinos, which is given by the seesaw relation (11), the relevant observables will only depend upon the effective neutrino scale

$$\mu_\nu = \frac{\mu_D^2}{\mu_M}. \quad (30)$$

For the Dirac and Majorana models, notice that the step function imposes a hard upper bound on the neutrino mass

spectrum,  $m_\nu \leq \mu_\nu$  or  $\Sigma m_\nu \leq \sqrt{3}\mu_\nu$ , but there is no such bound for the seesaw model.

The normalization factors  $\mathcal{N}$  are obtained by imposing  $\int d\mathbb{P} = 1$  over the space of random matrices. For the Dirac model  $\int d\mathbb{P}$  equals  $\mathcal{N}$  times the volume of the 18-dimensional ball<sup>3</sup> with radius  $\mu_\nu$ ; for the Majorana model it equals  $\mathcal{N}$  times the volume of a 12-dimensional ellipsoid; and for the seesaw model  $\int d\mathbb{P}$  is the product of these two. Solving for the normalization factor gives:

$$\mathcal{N} = \begin{cases} \frac{1}{3}\Gamma(10)\mathcal{V}_2^{-1}\mu_\nu^{-18}, & \text{D} \\ \frac{16}{3}\Gamma(7)\mathcal{V}_1^{-1}\mu_\nu^{-12}, & \text{M}, \\ \frac{16}{9}\Gamma(7)\Gamma(10)\mathcal{V}_1^{-1}\mathcal{V}_2^{-1}\mu_M^{-12}\mu_D^{-18}, & \text{S} \end{cases} \quad (31)$$

where  $\Gamma(z)$  denotes the Gamma function, and where  $\mathcal{V}_1$  and  $\mathcal{V}_2$  were given below Eq. (20). The high power of  $\mu_\nu$  appears to compensate the mass dimensions of  $d^{12}M_\nu$ ,  $d^{18}M_\nu$ , and  $d^{18}M_Dd^{12}M_M$  in the Majorana, Dirac, and seesaw models, respectively; see Eq. (12).

### D. Methodology

We have performed an extensive semianalytical and numerical analysis of the probability distributions appearing in Eqs. (28) and (29). In the remainder of this section we describe the numerical techniques, and in the following section we describe the results for  $\Sigma m_\nu$ .

A semianalytical analysis is computationally feasible for the (M) and (D) models. Recall that the probability distribution over the neutrino mass sum and squared splittings was given by Eq. (25) where  $p_\nu(\hat{M}_\nu)$  is given by Eq. (28). Upon fixing the mass splittings to their best-fit measured values, the distribution over the neutrino mass sum is given by Eq. (27) for both NO and IO cases. These integrals are evaluated numerically with Mathematica.

In addition we perform a fully numerical analysis for all three models: (M), (D), and (S). Here we implement the mass splitting measurements as a Gaussian likelihood. To perform the numerical integrals we Markov Chain Monte Carlo (MCMC) sample the full space of random matrices by means of the emcee algorithm [43]. We analyze the resulting distributions with the GETDIST package [44]. To compute the overall normalization of the probability densities we use the algorithm described in [45] and direct nested sampling integration with multineest [46,47]. We shall hereafter refer to the normalization of the probability densities as the evidence,  $\mathcal{E}$ .

For the (M) and (D) models the semianalytical analysis provides a test of the accuracy of the MCMC integration, which is the only technique that we use to study the (S) model. Overall, we find excellent agreement for the (M) and (D) models (see Fig. 7 below), and these tests further

<sup>3</sup>The volume of the  $n$ -dimensional unit ball is  $\pi^{n/2}/\Gamma(n/2 + 1)$ .

validate the pipeline that we apply to the (S) model. In all three models the precision of the results is also estimated with several split-tests on the MCMC samples. In particular we require stability of the results to: residual burn-in samples by discarding relevant fractions of the initial samples; subdivision of the samples in uncorrelated and correlated subsamples; and convergence of the results by discarding relevant fractions of the last samples.

The difference between the evidence calculations for the semianalytical and numerical cases is  $\Delta \log_{10} \mathcal{E} = 0.2(0.2)$  in the (M) model NO (IO) and  $\Delta \log_{10} \mathcal{E} = 0.003(0.3)$  in the (D) case for NO (IO). In these cases, the semianalytical calculation is the more accurate and we use these in our results below. This should be compared with the numerical error reported by multineest that reads  $\Delta \log_{10} \mathcal{E} = 0.1(0.1)$  in the (M) case and  $\Delta \log_{10} \mathcal{E} = 0.01(0.4)$  in the (D) case. In the (S) case, where the semianalytical results are not available, the multineest estimate of the evidence error  $\Delta \log_{10} \mathcal{E} \sim 0.2(0.3)$  and the nested sampling and importance nested sampling estimates are within this error estimates.

### E. Implications for $\Sigma m_\nu$

Let us now discuss the numerical results. We first investigate the probability distributions over the summed neutrino masses and the squared mass splittings that are predicted from a flat-linear distribution on individual matrix elements (28) and (29). At this point we do not fix the squared mass splittings to equal their measured values, but we explore the parameter space more broadly.

Figure 2 shows the joint distribution of the sum of the neutrino masses against the distribution of one squared mass splitting after marginalizing over the other squared mass splitting for both the NO and IO cases from the numerical analysis. Masses are expressed in units of the neutrino energy scale  $\mu_\nu$  since, without reference to external data to fix the energy scale, the prior distribution is invariant against rescaling. The predicted distributions are strongly influenced by two boundaries in the parameter space: the first boundary requires the lightest neutrino mass to be positive ( $m_L > 0$ ), which implies

$$\Sigma m_\nu \geq \Sigma m_\nu^{\min} \equiv \sqrt{\Delta m_{HM}^2 + \Delta m_{ML}^2} + \sqrt{\Delta m_{ML}^2}$$

$$\Sigma m_\nu^{\min} \simeq \begin{cases} 0.05895 \pm 0.00041 \text{ eV}, & \text{NO} \\ 0.09919 \pm 0.00081 \text{ eV}, & \text{IO} \end{cases}, \quad (32)$$

where we have used the mass splitting measurements from Sec. III F. The second boundary is imposed by the step function in Eq. (28), which implies  $\Sigma m_\nu \leq \sqrt{3}\mu_\nu$  for the (M) and (D) models.

Flavor oscillation measurements determine the squared mass splittings, which corresponds to selecting a horizontal section of the parameter space in Fig. 2. The (S) model, shown on the right panels of Fig. 2, displays a strong

degeneracy: for any  $\mu_\nu$ , once the squared mass splittings are fixed, the most probable value of  $\Sigma m_\nu$  can be read off from the equality in Eq. (32), which corresponds to  $m_L = 0$ . For the (D) and (M) models, on the other hand, fixing  $\Delta m^2$  leads to a distribution over  $\Sigma m_\nu$  that peaks at approximately  $\mu_\nu$ , regardless of the chosen value for  $\Delta m^2$ .

To better understand the distribution over  $\Sigma m_\nu$  at fixed splittings, it is useful to consider Fig. 3, which shows the distribution over  $\Sigma m_\nu$  for a fixed  $\mu_\nu$  with the two squared mass splittings set equal to their best-fit measured values. The distribution  $\pi(\Sigma m_\nu)$  from the semianalytical analysis is observed to rise linearly and peak at  $\Sigma m_\nu = \sqrt{3}\mu_\nu$ . The linearly rising distribution can be understood from the discussion around Eq. (26) where we argued that  $P \propto p(\hat{M}_\nu)(\Sigma m_\nu/\mu_\nu)$ , which grows linearly with  $\Sigma m_\nu$  for flat  $p(\hat{M}_\nu)$ . This argument also implies that NO and IO occur with equal probability if the mass splittings are fixed to their measured values, and  $\mu_\nu$  is fixed to a larger value. However, the evidence  $\mathcal{E}_{\text{NO}} + \mathcal{E}_{\text{IO}}$  drops substantially from  $5.6 \times 10^3$  at  $\mu_\nu = 0.1$  eV to  $1.6 \times 10^{-8}$  at  $\mu_\nu = 1.0$  eV for the (M) model, and it drops from  $8.4 \times 10^1$  to  $2.4 \times 10^{-16}$  for the (D) model. The additional suppression for the (D) model is a result of eigenvalue repulsion, and this behavior can be understood analytically. The probability density over the summed masses and squared splittings is given by Eq. (25) where the Jacobian determinant  $\mathcal{J}_\nu$  is given by Eq. (10); see also Eq. (26). The phenomenon of eigenvalue repulsion is manifest in  $\mathcal{J}_\nu$  through the factors of  $(m_{\nu i}^2 - m_{\nu j}^2)$ . The mass dimension of these factors is compensated by factors of  $\mu_\nu^{-2}$  through the normalization (31). Taking  $\mu_\nu$  to be much larger than the scale of the measured mass splittings results in a suppression that makes such models very improbable; the suppression factors are  $\sim (\delta m^2/\mu_\nu^2)^2 (\Delta m^2/\mu_\nu^2)^4$  for (D) and  $\sim (\delta m^2/\mu_\nu^2) (\Delta m^2/\mu_\nu^2)^2$  for (M). Thus the overall probability of satisfying the observed mass splittings is extremely small unless  $\mu_\nu$  is comparable to the larger splitting.

