

Least-informative priors for $0\nu\beta\beta$ decay searches

Frank F. Deppisch^{*} and Graham Van Goffrier[†]

*Department of Physics and Astronomy, University College London,
Gower Street, London WC1E 6BT, United Kingdom*

 (Received 29 March 2021; accepted 2 September 2021; published 27 September 2021)

Bayesian parameter inference techniques require a choice of prior distribution, which can strongly impact the statistical conclusions drawn. We discuss the construction of least-informative priors for neutrinoless double beta decay searches. Such priors attempt to be objective by maximizing the information gain from an experimental setup. In a parametrization using the lightest neutrino mass m_1 and an effective Majorana phase parameter Φ , we construct such a prior using two different approaches and compare them with the standard flat and logarithmic priors in m_1 .

DOI: [10.1103/PhysRevD.104.055040](https://doi.org/10.1103/PhysRevD.104.055040)

I. INTRODUCTION

Neutrinoless double beta ($0\nu\beta\beta$) decay is a hypothetical process of crucial interest due to its sensitivity both to the neutrino mass scale and to lepton-number violation. Direct searches for the decay, alongside neutrino oscillation studies and other probes of neutrino masses such as cosmological bounds [1] and single-beta decay measurements [2], are key to the improvement of our understanding of neutrinos. While a measurement of the $0\nu\beta\beta$ decay rate has not yet been made, upper bounds have been placed on the effective $0\nu\beta\beta$ mass $m_{\beta\beta}$, from which constraints on the neutrino mass scale and Majorana phases may be inferred.

Lacking evidence to the contrary, early formulations of the Standard Model (SM) took neutrinos to be massless. However, data suggesting the occurrence of neutrino flavor oscillations has accumulated since the 1960s, beginning with the Homestake Experiment [3] and culminating with the combined observation of oscillations for atmospheric neutrinos by Super-Kamiokande (SK), for solar neutrinos by the Sudbury Neutrino Observatory [4], and for reactor antineutrinos by the KamLAND-Zen Experiment [5]. From this, it is clear that oscillations necessitate a nonzero mixing angle as well as a nonzero mass difference; i.e., no more than one neutrino mass may be zero. For a model describing the three flavors ν_e , ν_μ , and ν_τ of neutrinos in the SM, mixing between weak and mass eigenstates is described by the Pontecorvo-Maki-Nakagawa-Sakata

(PMNS) matrix, and again nonzero mixing angles and two nonzero mass eigenvalues are required, with the additional possibility of CP violation.

The realization that at least two generations of neutrinos are massive leads to natural questions: what is the mechanism responsible for the neutrino masses, and how can all three masses be measured? Oscillation itself lends evidence to the latter question, as measurements have been made of all PMNS matrix elements as well as mass-squared differences $\Delta m_{21}^2 \sim 7.4 \times 10^{-5} \text{ eV}^2$ and $|\Delta m_{31}^2| \sim 2.5 \times 10^{-3} \text{ eV}^2$ [6].¹ The sign of Δm_{21}^2 is known, due to solar matter effects on oscillation [8], leading to two candidate mass hierarchies: the normal ordering (NO) $m_1 < m_2 < m_3$ and the inverse ordering (IO) $m_3 < m_1 < m_2$. Throughout this paper, we take these oscillation parameters to have their best-fit values from NuFIT4.0 + SK [9], for which uncorrelated errors are also accounted for in simulation.

Further information on neutrino parameters is provided by cosmological observations, as massive neutrinos play the unique role of both radiation during the early baryon acoustic oscillation epoch and hot dark matter during later formation of large-scale structure. With the reasonable assumption of equal number densities for all flavors, the sum of masses $\sum m_i$ is proportional to the total energy density of neutrinos in nonrelativistic eras. As a result, this quantity is observable in the redshift fluctuations of the cosmic microwave background, which are controlled by energy density at last scattering via the integrated Sachs-Wolfe effect as well as in a suppression of structural matter fluctuations due to freestreaming neutrinos. The latest fits from Planck observatory data [1] place an upper bound $\sum m_i < 0.12 \text{ eV}$ at 95% confidence, and efforts in this direction have great potential for further precision.

^{*}f.deppisch@ucl.ac.uk

[†]graham.vangoffrier.19@ucl.ac.uk

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Funded by SCOAP³.

¹A more recent global fit of oscillation parameters, which does not significantly impact on our results, may be found in Ref. [7].

Our focus is on $0\nu\beta\beta$ decay, which is a hypothetical nuclear transition $2n \rightarrow 2p^+ + 2e^-$ [10,11]. Total electron number is violated by two units, and the process is therefore not permitted in the SM with zero neutrino masses. It instead proceeds if the SM Lagrangian is extended by a Majorana mass term of the form $-\frac{1}{2}m\bar{\nu}_L^C\nu_L$. The decay process is sensitive to Majorana neutrinos with an observable effective mass,

$$m_{\beta\beta} = \sum_{i=1}^3 U_{ei}^2 m_i, \quad (1.1)$$

where PMNS matrix U includes Majorana phases that are unobservable in oscillation experiments.

The focus of the present work is the development of computational techniques for data-driven Bayesian inference on the $0\nu\beta\beta$ parameter space. Bayesian methodologies such as Markov chain Monte Carlo (MCMC) require a choice of prior distribution, which can strongly influence derived bounds. While it is standard to apply flat priors to bounded parameters such as Majorana phases, for unbounded parameters such as neutrino masses, there is less of a consensus. Recent Bayesian analyses have either preferred log-flat priors for their scale invariance [12] or have considered both flat and log-flat priors [13] in order to demonstrate the strong dependence of quantities such as discovery probability on the prior selection. In this paper, we examine a class of least-informative priors (LIPs) and apply them to the analysis of $0\nu\beta\beta$ decay. LIPs are constructed by numerical maximization of the expected Kullback-Leibler divergence between posterior and prior distributions [14], which is taken to represent inferential information gain and therefore to indicate a minimum quantity of information contained in the prior distribution.

In the frequentist interpretation of statistics, a perfect theory, whose parameters have an unknown but fixed value, is believed to exist. Bayesian statistics, however, treats theories as having associated degrees of belief. Provided some data, this approach allows a practitioner to infer a sensible probability that a candidate theory is correct. Given that the primary goal of particle physics experimentation is to reject or accept candidate theories and refine their parameters, this Bayesian inference methodology is a powerful analysis tool. We consider the light neutrino exchange mechanism for $0\nu\beta\beta$ decay and use published or Poisson-estimated likelihood functions from cosmological observations of $\sum m_i$ and direct searches for $m_{\beta\beta}$ to derive bounds on the neutrino masses and Majorana phases.

In Sec. II, we briefly summarize the key aspects of neutrinoless double beta decay and how MCMC studies are being used to infer neutrino parameters. We also introduce an effective parameter which combines the effect of the Majorana phases. In Sec. III, we outline the theoretical foundations for least-informative priors, and we detail an

algorithm for generating LIPs applicable to our physics context. Our results are presented in Sec. IV. Conclusions and an outlook are featured in Sec. V.