Next we explore the relationship between the ordering, the neutrino energy scale  $\mu_\nu$ , and the observed splittings. In Fig. 4 we show the distribution of the two splittings marginalized over  $\Sigma m_\nu$  or equivalently over  $m_L$ , in units of the neutrino mass scale  $\mu_\nu$  from the numerical analysis. The solid black line divides the regions of parameter space that correspond to the normally ordered mass spectrum (NO) and inverted ordered spectrum (IO). In general NO is preferred with respect to IO increasingly from (M) to (D) to (S), as it occupies more volume in parameter space. In the (M) case this volume factor corresponds to an odd ratio of 4:1 in favor of NO, in the (D) case of 6:1 in favor of NO and 22:1 in the (S) case.

By studying Fig. 4 we can also understand the effect of imposing the mass splitting measurements. Fixing the squared mass splittings to equal their measured values selects two solutions in the  $(\Delta m_{HM}^2, \Delta m_{ML}^2)$  space, which

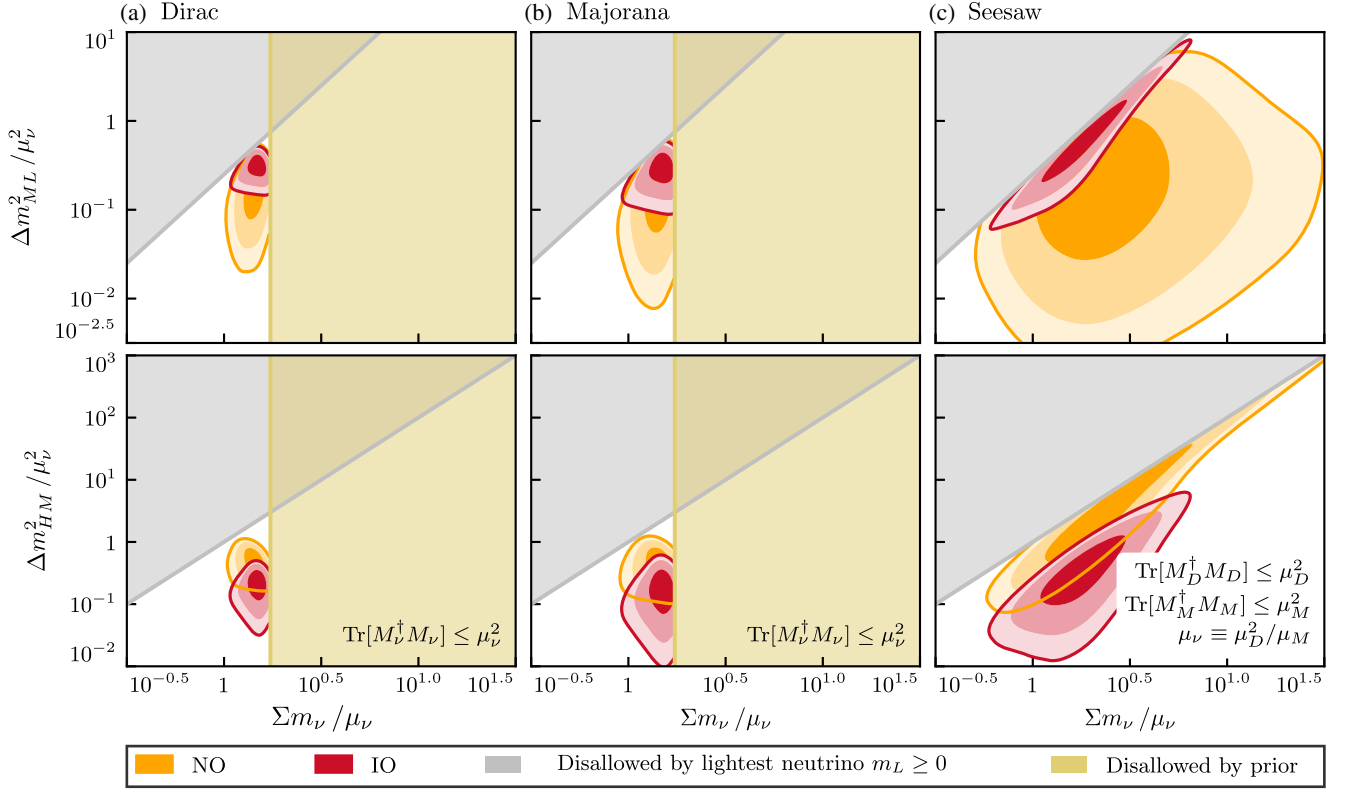


FIG. 2. The probability distribution over the mass sum,  $\Sigma m_\nu = m_L + m_M + m_H$ , and one of the squared mass splittings,  $\Delta m_{HM}^2 = m_H^2 - m_M^2$  and  $\Delta m_{ML}^2 = m_M^2 - m_L^2$ , after marginalizing over the other splitting. The three panels correspond to different models for the origin of neutrino mass: Left, we sample the Dirac neutrino mass matrix  $M_\nu$  with a flat distribution on its matrix elements and cutoff  $\text{Tr}[M_\nu^\dagger M_\nu] \leq \mu_\nu^2$  (28); Middle, we sample the (symmetric) Majorana neutrino mass matrix  $M_\nu$  with a flat distribution on its matrix elements and cutoff  $\text{Tr}[M_\nu^\dagger M_\nu] \leq \mu_\nu^2$  (28); Right, we sample the high-scale Dirac and Majorana mass matrices,  $M_D$  and  $M_M$ , with a flat distribution on their matrix elements and separate cutoffs  $\mu_D$  and  $\mu_M$  (29), and then we use the seesaw relation (11) to calculate the neutrino mass matrix, which only depends on the ratio  $\mu_\nu = \mu_D^2/\mu_M$ . The orange (red) region indicates models with normal (inverted) mass ordering for which  $\Delta m_{HM}^2 > \Delta m_{ML}^2$  ( $\Delta m_{HM}^2 < \Delta m_{ML}^2$ ). Requiring  $m_L > 0$  forbids the parameter space where  $\Sigma m_\nu < \sqrt{\Delta m_{HM}^2 + \Delta m_{ML}^2} + \sqrt{\Delta m_{ML}^2}$ , which corresponds to the upper-left gray triangle. Additionally the prior (28) forbids  $\Sigma m_\nu > \sqrt{3}\mu_\nu$  in the Dirac and Majorana cases, which corresponds to the yellow rectangle in the left and middle panels. The darker and lighter shades correspond respectively to the 68% C.L., 95% C.L. and 99.7% C.L. regions.

extend to two lines in the scaled space ( $\Delta m_{HM}^2/\mu_\nu^2$ ,  $\Delta m_{ML}^2/\mu_\nu^2$ ) indicated by the dashed line for NO and the dot-dashed line for IO. Where these lines intersect the high probability regions determines the best values of  $\mu_\nu$  for explaining the splittings. Now we see that taking  $\Sigma m_\nu \gg \sqrt{|\delta m^2|}$ ,  $\sqrt{|\Delta m^2|}$  selects a model in the low probability tail, which agrees with the behavior seen in Fig. 3 that we discussed in the previous paragraph. Thus even though such a case would produce a  $\Sigma m_\nu$  distribution that peaks at  $\sqrt{3}\mu_\nu$  for (D) and (M), it would be a highly unlikely realization of the underlying matrix elements. Instead the probability is maximal for a value of  $\mu_\nu$  that is comparable to scale of the larger mass splitting:  $\mu_\nu = 0.02\text{--}0.1$  eV depending on the scenario. For smaller values of  $\mu_\nu$  the probability decreases, and it becomes identically zero in the (D) and (M) models because the prior (28) requires  $m_\nu \leq \mu_\nu$ .

The key point in the preceding discussion is that in the context of a specific neutrino mass generation mechanism the mass splittings are informative on the neutrino energy scale  $\mu_\nu$ , because the measured mass splittings set a preferred scale for the mass matrix elements, and thus  $\mu_\nu$  should not be chosen *a priori*. Instead we take it as a hyperparameter and allow the mass splittings to determine its posterior probability distribution. We assume a wide, uninformative, flat-linear prior on  $\mu_\nu \in (0, 1)$  eV since we have no strong theoretical preference for one value of the neutrino energy scale over another. Figure 5 shows the resulting distribution of  $\mu_\nu$  after marginalizing over the other model parameters from the numerical analysis. For all three models and both hierarchies, the distribution is sharply peaked around a maximum value that depends on the mechanism and hierarchy. The confidence levels on the neutrino scale are reported in Table I.

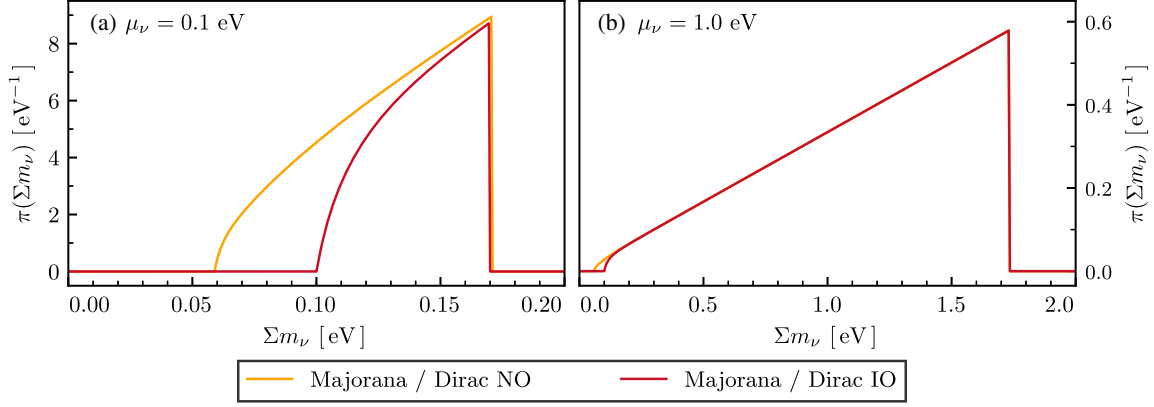


FIG. 3. The probability distribution over the sum of neutrino masses after the squared mass splittings are fixed to their best-fit measured values, and the neutrino mass scale is fixed to be  $\mu_\nu = 0.1$  eV ( $1.0$  eV) in the left (right) panel. The shape of the probability distributions are identical for both the Majorana and Dirac models, but the shapes are distinguished for normal and inverted ordering. For a fixed  $\mu_\nu$  the most probable value is  $\Sigma m_\nu = \sqrt{3}\mu_\nu$ , and increasing  $\mu_\nu$  moves the peak to larger  $\Sigma m_\nu$ , but at the same time it reduces the overall probability of obtaining the observed squared mass splittings (not shown on the figure); i.e., the evidence decreases rapidly with increasing  $\mu_\nu$  (see discussion in the text).

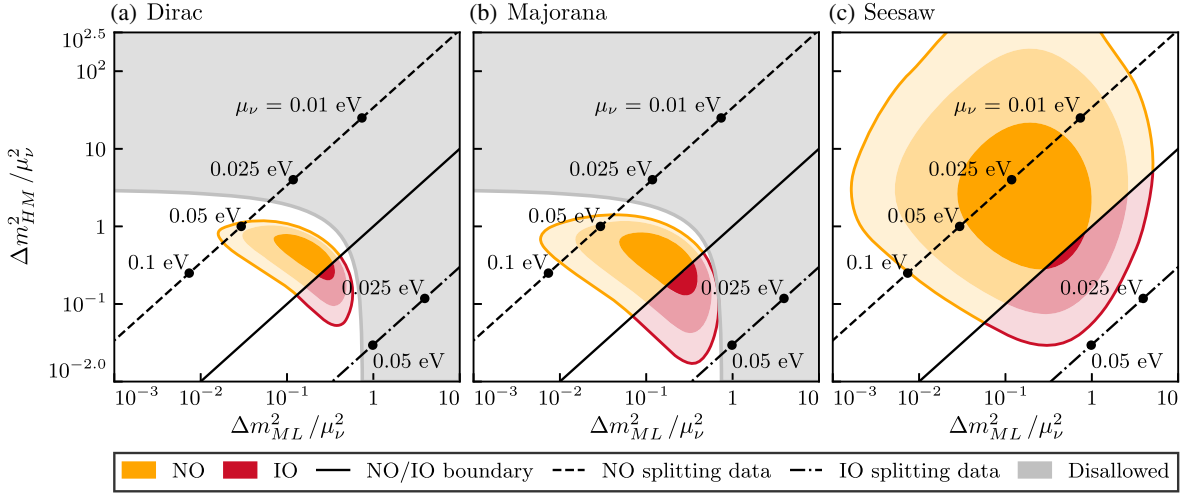


FIG. 4. The probability distribution over the squared mass splittings,  $\Delta m_{HM}^2$  and  $\Delta m_{ML}^2$ , after marginalizing over the mass sum,  $\Sigma m_\nu$ . The notation and shading are defined in the caption of Fig. 2. The solid black line divides models with normal ordering ( $\Delta m_{ML}^2 < \Delta m_{HM}^2$ ) from those with inverted ordering ( $\Delta m_{HM}^2 < \Delta m_{ML}^2$ ). Along the black dashed and dot-dashed lines we fix  $\Delta m_{HM}^2$  and  $\Delta m_{ML}^2$  to equal their measured values (see Sec. III F), and we slide the value of  $\mu_\nu$ .

As discussed in the Appendix the sharpness of the  $\mu_\nu$  peak makes this distribution extremely stable against different prior choices for  $\mu_\nu$ ; for instance, raising the upper limit on  $\mu_\nu$  from 1 eV does not affect the resultant distribution on  $\Sigma m_\nu$ . For the seesaw model one can also specify the hyperprior on the cutoffs of the high-scale matrices,  $\mu_D$  and  $\mu_M$  from Eq. (29), and let the corresponding distribution on  $\mu_\nu$  be inferred from Eq. (30). As long as the resulting distribution on  $\mu_\nu$  is uninformative, the distributions over  $\Sigma m_\nu$  are unaffected. However, the evidences are affected, which complicates model comparison between (S) and either (D) or (M).

After marginalizing over  $\mu_\nu$ , the distribution over  $\Sigma m_\nu$  from the numerical analysis is shown in Fig. 6. This is

dramatically different for the (D) and (M) models from the distribution with  $\mu_\nu$  held fixed, which we showed in Fig. 3. The falling distribution at high  $\Sigma m_\nu$  comes from the combination of the  $\Sigma m_\nu \leq \sqrt{3}\mu_\nu$  limit and the decreasing probability at high  $\mu_\nu$  shown in Fig. 5, whereas the peak comes from the fact that for all  $\mu_\nu$  consistent with the splittings, the  $\Sigma m_\nu$  distribution rises near the minimum. In contrast the distribution over  $\Sigma m_\nu$  for the (S) model falls off strongly at large  $\Sigma m_\nu$  for both fixed  $\mu_\nu$  and marginalized  $\mu_\nu$ .