II. NEUTRINOLESS DOUBLE BETA DECAY AND NEUTRINO PARAMETER INFERENCE

Thinking of the three terms in Eq. (1.1) as complex numbers, two relative phases $-\pi \leq \alpha, \beta < \pi$ are sufficient to describe $m_{\beta\beta}$. Explicitly writing out the PMNS mixing matrix elements, this yields

$$\begin{aligned} m_{\beta\beta} &= c_{12}^2 c_{13}^2 m_1 + s_{12}^2 c_{13}^2 m_2 e^{i\alpha} + s_{13}^2 m_3 e^{i\beta} \\ &= A + B e^{i\alpha} + C e^{i\beta}, \end{aligned} \quad (2.1)$$

where s_{ij}^2 and c_{ij}^2 are shorthand for $\sin^2 \theta_{ij}$ and $\cos^2 \theta_{ij}$, respectively. The real-valued coefficients A , B , and C are functions of the neutrino masses and mixing parameters. The neutrino masses are not independent of each other, with the mass-squared splitting values fixed by oscillations, where the lightest mass m_l can be chosen as a free parameter; $m_l = m_1$ if the neutrinos are normally ordered and $m_l = m_3$ if they are inversely ordered. $0\nu\beta\beta$ experiments measure the half-life $T_{1/2}({}^A\text{X})$ of the decay of an isotope ${}^A\text{X}$ [12], which is connected to the effective double beta mass as

$$T_{1/2}^{-1}({}^A\text{X}) = |m_{\beta\beta}|^2 G_{0\nu} |\mathcal{M}_{0\nu}|^2. \quad (2.2)$$

Here, $G_{0\nu}$ is a phase-space factor encoding the leptonic part of the process including Coulomb effects between the nucleus and outgoing electrons, and $\mathcal{M}_{0\nu}$ is the nuclear matrix element (NME) of the underlying nuclear transition. Both depend on the isotope in question. Because of the large size of the relevant nuclei considered and correlations between nucleon states, the $0\nu\beta\beta$ NMEs are challenging to calculate, and disagreement between different nuclear models is a significant source of theoretical error in $0\nu\beta\beta$ studies [15]. A large number of experiments have succeeded in placing increasingly restrictive lower bounds on the $0\nu\beta\beta$ decay half-life. Recently, the KamLAND-Zen and GERDA experiments have determined the limits $T_{1/2}({}^{136}\text{Xe}) > 1.1 \times 10^{26}$ yr [16] and $T_{1/2}({}^{76}\text{Ge}) > 1.8 \times 10^{26}$ yr [17] at a 90% confidence level.²

A. Effective Majorana phase parameter

The parametrization of $0\nu\beta\beta$ decay arising from the neutrino mass matrix, while physically natural, introduces unnecessary complications into our inference. The Majorana phases entering it cannot be determined

²A Bayesian analysis is also performed by the GERDA Collaboration, obtaining a slightly weaker bound of $T_{1/2}({}^{76}\text{Ge}) > 1.4 \times 10^{26}$ yr for the same experimental data [17].

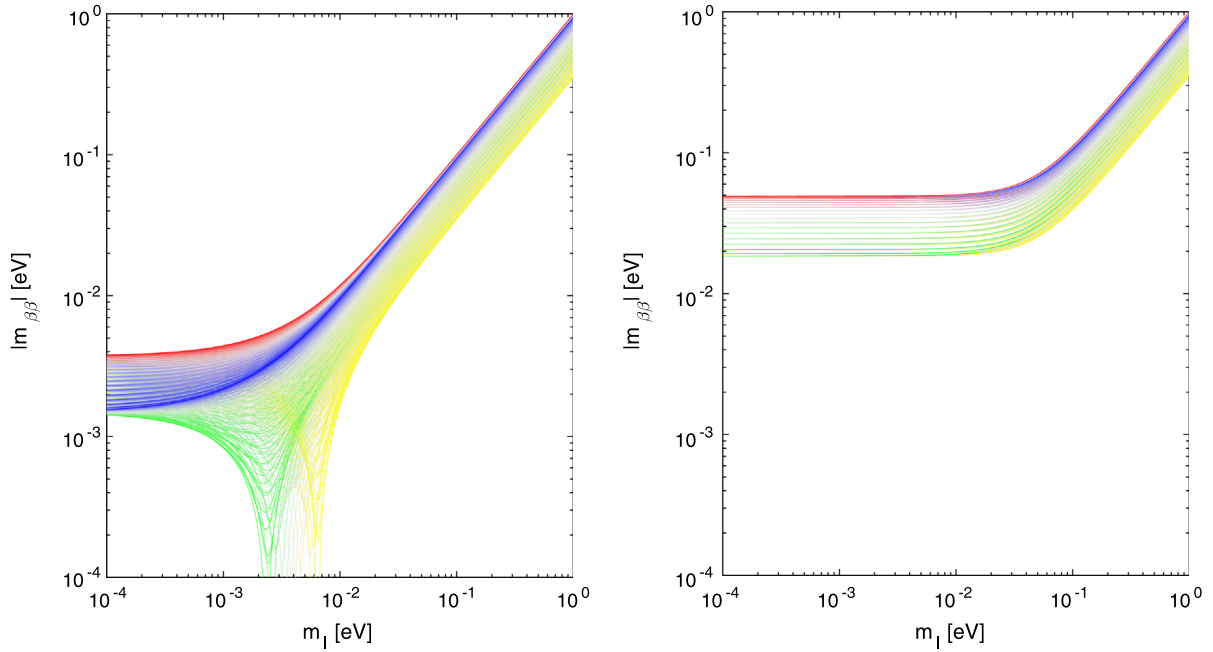


FIG. 1. Effective mass $|m_{\beta\beta}|$ as function of the lightest neutrino mass m_1 for constant sets of Majorana phases (α, β) , in the NO case (left) and IO case (right). The trajectories are color coded, interpolating (toroidally) between the four corners $(\alpha, \beta) = (0, 0)$ (red), $(0, \pi)$ (blue), $(\pi, 0)$ (green), and (π, π) (yellow).

individually [18], and a combination of both leads to the band structure observed. Figure 1 shows a continuum of $|m_{\beta\beta}|$ trajectories against the lightest neutrino mass m_1 for constant choices of the Majorana phases (α, β) , where both α and β are uniformly scanned over the range $[-\pi, \pi]$. In a two-dimensional (2D) rectilinear uniform distribution, volume effects imply that a point is more likely than not to be found near the boundary of the rectangle, and in particular near the corners. For both types of ordering, this and the functional dependence result in four bands of markedly high density, which are often but not always located near the theoretical limits on $|m_{\beta\beta}|$.

The key distinguishing feature of the NO case is its funnel, an interval $[m'_1, m''_1]$ within which it is possible to

choose α and β such that $m_{\beta\beta} = 0$ is attained. As shown in Eq. (2.1), we can think of $m_{\beta\beta}$ as the sum of three complex numbers $A + Be^{i\alpha} + Ce^{i\beta}$, where $A \approx 0.67m_1$, $B \approx 0.30m_2$, and $C \approx 0.02m_3$. This is depicted in Fig. 2 and in the IO case m_3 is too small for the sum to reach 0. In the NO case, many such closed triangles are found. Funnel edges m'_1 and m''_1 occur at $(\alpha, \beta) = (\pi, 0)$ and (π, π) , respectively, and are analytic over the neutrino masses and mixing angles. Using best-fit values from NuFITv4.0 + SK [6,9], $m'_1 = 2.330$ meV and $m''_1 = 6.535$ meV. For any $m_1 \in [m'_1, m''_1]$, there is a unique Majorana phase pair (α, β) which satisfies $m_{\beta\beta} = 0$. This choice of phase pair is extremely fine-tuned, leading to a statistical inaccessibility of the funnel in Fig. 1 (left) where $m_{\beta\beta} \lesssim 10^{-3}$; within the funnel interval in m_1 ,