The 68% confidence levels on the neutrino mass sum and other parameters from the numerical analysis are reported in Table I. Evidences are reported in Table II and come from the semianalytical calculation for (M) and (D) and the numerical multinest calculation for (S).

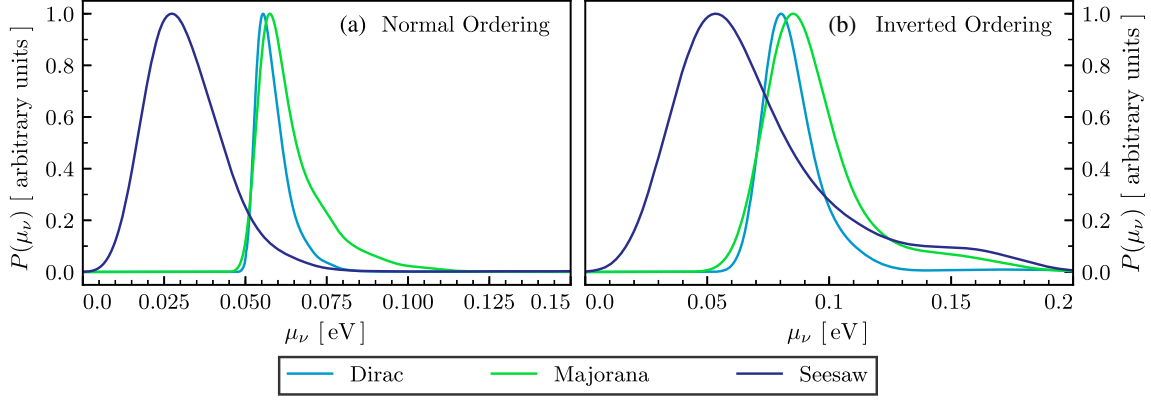


FIG. 5. The probability distribution over the neutrino energy scale  $\mu_\nu$  after marginalizing over the mass sum and imposing the mass splittings measurements. In the left (right) panel we implement the mass splitting measurement with normal (inverted) ordering.

We notice that once splitting measurements are factored into our study, the preference for NO with respect to IO becomes higher; the evidence increases by  $\Delta \log_{10} \mathcal{E} = 2.1$  for (D),  $\Delta \log_{10} \mathcal{E} = 1.3$  for (M), and  $\Delta \log_{10} \mathcal{E} = 2.7$  for (S). These odds are easily understood from Figs. 4 and 5. Consider first the (D) model with the splittings fixed to their measured values and NO assumed (dashed line in Fig. 4). Figure 5 shows that the most probable value for the neutrino energy scale is  $\mu_\nu \approx 0.06$  eV, which lies between the  $2\sigma$  (95% C.L.) and  $3\sigma$  (99.7% C.L.) contours in Fig. 4. On the other hand, if IO is assumed (dot-dashed line), the most probable value is  $\mu_\nu \approx 0.08$  eV from the right panel of Fig. 5, and it lies outside of the  $3\sigma$  contour and closer to the  $4\sigma$  (0.99993% C.L.) one in Fig. 4. The probability of these two events explains the approximately 130:1 odds favoring NO over IO that was reported above. Consider next the (M) model for which the most probable neutrino energy scales are  $\mu_\nu \approx 0.06$  eV for NO and  $\mu_\nu \approx 0.08$  eV for IO, which correspond to points in Fig. 4 that are at the  $2\sigma$  and  $3\sigma$  levels, respectively. The relative probability of these two events explains the 19:1 odds in favor of NO for (M). In (S) the maximal  $\mu_\nu$  occurs at  $\mu_\nu \approx 0.025$  eV for NO and at  $\mu_\nu \approx 0.05$  eV for IO, and the corresponding points in Fig. 4 are at approximately the  $0.5\sigma$  level (NO) and slightly above  $3\sigma$  (IO). This explains the odds of approximately 470:1 in favor of NO for (S).

The evidence results in Table I also allows us to comment on the odds of different models summed over the two hierarchies. Given equal prior odds of S:M:D, conditioning to the mass splitting data gives us odds of 830:3.3:1.

## F. Fitting formulas

In this section we present an empirical fitting formula that approximates the probability distribution over  $\Sigma m_\nu$  for practical applications. This is constructed from three building blocks.

The first component is a smooth cutoff at the minimal value of the sum of the masses given by Eq. (32):  $x_{\min} = 0.05895$  eV ( $x_{\min} = 0.09919$  eV) with width  $\sigma_{\min} = 0.00041$  eV ( $\sigma_{\min} = 0.00081$  eV) for NO (IO). The cutoff is implemented as a smoothed step function:

$$\text{Step}(x, \mu, \sigma) \equiv \frac{1}{2} \left[ 1 + \text{Erf} \left( \frac{x - \mu}{\sqrt{2}\sigma} \right) \right], \quad (33)$$

where  $\mu$  is the value at which the function transitions from zero to one, and  $\sigma$  is the width of the transition.

The second component is a skewed Gaussian distribution that describes the behavior of the peak of the distribution. The center of the Gaussian is denoted by  $\bar{x}$ , its width is  $\sigma_{\bar{x}}$ , and the skewness is  $\alpha$ .

TABLE I. Marginalized 68% C.L. results on the neutrino prior distribution. Notice that the distributions that are used to extract these bounds are markedly non-Gaussian. Higher confidence regions can be easily computed from the fitting forms in Sec. IV F.

Model	$\Sigma m_\nu$ [eV]	$\mu_\nu$ [eV]	$m_L$ [eV]	$m_M$ [eV]	$m_H$ [eV]
Dirac NO	$0.069^{+0.01}_{-0.007}$	$0.055^{+0.006}_{-0.003}$	$0.008^{+0.007}_{-0.004}$	$0.01^{+0.005}_{-0.002}$	$0.0509^{+0.001}_{-0.0008}$
Majorana NO	$0.07^{+0.02}_{-0.01}$	$0.057^{+0.009}_{-0.006}$	$0.009^{+0.009}_{-0.006}$	$0.011^{+0.008}_{-0.002}$	$0.051^{+0.002}_{-0.001}$
Seesaw NO	$0.06^{+0.004}_{-0.002}$	$0.03^{+0.01}_{-0.01}$	$0.0^{+0.003}_{-0}$	$0.0087^{+0.0006}_{-0.0005}$	$0.0505^{+0.0005}_{-0.0005}$
Dirac IO	$0.11^{+0.01}_{-0.009}$	$0.074^{+0.007}_{-0.006}$	$0.01^{+0.009}_{-0.006}$	$0.05^{+0.002}_{-0.001}$	$0.051^{+0.002}_{-0.001}$
Majorana IO	$0.12^{+0.02}_{-0.01}$	$0.079^{+0.01}_{-0.009}$	$0.015^{+0.01}_{-0.01}$	$0.051^{+0.003}_{-0.002}$	$0.051^{+0.003}_{-0.002}$
Seesaw IO	$0.102^{+0.009}_{-0.006}$	$0.05^{+0.03}_{-0.02}$	$0.0^{+0.01}_{-0}$	$0.05^{+0.002}_{-0.002}$	$0.05^{+0.002}_{-0.002}$

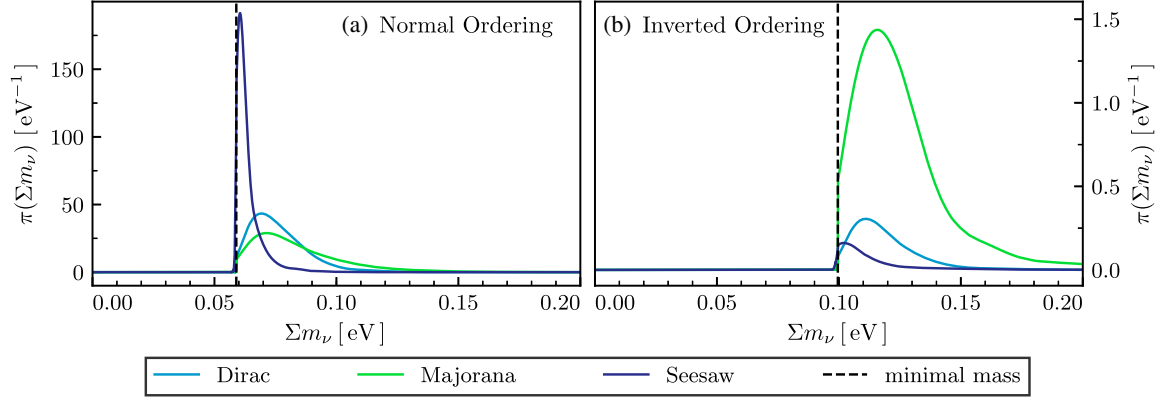


FIG. 6. The probability distribution over the neutrino mass sum after marginalizing over the neutrino energy scale  $\mu_\nu$  and imposing the squared mass splittings measurements. In the left (right) panel we implement the mass splitting measurement with normal (inverted) ordering. The vertical black dashed line indicates the minimal value of  $\Sigma m_\nu$  consistent with the mass splittings (32), which is obtained by setting the mass splittings to their best-fit measured value and setting the lightest neutrino mass to zero.

The third component is a power law cutoff at large  $\Sigma m_\nu$ . To prevent this component from dominating at small mass, it is cut off using a smooth transition centered at  $\hat{x}$  with width  $\sigma_{\hat{x}}$ . The exponent of the power law cutoff is taken to be  $p = -16, -10$ , and  $-12$  for the Dirac, Majorana, and seesaw models, respectively.

Joining these pieces together we obtain a fitting function for the prior  $\pi(\Sigma m_\nu)$ :

$$\begin{aligned}
 P_{\text{fit}}(x) &= N \cdot \text{Step}(x, x_{\text{min}}, \sigma_{\text{min}}) \\
 &\times \left[ \frac{\sqrt{2}}{\sqrt{\pi\sigma_{\hat{x}}^2}} \text{Step}(x, \bar{x}, \alpha) \exp\left(-\frac{(x - \bar{x})^2}{2\sigma_{\hat{x}}^2}\right) \right. \\
 &\left. + \mathcal{A} \text{Step}(x, \hat{x}, \sigma_{\hat{x}}) \left(\frac{x}{\hat{x}}\right)^p \right]. \quad (34)
 \end{aligned}$$

The relative amplitude between the exponential and power law behavior is given by  $\mathcal{A}$ . The coefficient  $N$  ensures that the fitting function is normalized such that  $\int_0^\infty dx P_{\text{fit}}(x) = 1$ .

Table II summarizes the parameters of the fitting formulas for the three different models considered in Sec. IV C (Dirac, Majorana, seesaw) and the two spectrum orderings (NO, IO). For the (D) and (M) models we have fit to the semi-analytical results, since they are more accurate, and

for the (S) model we fit to the numerical results. A reference implementation of the fitting formulas can be found at [https://github.com/mraveri/Neutrino\\_Prior](https://github.com/mraveri/Neutrino_Prior). For all six cases we observe a very good agreement between the semi-analytical results, the numerical results, and the fitting formula; this agreement is illustrated in Fig. 7. In the (M) and (D) cases, the fitting formula agrees with the semi-analytical results at 1% on the 95% C.L. intervals and at 5% on the 0.99993 C.L. ( $4\sigma$ ) intervals; in the (S) case, the fitting formula agrees with the numerical results at 1% on the 95% C.L. intervals and at 10% on the 0.99993 C.L. ( $4\sigma$ ) tails, but note that the numerical results themselves become increasingly poorly sampled in the tails.

Table II summarizes the parameters of the fitting formulas for six different combinations of orderings and mass models. The reader may choose to combine these fitting functions by marginalizing over orderings or models. The (totally) marginalized probability distribution is calculated as

$$P_{\text{fit,tot}}(x) = \frac{\sum_i \mathcal{E}_i \Pi_i P_{\text{fit}}(x|i)}{\sum_i \mathcal{E}_i \Pi_i}, \quad (35)$$

where  $P_{\text{fit}}(x|i)$  is given by Eq. (34) for one of the rows of Table II, the evidences  $\mathcal{E}_i$  appear in Table II, and we

TABLE II. Parameters for the fitting formula discussed in Sec. IV F. We also report in the first column the model prior volume, without considering splitting data, and, in the second column, the overall normalization of the prior distributions once splitting data are included and we marginalize over  $\mu_\nu$ .

Model	$\log_{10} \mathcal{E}$	$\log_{10} N$	$\bar{x}$ [eV]	$\sigma_{\hat{x}}$ [eV]	$\alpha$ [eV]	$\log_{10} (\mathcal{A} \text{ eV})$	$\hat{x}$ [eV]	$\sigma_{\hat{x}}$ [eV]	$p$
Dirac NO	3.08	-0.006035	0.06109	0.01700	0.002484	-0.6430	0.1315	0.01603	-16
Majorana NO	3.58	-0.01028	0.06164	0.02416	0.003215	0.1945	0.1336	0.01251	-10
Seesaw NO	6.00	26.91	-0.4099	0.04301	0.01624	-26.25	0.0779	0.007016	-12
Dirac IO	0.96	-0.01345	0.1033	0.01792	0.002936	-0.2773	0.1713	0.01876	-16
Majorana IO	2.31	-0.02738	0.1039	0.02490	0.003634	0.4442	0.1705	0.01359	-10
Seesaw IO	3.33	65.07	0.0000	0.01851	-0.006123	-62.94	0.09837	0.003094	-12

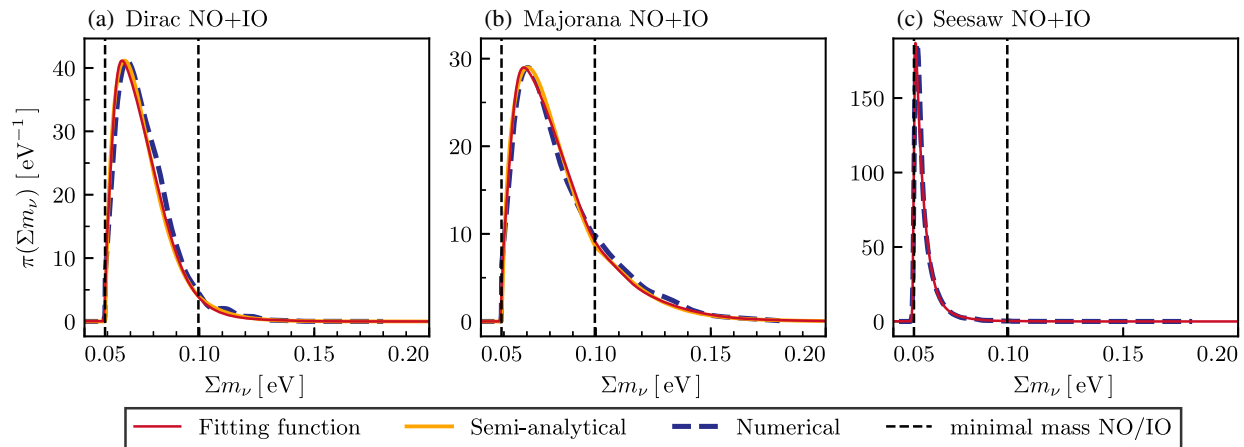


FIG. 7. The probability distribution over the neutrino mass sum after marginalizing over the neutrino energy scale  $\mu_\nu$ , marginalizing over the ordering of the mass spectrum, and imposing the mass splitting measurements. For the Dirac and Majorana models, there is an excellent agreement between the semi-analytical calculation and the fully numerical MCMC results; a semi-analytical calculation was not performed for the seesaw model. In all three cases, the empirical fitting function also matches extremely well.

allowed for the possibility of having a prior  $\Pi_i$  on the  $i$ th model. Note that within each mass model (Dirac, Majorana, or seesaw), we have assumed  $\Pi_{\text{IO}}/\Pi_{\text{NO}} = 1$ . Then to obtain the prior  $\pi(\Sigma m_\nu)$  appearing in Fig. 6, it is necessary to scale the fitting function by  $\mathcal{E}_{\text{NO}}/(\mathcal{E}_{\text{NO}} + \mathcal{E}_{\text{IO}})$  for the NO distribution and by  $\mathcal{E}_{\text{IO}}/(\mathcal{E}_{\text{NO}} + \mathcal{E}_{\text{IO}})$  for the IO distribution.