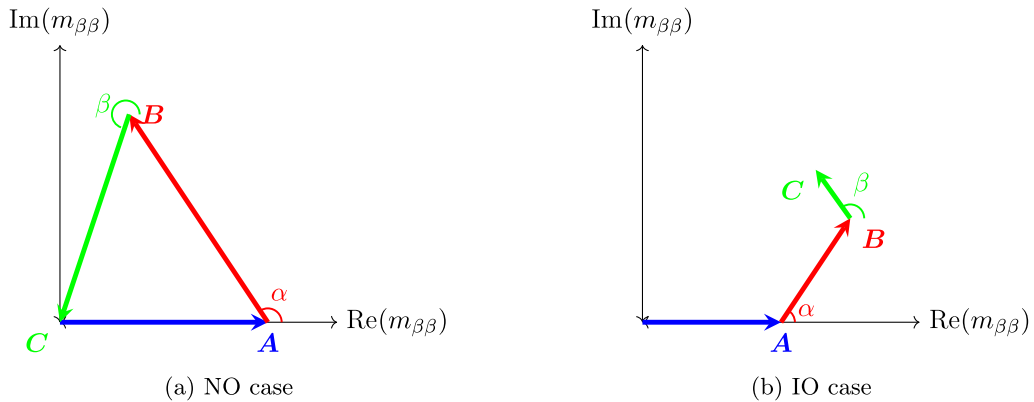


FIG. 2. Visualization of $m_{\beta\beta}$ as the sum of three complex numbers in examples of NO (left) and IO (right) scenarios. The NO example depicts a case where $m_{\beta\beta} = 0$.

the fraction of (α, β) parameter space with $|m_{\beta\beta}| < 10^{-3}, 10^{-4}$, and 10^{-5} eV is $\approx 10^{-1}, 10^{-3}$ and 10^{-5} , respectively.

The high-density banding and the inaccessibility of the funnel in the $|m_{\beta\beta}| - m_l$ parameter space due to the variation of the Majorana phases are a consequence of the parametrization chosen. In Bayesian language, we may also say that the corresponding flat prior on the phases is an arbitrary choice, given our lack of knowledge on the phases. Moreover, the two Majorana phases are somewhat redundant, and what truly matters is the range of values in $|m_{\beta\beta}|$ that can occur for a given m_l . For phase angles and parameters derived from phase angles, a flat prior is the standard choice [12] and may be most easily described by a linear parametrization. To capture the effects of the Majorana phases, we therefore introduce an effective phase parameter $0 \leq \Phi \leq 1$, which interpolates linearly in $|m_{\beta\beta}|$ between the boundaries of this permissible region. From Eq. (2.1), these boundaries occur for pairs of 0 and π Majorana phases, which leads to the definitions

$$|m_{\beta\beta}^{\text{NO}}| = \begin{cases} B + (A + C)(2\Phi - 1), & m_l \leq m_l', \\ (A + B + C)\Phi, & m_l' < m_l \leq m_l'', \\ A + (B + C)(2\Phi - 1), & m_l > m_l'', \end{cases} \quad (2.3)$$

for the NO case and

$$|m_{\beta\beta}^{\text{IO}}| = A + (B + C)(2\Phi - 1) \quad (2.4)$$

in the IO case, where A , B , and C depend explicitly upon m_l . In this parametrization, $|m_{\beta\beta}|$ changes linearly as Φ is varied linearly. This includes the funnel region of the NO case.

B. Bayesian methodology

A model (H, θ) is defined by both selecting a theory H and a vector of values θ in the space of continuous parameters Θ_H spanning that theory. The following probability densities may be defined:

- (i) $\pi(\theta) \equiv P(H, \theta)$ is the prior belief in the model (H, θ) before data collection;
- (ii) $L_x(\theta) \equiv P(x|H, \theta)$ is the likelihood of observing data x if the parametrized model (H, θ) is assumed to be true;
- (iii) $p(\theta|x) \equiv P(H, \theta|x)$ is the posterior belief in the model (H, θ) given observed data x .

While the posterior and prior distributions are probability densities over Θ_H , and therefore normalizable on this domain, the likelihood is instead a probability density over the space of measurable data X . Bayes' theorem relates these quantities in a manner analogous to statistical mechanics, where the posterior gives the probability density for the model to “occupy” state θ out of all possible parameter choices,

$$p(\theta|x) = \frac{L_x(\theta)\pi(\theta)}{\int_{\Theta_H} d\theta' L_x(\theta')\pi(\theta')} \equiv \frac{L_x(\theta)\pi(\theta)}{M_x^H}, \quad (2.5)$$

where the normalisation factor M_x^H is known as the marginal likelihood. In practice, the prior $\pi(\theta)$ is an educated guess, perhaps taking preceding experimental information into account, but presumed to be incomplete. As measured data become available, the prior probability is updated according to Bayes' theorem, and each calculated posterior probability becomes the new prior. Given enough data, this process converges to the true best-fit model regardless of error in the prior.³

In this paper, we take $0\nu\beta\beta$ decay to be mediated by light neutrino exchange as described above. Our model parameter space Θ_H is represented by the lightest neutrino mass m_l and the effective Majorana phase parameter Φ , $(m_l, \Phi) \in \Theta_H$. For simplicity, the other neutrino oscillation parameters are fixed at their best-fit values, and the mass ordering scenario, NO or IO, is considered to be known. The natural hypothesis to consider is the observation of a certain number of signal events n in a $0\nu\beta\beta$ experiment, and so our data space X is represented by the possible counts $n = 0, 1, 2, \dots$. This framework is easily applied to the comparison of multiple hypotheses, e.g., NO vs IO neutrino mass hierarchies, by computing marginal likelihoods for both models given the same observed data. The ratio of these quantities $K = M_x^{\text{NO}}/M_x^{\text{IO}}$ is the Bayes' factor of the NO hierarchy versus the IO hierarchy.

Data available from current or upcoming $0\nu\beta\beta$ decay experiments in isolation are insufficient to claim convergence of the posterior distribution—instead, predictions and bounds on model parameters are sought. All such quantities are expressible as posterior integrals [19], which we calculate using samples obtained by MCMC with the Metropolis-Hastings algorithm [20].

III. LEAST-INFORMATIVE PRIORS

We now consider the impact of prior selection on Bayesian inference from $0\nu\beta\beta$ decay. In the limit of perfectly precise measurements of observables which fully cover the given parameter space, the bias introduced by a prior vanishes. Unfortunately, this is not the situation for any real experimental outcome, and any assumption implicitly made by a prior distribution must contribute to the posterior, as demonstrated for $0\nu\beta\beta$ in particular by Ref. [21]. It is therefore advantageous, to the practitioner who wishes to avoid inferential bias, to choose priors which assume the least about the outcome of the experiment, and are in this sense “uninformative.” In this section, we study

³An exception is priors using Dirac δ -function shapes imprinting on the posterior, $p(\theta|x) = \pi(\theta) = \delta(\theta - \hat{\theta})$, thereby leading to an incorrect result regardless of data. Similarly, sharp prior distributions tend to converge very slowly.

LIPs, which are intended to maximize the expected information gained through measurement and inference and develop a methodology for the case at hand.

A. Theoretical construction of reference priors

It is of import to first cite Ref. [22] for a justification of the overall methodology of this paper, that is, to conduct Bayesian inference with a diverse set of priors in order to arrive at the fullest picture of what our data imply. The situation in scientific experiment is not so different from that in psychological studies of human behavior, where personal knowledge or opinions play a nontrivial role in even the most rational decision making. By employing priors which take account of some full range of acceptable prior beliefs which an experimenter might hold, we can both gain confidence in inferences which hold broadly across the considered priors and quantify the variation of inferred bounds or measurements between priors.