In Fig. 7 we show the distribution of  $\Sigma m_\nu$  after marginalizing over the two hierarchy to form the total distribution in the (D,M,S) models. We also showcase here the agreement between different ways of obtaining such distributions by plotting the fitting formula against the semianalytical results and the numerical ones. This total prior can then be used in cosmological parameter estimation.

As an example let us return to the test case considered in Sec. II. Recall that this test case involved cosmological data that constrained  $\Sigma m_\nu$  to an uncertainty of  $\sigma = 0.3$  eV around zero. If we repeat the same calculation for the priors in Table IV F instead, we obtain the following 95% CL upper limits:  $\Sigma m_\nu < 0.0975$  eV for (D), 0.126 eV for (M), and 0.0730 eV for (S) after marginalizing over the ordering with Eq. (35). Since the distributions are sharply peaked toward small masses, the inferred 95% CL upper limits are stronger than the ones we obtained for the flat-linear and flat-log priors in Sec. II.

## V. CONCLUSION

In this work we have discussed the choice of priors on the sum of neutrino masses  $\Sigma m_\nu$  for cosmological data analyses. Whereas it is customary to assume a flat prior on  $\Sigma m_\nu$ , which is an *ad hoc* choice, we have instead argued that the physically motivated choice is to specify the prior at the level of the neutrino mass matrix  $M_\nu$ . In this regard, our study extends earlier work on neutrino flavor model building with the anarchy hypothesis and applies that formalism to cosmological observables. Specifically, we

focus on the basis-independent anarchy hypothesis (BAH) that assigns equal probability to any matrices that can be related by a change of basis. Subject to the BAH restriction, simple priors on  $M_\nu$  include a flat or Gaussian distribution on its individual matrix elements.

One of the main conclusions of our work, is that these simple implementations of the BAH generally disfavor the degenerate regime where the neutrino mass scale is much larger than the mass splittings,  $\Sigma m_\nu \gg \sqrt{|\delta m^2|}$ ,  $\sqrt{|\Delta m^2|}$ . This is because the probability distribution over elements of the neutrino mass matrix  $M_\nu$  necessarily selects a scale  $\mu_\nu$ , e.g. the cutoff on a flat distribution or the variance of a Gaussian distribution. Due to the phenomenon of eigenvalue repulsion, high probability spectra have  $\Sigma m_\nu \sim \sqrt{|\delta m^2|} \sim \sqrt{|\Delta m^2|} \sim \mu_\nu$ , and in particular, degenerate spectra are especially unlikely. The repulsion effect can be understood from the Jacobian determinant that relates the distribution over (mass) matrix elements to the distribution over (mass) eigenvalues. This determinant [see Eq. (10) or (26)] is proportional to  $|\delta m^2|$  and  $|\Delta m^2|$ , which makes the probability density proportional to  $|\delta m^2|/\mu_\nu^2$  and  $|\Delta m^2|/\mu_\nu^2$ . Then in order for the measured mass splittings to be a probable realization, it is necessary to take  $\mu_\nu \sim \sqrt{|\Delta m^2|} \sim 0.05$  eV, see Fig. 5, which also implies  $\Sigma m_\nu \sim 0.05$  eV, see Fig. 6. Consequently, simple basis-independent priors at the level of the neutrino mass matrix translate into distributions over  $\Sigma m_\nu$  that peak around the smallest value allowed by the measured mass splittings (32), roughly 0.06 eV for normal ordering and 0.10 eV for inverted ordering.

Using the same reasoning as above, one can see that a flat distribution over  $\Sigma m_\nu$ , which is often assumed in cosmological studies, is highly improbable for *simple* basis-independent priors on  $M_\nu$ . In other words, this *ad hoc* prior assumption is not well motivated from fundamental

physical principles described by random, anarchical neutrino mass matrices.

In order to obtain qualitatively different results, it is necessary to choose a prior on  $M_\nu$  that is not *simple*. As we discuss in Sec. IV A, one can counterbalance the effects of eigenvalue repulsion by choosing the prior on the neutrino mass matrix to be the reciprocal of the Jacobian determinant. Although this leads to a wider tail in the  $\Sigma m_\nu$  distribution (see Appendix), now the prior is expressed as a very complicated function of the neutrino mass matrix, which undercuts the physical motivation for the anarchy hypothesis.

Our main quantitative results appear in Sec. IV where we focus on a particular prior distribution that is flat in the individual elements of the neutrino mass matrix out to a scale  $\mu_\nu$  for the Dirac and Majorana models, and it is flat in the separate high-scale Dirac and Majorana matrices for the seesaw model. We derive the distributions of the sum of neutrino masses and the squared mass splittings, and our results are discussed in Sec. IV C. Most notably, the predicted distribution over the sum of neutrino masses  $\Sigma m_\nu$  appears in Fig. 6. For all three models (Dirac, Majorana, and seesaw) the distribution is sharply peaked close to the lowest value allowed by the measured neutrino mass splittings, but the models are notably distinguished by the behavior in the tail of the distribution at high  $\Sigma m_\nu$ . We find that this prior prefers the mass spectrum with normal ordering over inverted ordering with odds 130:1 for (D), 19:1 for (M), and 470:1 for (S); see also Fig. 4 [cf., Ref. [9]]. We thoroughly tested the stability of these results under the various assumptions that are used to build the prior distributions; for a detailed discussion see Appendix.

The reader is encouraged to apply the prior probabilities  $\pi(\Sigma m_\nu)$  appearing in Fig. 6 to cosmological data analyses of other cosmological parameters, and an empirical fitting formula is available in Sec. IV F. Since the distributions in Fig. 6 are sharply peaked, this offers some justification for cosmological studies that simply fix  $\Sigma m_\nu$  to equal the minimal value consistent with the measured squared mass splittings (32), roughly 0.06 eV for NO and 0.1 eV for IO. In fact the priors that we have presented here may be useful from a phenomenological perspective if one seeks to have a prior that favors minimal mass but also allows for the possibility that a true preference by the data may drive the fit to larger  $\Sigma m_\nu$ . For practical applications, we recommend the reader start with the prior for the Majorana model, since it is least sharply peaked.

Let us close by discussing the impact of our analysis for cosmological probes of neutrino mass and their potential implications for neutrino mass models. Given the current sensitivity of the cosmological measurements, the priors  $\pi(\Sigma m_\nu)$  discussed here are presently more informative than the data. In this sense the priors can be viewed as providing targets for future experimental searches. In particular, our priors define several challenging objectives of increasing

experimental sensitivity, which can be inferred from marginalized distributions in Figs. 6 and 7 and their corresponding fitting functions in Sec. IV F. If you detect  $\Sigma m_\nu \gtrsim 0.13$  eV with sufficient experimental accuracy<sup>4</sup> then you rule out the Dirac model (i.e., it is disfavored at greater than 99.7% confidence); if you detect  $\Sigma m_\nu \gtrsim 0.18$  eV then you rule out the Majorana model; and if you detect  $\Sigma m_\nu \gtrsim 0.10$  eV then you rule out the seesaw model. By the time when the cosmological measurements reach this level of sensitivity to  $\Sigma m_\nu$ , the ordering of the neutrino mass spectrum may already be known from terrestrial experiments. If the neutrino mass spectrum has normal (inverted) ordering, and you detect  $\Sigma m_\nu \gtrsim 0.12$  eV (0.18 eV) then you rule out the Dirac model; if you detect  $\Sigma m_\nu \gtrsim 0.17$  eV (0.23 eV) then you rule out the Majorana model; and if you detect  $\Sigma m_\nu \gtrsim 0.097$  eV (0.17 eV) then you rule out the seesaw model. Recall however that the odds for normal versus inverted ordering in each case already disfavors any mass value for the latter case.