However, even if we are persuaded that no single prior can offer a complete understanding of any statistical inference, it becomes necessary to establish a *reference prior* against which the performance of all other priors might be consistently compared. A general procedure is developed in Ref. [14], which may be applied to any inference, for identifying such a prior as the solution to an optimization problem over information gain: an LIP. This procedure is summarized as follows, applied separately for each parameter θ_i with all others fixed.

First, taking only the requirement that information gained from multiple measurements is additive, the information contained in a distribution $P(\theta_i)$ is [23,24]

$$I = \int_{P(\theta_i) \neq 0} d\theta_i P(\theta_i) \log P(\theta_i), \quad (3.1)$$

which is familiar in physics as the Boltzmann-Gibbs entropy for a continuous collection of states, up to a constant factor [25]. We then take an experiment E , which measures data x , with likelihood function $L_x(\theta_i)$. Following Ref. [24], Ref. [26] calculates the expected information gain of prior $\pi(\theta_i)$ as

$$\begin{aligned} I\{E, \pi\} &= \int dx \int d\theta_i L_x(\theta_i) \pi(\theta_i) \log \left[\frac{p(\theta_i|x)}{\pi(\theta_i)} \right] \\ &= \int dx M_x \int d\theta_i p(\theta_i|x) \log \left[\frac{p(\theta_i|x)}{\pi(\theta_i)} \right], \end{aligned} \quad (3.2)$$

where $p(\theta_i|x)$ is the posterior given by Bayes' theorem, and $M_x = \int d\theta_i L_x(\theta_i) \pi(\theta_i)$ is the marginal likelihood of data x . The inner integral on the second line of Eq. (3.2) is known as the Kullback-Leibler divergence $K[p, \pi]$ between $p(\theta_i|x)$ and $\pi(\theta_i)$ [26], of which $I\{E, \pi\}$ is therefore an expectation value over the data space. Whether phrased as a prior or posterior integral, this quantity depends strongly on

the choice of prior, which appears implicitly in the inference of the posterior and as the measure over θ_i in the marginal likelihood. Letting $E(k)$ indicate k independent replications of experiment E , the quantity $I\{E(\infty), \pi\}$ describes the vagueness of prior $\pi(\theta_i)$ [14], as an infinite quantity of well-defined experiments must arrive at the same precise measurement, and so a greater expected information gain through inference implies that more information was missing to begin with.

The prior π which maximizes $I\{E(\infty), p\}$ cannot simply be selected because an infinite quantity of information is needed to measure any parameter exactly. Instead, a limit must be taken as the number of measurement repetitions k approaches ∞ . For a given k , the reference prior $\pi_k(\theta)$ is defined as that among all permissible priors [27] which maximizes $I\{E(k), \pi\}$, where permissibility is defined using boundedness and consistency arguments over compact subsets of the parameter space. Given any measurement x , the reference posterior $p_k(\theta_i|x)$ corresponding to prior $\pi_k(\theta)$ is calculable by Bayes' theorem, and assuming compactness on the set of possible posteriors, the limit $p(\theta_i|x) = \lim_{k \rightarrow \infty} p_k(\theta_i|x)$ is well defined. Because of consistent validity of Bayes' theorem across the measurement domain, a prior $\pi(\theta_i)$ proportional to $p(\theta_i|x)/L_x(\theta_i)$ is then a well-defined LIP, which is independent of x .

Obtaining the LIP via a posterior limit of course does not feel very efficient, but so long as certain regularity conditions are met [27], a limiting sequence among priors which still maximizes Eq. (3.2) may be found:

$$\pi_k(\theta_i) \propto \exp \left\{ \int dx L_x(\theta_i) \log [p^*(\theta_i|x^k)] \right\}. \quad (3.3)$$

Here, x^k is a collection of data from k repeated measurements, $\pi^*(\theta)$ is an initialization prior chosen among any in the permissible set, and k is taken to be large enough that the posterior $p^*(\theta|x^k)$ induced by prior $\pi^*(\theta)$ is dominated by the characteristics of the likelihood rather than by that prior.

A subtlety of this construction is that a direct limit $\lim_{k \rightarrow \infty} \pi_k$ can be poorly behaved at singularities and boundaries, resulting in a comblike LIP. In such cases, the LIP can instead be defined as a conditioned limit at some well-behaved parameter point $\theta_{i,0}$ [27],

$$\pi(\theta_i) = \lim_{k \rightarrow \infty} \frac{\pi_k(\theta_i)}{\pi_k(\theta_{i,0})}. \quad (3.4)$$

In addition to maximizing the expected inferential information gain, the generated LIP enjoys simple Jacobian transformation under reparametrizations of θ [14].

B. Implementation of LIP algorithm for $0\nu\beta\beta$

The application of LIPs to neutrino oscillation experiments is explored in Ref. [26]. When an experiment is such

that the asymptotic posterior is well approximated by a Gaussian distribution (a condition known as asymptotic posterior normality), it can be shown that the LIP is simply the multivariate Jeffreys prior of the likelihood [14]. This assumption holds and significantly simplifies computation for oscillation studies of the neutrino mass splittings and cosmological studies of the sum-of-masses Σ . However, in $0\nu\beta\beta$ decay, the observable of interest $|m_{\beta\beta}|$ may be asymptotically small, or may even vanish if neutrinos are not Majorana particles. Asymptotic posterior normality can therefore not be said to hold for any $0\nu\beta\beta$ decay search, and instead, we follow the computational procedure set out by Berger [27] for solving Eqs. (3.3) and (3.4) in full generality.

So long as an amenable likelihood model is chosen, the LIP computation expressed in Eq. (3.3) is well suited to a sampling procedure. We employ Berger’s method of selecting m sets of k likelihood samples each: for each set, Bayesian inference is made using a fiducial flat prior and the product of the k likelihoods, as the samples are treated as occurring from independent experiments. The resultant posterior distribution is used to compute Eq. (3.3), where the numerical integration is performed by averaging over the m sample sets.

Following Ref. [13], we define our $0\nu\beta\beta$ measurement model to be a Poisson counting experiment, where the likelihood of observing n counts given a background expectation λ and signal expectation ν is

$$L_n(\nu) = e^{-\lambda-\nu} \frac{(\lambda + \nu)^n}{n!}. \quad (3.5)$$

The number of expected signal events ν is related to the $0\nu\beta\beta$ decay half-life by

$$\nu = \frac{\mathcal{E} N_A \log 2}{m_{\text{iso}}} T_{1/2}^{-1}, \quad (3.6)$$

where N_A is Avogadro’s constant, m_{iso} is the molar mass of the enriched isotope used in detection, and \mathcal{E} is the sensitive exposure, also accounting for detection efficiency. In our simulations for the LEGEND-200 ^{76}Ge experiment, we take one year of run time with $m_{\text{iso}} = 75.921$ u and $\lambda = 1.7 \times 10^{-3}$ cts/(kg · yr) · \mathcal{E} with sensitive exposure $\mathcal{E} = 119$ kg · yr [12]. The half-life $T_{1/2}(|m_{\beta\beta}|)$ also depends on the isotope through the phase-space factor and nuclear matrix element, where we use the values $G_{0\nu} = 3.04 \times 10^{-26}$ yr $^{-1}$ eV $^{-2}$ and $|\mathcal{M}_{0\nu}| = 4.32$ [28].

Berger’s algorithm discussed above applies only to a single parameter θ_i , and therefore in a multiparameter problem such as ours, the LIP must be obtained sequentially. Following Ref. [26], at step j in the iteration, the prior on θ_j is computed using fixed values of all $\theta_{i>j}$, written $\pi(\theta_j|\theta_{i>j})$. The likelihood function is then marginalized by parameter θ_j ,

$$L_x(\theta_{i>j}) = \int d\theta_j \pi(\theta_j|\theta_{i>j}) L_x(\theta_{i\geq j}). \quad (3.7)$$

Note that this procedure must be repeated for each combination of parameter values for which we seek to know the LIP, placing a strong bottleneck on achievable precision. At the end of the iteration, the total LIP is given by the product

$$\pi(\theta) = \prod_{j \leq n} \pi(\theta_j|\theta_{i>j}). \quad (3.8)$$

For a nonseparable likelihood function, this depends on the ordering of parameters [26], with more impactful parameters customarily ordered first; we therefore take $\theta_1 \equiv m_l$ and $\theta_2 \equiv \Phi$. The resultant two-stage LIP algorithm is summarized in Fig. 6 in the Appendix. A “free-phi” approximation is also considered, in which the above multiparameter procedure is still followed, but the Φ -likelihood $L_n^\Phi(\Phi)$ is computed over a flat m_l prior, thereby removing costly interpolation evaluations.

From a practical standpoint, the number of repetitions m impacts the precision of the final LIP, while the sample quantity k affects its accuracy. If m is too small, the prior may be noisy but still accurate, while a k far from the convergence region could lead to a prior which is far from least informative. We select $m = 100$ for our simulations, sufficiently large to be near convergence but small enough to avoid precision errors as the product of likelihoods dips near 10^{-200} . The outer loops of the algorithm are parallelizable, and so a 16-core Message Passing Interface implementation of the algorithm leads to significant speed-up, allowing for k up to 2000 with approximately 12 h run times.

IV. RESULTS

A. Generated LIPs for LEGEND-200

To illustrate the above procedure, we choose the future $0\nu\beta\beta$ decay experiment LEGEND-200 [29] as the basis for a measurement example. LEGEND-200 plans to use 175 kg of the isotope ^{76}Ge to achieve a 3σ sensitivity to half-lives greater than 10^{27} yr. This corresponds to an expected number of background events of $\lambda = 1.7 \times 10^{-3}$ cts/(kg · yr) · \mathcal{E} with sensitive exposure $\mathcal{E} = 119$ kg · yr [12] taken over one year of run time as an example to illustrate our algorithm. Here, the sensitive exposure \mathcal{E} is defined as the product of the total exposure with fiducial volume and signal detection efficiencies, also accounting for a 2σ region of interest around the decay energy. Two configurations of the LIP algorithm were considered: the full two-parameter integration discussed above and specified in Fig. 6 in Appendix and the free-phi approximation.

The results of our simulation, with 2×10^5 likelihood draws at each parameter point (m_l, Φ) across a grid with resolutions $\Delta\Phi = 0.1$ and $\Delta \log_{10}(m_l) = 0.2$, are shown in

Fig. 3. The top left and top right plots show the LIP $\pi(m_l, \Phi)$ in the NO scenario calculated using the full simulation and the free-phi approach, respectively. They demonstrate that the free-phi approximation is generally valid throughout the parameter space. The greatest deviation occurs for large values of both parameters, a region where relevant likelihoods tend to be very low, and the impact upon posterior inference is therefore expected to be negligible. It should be noted that slow convergence at parameter boundaries causes erroneous growth of the raw LIP, producing boundary walls which are locally smoothed in postprocessing to produce the visuals throughout this work and before MCMC or information computations are performed. Signal-to-noise ratio (SNR) analysis of repeated trials showed a minimum SNR of 1000, or 30 dB, between individual generated LIPs and averages of three LIPs, as a

reference. Practically, this corresponds to a noise amplitude of at most ± 0.05 everywhere in the LIP distribution.

In the right column of Fig. 3, LIPs computed for NO (top) and IO (bottom) neutrino mass orderings are compared and seen to be structurally similar. Both priors feature a near-linear increase with Φ and a significant trough in m_l between 10^{-2} and 10^{-1} eV. On either side of this trough, the prior is nearly flat in m_l , with higher density for $m_l > 0.1$ eV. Note that these functions are distributions over m_l , plotted on a logarithmic scale, rather than distributions over $\log(m_l)$, which would feature the presence of an additional factor $1/m_l$ from the Jacobian transformation.

It is significant that a predominantly flat prior in m_l emerges from a first-principles Bayesian simulation, perhaps indicating the naturalness of a flat prior for unknown particle masses. The trough may be understood physically as the region of

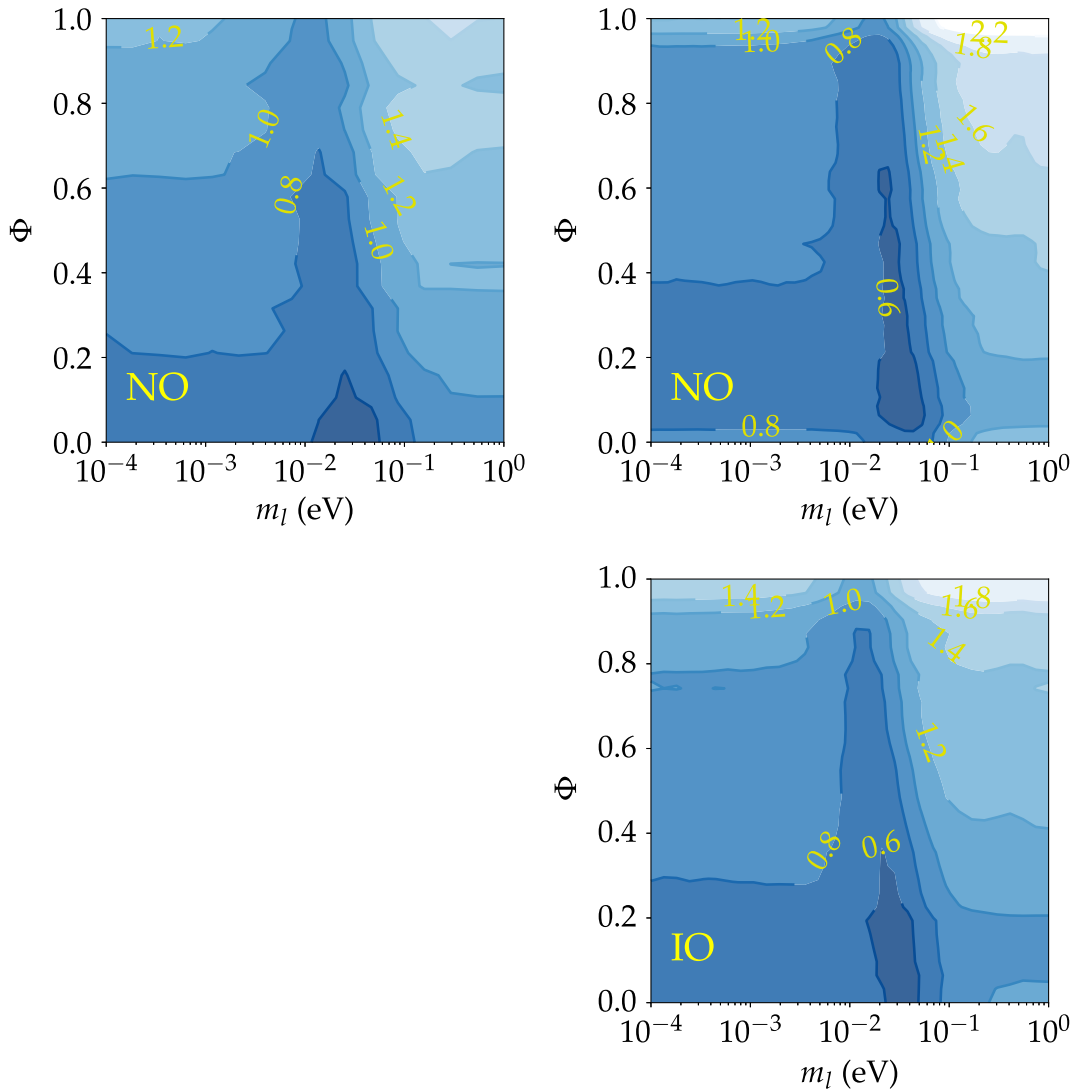


FIG. 3. Least-informative priors $\pi(m_l, \Phi)$ in terms of the lightest neutrino mass m_l and the effective Majorana phase parameter Φ based on a likelihood using the LEGEND-200 experiment [expected background events $\lambda = 1.7 \times 10^{-3}$ cts/(kg \cdot yr) \cdot \mathcal{E} with sensitive exposure $\mathcal{E} = 119$ kg \cdot yr]. The LIPs in the NO scenario are computed using the full algorithm (top left) and free-phi integration (top right). The bottom right plot shows the LIP for IO using free-phi integration.