In this way, increasing accuracy to  $\Sigma m_\nu$  from cosmological observations will test the hypothesis that neutrino masses arise from unspecified high energy physics that can be described by effectively random mass matrices distributed such that matrices related by a change of flavor basis are equally probable. As we have explored in this work, the phenomenon of eigenvalue repulsion in the simplest implementation of the basis-independent anarchy hypothesis requires  $\Sigma m_\nu$  to be not much larger than the scale of the larger squared mass splittings,  $\sqrt{\Delta m^2} \sim 0.05$  eV, which leads to a tension with the data if  $\Sigma m_\nu$  is measured to be larger (see previous paragraph). If next-generation experiments measure a large  $\Sigma m_\nu$ , what does this imply for the basis-independent anarchy hypothesis? In such a situation, one is forced to abandon the simplest implementation of the BAH that we consider in this work. One can nevertheless construct potentially viable models if the prior on the mass matrix is chosen to counterbalance the eigenvalue repulsion, but such baroque priors are not in the *spirit* of the anarchy hypothesis and they undercut its physical motivation.

## ACKNOWLEDGMENTS

We are grateful to Jonathan Braden, Stephen Feeney, David J. E. Marsh, Eugene Lim, Hitoshi Murayama, Serguey Petcov, and Lian-Tao Wang for discussions. A. J. L. and W. H. are supported at the University of Chicago by the Kavli Institute for Cosmological Physics through Grant No. NSF PHY-1125897 and an endowment

<sup>4</sup>Here we mean that the experimental uncertainty is small in the sense that  $\sigma(\Sigma m_\nu) \ll (\Sigma m_\nu^{99.7} - \Sigma m_\nu^{\min})/3$  where  $\Sigma m_\nu^{\min}$  is given by Eq. (32) and  $\Sigma m_\nu^{99.7}$  is the quoted 99.7% CL threshold that gives the  $3\sigma$  Gaussian equivalent width of the prior distribution from the minimum. Otherwise, the experimental error should be combined appropriately into a joint significance.



from the Kavli Foundation and its founder Fred Kavli. M. R. and W. H. are supported by U.S. Dept. of Energy Contract No. DE-FG02-13ER41958. WH was additionally supported by NASA Grant No. ATP NNX15AK22G and the Simons Foundation.

### APPENDIX: STABILITY OF THE PRIOR ON $\Sigma m_\nu$

In order to assess the stability of our main results, which appear in Sec. IV C and specifically Fig. 6, we enumerate here the various assumptions, and we test the effect of relaxing or modifying each assumption. Overall we find that the prior  $\pi(\Sigma m_\nu)$  is very robust to what we called *simple* priors in the main text.

In calculating  $\pi(\Sigma m_\nu)$ , which appears in Fig. 6, we have made two key assumptions: the probability distribution is assumed to be flat in the matrix elements with a cutoff at  $\text{Tr} M_\nu^\dagger M_\nu = \mu_\nu^2$  (28); the cutoff  $\mu_\nu$  is marginalized with a flat prior from 0 eV to 1 eV.

Let us first discuss how our results depend upon the prior on the neutrino energy scale  $\mu_\nu$ . As we saw in Fig. 5, the distribution over  $\mu_\nu$  is sharply peaked. As such it is difficult to change the distribution  $\pi(\Sigma m_\nu)$  with relatively uninformative priors on  $\mu_\nu$ . To demonstrate this point we show in Fig. 8 the distribution  $\pi(\Sigma m_\nu)$  calculated from various different priors on  $\mu_\nu$ , namely (1.a) a flat prior on  $\mu_\nu \in (0 \text{ eV}, 1 \text{ eV})$ , (1.b) a flat prior on  $\mu_\nu \in (0 \text{ eV}, 10 \text{ eV})$ , (1.c) a flat prior on  $\log \mu_\nu \in (\log 0.1 \text{ eV}, \log 1 \text{ eV})$ , and (1.d) a flat prior on  $\log \mu_\nu \in (\log 0.01 \text{ eV}, \log 10 \text{ eV})$ . In these cases the evidences for the models read:  $\log_{10} \mathcal{E}(a) = 3.1$ ,  $\log_{10} \mathcal{E}(b) = 2.1$ ,  $\log_{10} \mathcal{E}(c) = 0.14$ , and  $\log_{10} \mathcal{E}(d) = 3.5$ . As we can see in Fig. 8 (1.a, 1.b, 1.d), once the  $\mu_\nu$  prior encompasses the region where the  $\mu_\nu$  posterior peaks, without adding significant curvature, the prior distributions  $\pi(\Sigma m_\nu)$  are almost identical. Case (1.a) is a factor ten preferred to (1.b) as the latter has wider prior on  $\mu_\nu$  that do not contain significant posterior. Case (1.d) is a factor two preferred with respect to case (1.a) as it assigns more weight to the region of smaller  $\mu_\nu$  where the posterior is peaked. When the prior on  $\mu_\nu$  excludes the region of maximum posterior, as in Fig. 8 (1.c), and selects out a particular value of  $\mu_\nu$ , the shape of the distribution is changed but the model has overall smaller probability with odds 800:1 with respect to case (1.a). In this case the hyperprior on  $\mu_\nu$  counteracts the natural tendency from eigenvalue repulsion that favors  $\Sigma m_\nu \sim \sqrt{|\Delta m^2|} \sim \mu_\nu \sim 0.05 \text{ eV}$  but also makes the observed splittings relatively unlikely; see Sec. V. While the figure only shows the (D) model, similar results hold for the (M) and the (S) models, relying on the peaked structure of the  $\mu_\nu$  posterior. In other words, the prior distribution  $\pi(\Sigma m_\nu)$  is insensitive to the prior on  $\mu_\nu$  as long as these priors on  $\mu_\nu$  have negligible curvature around the region of parameter space where the distribution of  $\mu_\nu$  peaks. Notice that in the (S)

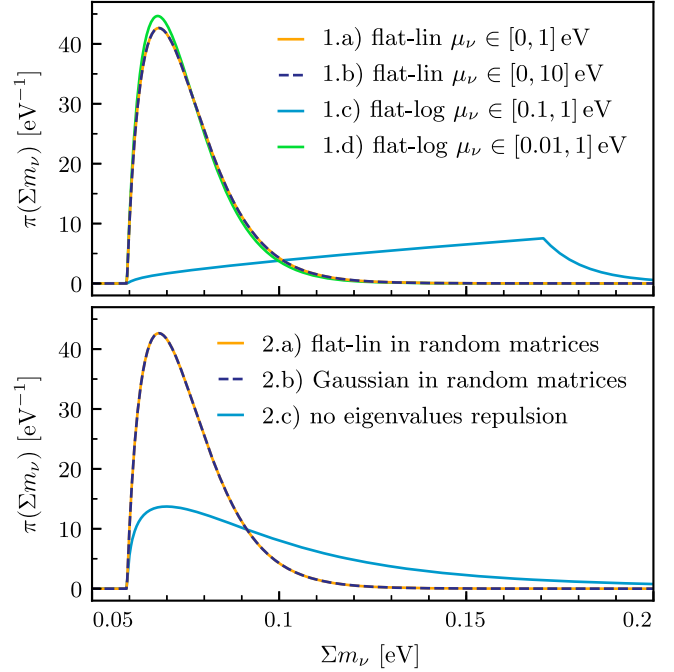


FIG. 8. Stability of the prior  $\pi(\Sigma m_\nu)$  against the various assumptions.

model a flat-log prior on the effective neutrino energy scale corresponds to a log prior on both the seesaw energy scales [ $\mu_D$  and  $\mu_M$  in Eq. (29)].