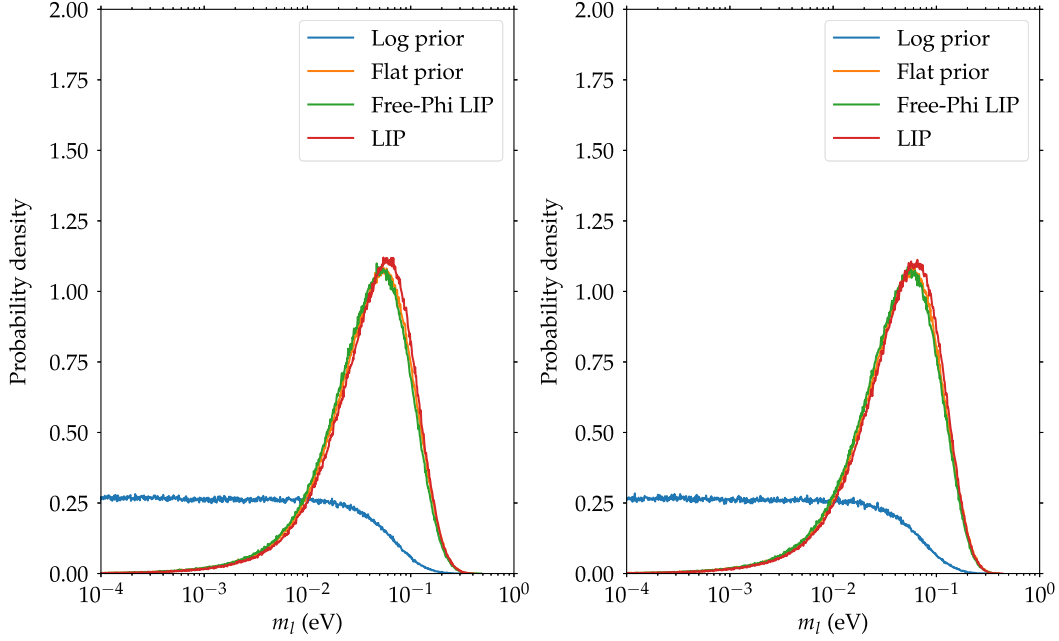


FIG. 4. Posterior distributions over m_l , computed from marginalized MCMC samples for the projected LEGEND-200 likelihood for NO (left) and IO (right) in the case of nonobservation (count $n = 0$). The curves correspond to the different priors used: log prior (log-flat m_l and flat Φ , blue), flat prior (flat m_l and flat Φ , orange), free-phi LIP (LEGEND-200 LIP with free-phi integration, green), and LIP (full LEGEND-200 LIP, red).

parameter space where LEGEND-200 has the greatest propensity to make either a measurement or an exclusion; the LIP therefore reduces the weight in this region so that any inferences made can be said to more fully data driven.

B. Comparison of inferred bounds on m_l

We evaluate the performance of the computed LIPs by utilizing them (after bicubic spline interpolation) in

our MCMC analysis, equipped with the effective parametrization (m_l, Φ) . Only the experimental likelihood from LEGEND-200 is included, modeled with a Poisson distribution following Ref. [13] with $\lambda = 1.7 \times 10^{-3}$ cts/(kg·yr)· \mathcal{E} with sensitive exposure $\mathcal{E} = 119$ kg·yr [12].

We consider first the case where LEGEND-200 registers $n = 0$ signal events during its run. We include 10^7 MCMC samples, using a Gaussian proposal distribution of width

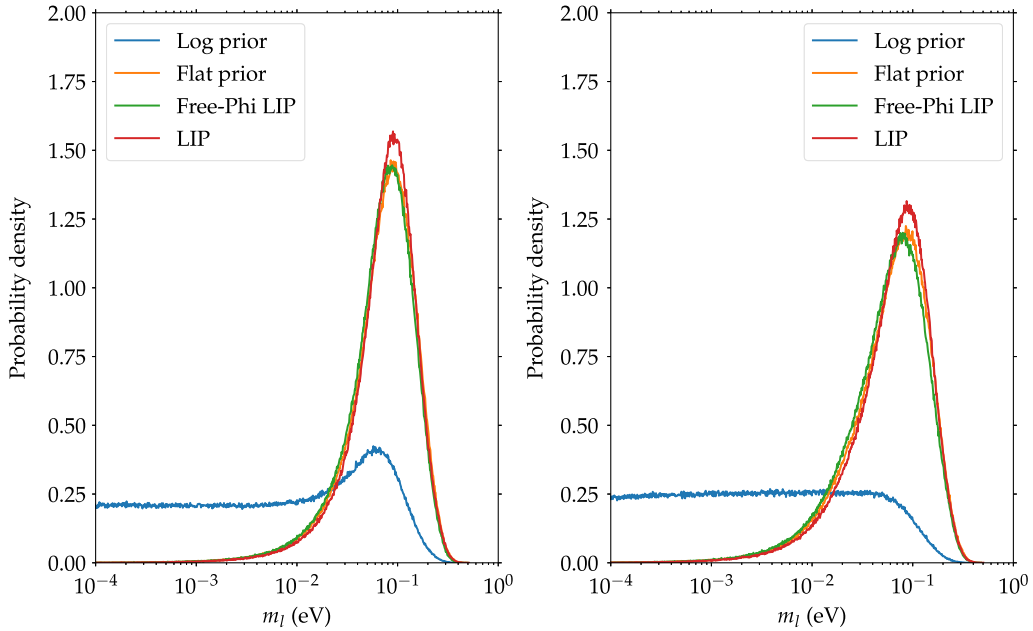


FIG. 5. As Fig. 4, but in the case of observation of one signal event, $n = 1$.

TABLE I. Kullback-Leibler divergences (in bits) for different priors, computed from MCMC LEGEND-200 posteriors in the case of NO and IO as well as observed counts $n = 0, 1$, as indicated. The entries in bold denote the highest divergence achieved among the different priors used.

Prior	NO, $n = 0$	IO, $n = 0$	NO, $n = 1$	IO, $n = 1$
Log prior	14.64	14.63	15.90	15.23
Flat prior	20.63	20.61	20.63	20.73
Free-phi LIP	21.92	21.94	21.76	21.81
LIP	22.08	22.09	21.90	22.04

10^{-2} in both m_l and Φ . The resulting posterior distributions for flat (in both m_l and Φ) and log-flat (flat in $\log m_l$ and Φ) priors as well as both LIP implementations are shown in Fig. 4 for NO (left) and IO (right). A further confirmation of the strong match between free-phi and full LIP calculations is gained, and both are seen to lead to very similar inferences as the flat prior. In the NO case, these three priors exclude $m_l > 0.11$ eV at a 90% credibility level,⁴ while the log-flat prior excludes $m_l > 0.03$ eV; in the IO case, these upper bounds are expectedly slightly higher, though within simulation error.

Next, we set the observed count to $n = 1$ and repeat the inference procedure, with results shown in Fig. 5. Again, both calculated LIPs give similar posteriors to the flat prior. In the NO case, Fig. 5 (left), these three priors result in $m_l = 90 \pm 50$ meV. In the IO case, Fig. 5 (right), the flat and LIP measurements give slightly lower values, but with comparable precision. However, the log-flat prior fails to make any measurement in either hierarchy, instead placing an exclusion at 90% credibility on $m_l > 75$ meV in the NO case and on $m_l > 45$ meV in the IO case, as the measurement falls squarely within the region extending to low m_l which log-flat priors are intended to probe. It is not surprising that LEGEND-200 fails to distinguish between the neutrino-mass hierarchies, as it does not probe sufficiently small $|m_{\beta\beta}|$ where the allowed parameter regions notably diverge.

C. Information content of inferences

However, the value of LIPs does not lie in their propensity to give a conservative bound or measurement (which the flat prior already achieves) but in their trustworthiness as a reference prior with minimized bias. In Table I, we report the Kullback-Leibler divergences of each MCMC posterior against its prior, using the same LEGEND-200 likelihood as above. Error propagation was performed by considering a noisy LIP of the form $\pi(\theta) = \pi_{\text{true}}(\theta) \pm n_\pi(\theta)$, which in the limit of small noise (and assuming that variation in the posterior is dominated by variation in the prior) corresponds to noisy Kullback-Leibler divergence,

⁴Bayesian credibility intervals may be thought of as analogous to frequentist confidence intervals.

$$K[p, \pi] = K[p, \pi_{\text{true}}] \pm \frac{1}{\log(2)} \int d\theta p(\theta) \frac{n_\pi(\theta)}{\pi_{\text{true}}(\theta)}, \quad (4.1)$$

where the integrated error may be interpreted as the posterior expectation value of the inverse of the signal-to-noise ratio distribution for $\pi(\theta)$. Computation of this integral for the worst-case noise distribution mentioned in Sec. IV. 1 gave an error of ± 0.03 bits, leading us to quote our divergence values to 0.01 bit precision.

The results give a numerical confirmation of the similarity between free-phi and full LIP computations, whose divergences consistently fall within 0.25 bits. In all configurations, the LIPs outperform both standard priors in information gain, as expected from their construction.

V. CONCLUSION

Bayesian parameter inference is a common tool to constrain or determine parameters in particle physics. An inherent issue in this context is the choice of a prior distribution over the model parameters. We have here focused on the neutrino parameter space relevant to $0\nu\beta\beta$ decay searches, specifically the lightest neutrino mass m_l and an effective Majorana phase parameter Φ encapsulating the effect of the Majorana phases in the lepton mixing matrix. Given that $0\nu\beta\beta$ decay has not been observed yet, prior distributions are expected to have a strong impact on the conclusions drawn.

We have adapted an algorithm for computing least-informative priors for a given experiment via likelihood sampling to the case of $0\nu\beta\beta$ direct searches, resulting in exact and approximate parallelized implementations. The LIPs were seen to take the form of a flat- m_l , linear- Φ distribution broken by a trough between $m_l = 10^{-2}$ and 10^{-1} eV. We demonstrated that for the proposed 200 kg ^{76}Ge LEGEND experiment these priors give similar posterior bounds to the usually adopted flat prior for both neutrino orderings, and in both observation and nonobservation scenarios. Furthermore, the LIPs were seen in nearly all cases to outperform both standard flat and logarithmic priors as far as their information gain during MCMC inference is concerned. This supports the functionality of the adapted algorithm and strengthens the argument for the use of LIPs as reference priors for $0\nu\beta\beta$ decay searches.

Natural extensions of this work include a study of the variation in LIP performance across proposed experiments of diverse background levels and exposures, simulation of LIPs for the usual parametrization using two Majorana phases, and research toward the construction of a prior which is jointly least informative over both the $0\nu\beta\beta$ observable $|m_{\beta\beta}|$ and the cosmology observable $\sum m_i$.

ACKNOWLEDGMENTS

The authors would like to thank Matteo Agostini for useful discussion. The authors acknowledge support from the United Kingdom Science and Technology Facilities Council (STFC) via the Consolidated Grants No. ST/P00072X/1 and No. ST/T000880/1. G. V. G. also acknowledges support from the UCL Centre for Doctoral Training in Data Intensive Science funded by STFC.