Let us next discuss how our results depend upon the prior over neutrino mass matrix elements. We also show in Fig. 8 the distribution  $\pi(\Sigma m_\nu)$  calculated from various different priors on  $M_\nu$ , namely (2.a) the flat prior  $p_\nu(\hat{M}_\nu|\mu_\nu) \propto \Theta(\mu_\nu^2 - \text{Tr}[\hat{M}_\nu^\dagger \hat{M}_\nu])$  from Eq. (28), (2.b) a Gaussian prior  $p_\nu(\hat{M}_\nu|\mu_\nu) \propto \exp[-\text{Tr}[M_\nu^\dagger M_\nu]/2\mu_\nu^2]$ , and (2.c) a prior that balances eigenvalue repulsion  $p_\nu(\hat{M}_\nu|\mu_\nu) \propto |\mathcal{J}_\nu(\hat{M}_\nu)|^{-1} (\det M_\nu)^{2/3}$  that was discussed in Sec. IV A. We marginalize each distribution over the neutrino energy scale with a flat distribution  $\mu_\nu \in (0, 1) \text{ eV}$ . For these cases the evidences are  $\log_{10} \mathcal{E}(a) = 3.1$ ,  $\log_{10} \mathcal{E}(b) = 2.4$ ,  $\log_{10} \mathcal{E}(c) = 4.1$ . The distribution  $\pi(\Sigma m_\nu)$  is nearly identical for the flat and Gaussian priors, although the evidence prefers the flat case. On the other hand, the third prior balances the eigenvalue repulsion factors in  $\mathcal{J}_\nu \sim |m_{\nu i}^2 - m_{\nu j}^2|$  by taking  $p_\nu \propto |\mathcal{J}_\nu|^{-1}$ , which increases the evidence, and the additional factor of  $p_\nu \propto (\det M_\nu)^{2/3}$  enhanced the high mass tail. In summary, the results presented already in Sec. IV C are robust against changing the prior on  $M_\nu$  due to the presence of eigenvalue repulsion, which prefers the mass scale of the prior to lie close to the mass scale of the measured neutrino mass splittings. In order to obtain qualitatively different results, it is necessary to choose a prior on  $M_\nu$  that counterbalances the eigenvalue repulsion.

- [1] P. A. R. Ade *et al.* (Planck Collaboration), *Astron. Astrophys.* **571**, A16 (2014).
- [2] P. A. R. Ade *et al.* (Planck Collaboration), *Astron. Astrophys.* **594**, A13 (2016).
- [3] S. Riemer-Sorensen *et al.*, *Phys. Rev. D* **85**, 081101 (2012).
- [4] N. Palanque-Delabrouille *et al.*, *J. Cosmol. Astropart. Phys.* **02** (2015) 045.
- [5] A. J. Cuesta, V. Niro, and L. Verde, *Phys. Dark Universe* **13**, 77 (2016).
- [6] K. A. Olive *et al.* (Particle Data Group), *Chin. Phys. C* **38**, 090001 (2014).
- [7] S. F. King, A. Merle, S. Morisi, Y. Shimizu, and M. Tanimoto, *New J. Phys.* **16**, 045018 (2014).
- [8] D. Sivia and J. Skilling, *Data Analysis: A Bayesian Tutorial*, Oxford science publications (Oxford University Press, New York, 2006).
- [9] F. Simpson, R. Jimenez, C. Pena-Garay, and L. Verde, *J. Cosmol. Astropart. Phys.* **06** (2017) 029.
- [10] T. Schwetz, K. Freese, M. Gerbino, E. Giusarma, S. Hannestad, M. Lattanzi, O. Mena, and S. Vagnozzi, [arXiv:1703.04585](https://arxiv.org/abs/1703.04585).
- [11] L. J. Hall, H. Murayama, and N. Weiner, *Phys. Rev. Lett.* **84**, 2572 (2000).
- [12] N. Haba and H. Murayama, *Phys. Rev. D* **63**, 053010 (2001).
- [13] A. de Gouvea and H. Murayama, *Phys. Lett. B* **573**, 94 (2003).
- [14] L. J. Hall, M. P. Salem, and T. Watari, *Phys. Rev. Lett.* **100**, 141801 (2008).
- [15] L. J. Hall, M. P. Salem, and T. Watari, *Phys. Rev. D* **76**, 093001 (2007).
- [16] L. J. Hall, M. P. Salem, and T. Watari, *Phys. Rev. D* **79**, 025010 (2009).
- [17] A. de Gouvea and H. Murayama, *Phys. Lett. B* **747**, 479 (2015).
- [18] X. Lu and H. Murayama, *J. High Energy Phys.* **08** (2014) 101.
- [19] J.-F. Fortin, N. Giasson, and L. Marleau, *Phys. Rev. D* **94**, 115004 (2016).
- [20] K. S. Babu, A. Khanov, and S. Saad, *Phys. Rev. D* **95**, 055014 (2017).
- [21] J.-F. Fortin, N. Giasson, and L. Marleau, *J. High Energy Phys.* **04** (2017) 131.
- [22] L. Susskind, [arXiv:hep-th/0302219](https://arxiv.org/abs/hep-th/0302219).
- [23] J. Bergstrom, D. Meloni, and L. Merlo, *Phys. Rev. D* **89**, 093021 (2014).
- [24] S. Hannestad and T. Tram, [arXiv:1710.08899](https://arxiv.org/abs/1710.08899).
- [25] E. Di Valentino, E. Giusarma, O. Mena, A. Melchiorri, and J. Silk, *Phys. Rev. D* **93**, 083527 (2016).
- [26] S. Hannestad and T. Schwetz, *J. Cosmol. Astropart. Phys.* **11** (2016) 035.
- [27] M. Gerbino, M. Lattanzi, O. Mena, and K. Freese, *Phys. Lett. B* **775**, 239 (2017).
- [28] E. Giusarma, M. Gerbino, O. Mena, S. Vagnozzi, S. Ho, and K. Freese, *Phys. Rev. D* **94**, 083522 (2016).
- [29] M. Gerbino, K. Freese, S. Vagnozzi, M. Lattanzi, O. Mena, E. Giusarma, and S. Ho, *Phys. Rev. D* **95**, 043512 (2017).
- [30] F. Couchot, S. Henrot-Versillé, O. Perdureau, S. Plaszczynski, B. R. D'orfeuil, M. Spinelli, and M. Tristram, *Astron. Astrophys.* **606**, A104 (2017).
- [31] S. Vagnozzi, E. Giusarma, O. Mena, K. Freese, M. Gerbino, S. Ho, and M. Lattanzi, *Phys. Rev. D* **96**, 123503 (2017).
- [32] H. Jeffreys, *Proc. R. Soc. A* **186**, 453 (1946).
- [33] R. E. Kass and L. Wasserman, *J. Am. Stat. Assoc.* **91**, 1343 (1996).
- [34] G. W. Anderson, A. Guionnet, and O. Zeitouni, *An Introduction to Random Matrices*, Cambridge Studies in Advanced Mathematics (Cambridge University Press, Cambridge, England, 2009).
- [35] P. Minkowski, *Phys. Lett.* **67B**, 421 (1977).
- [36] R. N. Mohapatra and G. Senjanovic, *Phys. Rev. Lett.* **44**, 912 (1980).
- [37] M. Gell-Mann, P. Ramond, and R. Slansky, *Conf. Proc.* **C790927**, 315 (1979).
- [38] T. Yanagida, *Prog. Theor. Phys.* **64**, 1103 (1980).
- [39] R. N. Mohapatra and G. Senjanovic, *Phys. Rev. D* **23**, 165 (1981).
- [40] J. Schechter and J. Valle, *Phys. Rev. D* **22**, 2227 (1980).
- [41] H. K. Dreiner, H. E. Haber, and S. P. Martin, *Phys. Rep.* **494**, 1 (2010).
- [42] F. Capozzi, E. Lisi, A. Marrone, D. Montanino, and A. Palazzo, *Nucl. Phys.* **B908**, 218 (2016).
- [43] D. Foreman-Mackey, D. W. Hogg, D. Lang, and J. Goodman, *Publ. Astron. Soc. Pac.* **125**, 306 (2013).
- [44] Getdist website, <https://github.com/cmbant/getdist>.
- [45] A. Heavens, Y. Fantaye, A. Mootooyaloo, H. Eggers, Z. Hosenie, S. Kroon, and E. Sellentin, [arXiv:1704.03472](https://arxiv.org/abs/1704.03472).
- [46] F. Feroz, M. P. Hobson, and M. Bridges, *Mon. Not. R. Astron. Soc.* **398**, 1601 (2009).
- [47] J. Buchner, A. Georgakakis, K. Nandra, L. Hsu, C. Rangel, M. Brightman, A. Merloni, M. Salvato, J. Donley, and D. Kocevski, *Astron. Astrophys.* **564**, A125 (2014).