APPENDIX: LIP ALGORITHM FLOWCHART

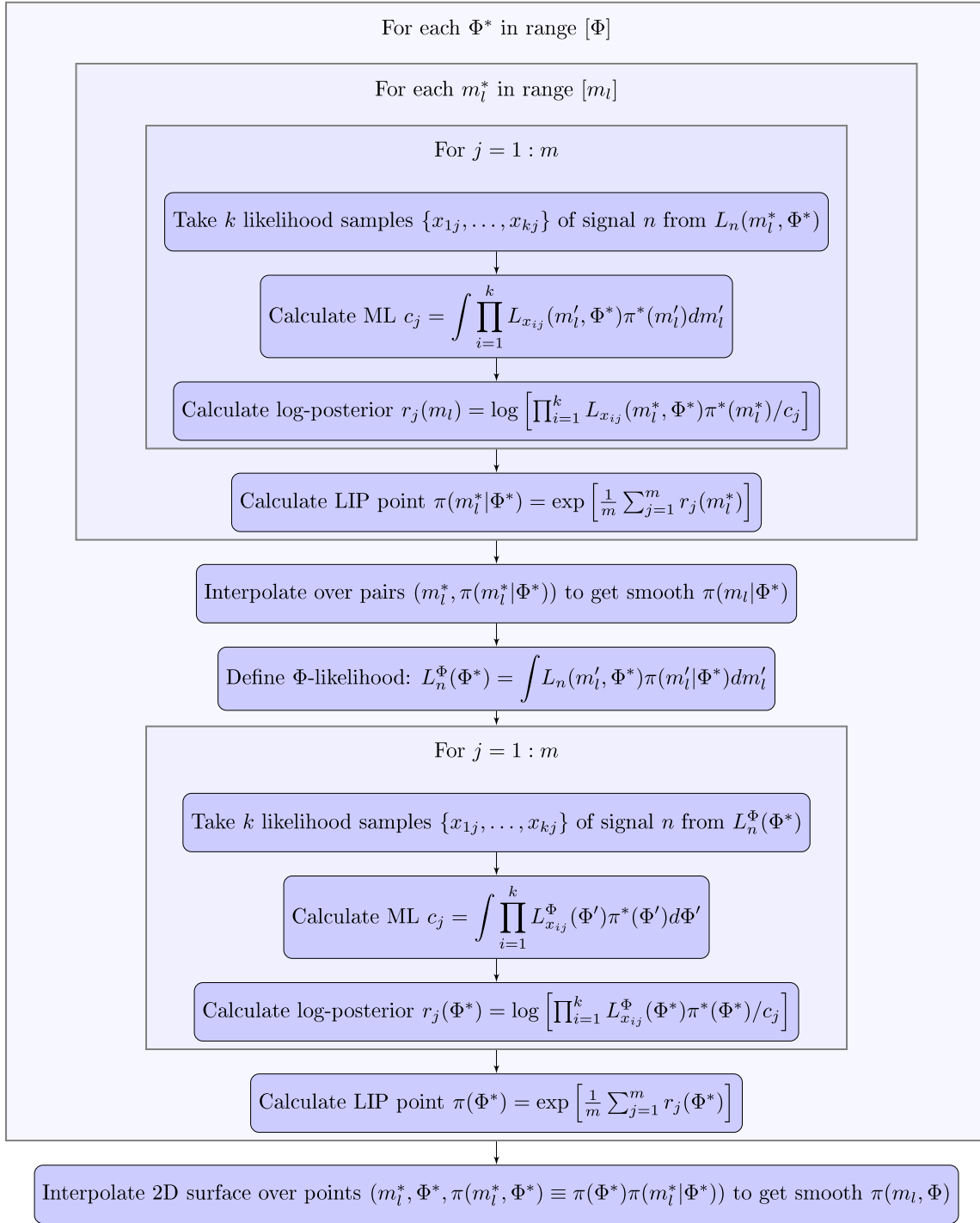


FIG. 6. Algorithm for generating a least-informative prior on the (m_l, Φ) parameter space, given experimental settings λ , m_{iso} , and \mathcal{E} , and fiducial priors $\pi^*(m_l^*)$ and $\pi^*(\Phi^*)$.

- [1] N. Aghanim, Y. Akrami, M. Ashdown, J. Aumont, C. Baccigalupi, M. Ballardini *et al.*, Planck 2018 results. VI. Cosmological parameters, *Astron. Astrophys.* **641**, A6 (2020).
- [2] S. Mertens *et al.* (K. Collaboration), Status of the KATRIN experiment and prospects to search for keV-mass sterile neutrinos in tritium β -decay, *Phys. Procedia* **61**, 267 (2015).
- [3] B. T. Cleveland, T. Daily, J. Raymond Davis, J. R. Distel, K. Lande, C. K. Lee *et al.*, Measurement of the solar electron neutrino flux with the homestake chlorine detector, *Astrophys. J.* **496**, 505 (1998).
- [4] Q. R. Ahmad *et al.* (SNO Collaboration), Measurement of the Rate of $\nu_e + d \rightarrow p + p + e^-$ Interactions Produced by ^8B Solar Neutrinos at the Sudbury Neutrino Observatory, *Phys. Rev. Lett.* **87**, 071301 (2001).
- [5] K. Eguchi, S. Enomoto, K. Furuno, J. Goldman, H. Hanada, H. Ikeda *et al.*, First Results from Kamland: Evidence for Reactor Antineutrino Disappearance, *Phys. Rev. Lett.* **90**, 021802 (2003).
- [6] <http://www.nu-fit.org/>, NuFIT4.0.
- [7] P. F. de Salas, D. V. Forero, S. Gariazzo, P. Martínez-Miravé, O. Mena, C. A. Ternes *et al.*, 2020 global reassessment of the neutrino oscillation picture, *J. High Energy Phys.* **02** (2021) 071.
- [8] P. de Salas, D. Forero, C. Ternes, M. Tortola, and J. Valle, Status of neutrino oscillations 2018: 3σ hint for normal mass ordering and improved CP sensitivity, *Phys. Lett. B* **782**, 633 (2018).
- [9] I. Esteban, M. C. Gonzalez-Garcia, A. Hernandez-Cabezudo, M. Maltoni, and T. Schwetz, Global analysis of three-flavour neutrino oscillations: Synergies and tensions in the determination of θ_{23} , δ_{CP} , and the mass ordering, *J. High Energy Phys.* **01** (2019) 106.
- [10] F. F. Deppisch, M. Hirsch, and H. Pas, Neutrinoless double beta decay and physics beyond the standard model, *J. Phys. G* **39**, 124007 (2012).
- [11] M. J. Dolinski, A. W. P. Poon, and W. Rodejohann, Neutrinoless Double-Beta Decay: Status and Prospects, *Annu. Rev. Nucl. Part. Sci.* **69**, 219 (2019).
- [12] M. Agostini, G. Benato, and J. Detwiler, Discovery probability of next-generation neutrinoless double-beta decay experiments, *Phys. Rev. D* **96**, 053001 (2017).
- [13] A. Caldwell, A. Merle, O. Schulz, and M. Totzauer, Global Bayesian analysis of neutrino mass data, *Phys. Rev. D* **96**, 073001 (2017).
- [14] J. M. Bernardo, Reference posterior distributions for bayesian inference, *J. R. Stat. Soc.* **41**, 113 (1979), <https://people.eecs.berkeley.edu/~jordan/sail/readings/bernardo-1979.pdf>.
- [15] H. Päs and W. Rodejohann, Neutrinoless double beta decay, *New J. Phys.* **17**, 115010 (2015).
- [16] A. Gando, Y. Gando, T. Hachiya, A. Hayashi, S. Hayashida, H. Ikeda *et al.* (KamLAND-Zen Collaboration), Search for Majorana Neutrinos Near the Inverted Mass Hierarchy Region with Kamland-Zen, *Phys. Rev. Lett.* **117**, 082503 (2016).
- [17] M. Agostini, G. R. Araujo, A. M. Bakalyarov, M. Balata, I. Barabanov *et al.* (G. Collaboration), Final Results of GERDA on the Search for Neutrinoless Double- β Decay, *Phys. Rev. Lett.* **125**, 252502 (2020).
- [18] F. Deppisch, H. Pas, and J. Suhonen, Double beta decay versus cosmology: Majorana CP phases and nuclear matrix elements, *Phys. Rev. D* **72**, 033012 (2005).
- [19] J. Speagle, A conceptual introduction to Markov Chain Monte Carlo methods, [arXiv:1909.12313](https://arxiv.org/abs/1909.12313).
- [20] S. Brooks, A. Gelman, G. Jones, and X.-L. Meng, *Handbook of Markov Chain Monte Carlo*, 1st ed. (Chapman and Hall/CRC, 2011), <https://doi.org/10.1201/b10905>.
- [21] S. Gariazzo, M. Archidiacono, P. F. de Salas, O. Mena, C. A. Ternes, and M. Tórtola, Neutrino masses and their ordering: Global data, priors and models, *J. Cosmol. Astropart. Phys.* **03** (2018) 011.
- [22] J. Dickey, Scientific reporting and personal probabilities: Student's hypothesis, *J. R. Stat. Soc.* **35**, 285 (1973).
- [23] C. E. Shannon, A mathematical theory of communication, *Bell Syst. Tech. J.* **27**, 623 (1948).
- [24] D. V. Lindley, On a measure of the information provided by an experiment, *Ann. Math. Stat.* **27**, 986 (1956).
- [25] E. T. Jaynes, Gibbs vs boltzmann entropies, *Am. J. Phys.* **33**, 391 (1965).
- [26] A. F. Heavens and E. Sellentin, Objective bayesian analysis of neutrino masses and hierarchy, *J. Cosmol. Astropart. Phys.* **04** (2018) 047.
- [27] J. O. Berger, J. M. Bernardo, and D. Sun, The formal definition of reference priors, *Ann. Statist.* **37**, 905 (2009).
- [28] A. Faessler, G. L. Fogli, E. Lisi, V. Rodin, A. M. Rotunno, and F. Simkovic, QRPA uncertainties and their correlations in the analysis of $0\nu\beta\beta$ decay, *Phys. Rev. D* **79**, 053001 (2009).
- [29] J. Myslik (LEGEND Collaboration), LEGEND: The large enriched germanium experiment for neutrinoless double-beta decay, in *Proceedings of the 13th Conference on the Intersections of Particle and Nuclear Physics* (2018) [[arXiv:1810.00849](https://arxiv.org/abs/1810.00849)].